Systematic Investigation into EndoRiftia Flagellar System Suggests Possibility of System Co-Option for Additional Functionality

Systematic Investigation into Endoriftia Flagellar System Suggests Possibility of System Co-Option for Additional Functionality

Trent LeMaster

A Thesis in the Field of Biology

for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

May 2021

Abstract

The overarching goal of this study is to contribute to the development of the *Candidatus* Endoriftia persephone-*Riftia pachyptila* symbiotic association as a biological model system with a more specific goal being to systematically investigate the symbionts flagellar system. While long thought to possess the flagellum for motility during its free-living life stage, no systematic investigation into the system and functions were discovered (Klose et al., 2015; Millikan et al., 1999). This is a significant gap in efforts to understand this association, particularly since endosymbionts frequently co-opt/re-functionalize their flagellar system for roles in symbiosis association mechanisms, often via additional functionalities distinct from motility itself (Aschtgen et al., 2019; Chaban et al., 2015). Furthermore, in ongoing energy metabolism experiments, flagellar genes were observed to be actively expressed in the *Riftia* endosymbionts(J. H. Mitchell et al., 2019). Assuming this is not about motility given the microbes' habitat at the time of transcript collection, such active expression seems to indicate the *Ca*. Endoriftia persephone flagellar system may have additional functionality during endosymbiosis. To elucidate this further, this study systematically investigated the system itself and expression across various environmental treatments, finding it complete with no obvious global pattern between treatments. However, one comparison did show clear patterns of flagellar gene upregulation and subsequent analysis of these expressed components suggested F-T3SS mediated export of non-flagellar effector proteins as the additional function. This was further supported by the prediction of large amounts of candidate type III secreted effectors in *Ca*. Endoriftia persephone while also demonstrating a lack of the NF-T3SS.

Acknowledgments

I would not have been able to complete this project were it not for the support of many people during this pursuit of my goals. Firstly, thank you to Dr. Peter Girguis, my thesis director and lab Principal Investigator, who has been an incredible source of mentorship, inspiration, and generosity. While others may have just advised on the thesis exclusively, I immensely appreciate the efforts you made to help me grow not only as a scientist but as a person. I also wish to thank Jessica Mitchell for her supervision and mentorship throughout the process. I am extremely grateful for your guidance, and particularly want to thank you for all your assistance, especially given your own dissertation obligations. I also want to say a special thank you to all people who existed as members of the Girguis lab during my time here. Never before have I come into an already established group of people, yet been so completely and genuinely welcomed.

I also want to thank Dr. James Morris, my thesis advisor, who has been immensely helpful throughout the process from start to finish. Finally, to my loving family and supportive friends, I wish to thank you for all aspects of support you provided to me during this process. The whole project would have been significantly more difficult without your consistent support.

Table of Contents

List of Primary Figures

Chapter I

Introduction

Understanding Symbiosis

Multicellular complex life arose initially in a world dominated by microorganisms. This setting, along with the subsequent interactions between them, has profoundly impacted animal evolution and lifestyles. Symbiosis was originally defined in 1879 as the “living together of two dissimilar organisms, usually in intimate association and usually to the benefit of at least one partner” (deBary, 1879; Sapp, 2002). Symbiosis would then accommodate three types of interactions depending on the effect on the host's fitness: parasitism (-), commensalism (neutral), and mutualism (+) (Mushegian & Ebert, 2016). However, recent work since then has revealed how thin the line between pathogenic and mutualistic relationships can be suggesting that knowledge from one type of interaction may be applicable in the other (Mushegian & Ebert, 2016; Wiles & Guillemin, 2019). This study will focus on microbial-eukaryotic associations and generally use symbiosis for when one partner benefits (increased fitness), and the other partner's net fitness is either not reduced or also improved (mutualism and commensalism) unless otherwise stated. Whereas pathogenic associations/symbiosis will be used when referring to the parasitism associations (see Appendix: terms).

Historically, microbial symbiosis research focused almost entirely on microbial pathogenic/parasitism associations, with a few exceptions to be discussed later, predominantly due to their significant negative impacts on human health or relevant species. However, many recent discoveries worldwide illuminate the important influence

of the microbiome on human and animal health have also elucidated that mutualistic/commensalism associations are potentially just as ubiquitous as parasitic ones and subsequently have gained significant research traction (Webster, 2014). This changing trend has been further encouraged by advances in NGS, which have provided access to previously unavailable data. This has allowed for meta-based approaches that dramatically accelerated research in the field of microbial symbiosis (Mcfall-Ngai, 2008).

However, the incredible complexity of symbiotic associations, and the fact that many endosymbionts are unculturable, has continued to be a barrier to understanding (Webster, 2014). Therefore, it is important to develop uncomplicated model systems of beneficial associations, as increasing our understanding of such processes and the mechanisms involved also provides insights into other types of symbiotic interactions (Webster, 2014). As a result, symbiotic associations between often obligate marine invertebrates and their chemoautotrophic symbionts were historically identified as well-suited targets for the development of such model systems.

Riftia pachyptila Symbiotic Association

One of the most conspicuous representatives of these associations, and the first animal in which chemoautotrophic symbionts were discovered, is the giant tube worm *Riftia pachyptila* (short *Riftia*), which thrives around deep-sea hydrothermal vents of the East Pacific. *Riftia*, with no mouth or gut, in fact entirely depends on an obligate nutritional association with a thiotrophic gammaproteobacterial symbiont currently known as *Candidatus* Endoriftia persephone (short: Endoriftia). Endoriftia have two primary life stages, that of a chemoautotrophic endosymbiont in *Riftia pachyptila* and other vestimentiferan tubeworms, and that of a free-living and/or surface-associated

microbe with proposed heterotrophic capabilities (Harmer et al., 2008; Y. Li et al., 2018). Location-wise, Endoriftia is enclosed in host cells called bacteriocytes, that are themselves sealed in a multi-lobule highly vascularized organ, the trophosome, located in the trunk of the adult host's body (Appendix: Supplemental Figures). Nourishment of the gutless host by the symbiont is obtained through the digestion of symbiont cells (releasing fixed organic carbon) that oxidize reduced sulfides such as hydrogen sulfide to fix CO2 via autotrophic carbon fixation to get their organic carbon compounds (Bright & Lallier, 2010).The symbiont being enclosed in host cells thereby depend on the unique blood of their host that can transport the substrates required for chemosynthesis (HS-, O2 and CO2) as well as all other essential nutrients from the hydrothermal vent environment to the bacteriocytes where it is then uptaken by the symbionts. See (Bright & Lallier, 2010) for a more comprehensive review of the anatomy and (Gardebrecht et al., 2012; Hinzke et al., 2020; Y. Li et al., 2018; J. H. Mitchell et al., 2019)for metabolic details.

In contrast to the hosts' complete dependence on the symbiont, the symbiont is facultative in the association. The symbiont is horizontally transmitted with each new generation of tubeworm larva being symbiont free until independently acquiring the bacterial symbionts from a free-living population in the environment (Bright & Bulgheresi, 2010). Additionally, while *Riftia* seems to exclusively associate with Endoriftia, Endoriftia in contrast exhibits low partner fidelity being positively identified as the chemosynthetic symbiont of at least three other vent tubeworm species, *T. jerichonana, R. piscesae,* and *O. alvinae* (Perez & Juniper, 2016). This is remarkable considering that these different species can be separated by thousands of kilometers of non-continuous habitats and often exist in completely different types of hydrothermal

vent habitats, and so also suggests a remarkable ability to survive and spread in the open ocean (Bright & Bulgheresi, 2010; Klose et al., 2015; Perez & Juniper, 2016). All of the above would thereby suggest significant motility capabilities, with the most common explanation being flagellum-directed motility, and also a diverse number of biological "tools" for the sensing and response to their surrounding environment (Y. Li et al., 2018).

However, while a functional flagellin gene was identified in *Riftia* previously, was successfully heterologous expressed (Millikan et al., 1999), and genes associated with flagellum have been repeatedly identified in the genome (Y. Li et al., 2018), there is still no clear understanding of the system. This lack of understanding or targeted study seems a bit odd given the clear importance the system would have in Endoriftia, especially considering the demonstrated importance of the flagellum system in the establishment and maintenance of a wide variety of pathogenic and symbiotic associations (Aschtgen et al., 2019; Chaban et al., 2015; Raina et al., 2019). Regardless, the long-held understanding by *Riftia*-Endoriftia researchers was that Endoriftia did have the capability for flagellum movement in the free-living life stage, but were specifically stated to not produce flagella as endosymbionts (Gardebrecht et al., 2012; Y. Li et al., 2018)

Overall, the host's absolute and exclusive dependency on a specific symbiont, along with its long history of study, has resulted in the *Riftia*-Endoriftia association being one in-progress model system for the study of host-microbe interactions in symbiosis. Although there is still much work to be done before the association can truly be considered a "model system" and this study hopes to contribute to that process. Specifically, there are particular gaps in detailed understanding when it comes to the free-

living stage of Endoriftia, non-metabolic host-microbe interactions, symbiotic maintenance mechanisms, and mechanisms of microbe escape after host death.

Prior Work & Focus of this Study

Before the writing of this study, Jessica Mitchell and the Girguis lab had designed experiments to study energy metabolism in this association. This study used an engineered reactor and HPRS system uniquely designed and implemented by the Girguis lab, which uniquely allowed for the incubation of hydrothermal vent animals at temperatures, pressures, and water chemistries that reflected in situ conditions (J. H. Mitchell et al., 2019)(Appendix: Supplemental Figures). RNA-seq analysis of the resulting data was then performed and the results provided data where metabolic and translational activity can be inferred using transcript abundance as a proxy for gene expression, rather than just for identification results as is the consequence of most prior Endoriftia studies being direct meta-based. In contrast, this study's use of the specially engineered reactor vessels and HPRS system designed by the Girguis lab (Appendix: Supplemental Figures), provided an unprecedented source of transcript data reflecting symbiont changes under varied conditions (J. H. Mitchell et al., 2019).

As a result, we, therefore, had a unique opportunity to leverage a series of treatments to look at the association's response to different environmental conditions replicating various possible vent conditions. Using some of the preliminary results from the dataset of differentially expressed genes(unpublished), I noticed that flagellum genes often showed significant differential expression across some of the created comparisons. This was unexpected given that the symbionts, during which this data was collected, would be contained within the bacteriocytes with no need to swim around inside the host

making it particularly odd that these flagellar genes showed such activity. While these experiments were not designed with flagellar expression in mind, they were sufficiently diverse in the scope of environmental conditions that the observations shouldn't be an outlier or artifact. Therefore, we chose to use the transcript data from these experiments to examine patterns of flagellar gene expression across these different conditions. While initial downregulation could be potentially partially explained as a matter of the flagellum being needed in free-living and then gradually lost as an endosymbiont, the observed incidence of significant up-regulation would seemingly suggest more to the story. Therefore, an investigation into the Endoriftia flagellar system as well as plausible explanations for such an observation was designated as the target of this study.

Flagellar Motility and Additional Functionalities Beyond Motility

Bacterial motility is an extremely intriguing topic from various scientific aspects, especially considering that successful infection/colonization by the majority of symbiotic bacterial, not just pathogens, is often at least partially dependent on active chemotaxis and motility (Nakamura & Minamino, 2019; Raina et al., 2019; Terashima et al., 2008). The most widespread of the various bacterial motility mechanisms is the flagellum machine, which provides efficient means of movement allowing bacteria to successfully colonize and compete with other microorganisms within their respective environments.

Structurally, while the specifics that code for flagellum pieces can vary greatly, there are generally three primary parts in all complete functional flagellum made up of similar or closely related genes. These include the basal body (motor/stator & Export Apparatus), the hook (joint), and the filament. Generally, the basal body is subsequently divided into two distinct major parts, the stator/motor system which drives rotation of the

propeller (hook-filament), and the flagellar export apparatus, or flagellar type III secretion system (F-T3SS). There is generally some sort of master regulator (often but not exclusively flhDC), that activates transcription of class II flagellar genes which transcript the basal body and F-T3SS, which subsequently lead to everything else (Fukumura et al., 2017). Of particular importance to be aware of for this study is the core of this F-T3SS, known as the export apparatus and composed of five core components (flip, fliQ, fliR, flhB, and flhA) known to be required for translocation of the principal flagellar components and at times other non-flagellar proteins outside the exterior cell membrane (Kuhlen et al., 2018).

However, while flagellum driven motility and subsequent derivative motility functions of the flagellum are understood as key virulence factors in various bacteria, many bacteria have capitalized on flagellum structure and function by adapting it to other various roles in the infection process (Aschtgen et al., 2019; Chaban et al., 2015). This seems like a logical natural development, especially when one considers that the production and functioning of flagella are considered highly energy-intensive and therefore additional functionality would further provide explanation or justification for the high expense (De Maayer et al., 2020). There are a wide variety of additional functions the flagellum has been shown to play a role in numerous symbiosis mechanisms of various types, which will be summarized only briefly below. See (Aschtgen et al., 2019; Belas, 2014; Chaban et al., 2015; Dongre et al., 2018; Green et al., 2019) for more examples and details.

Briefly, once the target site is reached (typically via flagellar motility/chemotaxis), the flagellum has been adapted for additional functions that

subsequently have been shown to play various roles in many key symbiosis establishment mechanisms including the adherence to surfaces, differentiation into biofilms, initial penetration into exterior tissue, host-communication in a specificity role, and mechanisms to trick the host into assisting subsequent penetration (Aschtgen et al., 2019; Chaban et al., 2015). Additionally, the flagellar function has been correlated to other persistence/maintenance mechanisms via Flagellum Type III secretion of immune modulation effector proteins or flagellar regulated differentiation into biofilms to hide from host produced immune or antibiotic proteins (Aschtgen et al., 2019; Chaban et al., 2015; Dongre et al., 2018; B. Li et al., 2017; Raymond et al., 2013). Therefore, while specific examples of such co-option are limited due to the historic focus on pathogens, the evidence does generally suggest that this co-opting for additional flagellum functionality is similarly vital in some beneficial associations symbiosis mechanisms. One key example being the V. fischeri-Bobtail squid symbiosis where researchers demonstrated the symbionts' flagella, in addition to their motility/chemotaxis function, played roles in biofilm formation, symbiont specificity, and the induction of host development, all of which were required to establish a sustained symbiosis (Aschtgen et al., 2019). Overall, it has become clear that regardless of whether the symbiosis is pathogenic or beneficial, flagella are vital and multi-functional life machinery that not only move symbionts towards their colonization site, but also can play multiple roles in communicating with the host, impacting microbial behaviors, and other facets of host interaction.

Bacteria Type III Secretion Systems (F-T3SS & NF-T3SS)

Bacteria often rely heavily on the use of their specialized protein secretion systems to form and maintain mutualistic, or pathogenic associations with eukaryotic hosts. In particular, the type III system (T3SS) is highly correlated to intracellular bacterial infection and/or persistence success and the method by which some bacterial pathogens intracellular niche is determined (Diepold & Armitage, 2015; Du et al., 2016; Raymond et al., 2013). Generally, there are two distinct subtypes of bacterial type III secretion systems, the injectosome (NF-T3SS) and the flagellar export apparatus (F-T3SS), and evidence have demonstrated that the NF-T3SS likely evolved from the F-T3SS (Abby & Rocha, 2012).

However, while the more recent, specialized injectosome system is an important secretion apparatus in many bacteria, many pathogens still use the flagellar T3SS to export several effector proteins/virulence factors that maintain the symbiotic association, in addition to the principal flagellar components (Chaban et al., 2015; Dongre et al., 2018; Konkel et al., 2004; Young et al., 1999). Additionally, to name just a few examples, effectors secreted from both of the T3SS have been described to facilitate entry into host cells (Konkel, Klena et al. 2004), interrupting vesicle transport and endocytic trafficking (Raymond et al., 2013), induce apoptosis of host cells (Poly et al., 2007), and manipulation of host immune response (Bliska et al., 2013; D'Haeze & Holsters, 2004), all of which promote bacterial colonization, survival, and replication.

In contrast to the understanding of this process in pathogenic associations, there are few studies with direct experimental evidence of such behavior in beneficial associations. Although recent research has demonstrated that endosymbionts, regardless of the fitness effect on host(mutualist/pathogen), share similarities in symbiosis

maintenance/persistence mechanisms including virulence determinants, host-interaction mechanisms, and even utilization of similar/same machinery to facilitate invasion and survival in host cells (Aschtgen et al., 2019; Hentschel et al., 2000; Pérez-Brocal et al., 2011; Singh et al., 2013; Wiles & Guillemin, 2019) Therefore, incidents of the flagellum being co-opted for additional functionality in pathogenic associations, such F-T3SS mediated export of host-effector proteins, would likely not be limited to pathogenic associations. One of the few direct examples being the mutualism involving the endosymbiont *X*. *nematophila* with the nematode *Steinernema carpocapsae*, as the symbiont was reported to secrete, via their F-T3SS, an extracellular lipase effector XlpA (Park & Forst, 2006), as well as a xenocin cytotoxin that was used to kill competing microbes (Singh et al., 2013).

