8-13-2014

Proteomic Characterization of Host-Microbe Interactions in the *Euprymna scolopes* Light Organ Symbiosis with *Vibrio fischeri*

Tyler R. Schleicher
University of Connecticut - Storrs, tyler.schleicher@gmail.com

Proteomic Characterization of Host-Microbe Interactions in the *Euprymna scolopes* Light Organ Symbiosis with *Vibrio fischeri*

Tyler R. Schleicher, Ph.D.
University of Connecticut, 2014

The squid, *Euprymna scolopes*, and the bacterium, *Vibrio fischeri*, form a highly specific relationship that is used to study how beneficial bacteria interact with animal hosts. Even though the light organ of *E. scolopes* is exposed to the outside marine environment, it can only be colonized by *V. fischeri*. For this reason *E. scolopes* must carefully control the light organ microenvironment. One mechanism that the squid utilizes to regulate this symbiosis is a daily expulsion of *V. fischeri* from the light organ. The squid expels the symbiont, in addition to some host tissue, as a thick exudate that was collected and analyzed by high-throughput mass spectrometry (MS)-based proteomics. As expected, many of the *V. fischeri* proteins identified exemplify the lifestyle of the symbiont in the light organ including quorum sensing and the management of oxidative stress. The function of one of the *V. fischeri* proteins identified, a bacterial immunoglobulin-like surface protein, was further characterized by creating a mutant deficient in this protein. Host proteins within the light organ also highlighted the oxidative microenvironment, but proteins associated with innate immunity were prominent as well. The innate immune response, including macrophage-like hemocytes, of *E. scolopes* is considered to provide an additional regulatory component to this host-microbe relationship. In the absence of adaptive immunity, the hemocytes still have the ability to differentiate *V. fischeri* from non-symbiotic bacteria. To get a better understanding of the molecular mechanisms involved in determining host-symbiont specificity, two quantitative proteomic techniques (iTRAQ and label-

free spectral counting) were used to characterize the influence of light organ colonization on the squid hemocytes. These studies demonstrated that the presence of *V. fischeri* in the light organ can alter the composition of the hemocyte proteome by affecting proteins related to the cytoskeleton, adhesion, and innate immunity. Further characterization of symbiont recognition by the host involved the analysis of the transcript distribution and protein localization of two carbohydrate-binding galectins. These studies represent the first application of a variety of MS-based proteomic techniques to the squid-vibrio symbiosis. In addition, the data generated from these analyses have laid the foundation for the characterization of many new target proteins, which may have important functions in regulating and shaping the outcome of host-microbe relationships in general.

Proteomic Characterization of Host-Microbe Interactions in the *Euprymna scolopes* Light Organ Symbiosis with *Vibrio fischeri*

Tyler R. Schleicher

B.A., Colby College, 2008

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree Of

Doctor Of Philosophy

at the

University of Connecticut

2014

APPROVAL PAGE

Doctor of Philosophy Dissertation

Proteomic Characterization of Host-Microbe Interactions in the *Euprymna scolopes* Light Organ Symbiosis with *Vibrio fischeri*

Presented by:

Tyler R. Schleicher

Major Advisor

_____

Spencer Nyholm

Associate Advisor

_____

Joerg Graf

Associate Advisor

_____

David Benson

Associate Advisor

_____

Daniel Gage

Associate Advisor

_____

Victoria Robinson

University of Connecticut
2014

Acknowledgements

First I would like to thank my parents, Ron and Tracy Schleicher. Luckily, my time in graduate school did not require too much of your financial support. However, I would not have had the opportunities I've had in my life without your guidance and help. For that I will always be grateful.

I knew I wanted to study microbiology, but when I came to the University of Connecticut I did not have a clue about which lab I should join. I would like to thank my advisor, Spencer Nyholm, for allowing me to work in his lab. In addition to helping me become an independent scientist, I am grateful for the opportunities to conduct and present my research in locations including Hawaii and England.

I would also like to thank my committee members, David Benson, Daniel Gage, Joerg Graf, and Victoria Robinson. Thank you for taking the time out of your busy schedules to listen to me, ask me difficult questions, and to help me complete my degree.

Over the past 6 years I couldn't have completed much of my dissertation without the help of many other graduate students, staff, and members of the Nyholm lab. There were many undergraduates that made the daily maintenance of the squid facility possible. I want to thank them all because without their help I would not have been able to finish my research. Particularly I want to thank Joe Raymond and Rachel Dobensky, as they also contributed to a lot to my research. Many departmental scientists helped me along the way too. I would like to thank Pascal Lapierre who helped me with my bioinformatics questions from time to time and Carol Norris and Steven Daniels for their help with microscopy. I would also like to thank Ala Shaqra of the Robinson lab for all of his help in preparing outer membranes from *V. fischeri*. I am also grateful

for the help of René Augustin, who provided his expertise on antimicrobial peptides to help me study the antimicrobial activity of histone H4. René contributed some of the data presented in Appendix II. I especially want to thank Andrew Collins, Allison Kerwin, Jacqui Benjamino, Sophie Colston, Andrea Suria, Sarah McAnulty, Bethany Rader, and Corey Bunce. I always appreciated the fact that everyone was available to talk and discuss science and/or frustrations. I would not have been able to finish graduate school without you guys!

Finally, I would like to thank my fiancé, Rachel. Without your daily support I would not have made it through the past 6 years. Thank you for always listening to me, reading my papers, and going with me to take care of the squid on weekends (even when it smelled like dead shrimp!).

Table of Contents

List of Figures

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

Appendix I:

Appendix II:

Appendix III:

List of Tables

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

Appendix I:

Appendix II:

Chapter 1: Introduction

Most host-microbe relationships are beneficial or benign. Animal associations with microorganisms are carefully regulated to ensure a successful relationship between the two organisms [1]. Understanding the host-microbe interactions that occur during these mutual associations is critical. Current studies demonstrate that microorganisms impact many aspects of host physiology including the development of the immune system and different disease pathologies [2]. However, host-microbe interactions can be very difficult to study. In some animals, the microbial communities are extremely complex and consist of hundreds of different bacterial species within a specific microenvironment, such as in the mammalian gut [3]. For other animals, the natural microbial community of certain tissues or organs can range from one to several types of microorganisms and provide us with much simpler models to study [1].

For the past 25 years the Hawaiian Bobtail squid, *Euprymna scolopes*, has been used as a model to study host-microbe interactions [4]. The squid-vibrio symbiosis has been important in identifying how bacteria influence host development and how animals achieve host-symbiont specificity [5-7]. The availability of genomic and transcriptomic information, in addition to the ability to raise each organism separately in the laboratory, further adds to the strength of this model [8-10]. *E. scolopes* is unique because it has a specialized light organ that can only be colonized by the bioluminescent bacterium *V. fischeri*. Juvenile *E. scolopes* hatch with an uncolonized light organ and acquire *V. fischeri* from the surrounding marine environment [6]. Once in the light organ, *V. fischeri* reaches a high cell density that is required to produce light. The squid use the luminescence produced by *V. fischeri* in an anti-predatory behavior called

counter-illumination [11]. In return, the bacteria receive a nutrient rich microenvironment for the duration of the squid's life.

The regulation required to maintain this binary relationship is complex. The host has developed both physical and cellular mechanisms to help control the interactions within the light organ. One method involves the daily expulsion of more than 95% of the symbiont population [6,12,13]. This mechanism removes a majority of *V. fischeri* from the light organ and also enriches the surrounding marine environment in order to inoculate new generations of juvenile squid. Previous studies have revealed that this process is preceded by many significant changes in gene expression for both the host and symbiont [10]. The material that is expelled from the light organ (exudate) is a mixture of symbiont and host cells [13]. The light organ exudate can be collected as it is expelled and offers the opportunity to discover proteins indicative of the host-microbe interactions occurring in the light organ at this critical transitional period.

Regulation of the light organ microenvironment is also mediated by macrophage-like hemocytes. Shortly after colonization hemocytes begin to traffic to the juvenile light organ in response to the symbionts [14]. These cells play an uncharacterized role in the light organ, but appear to remain present for the duration of this association. In the adult light organ hemocytes are also known to traffic into the symbiont-containing crypt spaces. Another aspect that adds to the intrigue of this cell type is the evidence that colonization influences the tolerance of the hemocytes to the *V. fischeri* [5,15]. In adults, when the symbiont population is removed from the light organ by treatment with antibiotics, hemocytes adhere to *V. fischeri* significantly more than hemocytes from colonized animals [5]. These hemocytes can easily be collected from the cephalic blood vessel of *E. scolopes* and provide the opportunity to characterize their role in this beneficial symbiosis [16].