An additional example lacking direct evidence was the case of the obligate a-motile Buchnera-aphid endosymbionts, who were shown to have hundreds of flagellum structures composed of only the basal body and export substructures covering the bacterial cell wall (Maezawa et al., 2006). While microbial symbionts and free-living microbes having numerous flagellar structures are not completely novel, see swarming style motility (Ghelardi et al., 2002), these symbionts lack the relevant genes for later stage flagellar substructures (Chong et al., 2019). Therefore, the continued existence and seemingly significant widespread expression of the retained flagellar genes would seem to suggest an additional functionality beyond motility. The authors specifically concluded that the most plausible explanation was that the structures likely serve as an export device (Maezawa et al., 2006), not only for flagellar proteins but also for other effector proteins involved in the maintenance of the symbiotic system (Schepers et al., 2021).

Prediction Approaches and Characteristics of Type III Secreted Effectors

Given the clear importance of type III secretion systems, particularly in regards to it being required for successful pathogenesis in a wide variety of pathogen responsible for human disease and/or economically important crops, efforts have been to both identify the system itself (injectosome) and predict which proteins are Type III secreted effector (T3SE) proteins. Numerous computational approaches have been developed to predict secreted effector proteins using methods that vary in the protein information used to identify the candidate effector proteins. However, T3SE prediction has struggled significantly due to the complexity of the system and the lack of any consensus among characteristics specific to type III secreted effectors (Dalio et al., 2018; Hobbs et al., 2016; Hui et al., 2020). This has resulted in the understanding that all T3SE prediction software are flawed to various degrees, with false positives (secretion signal identified in non-secreted proteins) and false negatives (documented secretion substrates not predicted as such) being common. An overview of the advantages and disadvantages of each option, curation/filtering approaches applied, the justification for the chosen software, and the methods used for ranking each software compared to the others are covered in more detail in the methods section.

Additionally, to evaluate and predict the potential role of a type III secretion system in a symbiotic association, an understanding of which T3SS is present is also of vital importance. Structurally, the bacterial flagellar-T3SS and the non-flagellar T3SS generally have a high degree of homology, but particularly in the previously discussed components that form the transmembrane export apparatus complex, and to a slightly lesser degree the components of the cytoplasmic ATPase complex (Abby et al., 2016).

Their similarity in these components is so high that various efforts have been made to develop ways to distinguish the F-T3SS from the NF-T3SS and even numerous software were designed specifically to assist in distinguishing between the two (Abby et al., 2016). The details of which approach was used in this study, along with the relevant methodology are described in more detail in the methods section.

Finally, while a comprehensive investigation into the Candidate Type III Secreted Effectors (C-T3SE) for functional predictions was beyond the scope of this study, preliminary efforts were made to compile an understanding of domains/motifs characteristic in type III secreted effectors, as well as from any host effectors specifically common to beneficial symbiotic associations, i.e., Symbiont Eukaryotic like domains (ELD). Prior studies had already demonstrated that Endoriftia possessed over 100 Eukaryotic-like protein structures, structures that are definitively associated with symbiont persistence mechanisms (Frank, 2019; Hinzke et al., 2019). Proteins with such structures are often also known as ELPs and are known to play major roles in host communication and/or manipulation in symbiosis as their ELDs are designed to bind and modulate various processes in their eukaryotic host (Frank, 2019). Additionally, while eukaryotic-like domains are distinguishing characteristics of bacterial secreted effector proteins, there is the important caveat that it is rare for an ELD to be characteristic of just one type of secretion system effector. Generally ended up with two groups of domains to check the candidates for. Group 1 consisted of (ELDs) from eukaryotic-like proteins (ELPs), with putative roles in mimicking or manipulating host cellular processes to promote colonization, intracellular survival, and other symbiosis maintenance roles (Frank, 2019; Hinzke et al., 2019). Group2 consisted of domains and motifs specifically

observed in known type III secreted effector proteins (Arnold et al., 2010; Hui et al., 2020; Raymond et al., 2013)

Chapter II.

Materials and Methods

Prior Work and Source Dataset

Specimen Collection and Data Source Experimental Conditions

The source of the count data for the differential expression portion of this study was obtained from the same *Riftia pachpytila* worms collected and processed by Girguis lab Ph.D. Candidate Jessica Mitchell as previously described in her recent publication (J. H. Mitchell et al., 2019). To obtain data regarding changes under different treatment conditions, a high-pressure respirometry system was used to allow incubation of hydrothermal vent animals at temperatures, pressures, and water chemistries that reflex in situ conditions. Various treatments(conditions) were then applied to various vessels to test and observe subsequent effects on the *Riftia pachyptila* symbiotic association (Appendix: Supplemental Figures). For interested parties, further details regarding the specifics of the High-pressure respirometry system, and animal collection locations are available in more detail in the previously referenced publication (J. H. Mitchell et al., 2019).

RNA-Seq Transcript Source

Responsive *Riftia* were chosen based on response to touch and trophosome tissue was quickly excised from three such subjects from each treatment for subsequent mRNA

transcript abundance analysis. All necessary RNA-seq sequencing and processing steps were performed as previously described (J. H. Mitchell et al., 2019). Additionally, in this ongoing experiment on energy metabolism, a differential expression results table was generated as an in-progress result containing various annotation information where each geneID representing a protein had a logFC value for each comparison of treatments (16 comparisons in total). Generation of the referenced results was performed in R using the workflow described previously but is similar to what will be described in a section below (J. H. Mitchell et al., 2019).

This Study Additional Dataset Creation

Source of Protein Set for Datamining Analysis

Starting set of proteins and CDS sequences for use in annotation and characterization efforts were from "endosymbiont of *Riftia pachyptila*", obtained from NCBI genome resource. More specifically, FASTA sequences for "endosymbiont of *Riftia pachyptila*", of the genome GenBank Assembly (GCA_000224455.2 (latest)) were obtained from the NCBI: FTP GenBank directory for endosymbiont of *Riftia pachyptila* (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/224/455/GCA_000224455.2_ASM22445v1/) under "Index of /genomes/all/GCA/000/224/455/GCA_000224455.2_ASM22445v1". The protein.faa (AA), genomic.fna (NT) was the primarily used files although the features and GenBank/GenPept/GFF files played supportive roles when needed.

Protein Annotation

Gene function annotation is a common and important first step to interpret genomic data and elucidate potential explanations or observed phenomenon. In order to investigate the completeness of the flagellar system and identify characteristics of the identified candidates for later interpretation, the amino acid sequences from the Endosymbiont were submitted to Blast KOALA, ghost KOALA(Kanehisa et al., 2016), KofamKOALA (Aramaki et al., 2020), KAAS (Moriya et al., 2007), and EggNOGmapper (Huerta-Cepas et al., 2019). In general, none of the programs were able to successfully match KOs to all present genes which led me to additional scanning and annotation efforts. Specifically, the proteins were also submitted to Interpro Scan and the Batch-CD NCBI: CDD scan tools to assist in later functional characterization and/or predictions (Lu et al., 2020; A. L. Mitchell et al., 2019). Interpro is a tool known to be useful for characterizing proteins and providing functional analysis of proteins by predicting domains and subsequently classifying them into families based on numerous other databases (A. L. Mitchell et al., 2019). Although the webserver limits the number of submissions and so the standalone version was downloaded (https://www.ebi.ac.uk/interpro/download/) before being set up in a Linux system for the subsequent domain scanning.

Endoriftia Flagellar System Investigation Part 1: Data-Mining Analysis

Identifying Flagellar Associated IDs

Using the annotation information obtained above, efforts were made to identify every present flagellum-associated gene. Specifically, KO identifiers for each gene were

isolated from other information in the combined file created previously and submitted to the KEGG pathway reconstructor tool to identify flagellum-associated hits (Kanehisa and Sato 2020). The “Cell Motility” category and more specifically, protein matches in the “bacterial motility proteins” (ko02035) section with subsections “bacterial chemotaxis” and “Flagellum assembly” category was downloaded for later use. This allowed for visualization, via the KEGG Mapper-Reconstruct Pathway tool and the flagellar reference map (Kanehisa & Sato, 2020), of what could already be identified via the various KO assignment software previously described (Figure 1). A Simple custom R script was then used to compile all descriptive information regarding the proteins obtained during the annotation step and then search through all information obtained regarding the proteins, name description information from the original download, and new descriptions/annotations sourced from the above methodology. In brief, searched for key annotation identifiers such as Flg (e.g., FlgM), Fli (e.g., FliA), flh (e.g., flhC), or flagella/flagellar/flagellum in protein name, annotation, description. Similar approaches were used to look for chemotaxis genes, although focused on ones with known links to flagellum assembly pathway or motility functionality in the KEGG or BRITE Models. They were manually curated afterward to ensure no obvious incorrect matches then combined with the above KEGG results to obtain a list of identified flagellar matches for later analyses. Additionally, the literature was also reviewed to obtain an understanding regarding the minimum set of core flagellum components/genes required for associated functions, focused on but not limited to endosymbionts (De Maayer et al., 2020; Echazarreta & Klose, 2019; R. Liu & Ochman, 2007; Toft & Fares, 2008). This

understanding was then paired with the Brite and KEGG pathway reference maps to demonstrate and visualize the Endoriftia flagellum system (Figure 1).

F-T3SS vs NF-T3SS Investigation

The previously reported observations indicating active changes in flagellar gene expression even while inside the host under healthy conditions led me to expect that the flagellar system may function in an additional role. As discussed in the introduction, secretion of type III effector proteins was hypothesized as the explanation for the observed activity. As a result, the proteins were submitted to Peffect (Goldberg et al., 2016) for a simple quick preliminary test. Out of this quick submission, only a limited number of proteins were analyzed due to computational limitations, and yet over half of the proteins were reported as type III effectors. These results seemed to suggest that secretion of Type III effectors may be occurring in Endoriftia. It is known that bacterial Type III secretion systems generally include two variations on core structures corresponding to the flagellum and injectosome designated as F-T3SS and NF-T3SS respectively(Appendix: Supplemental Figures). However, the structures are difficult to distinguish, and so the TXSScan tool under the MacSyFinder program umbrella was applied given it was specifically designed to predict the presence and degree of completion of each bacterial secretion system (Abby et al., 2016). MacSyFinder: TXSScan methodology is covered in more detail in the stated publication and is available for download at https://github.com/gem-pasteur/Macsyfinder_models. Attempts were subsequently made to use additional software designed for similar purposes to support these results(see table 1 from Abby et al., 2016). Of the other options found, SSpred(http://www.bioinformatics.org/sspred/html/sspred.html) was limited to a web

server with single sequence submissions, while T3DB (http://biocomputer.bio.cuhk.edu.hk/T3DB) and T346Hunter (http://bacterial-virulence-factors.cbgp.upm.es/T346Hunter) were no longer active (Martínez-García et al., 2015; Pundhir & Kumar, 2011; Y. Wang et al., 2012). Therefore, to further support the results from TXSScan, the necessary components were also searched for using classic blast search methodology while factoring in aspects of the methodology used from the above and additional relevant publications (Abby et al., 2016; Hobbs et al., 2016; Martínez-García et al., 2015).

Prediction of Candidate Type III Secreted Effectors(C-T3SE)

See Figure 6 for a broad overview of the following described work-flow. Despite the clear importance of and a dramatic increase in the number of available whole-genome sequences, it remains challenging for accurate prediction of T3SE given the inconsistency in characteristics among known T3SE proteins. It is known that all Type-III bioinformatic predictors are flawed in various ways, with various levels of false positive and false negative rates depending on the program in question (Hobbs et al., 2016; Hui et al., 2020). As a result, while they are considered viable and valuable as an investigative tool to inspire future research directions, it is generally understood that subsequent experimental validation is required to confidentiality declare them as type III secreted proteins (An et al., 2016; Dalio et al., 2018; Hui et al., 2020).

Applied Prediction Software

Many methods to predict T3SE proteins are available, with each publication that releases a new software also showing statistics regarding their accuracy over alternatives

(An et al., 2016). However as discussed it is known that the majority of current software tools often suffer problems of high false-positive rates, and the specific historic issues with T3SE prediction efforts are reviewed in extensively more detail elsewhere (Hui et al., 2020). To control for some of the known issues with T3SE prediction, the literature was reviewed regarding usage, accuracy, and known methods to identify false hits. Three different programs were chosen based on their accuracy and ensemble-based prediction methodologies. Specifically, Effective DB (Eichinger et al., 2016), Bastion3 (J. Wang et al., 2019), and T3Sepp (Hui et al., 2020) were used to identify potential candidates. Effective DB was also chosen over other comparable options (e.g., Bean2.0) as its report included the predicted eukaryotic-like domains (ELD) and Conserved Chaperone Binding domains which are likely to interact with host proteins and effector chaperones respectively (Eichinger et al., 2016). There was also additional software considered, but many were inactive at the time of this study, did not allow for large submissions, or were previously evaluated to be worse than the above methods. Additional details regarding the specifics of each program's methodology are available in their publications briefly discussed in the supplemental of this study.

Weighting and Curation of Prediction Results

While a detailed investigation and confirmation of these candidates were beyond the scope of this study. Some efforts were made to narrow down the list of potential candidates, based on known filtering criteria and weighting of the chosen software via reported efficacy (Dalio et al., 2018; Hui et al., 2020). Weighting idea is that based on the analysis performed by Bastion3 Makers and others, EffDbT3 was about 70-80% less in all reported test criteria (SN, sensitivity; SP, specificity; ACC, accuracy; PRE, precision)

when compared to Bastion3 Ensemble (J. Wang et al., 2019). Various T3Sepp publication comparison analyses showed T3Sepp to be more specific, accurate, and precise when compared to Bastion3. Also, T3SEpp showed increased specificity (71% [24/34]), precision (80% [39/49]), and accuracy (83% [63/76]) compared to those of Bastion3 and the other tools. Their accuracy stemming from their comprehensive approach using not only a sequence-based prediction, but also results from submission to signal (Almagro Armenteros et al., 2019), psortB (Yu et al., 2010), and TMHMM (Krogh et al., 2001) software. Bastion3 was also demonstrated to be a bit more sensitive than T3Sepp and so was ranked closer to T3Sepp than EffectiveT3 (Hui et al., 2020). As a result, the weighting methodology followed a similar trend where in terms of weight T3Sepp>Bastion3>>EffectiveT3. Prediction Weight Scores (PWS) were generated based on these understandings in order to weight the prediction scores from each software to highlight the more likely candidates from the possible false positives. An overview of the specific approach and methodology used to generate the prediction weight scores (PWS) is available in this study's supplemental. This final list of proteins included 150 Candidate T3SEs (Figure 6).

However, given the stated issues with T3SE prediction, even the authors of the more comprehensive T3Sepp approach are careful to state that their identified T3SE characteristics used to guide prediction and evaluation were not comprehensive with numerous exceptions (Hui et al., 2020). As a result, while some additional filtering efforts were performed as general supportive use(see supplemental), the comprehensive effort needed would be beyond the scope of this study and so was not directly used to further narrow down my list of candidates beyond the stated 150 set.

Endoriftia Flagellar System Investigation Part 2: Differential Expression Analysis

There are many steps involved in analyzing an RNA-seq experiment but it begins with sequencing reads. These are aligned to a reference genome, then the number of reads mapped to each gene can be counted which results in a table of counts usable for differential expression analyses. The details regarding the generation of this counts table, which is the starting point of my DE analysis, is described in the section "RNA-Seq Transcripts and Counts" above. Given the context of my study and some initial observations of flagellar DE trends (Figure 2A), the analysis was focused on count data obtained from the High Oxygen, no nitrogen samples(noN) with either SW (upper surface seawater~no sulfide), LS (Low Sulfide), or HS (High Sulfide) treatments (Figure S2). The analysis to be performed was also generally focused on two predominantly flagellar-associated IDs, categorized as either associated with Flagellar Assembly or Flagellar Chemotaxis. The additional supporting analysis also included the flagellar genes further divided into associated subsystems, or the identified C-T3SE IDs. All steps that follow were performed in R using custom R scripts and packages listed in the R session info while the workflow performed was derived from the methodology outlined in (Law et al., 2018).

Differential Expression Workflow

For RNA-seq data, it is vitally important to filter out genes or exons that are never detected or have very small counts across all libraries as they can interfere with statistical approximations in DGE pipelines and provide little evidence of biologically important differential expression since a gene must be expressed at some minimal level before it is translated (Ritchie et al., 2015). This study followed the general rule of thumb where,

genes are dropped if they can't possibly be expressed in all the samples for any of the conditions, and filtering is done with count-per-million (CPM) rather than raw counts, as the latter does not account for differences in library sizes between samples. Normalization is also required in RNA-seq data analysis to ensure that the expression distributions of each sample are similar across the entire experiment. TMM normalization is designed to eliminate composition biases between libraries (Robinson & Oshlack, 2010), and so was chosen for this analysis due to the significant discrepancies in our library size.

Once the data is normalized and the quality looks decent, testing for differentially expressed genes can then proceed. Limma was chosen based on suggestions by peers and because the hallmark of the Limma approach is the use of linear models to analyze entire experiments as an integrated whole rather than making piece-meal comparisons between pairs of treatments (Ritchie et al,. 2015). In the limma approach to RNA-seq, read counts are converted to log2-counts-per-million (logCPM) and the mean-variance relationship is modeled either with precision weights(voom) or with an empirical Bayes prior trend(limma-trend) (Ritchie et al., 2015). The voom, but specifically the voomWithQualityWeights function was applied because its approach of balancing the need to reduce noise without reducing its power of detecting meaningful changes was well-suited to our dataset (Ruijie Liu et al., 2015). This provided us voom-transformed counts which Limma subsequently uses to test for differential expression. To summarize, first, a linear model was fit using the lmFit function from Limma which estimates group means according to the design matrix, as well as gene-wise variances (Ritchie & Liu et al., 2015). Then a contrast matrix was created using the makeContrasts function from

edgeR designed for this purpose and the resulting contrasts are between treatments with certain aspects held constant (Chen et al., 2016). The comparisons predominantly used for this study were between the treatments of different sulfide levels (SW, LS, HS) with the constants being no nitrogen(noN) and high oxygen (HO) conditions. The contrast matrix was then applied to the fit object to obtain statistics and estimated parameters of the comparison we are interested in. Finally, the eBayes function was applied to the generated contrasts.fit object to perform empirical Bayes shrinkage on the variances, and estimate t-statistics and associated p-values (Mccarthy et al., 2012). Custom R functions were then used to organize and reformat the produced file into a comparison result table, containing the log-transformed data representing an estimate of log2fold expression change between treatments. The resulting table Values of P were adjusted for false discovery using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995), with a preliminary cutoff of 1, and later significant DE(sigDE) cutoff values being adj.p.value <= 0.05.

All logFC based figures were generated using the above results table and any differential expression visualization (e.g., logFC shown) in this study is only on the significantly DE genes (adj.p.value <= .05) unless otherwise specified. In contrast, the candidate characteristic figures(Figure 7) were also done using candidates found in the transcripts, but not with .05 adjusted p. value filtering as their candidate identity was independent of them being differentially expressed. A more detailed version of the differential expression workflow along with the R session info document is found in this study's supplemental document.

Chapter III

Results

Flagellar System in Endoriftia

All indications from our results suggest a complete and functional Flagellar Assembly Pathway (FAP) in Endoriftia, at least regarding the minimum core components considered necessary for motility function in the reference flagellum (Figure 1). The system being complete was generally expected as the endosymbionts exist in a free-living stage during which the flagellar system is likely vital for motility. Additionally, while not investigated in detail, previous publications studying siboglinid symbionts have claimed that all sequenced symbiont genomes had the full complement of genes indispensable for a functional flagellum, plus several accessories or duplicated components (Y. Li et al., 2018).

More specifically, all the previously discussed minimum primary components of the flagellar system were present, except for the core fliN and pseudo-core flhCD. However, fliN was subsequently detected in an alternative published Endoriftia protein set, *Candidatus* Endoriftia Persephone str. Guaymas (taxid:910259), as well as in the endosymbiont of *Tevnia Jerichonana* (vent Tica) (taxid:1049564) published set. Furthermore, outside of the various core and standard flagellum genes, there was a variety of additional flagellar system matches corresponding to the “others" section in the BRITE hierarchy (Kanehisa et al., 2021). The majority of which seemed to be regulators associated with either two-component system sensors/response and flagellum

biosynthesis. Except for fliN, all identified components were also identified specifically in the transcript dataset obtained and used for other aspects of this study (J. H. Mitchell et al., 2019). See Appendix for the figure from the Liu 2007 study displaying the identified core components among flagellated bacterial species.

Unexpected Flagellar Expression Suggests Additional Functionality

As previously discussed, in ongoing studies of energy metabolism conducted by a member of the Girguis lab it was observed that some flagellum genes were being differentially expressed across the treatments (J. H. Mitchell et al., 2019). While these experiments were not designed with flagellum expression in mind, they were sufficiently diverse in the scope of environmental conditions that we chose to use the obtained transcripts to investigate the flagellar system in Endoriftia. More specifically, the analysis sought to determine if the expression of flagellum-associated genes could be directly correlated in any clear way to the treatments looking at energy metabolism. As a result, an initial Differential Expression Analysis for all treatments was performed using the previously obtained RNA-seq reads (J. H. Mitchell et al., 2019). Specifically, the goals were to: (1) Determine if there was any pattern in terms of flagellar gene expression that could be correlated to any of the given treatments (Sulfur, Nitrogen, or Oxygen) corresponding to metabolic capabilities of Endoriftia, (2) Identify a singular relevant condition to focus on and then Visualize logFC changes in the created comparisons for subsequent analysis. While hydrogen analysis was also part of the initial energy metabolism study, these comparisons were not included due to a lack of statistical robustness when it came to the flagellum-specific counts.

Interestingly, the results from this investigation showed that many of the flagellum genes were significantly differentially expressed (adj. p.val <.05) in many of the treatment comparisons (Figure 2). Considering the circumstances of the symbiont at the time of collection, this was considered an unexpected result. Additionally, in contrast to initial expectations, there was seemingly no clear global pattern of flagellum gene expression across corresponding treatments (Figure 2). However, while there was no clear global pattern, there were some examples of drastic changes in the logFC of Flagellum genes that seemed to emphasize a clear strong effect in LS samples, particularly in the LSvsHS comparison, and so sulfide was targeted for additional analysis (Figure 2 and 3). A new differential expression analysis was subsequently performed but now oriented only for the chosen treatments of interest, specifically low sulfide (LS), high sulfide (HS), and near-surface sourced seawater [<=0 sulfide] (SW). The hope was that narrowing it to just this subset might elucidate a pattern between the various levels of sulfide and flagellum gene expression, as well as provide some hints to the cause of this expression.

Flagellar Gene Expression as a Function of Environmental Conditions

This analysis focused on specifically the sulfur treatments did indeed provide some further clarity regarding the flagella expression under these treatments. The resulting comparisons/contrasts obtained included LSvsHS, SWvsHS, and SWvsLS all of which were also under no Nitrogen and High Oxygen conditions. While interesting to observe, the results from an analysis similar to figure3 but with the flagellar subsystems showed no clear stand-out pattern distinct from the total flagellar expression that provided clarity on possible explanations for this unexpected activity (Figure 5).

Therefore, a subsequent analysis was performed to obtain the log fold change(logFC) of each specific flagellar gene across the chosen comparisons (Figure 4).

Interestingly, in the LSvsHS comparison(4A,4B), It was discovered that all flagellar genes identified significantly DE (adj.p.value <= .05) were all present and ubiquitously upregulated with only a few exceptions. Although it should be noted that these exceptions are genes that were indeed identified as flagellum associated, but are not "core" or part of the KEGG or BRITE "other" category because they are either global regulators in numerous other processes besides the FAP(e.g. rpoD,rpoN), or are not considered to be core to the standard flagellum(e.g. flrB). Overall, this almost ubiquitous expression might seemingly suggest that lower levels of sulfur are inducing this change. However, the other sulfur comparisons demonstrate that the correlation is seemingly not that direct in cause and effect(Figure 4). In fact, about a third of the same genes shown to be upregulated in LSvsHS show down-regulation in the SWvsLS comparison which seems counterintuitive(Figure 4A,4C). Similarly, in contrast to expectations given its seeming similarity to the LS vs HS comparison, the SW_vs_HS comparison showed even fewer sigDE flagellum genes with even less of a clear pattern(Figure 4D).

Flagellar may function as T3SS-exporter of non-flagellar likely host effector proteins

As we previously demonstrated, the Endoriftia flagellum system seems to be not only complete but also seems to be actively expressed under a variety of treatment conditions, even while still sealed inside the host. In addition, all of the differentially expressed flagellum genes in the chosen comparisons were almost entirely associated with only two flagellum categories, those being Chemotaxis and flagellum type III

secretion(F-T3SS). As a result, an investigation into the type III secretion capabilities of Endoriftia was subsequently performed to investigate our hypothesis.

While the flagellum Type III secretion system(F-T3SS) was already identified as present, a prior study had also stated that no type III secretion system was identified in any of the tested siboglibonoid symbionts, Endoriftia included (Y. Li et al., 2018). However, the authors of that study didn't seem to investigate anything specifically, and given the significant degree of commonality in F-T3SS and NF-T3SS structural components, specific evidence is needed (Abby et al., 2016; Kuhlen et al., 2018). Therefore, attempts at detection of any type III secretion system were performed using software (MacSyFinder: TXSScan) previously designed with the capability to distinguish flagellum from injectosome T3SS components (Abby & Rocha, 2017). Furthermore, a manual investigation was also performed using aspects of the same "forbidden" vs "mandatory" components logic previously established in the publication by the same authors (Abby et al., 2016). The results from this investigation concluded that no injectosome type III secretion system (NF-T3SS) was seemingly present within the Endoriftia genome(Appendix: Supplemental Figures)

Additionally, the FASTA sequences for the downloaded Endoriftia set of proteins were submitted to the three specifically chosen software designed for the prediction of Type III Secreted Effectors(T3SE). Why these specific software options were chosen as well as the methods used to control for false positives and negatives were performed using the methodology described further in the methods section. The initial results, after compilation and deletion of duplicates, resulted in over 700 Candidate Type III secreted effectors(C-T3SEs) predicted (Figure 6). While an initially intriguing result, the

previously reported high rates of false positives meant that subsequent curation was needed before making any significant inferences. After the remaining protein prediction scores were weighted based on the reported accuracy of each software, further curation was performed resulting in approximately 200 candidates. Additional filtering was performed using additional criteria (see methods for details), resulting in a final list of 150 candidate type III secreted effectors, or C-T3SE (Figure 6). However, while the final candidate list contained 150, also flagellar IDs and so were often excluded from further candidate analysis since they are known to be flagellar secreted thereby no longer being a "candidate", resulting in 135 candidates. Stopping at this level of filtering, while understood to be not comprehensive in possible ways to remove false hits, was determined to be a reasonable balanced stopping point to both limit false hits without removing actual candidates given the known inconsistencies with type III effector proteins. Additionally, the candidates were submitted to various domain scanning tools to assist in guiding future understanding of the potential effector type capabilities of these candidates if they are indeed actual type III secreted effectors. To facilitate this, a literature review of some known domains and motif patterns common to type III secreted effectors and/or effector proteins, in general, was also obtained to assist in identifying the best potential candidates for any future research. While there were some encouraging results showing commonality in domain or motif, a thorough understanding was limited and the results were not used as criteria for any subsequent curation efforts given the lack of clarity regarding the degree of commonality between F-T3SE vs NF-T3SE domains and motifs.

A broad overview of the predicted Candidates and their Characteristics

A detailed investigation of the 135 identified C-T3SEs is beyond the scope of this study, however, efforts were made to obtain various information and characteristics of the candidates. The idea being that this may assist in the choosing of a few specific targets for a later more detailed investigation into the candidate T3SE as well as guiding potentially more thorough functional investigations. Broadly, it was observed that the KEGG groups and COG aspects for these candidates seemed to generally follow previous publications demonstrating such groups across the entire Endoriftia genome. Specifically, this is referring to a large amount of Environmental Sensing and Signal transduction, as well as Defense and secondary metabolisms associated matches (Figure 7). Both of which were categories previously reported to be quite high in Endoriftia given the mutualistic association (Hinzke et al., 2019; Y. Li et al., 2018). While interesting, it should be stated that a not-small portion of the candidates identified was uncharacterized or hypothetical in functional understanding making detailed characterization of some difficult, and leading to possible exclusion from figure 7 if no matches were obtained. Interestingly, a significant amount of the C-T3SEs was classified as "chaperones, known host interactors, peptidases, inhibitors, and stress response" which was in line with expectations if these were indeed T3SEs. Furthermore, a not-insignificant portion of the candidates possessed domains previously identified as playing putative roles in symbiont-host interactions, both in terms of general ELP/effector domains and even some matches to specifically type III secreted effector characteristic domains.

Chapter IV

Discussion

In recent years it has become clear that flagella in beneficial symbiosis, similarly to pathogenic associations, are essential and multifunctional machinery that not only mediate movement of symbionts towards colonization sites but also play multiple roles in communicating with the host (Aschtgen et al., 2019; Chaban et al., 2015; Dongre et al., 2018; Raina et al., 2019; Schepers et al., 2021). Herein, we sought to investigate the flagellar system in Endoriftia, not only to confirm the presence of the required genes but hopefully to elucidate how it may or may not be influencing the beneficial association and other symbiont–host interactions.

Significance of Results

Overall, the results from the initial stage of this study demonstrate that the Endoriftia flagellar system was seemingly complete, at least in regards to the generally accepted core machinery based on mapping to the KEGG flagellar reference system and a previously identified set of 21-24 core components for minimum flagellar function (Kanehisa & Sato, 2020; R. Liu & Ochman, 2007)(Figure 1). The sole exception being the flagellar transcriptional activators FlhC and FlhD (flhDC) traditionally included as standard in the KEGG model. However, while flhDC is indeed a vital flagellum assembly pathway (FAP) regulator in some species (KEGG), it's also known to have highly variable distributions and was deliberately excluded from the Liu identified core set for

that very reason (R. Liu & Ochman, 2007). Various other non-standard flagellum proteins were also identified in this study, as well as in a prior study (Y. Li et al., 2018). This suggests the possibility for compensatory effects from one of the more atypical, variable, and less understood flagellum-associated regulators. Furthermore, while a complete explanation for the above missing pieces is not fully understood, the historic hints at Endoriftia flagellum (Millikan et al., 1999), the free-living life stage with extensive biogeographical capabilities (Gardebrecht et al., 2012; Perez & Juniper, 2016), and surprisingly rapid escape from a dying host (Klose et al., 2015) all support Endoriftia having a functional and vital flagellum. Therefore, one could conclude that either: 1) flhDC is not needed for Endoriftia Flagellum assembly, 2) some other product fulfills/compensates for the typical role of flhDC in the flagellum assembly process, or 3) flhDC is there but just undetected for some reason, potentially due to not being required for the undefined flagellum role during the endosymbiotic life stage.

In the previously discussed ongoing studies of energy metabolism conducted by a member of the Girguis lab (J. H. Mitchell et al., 2019), it was observed that some flagellum genes were being differentially expressed across the treatments. Although a complete flagellar system in Endoriftia was certainly expected, it was quite striking to find active expression of so many flagella-chemotaxis and flagellum assembly proteins, even while contained within the bacteriocyte host cells where no motility would be needed (Figure 2). Especially since metagenomic data from prior studies specifically discussed Endoriftia having flagella mediated movement in the free-living life stage, but were thought to not produce flagella as endosymbionts (Gardebrecht et al., 2012; Y. Li et al., 2018). It is also now known that the flagellar system of beneficial symbionts is

frequently adapted for roles beyond motility and often specifically oriented towards symbiosis persistence or maintenance mechanisms (Aschtgen et al., 2019; Chaban et al., 2015; Schepers et al., 2021; Singh et al., 2013). Subsequently, this could lead one to infer, especially given the high energetic costs of the system itself and the environment they live in, that such active expression of flagellar assembly genes would indicate the flagellum was co-opted for an additional function (De Maayer et al., 2020).

In contrast to initial expectations, the patterns of flagellar gene expression we were studying did not correlate to the patterns observed in genes related to energy metabolism (Figure 2). However, it was notable that over half of all the identified flagellar genes were significantly differentially expressed in multiple different comparisons (adj. p.val <= .05) (Figure 2). This trend was particularly prominent in the comparisons involving low sulfide (LS) treatments, where it was observed a particularly high amount of flagellum associated genes were differentially expressed compared to all other treatments and often with comparatively larger logFC as well (Figure 2, Figure 3). Furthermore, it was significant that almost every single flagellar gene present was upregulated in the LS_vs_HS_noN comparison (Figure 4B). However, if such a pattern was connected to lower sulfide levels, it was unexpected that in the SW_vs_LS_noN comparison, all the same, flagellum genes were either down-regulated, indicating higher expression under the less sulfide treatment, or not expressed at sig value (Figure 4C). Similarly, the SW_vs_HS comparison, also being a low vs higher comparison, had a negligible amount of flagellar-associated genes and seemed to exhibit no trend at all(Figure 4D).