The introduction of "omic" analyses has been a significant development in the study of host-microbe interactions [17-19]. These studies enable the generation of large amounts of data relevant to each organism involved in the relationship, helping to develop a more complete understanding about how this and other symbioses work. Transcriptomic studies are common and have been applied to many host-microbe associations including the squid-vibrio symbiosis [10,17,20-22]. Less common, but increasing in popularity, are proteomic analyses of host-microbe interactions. The emergence of novel mass spectrometry technologies now allows the identification of thousands of proteins from complex cellular lysates. Proteomics can be used to identify the presence or absence of proteins, quantify protein abundance, or even identify specific post-translational modifications [23]. In addition, the proteome corresponds to the functional molecules in a cell, and may be a better indicator of the interactions that are occurring between the host and symbiont. In comparison, transcriptional analyses can identify short-lived changes that may not be represented in the proteome [24]. Until recently there have been no mass spectrometry (MS)-based proteomic studies involving the squid-*V. fischeri* symbiosis. Previous proteomic analyses of the *E. scolopes* and *V. fischeri* association have been restricted to the comparison of protein differences during the development of colonization in the light organ using 2D-PAGE [25]. However, no proteins were identified from these studies.

This dissertation describes the first application of high throughput MS-based proteomics to study the *E. scolopes* and *V. fischeri* light organ symbiosis. Shotgun proteomics was used on both the light organ exudate and circulating hemocytes in an effort to identify host and symbiont proteins important to the interactions occurring during this association. In addition, quantitative proteomics was used to compare hemocytes from squid with colonized and uncolonized light organs to determine the factors involved in host-symbiont recognition. Finally, a *V. fischeri*

surface protein with immunoglobulin-like domains and two squid galectins were chosen as candidate proteins to analyze their contribution to the maintenance of this beneficial host-microbe relationship.

Chapter 2: Characterizing the host and symbiont proteomes in the association between the bobtail squid, *Euprymna scolopes*, and the bacterium, *Vibrio fischeri*

Abstract

The beneficial symbiosis between the Hawaiian bobtail squid, *Euprymna scolopes*, and the bioluminescent bacterium, *Vibrio fischeri*, provides a unique opportunity to study host-microbe interactions within a natural microenvironment. Colonization of the squid light organ by *V. fischeri* begins a lifelong association with a regulated daily rhythm. Each morning the host expels an exudate from the light organ consisting of 95% of the symbiont population in addition to host hemocytes and shed epithelial cells. We analyzed the host and symbiont proteomes of adult squid exudate and surrounding light organ epithelial tissue using 1D- and 2D-polyacrylamide gel electrophoresis and multidimensional protein identification technology (MudPIT) in an effort to understand the contribution of both partners to the maintenance of this association. These proteomic analyses putatively identified 1,581 unique proteins, 870 proteins originating from the symbiont and 711 from the host. Identified host proteins indicate a role of the innate immune system and reactive oxygen species (ROS) in regulating the symbiosis. Symbiont proteins detected enhance our understanding of the role of quorum sensing, two-component signaling, motility, and detoxification of ROS and reactive nitrogen species (RNS) inside the light organ. This study offers the first proteomic analysis of the symbiotic microenvironment of the adult light organ and provides the identification of proteins important to the regulation of this beneficial association.