Notably, after a deeper investigation of all the sigDE flagella associated genes (adj. p. val <= .05) upregulated in the LSvsHS comparison, almost all (with only a few exceptions) were associated with only a few specific subcategories in the flagellum system(Figure 5,4). Specifically, most differentially expressed flagellar genes were either regulators sometimes associated with F-T3SS assembly (e.g., flhF, flgN), previously demonstrated F-T3SS secretion chaperones (e.g., fliT, fliS, fliK), the F-T3SS cytoplasmic complex powering export (FliH, FliI, and FliJ), or the core genes known to build the export apparatus (EA) itself (e.g., FlhB, FliP, FliQ, FliR, and sometimes fliO|Z) (Figure 4A). Interestingly, while the cytoplasmic ATPase structure is not required for principal secretion (forming filament or needle) in either system (Wagner et al., 2018), they are often directly associated with successful non-principal secretion of effector and/or virulence proteins (Deng et al., 2017; Kuhlen et al., 2018). Additionally, as previously stated, studies on gene loss events in many endosymbionts showed a loss of flagellar motility genes, with only the proteins involved in protein export within the flagella assembly pathway (type III secretion system and the basal body) being not only kept but highly expressed, which suggested a change in the functional purpose of these structures (Chong et al., 2019; Schepers et al., 2021; Toft & Fares, 2008). Interestingly, while gene loss events are typically only common to obligate endosymbionts unlike Endoriftia, the implications are nonetheless intriguing given that the genes kept versus those lost seem to follow patterns similar to the DE flagellar genes vs not DE in our LSvsHS comparison (Figure 4B). All of which provides evidence suggesting that a plausible explanation of such expression is that the Endoriftia flagellar system has been co-opted for the export of non-flagellar proteins.

Subsequently, simple preliminary efforts were made to determine if any Type III secreted effectors could be identified in the Endoriftia genome given that the flagellar export apparatus is also a type III secretion system(F-T3SS). Initial results were encouraging as a quick preliminary test with a simple T3SE prediction software showed quite a significant amount of predicted type III secreted effectors. After subsequently reviewing the literature for the most updated and comprehensive T3SE prediction software, the Endoriftia set was subsequently submitted to the three identified programs. An unexpectedly large number of Candidate Type III Secreted Effectors was predicted considering the demonstrated lack of an injectosome, which seemingly supports the F-T3SS hypothesis (Figure 6). However, while T3SE prediction software is known to suffer from high rates of false positives and negatives, the fact that such a large number of candidates (Figure 6) were identified would suggest that some T3SEs are indeed present. However, to properly suggest that the flagellar secretion system is indeed the one responsible for exporting these effectors, we also had to determine if the typical injectosome (NF-T3SS) was also present. My subsequent investigations using MacSyFinder and direct searching also demonstrated that while the F-T3SS was complete, there was seemingly almost none of the mandatory NF-T3SS genes required for assembly of a functional injectosome structure present in Endoriftia. Therefore, such a quantity of T3SEs identified in an organism seemingly lacking a complete NF-T3SS would seem to suggest they are specifically F-T3SEs and some sort of F-T3SS mediated export functionality.

An in-depth investigation of the candidates and their characteristics were beyond the scope of this study,but preliminary results seemed encouraging in regards to their

effector identity. For example, the visualizations in which KEGG IDs were mapped to the candidates showed a high number of candidates mapped into categories, such as “defense and secondary metabolism”, “membrane trafficking transport”, by which some known effectors would also be grouped (Figure 7). Furthermore, as previously mentioned, there were also large amounts of eukaryotic-like domain proteins identified in Endoriftia suggested to function in host communication and/or manipulation roles (Frank, 2019; Hinzke et al., 2019; Y. Li et al., 2018). In fact, the majority of our candidates also possessed ELDs, and many of our candidates even had domains with “virulence” or “effector” in their name/description with some being exact matches to previously identified putative type III secreted effector domains (Arnold et al., 2010; Dean, 2011). Therefore, although limited in specifics, the evidence herein does certainly indicate at least some sort of effector driven active communication or manipulation by the symbiont, even if later shown to be exported via a different mechanism.

Overall, regardless of if the hypothesis of this study is supported (i.e., F-T3SS mediated export of non-flagellar proteins), it is significant that the results suggest the Endoriftia flagellum system is actively transcribed inside the host bacteriocytes where no motility is needed. Therefore, given the reported evidence and previously discussed examples, one could reasonably conclude that the flagellar system in Endoriftia, similar to many endosymbionts, was likely co-opted for additional functionality during its endosymbiotic life stage. Furthermore, the identity of the differentially expressed components, lack of injectosome, and large numbers of predicted T3SEs suggest that the co-option of the F-T3SS for effector export may be one of the more plausible possibilities.

Limitations

One simple limitation is that aspects of this study were greatly supported by the conclusions from the MacSyFinder: TXSScan software (Abby & Rocha, 2017), which was used to demonstrate the lack of an injectosome (NF-T3SS), as well as provide an additional level of support to the results demonstrating the flagellar system being present(Appendix: Supplemental Figures). The limitation being that while this software does seem to be accurate in a wide variety of cases, recent publications have demonstrated that, at least in the studied Shigella species, “all classes of type III secretion systems (F-T3SS or NF-T3SS) shared the same architecture at the level of the core structures", part of which is in contrast to the “mandatory” and “forbidden” concepts core to the prediction methodology of MacSyFinder: TXSScan (Abby et al., 2016; Kuhlen et al., 2018). Of course, this does not completely rule out the MacSyFinder methodology, given the limited subject size of the study, but does hint at the possibility that their methodology may not be ubiquitously viable.

The majority of other limitations associated with this study stem from aspects predominantly associated with our specific F-T3SS alternative function hypothesis. The first being the previously discussed in-consistency in signature characteristics specifically associated with T3SEs, which results in T3SE prediction software having higher than desired false positive and negative rates reported (Hobbs et al., 2016; Hui et al., 2020). Furthermore, even if one uses the newer software and assumes accuracy, the prediction methods may be limited in their capability to predict specifically F-T3SEs as a natural consequence of machine learning approaches. Specifically, because investigations into the various “training” datasets showed an expected discrepancy in the amount of F-T3SE

to NF-T3SE, thereby biasing the prediction and potentially limiting the accuracy of predicting flagellar secreted ones. Although a simple test we performed was indeed successful at identifying some(but not all) of the known specifically F-T3SEs (Chaban et al., 2015; Konkel et al., 2004; Singh et al., 2013)

Future Directions

Future Directions: F-T3SE Explanation Hypothesis

Future studies intending to explore this hypothesis further would be supported by differential expression style experiments similar to the ones in this study but instead designed specifically with flagellar expression in mind. Furthermore, adjustments to the treatment conditions with an effort to create comparisons more distinctly reflecting "dying", "normal healthy", and "optimal health" tubeworms may provide patterns that more clearly elucidate the possible additional function being exhibited. Speculating on some of the possible explanations for the expression trends being seemingly concentrated around the LSvsHS comparison. Prior studies have shown that bacteriocyte cells in the trophosome of dying hosts were "fuzzy", which led to speculation that maybe the F-T3SS is secreting effectors that lyse host cells to facilitate the observed remarkably quick escape and subsequently could be one role of this F-T3SS mediated export. Another idea was that since the majority of the components were concentrated specifically in the LSvsHS comparison (Figure 4), this could be indicative of the F-T3SS exporting something that signals the host it needs more "food".

Additionally, there are a wide variety of additional methods that could be used to further curate and rank the predicted candidate T3SEs that would support the stated

hypothesis but were beyond the scope of this study. Specifically, anyone seeking to further evaluate the predicted candidates identified in this study or any C-T3SE in general, should see these publications for a more comprehensive review of the issues with T3SE prediction and the various criteria used to predict, characterize, and evaluate T3SEs (Dalio et al., 2018; Hui et al., 2020). Briefly, most of the common criteria used include the presence/absence of signal peptides, transmembrane domains, effector characteristics domains, and localization software results. Additional evidence could be obtained through an application of more advanced statistical clustering methods to evaluate if any of the candidates showed a strong similarity in expression patterns with either the F-T3SS export apparatus or T3SE chaperones.

Future Directions: Additional Alternative Explanations

While a detailed investigation was beyond the scope of this study, a wide variety of alternative explanations for the unexpected expression of Endoriftia's flagellum were also considered, potentially serving as a signaling mechanism for further action. One simple idea is based on a role in host-bacteria signaling as it is known that components of the flagellar system are often warning signs for host immune systems signaling the presence of a bacterial infection (Chaban et al., 2015). A prior study also suggested that *Riftia* might employ histone-derived antimicrobial peptides to modulate the symbiont's cell division (Hinzke et al., 2019). Therefore, it could be theorized that Endoriftia flagellum activity in the host is functioning as a sort of health indicator letting the host know the "status" of its current host population thereby allowing it to only modulate the symbiont cell division when necessary.

Another compelling explanation regards the possibility of flagellar-mediated host immune defense, likely via some facilitatory role in biofilm development and escape. Previous evidence suggests that the mechano-sensing of surfaces by flagellum is often vital to the initialization of biofilms (Aschtgen et al., 2019; Belas, 2014; Chaban et al., 2015). Furthermore, many animal and plant hosts will generate reactive oxygen species to defend against pathogenic endosymbionts, and such behavior has been suggested as a possibility in *Riftia* as well (Hinzke et al., 2020). In support of this, flagellar-mediated biofilm development is one suggested method of defense endosymbiont use to counter reactive oxygen species (Belas, 2014; Chaban et al., 2015; Leid et al., 2009). Our study demonstrated that of the many significantly upregulated flagellar-associated gene groups in the LSvsHS comparison, a significant portion was environmental sensing and response (motor/stator) mechanisms associated with chemotaxis (Figure 4). This was not completely unexpected as flagellar mediation of biofilm development has been previously associated with these same chemotaxes and stator/motor genes. Therefore, given the potential host pressures, one could easily foresee a scenario where susceptible subpopulations of the symbiont establish biofilm in response to host modulation efforts. Additionally, this functionality could similarly play a role in initial attachment, or later dispersal out of the biofilm for later life stages and/or escape from a dead host.

Conclusion

The Endoriftia-*Riftia pachyptila* symbiotic association is an excellent target of development for a host-microbe interactions biological model system, and while the development process has only just started in comparison to known systems, the results from this study have proceeded to narrow that gap.

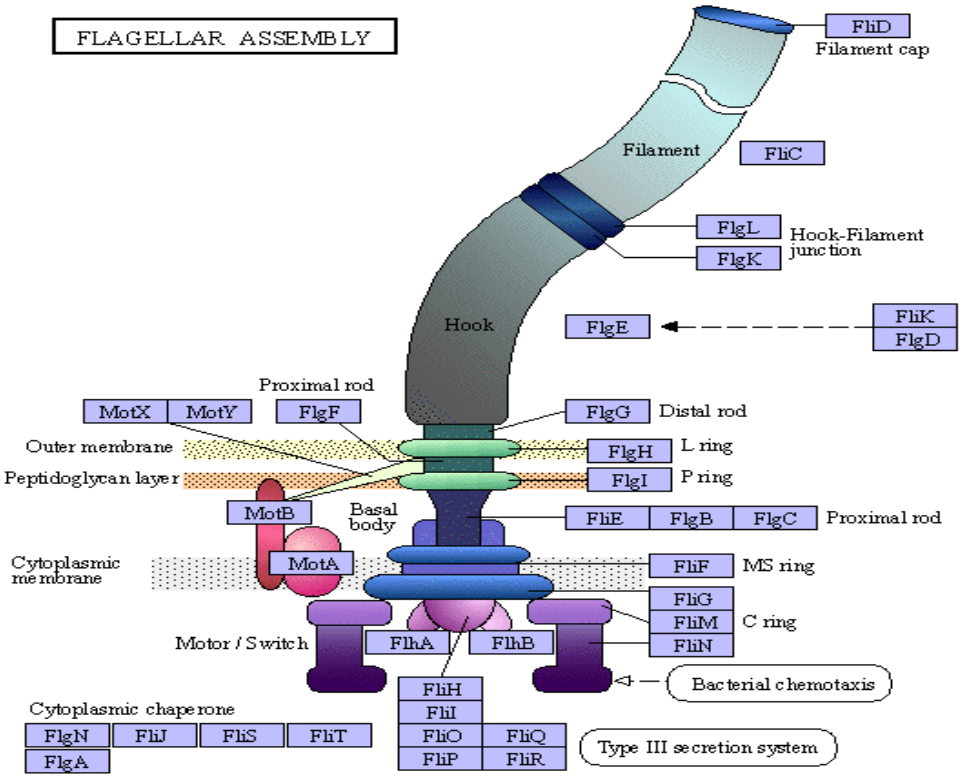
This study herein systematically looked at the Endoriftia flagellar system demonstrating it possessed all necessary components for principal function. Notably, in contrast to prior conclusions, we also observed active expression of Endoriftia flagellar components that are theoretically distinct from its motility function given the microbes' enclosed bacteriocyte habitat at the time of transcript collection. One implication of this being that such active expression may be indicative of the Endoriftia flagellar system being adapted for additional functions during endosymbiosis, especially given the high energetic costs associated with flagellar structures.

Together, the results from our study indicate that the flagellar system in endosymbiotic Endoriftia was likely adapted for an additional function, distinct from the motility function vital to its free-living stage. A hypothesis that becomes even more plausible when one considers the numerous examples of such behavior in a wide variety of endosymbionts in both pathogen and beneficial associations. Furthermore, our results show that one of the more plausible explanations of this additional function seems to be the F-T3SS mediated export of non-flagellar effector proteins. In support of this, our results also showed a significant amount of C-T3SE predicted, with many possessing domains matching known T3SE and other symbiont host effector domains. However, while it was certainly compelling evidence to find so many C-T3SEs in a system lacking a typical injectosome, the lack of experimental evidence and limitations of T3SE prediction software made us unable to confidently point to a protein and confidentially declare it a F-T3SE rather than just a candidate limiting this secondary conclusion.

Figures

Figure 1: Endoriftia Flagellar System Map

Visualization of the flagellar assembly reference map(top) with identified Flagellar Assembly Pathway genes identified in Endoriftia*(bottom).* ***Green*** *= Identified in KEGG,* ***Yellow*** *= Found via Blast,* ***Purple*** *= not identified in our* Endoriftia *data but identified in endosymbiont of Tevnia (vent Tica). Adapted from the current and previous version of the KEGG Flagellar Assembly Pathway Map*



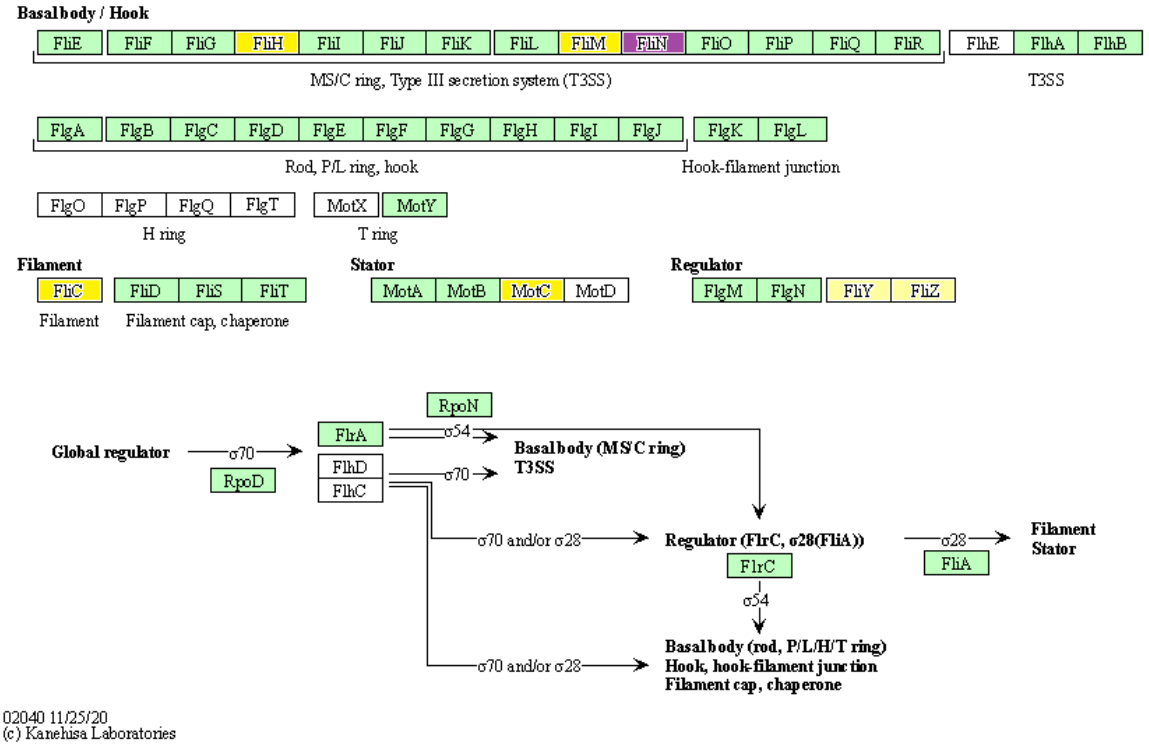
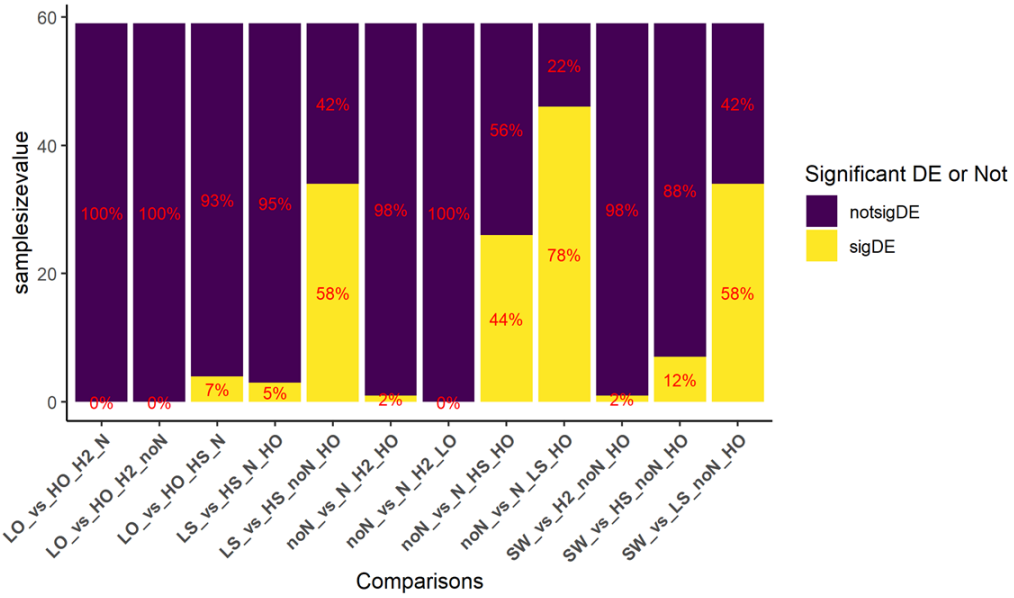


Figure 2: Percent SigDE vs notSigDE of FLAGs across comparisons

2A) Percent of Flagellar Assembly Proteins and 2B) Flagellar Chemotaxis proteins significantly differentially expressed across treatment comparisons. Treatment abbreviations: SW = Sea Water, LS = Low Sulfide, HS = High Sulfide, noN = no Nitrogen, HO = High Oxygen.

Percent SigDe vs notSigDE of Flagellar System Assembly Proteins



Percent SigDe vs notSigDE of Flagellar System Chemotaxis Proteins

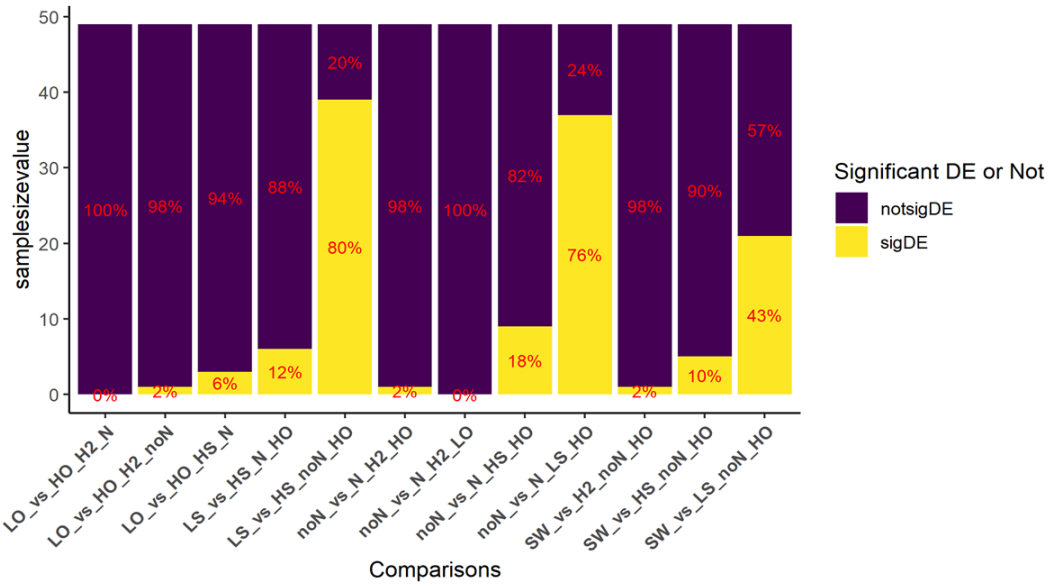
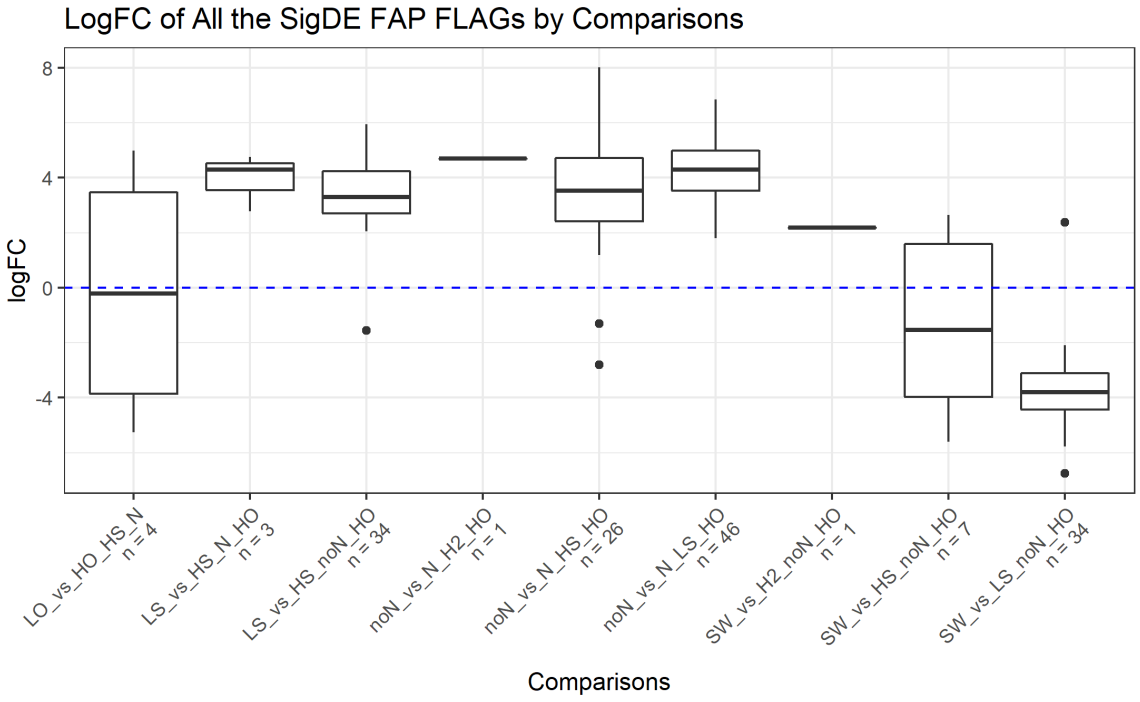


Figure 3: LogFC of all SigDE Flagellar proteins across comparisons

3A) logFC of Flagellar Assembly Proteins and 3B) logFC of Flagellar Chemotaxis Proteins 3C) logFC of both Flagellar Type IDs across chosen S comparisons. Treatment abbreviations: SW = Sea Water, LS = Low Sulfide, HS = High Sulfide, noN = no Nitrogen, HO = High Oxygen, H2 = Hydrogen treatment.



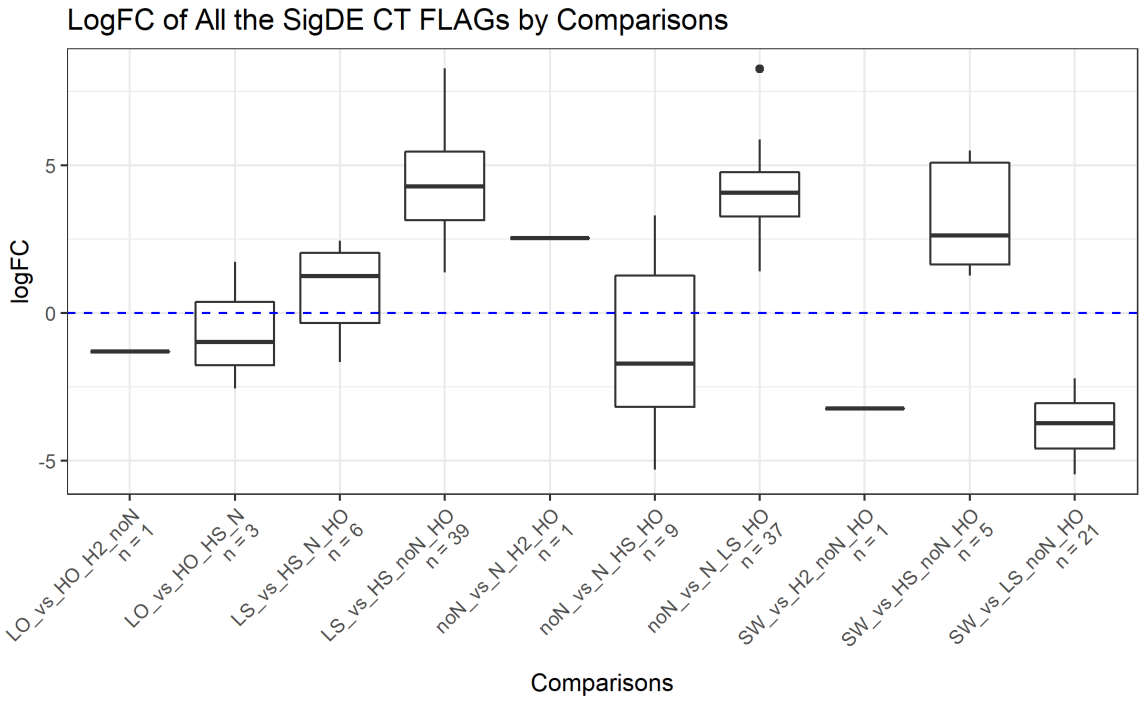
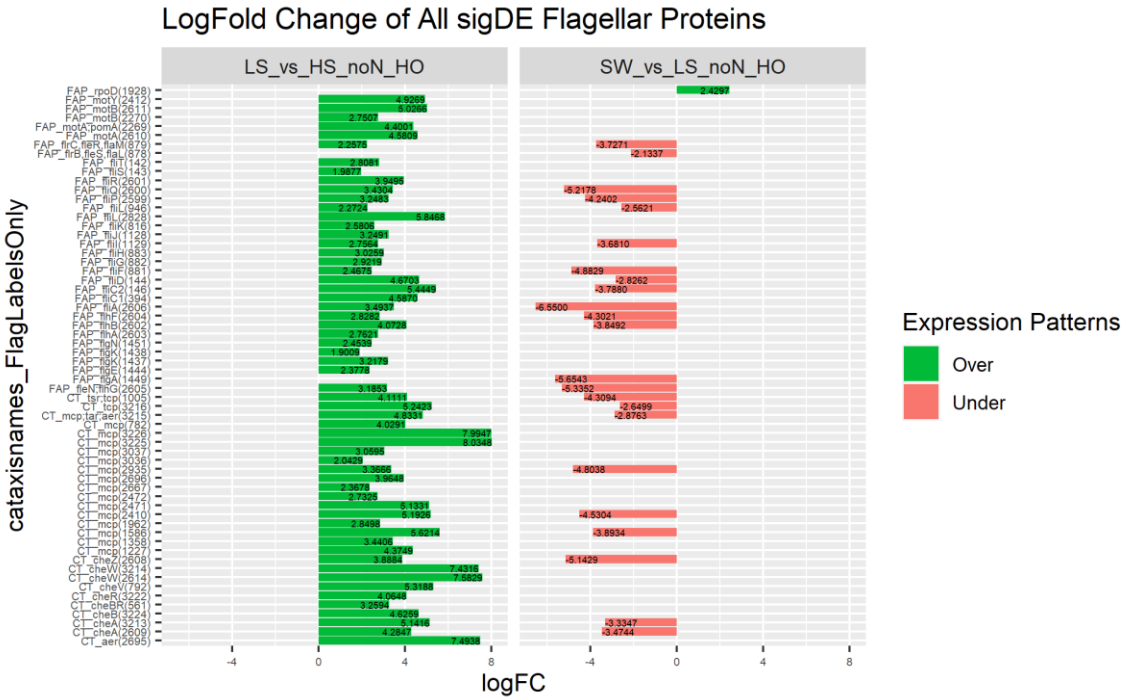
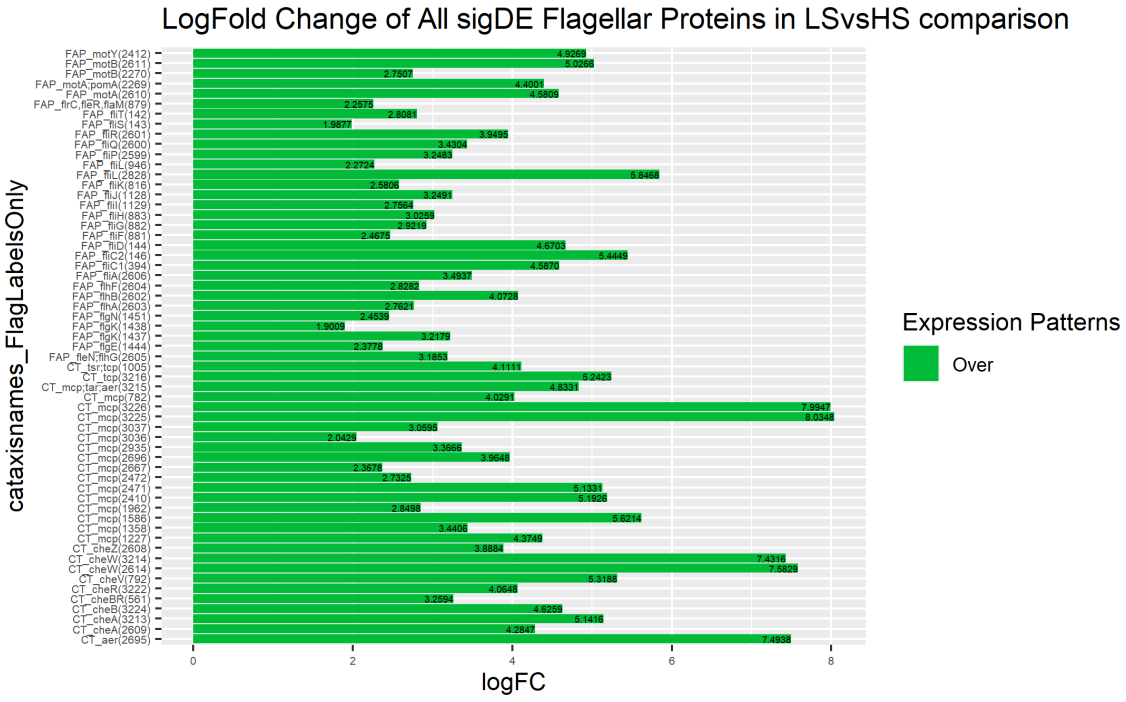


Figure 4: logFC of Individual Flagellum Components across comparisons

4A) LSvsHS_noN_HO faceted with SWvsLS_noN_HO, 4B) LSvsHS_noN_HO only, 4C) SWvsLS_noN_HO only. Treatment abbreviations: SW = Sea Water, LS = Low Sulfide, HS = High Sulfide, noN = no Nitrogen, HO = High Oxygen.