Introduction

The light organ symbiosis between the Hawaiian bobtail squid, *Euprymna scolopes*, and the bioluminescent bacterium, *Vibrio fischeri*, is used as a model association for understanding host-microbe interactions [6,26,27]. Hours after hatching from its egg case, the host is colonized when environmental *V. fischeri* take up residence in epithelia-lined crypt spaces located within a specialized light organ [6]. *V. fischeri* is the sole bacterium that colonizes the light organ and prior research has focused on understanding the mechanisms for establishing and maintaining the high degree of specificity between the partners [6,26-28]. While in the light organ, the bacteria are connected directly to the external environment through ciliated ducts and pores (Fig. 2.1). This conduit is important as it serves as an interface between the host and the environment and is used in a daily venting of the symbionts. The venting behavior is linked to the nocturnal foraging activities of the host. At night the light organ crypt spaces contain the highest densities of bacteria (10^9/adult squid; [13]), and the light provided by these symbionts is used to avoid predation [11]. At dawn the host expels 95% of its symbionts from the light organ, while entering a quiescent state in which it buries in the substrate [12,13]. The remaining bacteria repopulate the crypts ensuring a full complement of symbionts by the following nightfall. This venting mechanism helps regulate the symbiont population in the light organ as well as increases the concentration of *V. fischeri* in the immediate squid habitat, allowing future generations to be colonized [6,29].

The exudate of adult hosts emerges from the light organ pores as a thick paste-like substance that can be easily collected for experimental analyses (Fig. 2.1). This material represents the immediate microenvironment of the light organ crypts and is comprised of

symbiont cells and a mixed population of host cells (macrophage-like hemoctyes and shed epithelial cells), all surrounded by an acellular matrix [13].

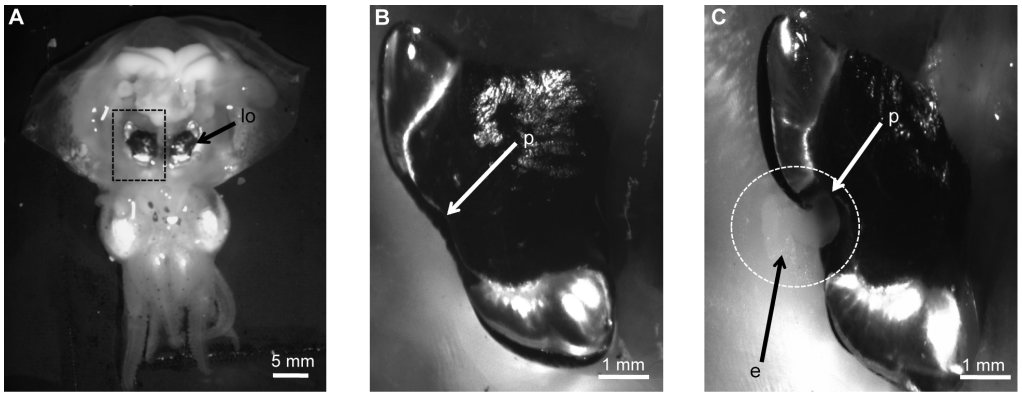


Figure 2. 1: Host and symbiont cells are expelled each morning as a thick exudate.

(A). A ventrally dissected adult squid reveals the bilobed light organ (lo), which is located in the center of the mantle cavity. (Dashed box highlights the region of light organ in B and C) **(B)**. One half of the light organ prior to expelling the light organ contents. **(C)**. One half of the light organ during the venting process. The exudate (e) emerges from a lateral pore (p) as an opaque paste (dotted circle).

In order to understand the host and symbiont contributions to this microenvironment, previous studies have focused on the cellular and biochemical components of the exudate [13,30]. Recent work has focused on changes in host and symbiont gene expression during the daily rhythm within the light organ [10]. Transcriptome analyses at different time points during the day/night revealed dynamic changes both metabolically and physiologically for the host and symbiont, and identified a large number of differentially expressed genes [10]. In addition, microscopy at these time points revealed that the crypt epithelium also undergoes morphological changes whereby apical surfaces are blebbed into the crypt spaces [10]. Many of these gene expression and cellular changes were most dramatic in the hours just before and after dawn, reflecting the dynamic turnover that occurs in the light organ upon venting.