LogFold Change of All sigDE Flagellar Proteins
LS_vs_HS_noN_HO
SW_vs_LS_noN_HO
cataxisnames_FlagLabelsOnly
logFC
Expression Patterns
Over
Under


LogFold Change of All sigDE Flagellar Proteins in LSvsHS comparison
cataxisnames_FlagLabelsOnly
logFC
Expression Patterns
Over

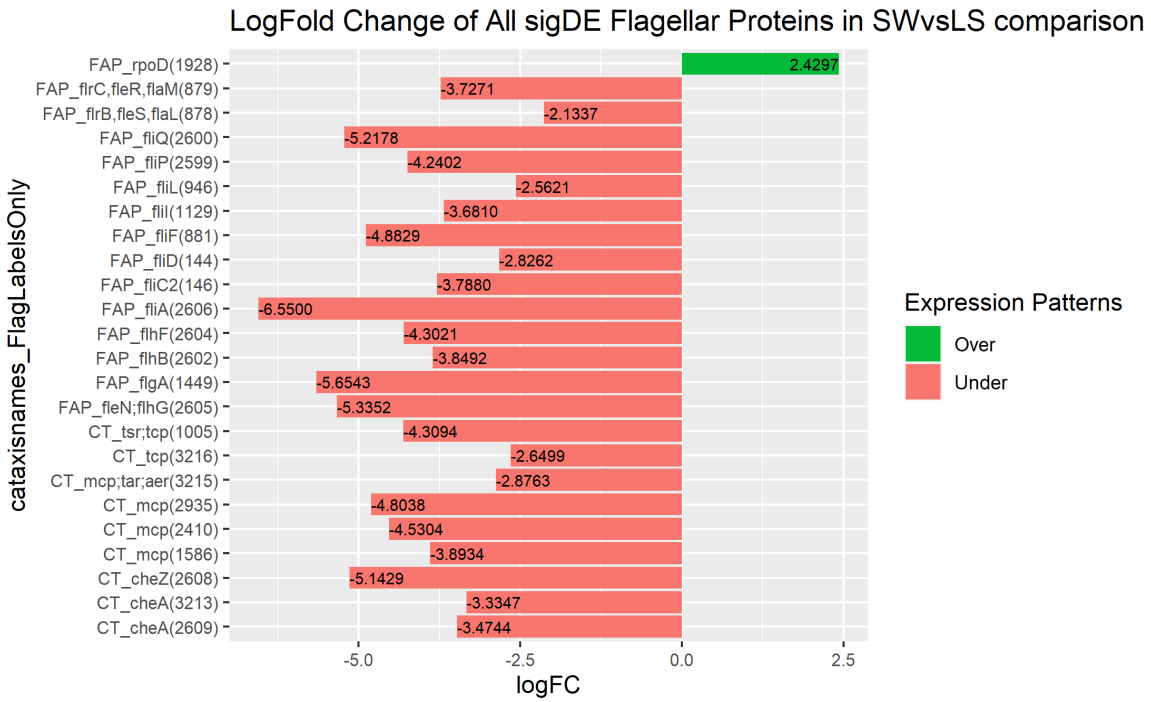
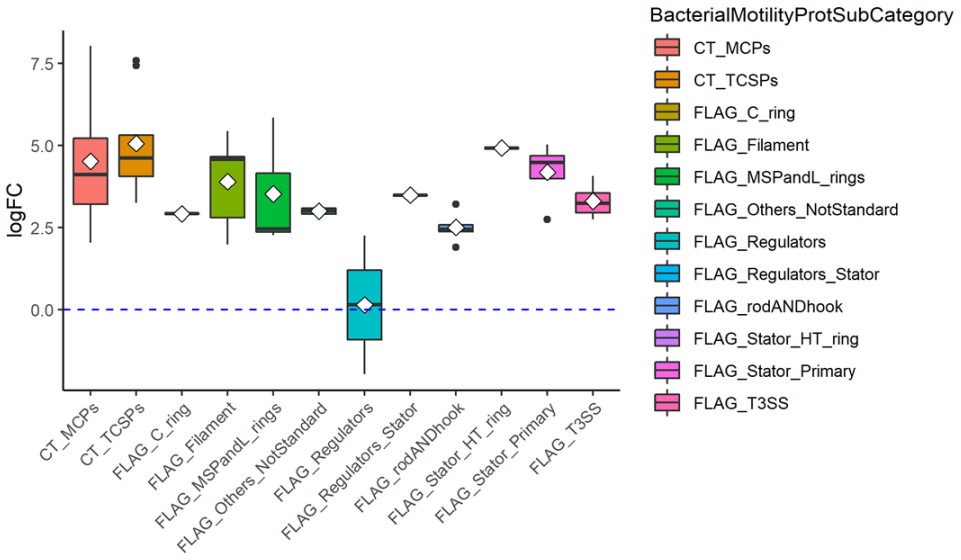


Figure 5: Flagellar Subsystem logFC across Comparisons

5A) LSvsHS_noN_HO 5B) SWvsLS_noN_HO. Groups are the Bacterial Motility Sub Category in KEGG:Brite. Specifically ko02035Bacterial motility proteins>Flagellar System: Chemotaxis(CT) and Flagellar Assembly Protein(FAP) Sub Categories.

LS_vs_HS_noN_HO logFC Trends of only SigDE Flagellar Genes, by SubCategory



SW_vs_LS_noN_HO logFC Trends of only SigDE Flagellar Genes, by SubCategory

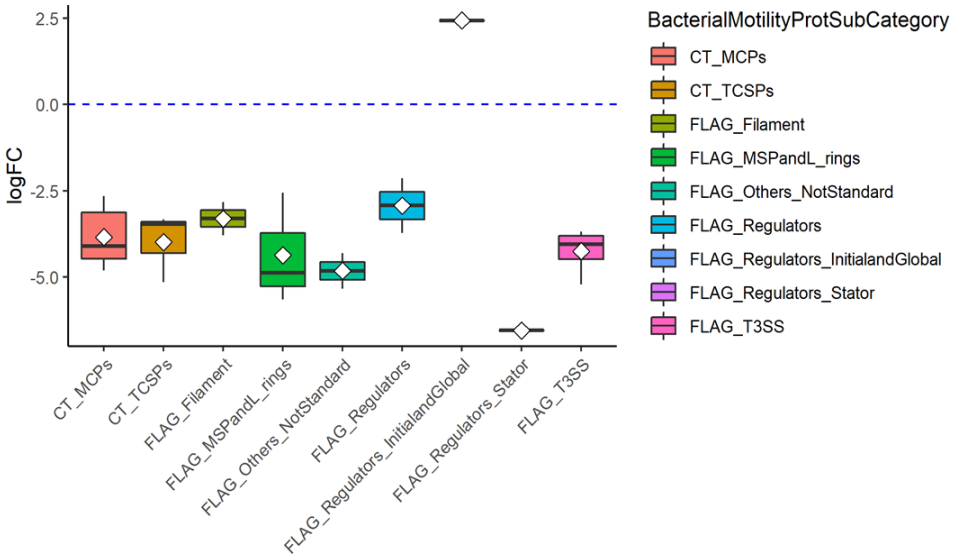


Figure 6: Candidate Selection Workflow

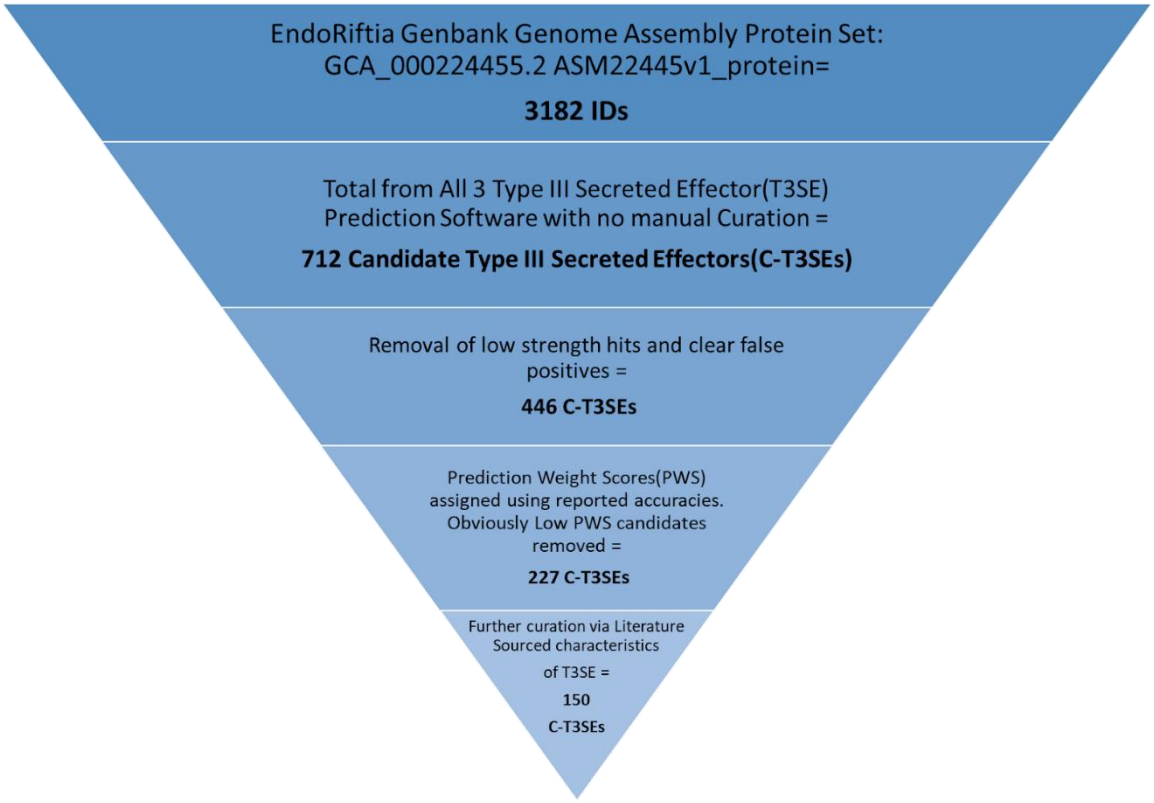
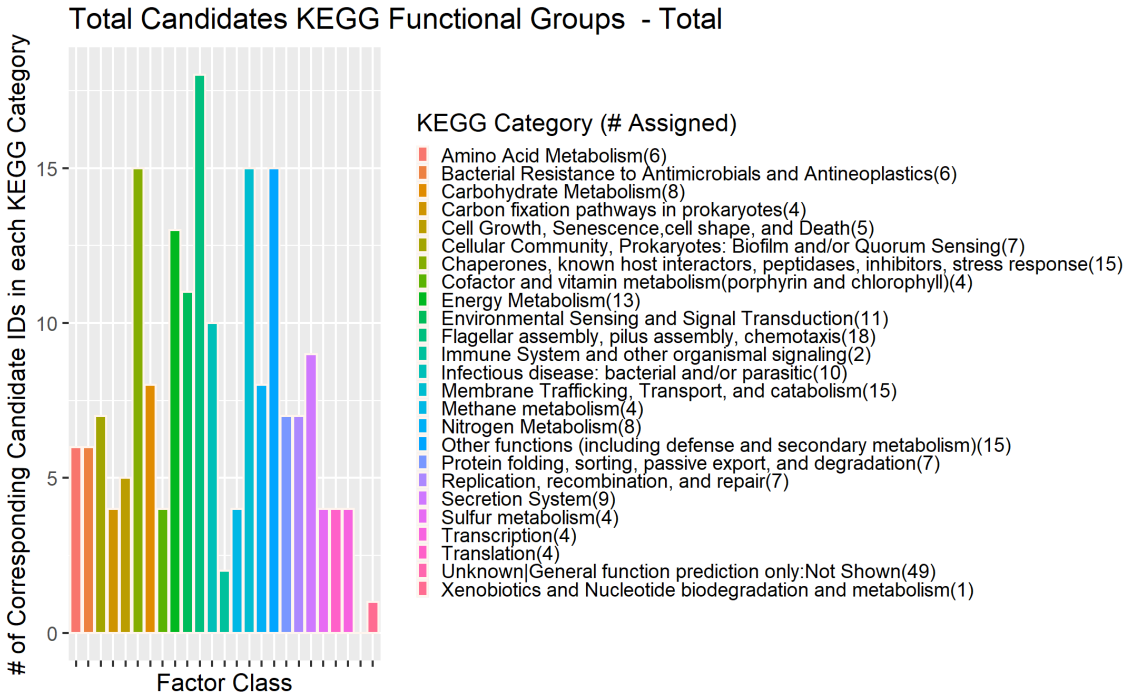
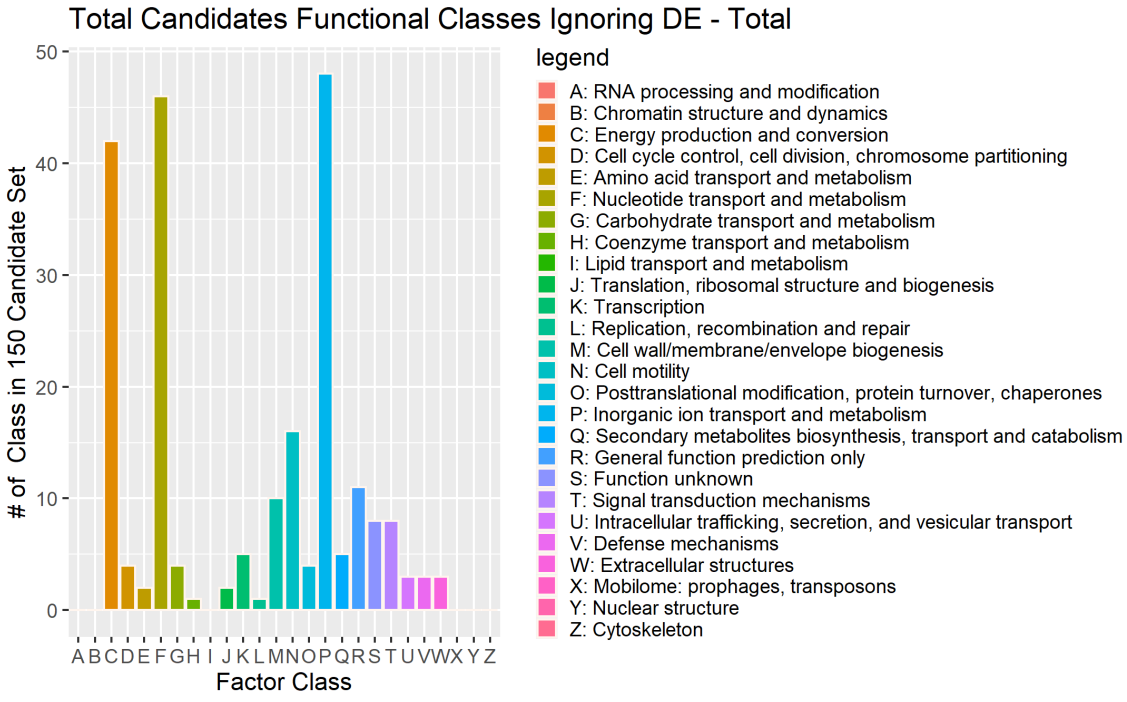


Figure 7: KEGG and COG assignments to Candidates (150 set)

7A) KEGG Assignments in Candidates, 7B) COG assignments in Candidates


Total Candidates KEGG Functional Groups - Total
# of Corresponding Candidate IDs in each KEGG Category
0
5
10
15
Factor Class
KEGG Category (# Assigned)
Amino Acid Metabolism(6)
Bacterial Resistance to Antimicrobials and Antineoplastics(6)
Carbohydrate Metabolism(8)
Carbon fixation pathways in prokaryotes(4)
Cell Growth, Senescence,cell shape, and Death(5)
Cellular Community, Prokaryotes: Biofilm and/or Quorum Sensing(7)
Chaperones, known host interactors, peptidases, inhibitors, stress response(15)
Cofactor and vitamin metabolism(porphyrin and chlorophyll)(4)
Energy Metabolism(13)
Environmental Sensing and Signal Transduction(11)
Flagellar assembly, pilus assembly, chemotaxis(18)
Immune System and other organismal signaling(2)
Infectious disease: bacterial and/or parasitic(10)
Membrane Trafficking, Transport, and catabolism(15)
Methane metabolism(4)
Nitrogen Metabolism(8)
Other functions (including defense and secondary metabolism)(15)
Protein folding, sorting, passive export, and degradation(7)
Replication, recombination, and repair(7)
Secretion System(9)
Sulfur metabolism(4)
Transcription(4)
Translation(4)
Unknown|General function prediction only:Not Shown(49)
Xenobiotics and Nucleotide biodegradation and metabolism(1)


Total Candidates Functional Classes Ignoring DE - Total
# of Class in 150 Candidate Set
0
10
20
30
40
50
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Factor Class
legend
A: RNA processing and modification
B: Chromatin structure and dynamics
C: Energy production and conversion
D: Cell cycle control, cell division, chromosome partitioning
E: Amino acid transport and metabolism
F: Nucleotide transport and metabolism
G: Carbohydrate transport and metabolism
H: Coenzyme transport and metabolism
I: Lipid transport and metabolism
J: Translation, ribosomal structure and biogenesis
K: Transcription
L: Replication, recombination and repair
M: Cell wall/membrane/envelope biogenesis
N: Cell motility
O: Posttranslational modification, protein turnover, chaperones
P: Inorganic ion transport and metabolism
Q: Secondary metabolites biosynthesis, transport and catabolism
R: General function prediction only
S: Function unknown
T: Signal transduction mechanisms
U: Intracellular trafficking, secretion, and vesicular transport
V: Defense mechanisms
W: Extracellular structures
X: Mobilome: prophages, transposons
Y: Nuclear structure
Z: Cytoskeleton

Appendix

Terms

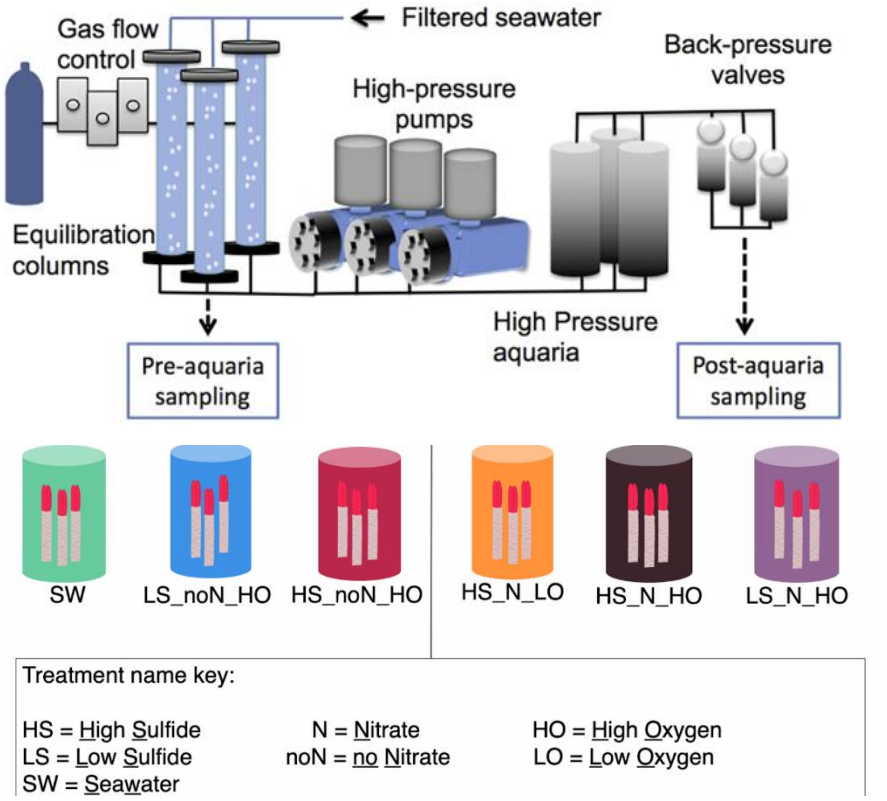
- F-T3SS|NF-T3SS: flagellar(F) and non-flagellar(NF) type III secretion systems, are two distinct variations on the bacterial type III secretion system which function as the export apparatus structures of the flagellum and injectosome systems respectively.
- Bacteriocyte: Specialized host cells that contain endosymbiotic organisms such as bacteria and fungi, which provide essential amino acids and other chemicals to their host.
- Vertical Transmission (Symbiosis): Vertical transmission of symbionts is the transfer of a microbial symbiont from the parent directly to the offspring
- Horizontal Transmission (Symbiosis): Horizontally transmitted symbionts are taken up from the environment anew by each host generation
- Obligate Relationship/Partner (Symbiosis): Obligate symbiosis suggests that at least one of the partners cannot complete its lifecycle on its own
- Facultative Relationship/Partner (Symbiosis): facultative symbiosis suggests that partners do not necessarily require one another to survive and/or reproduce
- Next-Generation Sequencing: Massive parallel sequencing or massively parallel sequencing is any of several high-throughput approaches to DNA sequencing using the concept of massively parallel processing; it is also called next-generation sequencing or second-generation sequencing. The rise of Next-

generation sequencing makes large-scale whole-genome sequencing (WGS) accessible and practical for the average researcher.

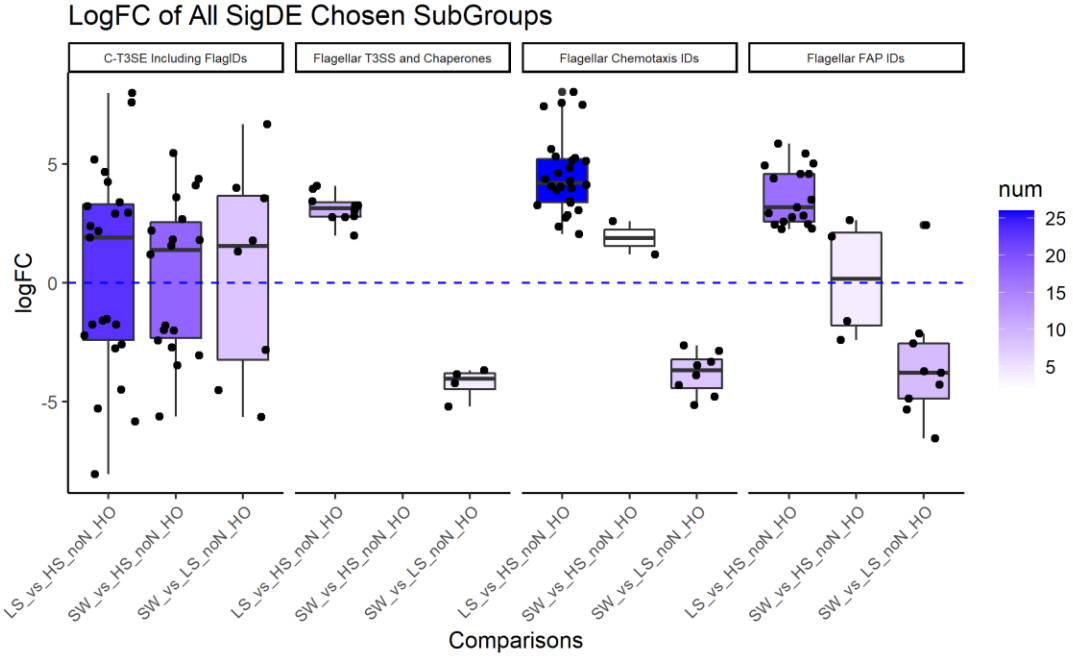
- Persistence (Symbiosis): continued capacity of an organism to form a particular symbiosis over evolutionary time
- Maintenance (Symbiosis): Mechanisms used in the process of maintaining or preserving the symbiotic association status
- Effectors & Substrates: Generally, Proteins secreted by secretion systems are globally known as 'substrates', but if the substrate mimics a host-cell function, like those from T3SS, T4SS, and T6SS, it is instead referred to as an 'effector'
- Parasitism and Pathogenicity: parasitism is intimately associated with pathogenicity since the ability of the parasite to invade and become established in the host generally results in the development of a diseased condition in the host

Supplemental Figures

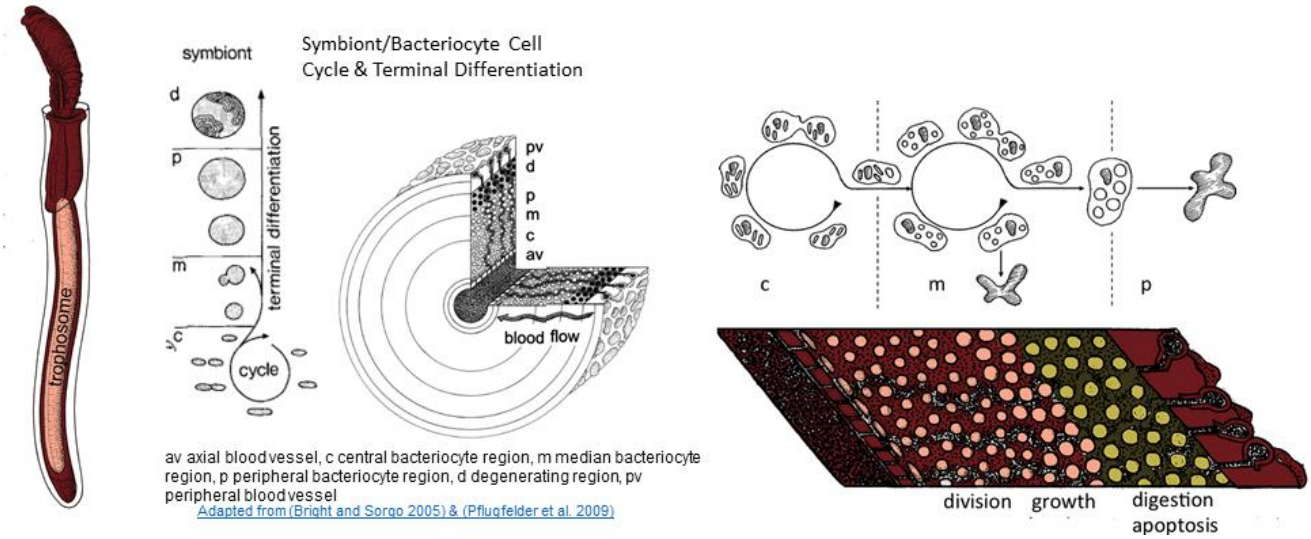
Supplemental Figure 1: HPRS system and Treatments(J. H. Mitchell et al., 2019)



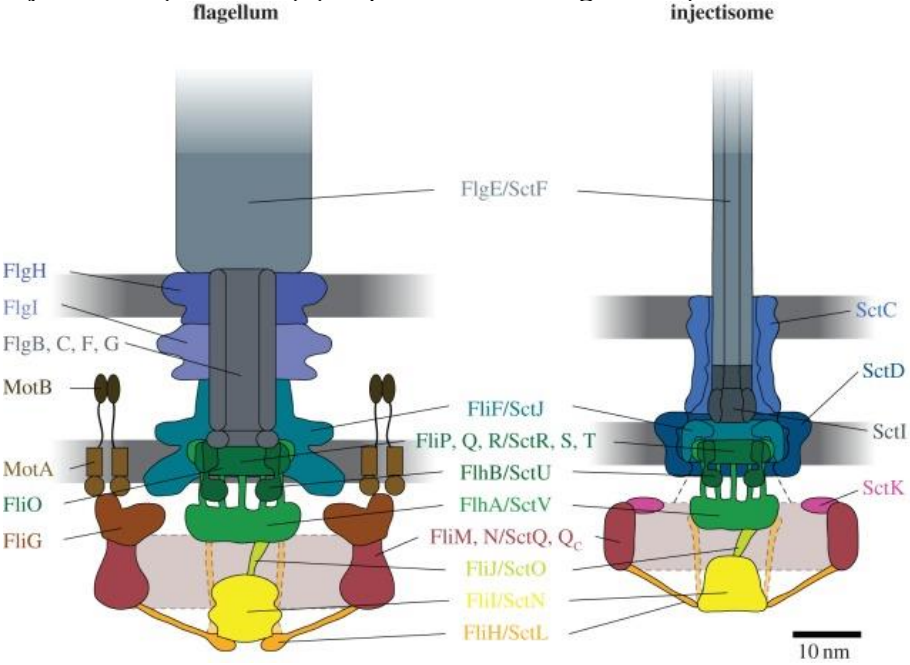
Supplemental Figure 2: LogFC patterns of All SigDe Chosen Subgroups



Supplemental Figure 3: *Riftia* Anatomy Digestion Overview



Supplemental Figure 4: Bacterial Type III Secretion Machinery- Flagellum(F-T3SS) vs Injectosome(NF-T3SS) (Diepold and Armitage, 2015)



Supplemental Figure 5: MacSyFinder:TXSScan: T3SS Methodology(S5A) and Endoriftia Results (S5B)

Supplemental 5A.): "The models of T3SS and flagellum were built based on a previous study(Abby 2014). Of the nine mandatory components for the T3SS, only the secretin is forbidden in the model of the flagellum. Conversely, three flagellum-specific components are forbidden in the T3SS model. Some genes are ubiquitous and specific to a system and can be defined as forbidden in models of other systems. This facilitates the discrimination between systems with homologous components. For example, the NF-T3SS-specific secretin may be declared as forbidden in the F-T3SS" (Abby 2016)

Supplemental 5B.):

F-T3SS Mandatory genes:	
Flg_sctJ_FLG	1
Flg_sctS_FLG	1
Flg_sctN_FLG	5
Flg_flgB	1
Flg_sctQ_FLG	(1/2)
Flg_sctT_FLG	1
Flg_sctU_FLG	1
Flg_fliE	1
Flg_sctR_FLG	1
Flg_flgC	2
Flg_sctV_FLG	1
Forbidden genes:	
T3SS_sctC	0

NF-T3SS Mandatory genes:	
T3SS_sctC	0
T3SS_sctJ	0
T3SS_sctN	3
T3SS_sctS	0
T3SS_sctR	0
T3SS_sctQ	0
T3SS_sctV	0
T3SS_sctU	0
T3SS_sctT	0
Forbidden genes:	
Flg_flgC	2
Flg_fliE	1
Flg_flgB	1

References

Abby, S. S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., & Rocha, E. P. C. (2016). Identification of protein secretion systems in bacterial genomes. *Scientific Reports, 6*(1), 23080. doi:10.1038/srep23080

Abby, S. S., & Rocha, E. P. C. (2012). The Non-Flagellar Type III Secretion System Evolved from the Bacterial Flagellum and Diversified into Host-Cell Adapted Systems. *PLoS Genetics, 8*(9), e1002983. doi:10.1371/journal.pgen.1002983

Abby, S. S., & Rocha, E. P. C. (2017). Identification of Protein Secretion Systems in Bacterial Genomes Using MacSyFinder. In *Methods in Molecular Biology* (pp. 1-21): Springer New York.

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., . . . Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology, 37*(4), 420-423. doi:10.1038/s41587-019-0036-z

An, Y., Wang, J., Li, C., Leier, A., Marquez-Lago, T., Wilksch, J., . . . Lithgow, T. (2016). Comprehensive assessment and performance improvement of effector protein predictors for bacterial secretion systems III, IV and VI. *Briefings in Bioinformatics*, bbw100. doi:10.1093/bib/bbw100

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*(7), 2251-2252. doi:10.1093/bioinformatics/btz859

Arnold, R., Jehl, A., & Rattei, T. (2010). Targeting effectors: the molecular recognition of Type III secreted proteins. *Microbes and Infection, 12*(5), 346-358. doi:10.1016/j.micinf.2010.02.003

Aschtgen, M.-S., Brennan, C. A., Nikolakakis, K., Cohen, S., Mcfall-Ngai, M., & Ruby, E. G. (2019). Insights into flagellar function and mechanism from the squid–vibrio symbiosis. *npj Biofilms and Microbiomes, 5*(1). doi:10.1038/s41522-019-0106-5

Belas, R. (2014). Biofilms, flagella, and mechanosensing of surfaces by bacteria. *Trends in Microbiology, 22*(9), 517-527. doi:10.1016/j.tim.2014.05.002

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical*

Society: Series B (Methodological), 57(1), 289-300. doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bliska, J. B., Wang, X., Viboud, G. I., & Brodsky, I. E. (2013). Modulation of innate immune responses byYersiniatype III secretion system translocators and effectors. *Cellular Microbiology*, n/a-n/a. doi:10.1111/cmi.12164

Bright, M., & Bulgheresi, S. (2010). A complex journey: transmission of microbial symbionts. *Nature Reviews Microbiology, 8*(3), 218-230. doi:10.1038/nrmicro2262

Bright, M., & Lallier, F. (2010). THE BIOLOGY OF VESTIMENTIFERAN TUBEWORMS. In Gibson, Rn, Atkinson, Rja, Gordon, & Jdm (Eds.), *OCEANOGRAPHY AND MARINE BIOLOGY: AN ANNUAL REVIEW, VOL 48* (Vol. 48, pp. 213-265): CRC PRESS-TAYLOR & FRANCIS GROUP.