In this study, we employed a number of techniques to characterize the host and symbiont proteomes of the adult light organ microenvironment at dawn when the association undergoes a dramatic reduction in symbiont population. To date, proteomic analyses of the squid-vibrio association are limited. A previous study used two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) to reveal numerous differences in the soluble proteins present in the light organs of juvenile aposymbiotic (uncolonized) and symbiotic (colonized) squid during the development of the symbiosis, however no proteins were identified [25]. Recent advances in proteomics, including multidimensional protein identification technology (MudPIT), have provided the tools to allow the identification of a large number of host and symbiont proteins in the squid-vibrio association for the first time [23,31]. MudPIT utilizes strong cation exchange chromatography (SCX) to separate peptides by charge prior to liquid chromatography tandem mass spectrometry (LC-MS/MS), thus increasing the number of identified peptides. In this study we utilized MudPIT, in addition to 1D- and 2D-PAGE, to describe both the host and symbiont proteomes in the light organ exudate and the surrounding host epithelial tissue. These analyses identified components of the host's innate immune system as well as numerous proteins involved in the detoxification of reactive oxygen species (ROS). Symbiont proteins detected were involved in stress responses, quorum sensing, motility, and two-component signaling pathways. Our data also highlight many proteins that are presently uncharacterized with regard to the squid-vibrio symbiosis. Identifying the host and symbiont proteins present in the light organ represents a first step to understanding key functional aspects of the association's molecular dialogue that is responsible for maintaining this highly specific relationship and complements a number of other molecular and genetic techniques that have been applied to this symbiosis.

Materials and Methods

Ethics statement

Euprymna scolopes is an invertebrate and is not regulated by animal care regulations in the United States. All field collection of research animals was done in accordance with state and federal regulations. The State of Hawaii does not require collection permits for this species outside of marine reserves. None of the animals collected for this study were caught/collected within a marine reserve or regulated area.

General methods

Adult animals were collected in shallow sand flats of Maunalua Bay, Oahu, HI by dip net and were either maintained in the laboratory in re-circulating natural seawater aquaria at the Hawaii Institute of Marine Biology or at the University of Connecticut with artificial seawater (ASW, Instant ocean) at 23°C. All animals were acclimated at least 48 hours under laboratory conditions and kept on an approximate 12 hr light/12 hr dark cycle before sample collection. *V. fischeri* strain ES114 was grown in saltwater tryptone (SWT) at 28°C as previously described [32].

Exudate and central core collection

Exudate was collected as previously described [13]. Briefly, adult squid were anesthetized in a 2% ethanol/seawater solution and ventrally dissected under red light within

minutes prior to dawn. A light stimulus (150 W halogen light) was used to induce venting behavior. Within 1 h, the squid had expelled the light organ contents, which were collected with a 10-µl disposable micropipette (Drummond Scientific Company) and stored on ice after the addition of a 1× protease inhibitor cocktail according to the manufacturer's protocol (Sigma Aldrich, P2714). Post-vented central cores were also dissected and removed from the light organ. All samples were flash frozen with liquid nitrogen and stored at −80°C until further analysis. No differences were detected between samples collected from animals maintained at either the Hawaii Institute of Marine Biology or at the University of Connecticut (data not shown).

Gel-based proteomic methods

Exudate sample preparation for 1D- and 2D-polyacrylamide gel electrophoresis

For PAGE applications, symbiont cells from freshly collected light organ exudate were separated from the soluble fraction, a source of host proteins, by centrifugation (Eppendorf 5810 R, 5,000 rpm, 10 minutes, 4°C). The symbiont pellet was washed three times with 0.22 µm filtered ASW to remove additional soluble proteins. Symbiont proteins were extracted by a modified method from Ho and Hsu [33]. Briefly, 10 consecutive liquid nitrogen freeze/thaw cycles were performed in the presence of a 1× protease inhibitor cocktail (Sigma Aldrich, P2714) with 80 mM Tris, pH 8.0 for cell lysis. After separation from the bacterial pellet, soluble host proteins were quantified (see below) and stored until further analysis. For cultured *V. fischeri*, cells were grown to early stationary phase [32] and proteins were extracted as described for the symbiont exudate pellet. Protein concentrations of separate symbiont exudate and host soluble

fractions, as well as culture-grown *V. fischeri*, were determined spectrophotometrically using the method of Whitaker and Granum [34] and/or a Bradford assay (Bio-Rad). Typically, protein extractions of exudate resulted in 10–20 μg of protein combined from the host soluble and symbiont pellet fractions. Comparison of 2D-PAGE gels from soluble proteins of culture-grown *V. fischeri* and the host soluble fraction of the exudate demonstrated that the soluble host fraction was devoid of bacterial proteins (data not shown).