Chaban, B., Hughes, H. V., & Beeby, M. (2015). The flagellum in bacterial pathogens: For motility and a whole lot more. *Seminars in Cell & Developmental Biology, 46*, 91-103. doi:10.1016/j.semcdb.2015.10.032

Chen, Y., Lun, A. T. L., & Smyth, G. K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research, 5*, 1438. doi:10.12688/f1000research.8987.2

Chong, R. A., Park, H., & Moran, N. A. (2019). Genome Evolution of the Obligate Endosymbiont Buchnera aphidicola. *Molecular Biology and Evolution, 36*(7), 1481-1489. doi:10.1093/molbev/msz082

D'Haeze, W., & Holsters, M. (2004). Surface polysaccharides enable bacteria to evade plant immunity. *Trends in Microbiology, 12*(12), 555-561. doi:10.1016/j.tim.2004.10.009

Dalio, R. J. D., Herlihy, J., Oliveira, T. S., Mcdowell, J. M., & Machado, M. (2018). Effector Biology in Focus: A Primer for Computational Prediction and Functional Characterization. *Molecular Plant-Microbe Interactions®, 31*(1), 22-33. doi:10.1094/mpmi-07-17-0174-fi

De Maayer, P., Pillay, T., & Coutinho, T. A. (2020). Flagella by numbers: comparative genomic analysis of the supernumerary flagellar systems among the Enterobacterales. *BMC Genomics, 21*(1). doi:10.1186/s12864-020-07085-w

Dean, P. (2011). Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiology Reviews, 35*(6), 1100-1125. doi:10.1111/j.1574-6976.2011.00271.x

deBary, A. (1879). *Die Erscheinung der Symbiose : Vortrag*. Strassburg: Verlag von Karl J. Trübner.

Deng, W., Marshall, N. C., Rowland, J. L., Mccoy, J. M., Worrall, L. J., Santos, A. S., . . . Finlay, B. B. (2017). Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology, 15*(6), 323-337. doi:10.1038/nrmicro.2017.20

Diepold, A., & Armitage, J. P. (2015). Type III secretion systems: the bacterial flagellum and the injectisome. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1679), 20150020. doi:10.1098/rstb.2015.0020

Dongre, M., Singh, B., Aung, K. M., Larsson, P., Miftakhova, R., Persson, K., . . . Wai, S. N. (2018). Flagella-mediated secretion of a novel Vibrio cholerae cytotoxin affecting both vertebrate and invertebrate hosts. *Communications Biology, 1*(1). doi:10.1038/s42003-018-0065-z

Du, J., Reeves, A. Z., Klein, J. A., Twedt, D. J., Knodler, L. A., & Lesser, C. F. (2016). The type III secretion system apparatus determines the intracellular niche of bacterial pathogens. *Proceedings of the National Academy of Sciences, 113*(17), 4794-4799. doi:10.1073/pnas.1520699113

Echazarreta, M. A., & Klose, K. E. (2019). Vibrio Flagellar Synthesis. *Frontiers in Cellular and Infection Microbiology, 9.* doi:10.3389/fcimb.2019.00131

Eichinger, V., Nussbaumer, T., Platzer, A., Jehl, M.-A., Arnold, R., & Rattei, T. (2016). EffectiveDB—updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Research, 44*(D1), D669-D674. doi:10.1093/nar/gkv1269

Frank, A. C. (2019). Molecular host mimicry and manipulation in bacterial symbionts. *FEMS Microbiology Letters, 366*(4). doi:10.1093/femsle/fnz038

Fukumura, T., Makino, F., Dietsche, T., Kinoshita, M., Kato, T., Wagner, S., . . . Minamino, T. (2017). Assembly and stoichiometry of the core structure of the bacterial flagellar type III export gate complex. *PLOS Biology, 15*(8), e2002281. doi:10.1371/journal.pbio.2002281

Gardebrecht, A., Markert, S., Sievert, S. M., Felbeck, H., Thürmer, A., Albrecht, D., . . . Schweder, T. (2012). Physiological homogeneity among the endosymbionts of Riftia pachyptila and Tevnia jerichonana revealed by proteogenomics. *The ISME Journal, 6*(4), 766-776. doi:10.1038/ismej.2011.137

Ghelardi, E., Celandroni, F., Salvetti, S., Beecher, D. J., Gominet, M., Lereclus, D., . . . Senesi, S. (2002). Requirement of flhA for Swarming Differentiation, Flagellin Export, and Secretion of Virulence-Associated Proteins in Bacillus thuringiensis. *Journal of Bacteriology, 184*(23), 6424-6433. doi:10.1128/jb.184.23.6424-6433.2002

Goldberg, T., Rost, B., & Bromberg, Y. (2016). Computational prediction shines light on type III secretion origins. *Scientific Reports, 6*(1), 34516. doi:10.1038/srep34516

Green, C. A., Kamble, N. S., Court, E. K., Bryant, O. J., Hicks, M. G., Lennon, C., . . . Stafford, G. P. (2019). Engineering the flagellar type III secretion system: improving capacity for secretion of recombinant protein. *Microbial Cell Factories, 18*(1). doi:10.1186/s12934-019-1058-4

Harmer, T. L., Rotjan, R. D., Nussbaumer, A. D., Bright, M., Ng, A. W., Dechaine, E. G., & Cavanaugh, C. M. (2008). Free-Living Tube Worm Endosymbionts Found at Deep-Sea Vents. *Applied and Environmental Microbiology, 74*(12), 3895-3898. doi:10.1128/aem.02470-07

Hentschel, U., Steinert, M., & Hacker, J. (2000). Common molecular mechanisms of symbiosis and pathogenesis. *Trends in Microbiology, 8*(5), 226-231. doi:10.1016/s0966-842x(00)01758-3

Hinzke, T., Kleiner, M., Breusing, C., Felbeck, H., Häsler, R., Sievert, S. M., . . . Markert, S. (2019). Host-Microbe Interactions in the Chemosynthetic Riftia pachyptila Symbiosis. *mBio, 10*(6). doi:10.1128/mbio.02243-19

Hinzke, T., Kleiner, M., Meister, M., Schlüter, R., Hentschker, C., Pané-Farré, J., . . . Markert, S. (2020). *Metabolic differences between symbiont subpopulations in the deep-sea tubeworm Riftia pachyptila*. Cold Spring Harbor Laboratory.

Hobbs, C. K., Porter, V. L., Stow, M. L. S., Siame, B. A., Tsang, H. H., & Leung, K. Y. (2016). Computational approach to predict species-specific type III secretion system (T3SS) effectors using single and multiple genomes. *BMC Genomics, 17*(1). doi:10.1186/s12864-016-3363-1

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., . . . Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research, 47*(D1), D309-D314. doi:10.1093/nar/gky1085

Hui, X., Chen, Z., Lin, M., Zhang, J., Hu, Y., Zeng, Y., . . . Wang, Y. (2020). T3SEpp: an Integrated Prediction Pipeline for Bacterial Type III Secreted Effectors. *mSystems, 5*(4). doi:10.1128/msystems.00288-20

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Research, 49*(D1), D545-D551. doi:10.1093/nar/gkaa970

Kanehisa, M., & Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Science, 29*(1), 28-35. doi:10.1002/pro.3711

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology, 428*(4), 726-731. doi:10.1016/j.jmb.2015.11.006

Klose, J., Polz, M. F., Wagner, M., Schimak, M. P., Gollner, S., & Bright, M. (2015). Endosymbionts escape dead hydrothermal vent tubeworms to enrich the free-living population. *Proceedings of the National Academy of Sciences, 112*(36), 11300-11305. doi:10.1073/pnas.1501160112

Konkel, M. E., Klena, J. D., Rivera-Amill, V., Monteville, M. R., Biswas, D., Raphael, B., & Mickelson, J. (2004). Secretion of Virulence Proteins from Campylobacter jejuni Is Dependent on a Functional Flagellar Export Apparatus. *Journal of Bacteriology, 186*(11), 3296-3303. doi:10.1128/jb.186.11.3296-3303.2004

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *Journal of Molecular Biology, 305*(3), 567-580. doi:10.1006/jmbi.2000.4315

Kuhlen, L., Abrusci, P., Johnson, S., Gault, J., Deme, J., Caesar, J., . . . Lea, S. M. (2018). Structure of the core of the type III secretion system export apparatus. *Nature Structural & Molecular Biology, 25*(7), 583-590. doi:10.1038/s41594-018-0086-9

Law, C. W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G. K., & Ritchie, M. E. (2018). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research, 5*, 1408. doi:10.12688/f1000research.9005.3

Leid, J. G., Kerr, M., Selgado, C., Johnson, C., Moreno, G., Smith, A., . . . Cope, E. K. (2009). Flagellum-Mediated Biofilm Defense Mechanisms of Pseudomonas aeruginosa against Host-Derived Lactoferrin. *Infection and Immunity, 77*(10), 4559-4566. doi:10.1128/iai.00075-09

Li, B., Yue, Y., Yuan, Z., Zhang, F., Li, P., Song, N., . . . Gu, L. (2017). Salmonella STM1697 coordinates flagella biogenesis and virulence by restricting flagellar master protein FlhD4C2 from recruiting RNA polymerase. *Nucleic Acids Research, 45*(17), 9976-9989. doi:10.1093/nar/gkx656

Li, Y., Liles, M. R., & Halanych, K. M. (2018). Endosymbiont genomes yield clues of tubeworm success. *The ISME Journal, 12*(11), 2785-2795. doi:10.1038/s41396-018-0220-z

Liu, R., Holik, A. Z., Su, S., Jansz, N., Chen, K., Leong, H. S., . . . Ritchie, M. E. (2015). Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Research, 43*(15), e97-e97. doi:10.1093/nar/gkv412

Liu, R., & Ochman, H. (2007). Stepwise formation of the bacterial flagellar system. *Proceedings of the National Academy of Sciences, 104*(17), 7116-7121. doi:10.1073/pnas.0700266104

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., . . . Marchler-Bauer, A. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research, 48*(D1), D265-D268. doi:10.1093/nar/gkz991

Maezawa, K., Shigenobu, S., Taniguchi, H., Kubo, T., Aizawa, S.-I., & Morioka, M. (2006). Hundreds of Flagellar Basal Bodies Cover the Cell Surface of the Endosymbiotic Bacterium Buchnera aphidicola sp. Strain APS. *Journal of Bacteriology, 188*(18), 6539-6543. doi:10.1128/jb.00561-06

Martínez-García, P. M., Ramos, C., & Rodríguez-Palenzuela, P. (2015). T346Hunter: A Novel Web-Based Tool for the Prediction of Type III, Type IV and Type VI Secretion Systems in Bacterial Genomes. *PLoS ONE, 10*(4), e0119317. doi:10.1371/journal.pone.0119317

Mccarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research, 40*(10), 4288-4297. doi:10.1093/nar/gks042

Mcfall-Ngai, M. (2008). Are biologists in 'future shock'? Symbiosis integrates biology across domains. *Nature Reviews Microbiology, 6*(10), 789-792. doi:10.1038/nrmicro1982

Millikan, D. S., Felbeck, H., & Stein, J. L. (1999). Identification and Characterization of a Flagellin Gene from the Endosymbiont of the Hydrothermal Vent TubewormRiftia pachyptila. *Applied and Environmental Microbiology, 65*(7), 3129-3133. doi:10.1128/aem.65.7.3129-3133.1999

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., . . . Finn, R. D. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research, 47*(D1), D351-D360. doi:10.1093/nar/gky1100

Mitchell, J. H., Leonard, J. M., Delaney, J., Girguis, P. R., & Scott, K. M. (2019). Hydrogen Does Not Appear To Be a Major Electron Donor for Symbiosis with the Deep-Sea Hydrothermal Vent Tubeworm Riftia pachyptila. *Applied and Environmental Microbiology, 86*(1). doi:10.1128/aem.01522-19

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research, 35*(Web Server), W182-W185. doi:10.1093/nar/gkm321

Mushegian, A. A., & Ebert, D. (2016). Rethinking “mutualism” in diverse host-symbiont communities. *BioEssays, 38*(1), 100-108. doi:10.1002/bies.201500074

Nakamura, S., & Minamino, T. (2019). Flagella-Driven Motility of Bacteria. *Biomolecules, 9*(7), 279. doi:10.3390/biom9070279

Park, D., & Forst, S. (2006). Co-regulation of motility, exoenzyme and antibiotic production by the EnvZ-OmpR-FlhDC-FliA pathway in Xenorhabdus nematophila. *Molecular Microbiology, 61*(6), 1397-1412. doi:10.1111/j.1365-2958.2006.05320.x

Pérez-Brocal, V., Latorre, A., & Moya, A. (2011). Symbionts and Pathogens: What is the Difference? In *Between Pathogenicity and Commensalism* (pp. 215-243): Springer Berlin Heidelberg.

Perez, M., & Juniper, S. K. (2016). Insights into Symbiont Population Structure among Three Vestimentiferan Tubeworm Host Species at Eastern Pacific Spreading Centers. *Applied and Environmental Microbiology, 82*(17), 5197-5205. doi:10.1128/aem.00953-16

Poly, F. D. R., Ewing, C., Goon, S., Hickey, T. E., Rockabrand, D., Majam, G., . . . Guerry, P. (2007). Heterogeneity of a Campylobacter jejuni Protein That Is Secreted through the Flagellar Filament. *Infection and Immunity, 75*(8), 3859-3867. doi:10.1128/iai.00159-07

Pundhir, S., & Kumar, A. (2011). SSPred: A prediction server based on SVM for the identification and classification of proteins involved in bacterial secretion systems. *Bioinformation, 6*(10), 380-382. doi:10.6026/97320630006380

Raina, J.-B., Fernandez, V., Lambert, B., Stocker, R., & Seymour, J. R. (2019). The role of microbial motility and chemotaxis in symbiosis. *Nature Reviews Microbiology, 17*(5), 284-294. doi:10.1038/s41579-019-0182-9

Raymond, B., Young, J. C., Pallett, M., Endres, R. G., Clements, A., & Frankel, G. (2013). Subversion of trafficking, apoptosis, and innate immunity by type III secretion system effectors. *Trends in Microbiology, 21*(8), 430-441. doi:10.1016/j.tim.2013.06.008

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research, 43*(7), e47-e47. doi:10.1093/nar/gkv007

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology, 11*(3), R25. doi:10.1186/gb-2010-11-3-r25

Sapp, J. (2002). *Evolution by association : a history of symbiosis*. New York, New York: Oxford University Press.

Schepers, M. J., Yelland, J. N., Moran, N. A., & Taylor, D. W. (2021). *Isolation of the Buchnera aphidicola flagellum basal body from the Buchnera membrane*. Cold Spring Harbor Laboratory.

Singh, P., Park, D., Forst, S., & Banerjee, N. (2013). Xenocin Export by the Flagellar Type III Pathway in Xenorhabdus nematophila. *Journal of Bacteriology, 195*(7), 1400-1410. doi:10.1128/jb.01532-12

Terashima, H., Kojima, S., & Homma, M. (2008). Chapter 2 Flagellar Motility in Bacteria. In *International Review of Cell and Molecular Biology* (pp. 39-85): Elsevier.

Toft, C., & Fares, M. A. (2008). The Evolution of the Flagellar Assembly Pathway in Endosymbiotic Bacterial Genomes. *Molecular Biology and Evolution, 25*(9), 2069-2076. doi:10.1093/molbev/msn153

Wagner, S., Grin, I., Malmsheimer, S., Singh, N., Torres-Vargas, C. E., & Westerhausen, S. (2018). Bacterial type III secretion systems: a complex device for the delivery of bacterial effector proteins into eukaryotic host cells. *FEMS Microbiology Letters, 365*(19). doi:10.1093/femsle/fny201

Wang, J., Li, J., Yang, B., Xie, R., Marquez-Lago, T. T., Leier, A., . . . Lithgow, T. (2019). Bastion3: a two-layer ensemble predictor of type III secreted effectors. *Bioinformatics, 35*(12), 2017-2028. doi:10.1093/bioinformatics/bty914

Wang, Y., Huang, H., Sun, M. A., Zhang, Q., & Guo, D. (2012). T3DB: an integrated database for bacterial type III secretion system. *BMC Bioinformatics, 13*(1), 66. doi:10.1186/1471-2105-13-66

Webster, N. S. (2014). Cooperation, communication, and co-evolution: grand challenges in microbial symbiosis research. *Frontiers in Microbiology, 5*. doi:10.3389/fmicb.2014.00164

Wiles, T. J., & Guillemin, K. (2019). The Other Side of the Coin: What Beneficial Microbes Can Teach Us about Pathogenic Potential. *Journal of Molecular Biology, 431*(16), 2946-2956. doi:10.1016/j.jmb.2019.05.001

Young, G. M., Schmiel, D. H., & Miller, V. L. (1999). A new pathway for the secretion of virulence factors by bacteria: The flagellar export apparatus functions as a protein-secretion system. *Proceedings of the National Academy of Sciences, 96*(11), 6456-6461. doi:10.1073/pnas.96.11.6456

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., . . . Brinkman, F. S. L. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics, 26*(13), 1608-1615. doi:10.1093/bioinformatics/btq249