1D-polyacrylamide gel electrophoresis of light organ exudate

Between 10 and 20 μg of exudate protein from either the host (soluble protein separated from bacterial pellet) or symbiont fraction (bacterial pellet) were resolved with 12.5% polyacrylamide gels (Bio-Rad). Electrophoresis was performed with a Hoeffer 250 mini-gel apparatus at 23 mA or a Bio-Rad Mini PROTEAN® Tetra cell at 200 V for 30 minutes. Gels were either stained with Bio-Rad Brilliant Blue Coomassie R-250 or a Bio-Rad Silver Stain Plus Kit (Bio-Rad). 1D-PAGE of both the soluble host fractions and the bacterial pellets were shown to be reproducible (n = 3, separate and fractionated pooled exudate samples for each; data not shown).

2D-polyacrylamide gel electrophoresis of light organ exudate

2-D PAGE was performed using the Amersham Pharmacia Biotech Multiphor II system as previously described [25]. 40 μg of pooled exudate protein from either the soluble host fraction or the bacterial pellet, originating from 2 or more adult squid or culture-grown *V.*

fischeri cells, were denatured 1:4 in 9 M urea, 1% DTT, 2% Pharmalyte 3–10, 0.5% Triton-X-100, 0.14% phenylmethylsulfonyl fluoride, loaded onto a first dimension gel strip with an immobilized pH gradient (4–7) and focused over a 20 hour period. Samples were then separated by molecular weight on pre-cast 12% to 14% polyacrylamide gradient gels (GE Healthcare Life Sciences). Gels were silver stained as previously described [25,35]. 2D-PAGE from the soluble host fractions and the bacterial pellets or culture-grown cells were deemed to be highly reproducible (n = 3, separate and fractionated pooled exudate samples; data not shown). For comparison, 2D gels were visually aligned and similarities and differences of the molecular weights and individual protein species were noted. Five spots of interest from the 2D-PAGE gel of the exudate bacterial pellet were excised and successively washed in 50% acetonitrile, 50% acetonitrile/50 mM NH_4HCO_3, and 50% acetonitrile/10 mM NH_4HCO_3. The five gel spots were then dried by speed vacuum (Eppendorf Concentrator 5301) and resuspended in 10 mM NH_4HCO_3. Digestion was completed with 0.1 μg trypsin (Promega, V5111) per each 15 mm^3 of gel in a final volume of 35 μl of 10 mM NH_4HCO_3 at 37°C for 24 hours. The digested samples were stored at −80°C until submission to the W. M. Keck Biotechnology Resource Laboratory, Yale University, for LC-MS/MS (See below, Mass spectrometry proteomics).

Mass spectrometry proteomic methods

Protein preparation for multidimensional protein identification technology and liquid chromatography tandem mass spectrometry

For MudPIT and LC-MS/MS, pooled host and symbiont fractions from freshly collected light organ exudate were combined and quantified as described above. Additionally, central cores were homogenized in the presence of a 1× protease inhibitor cocktail (Sigma Aldrich, P2714) with 80 mM Tris, pH 8.0 using a ground-glass homogenizer. Proteins from central cores were collected from the supernatant of the homogenate after centrifugation (Eppendorf 5810 R, 14,000 rpm, 30 minutes, 4°C) and quantified as described above. Extractions of the central core tissue resulted in approximately 20 µg of soluble protein per central core. Total protein from pooled exudate samples (50 µg, n = 7 squid and 100 µg, n = 7 squid) and pooled central core samples (40 µg, n = 3 squid) were precipitated with 10% trichloroacetic acid (Fisher Scientific) at 4°C overnight. The protein precipitates of the exudates and central cores were collected by centrifugation (Eppendorf 5810 R, 11,000× g, 30 minutes, 4°C) and washed twice with ice-cold acetone. The protein pellets were briefly air-dried and then solubilized in 25 µl of 8 M urea, 0.4 M ammonium bicarbonate, pH 8.0. Both samples were reduced and alkylated with 5 µl of 45 mM dithiothreitol (DTT; Acros Organics) at 37°C for 20 minutes and 5 µl of 100 mM iodoacetamide (Acros Organics) at room temperature in the dark for 20 additional minutes. Sequencing grade trypsin was added 1:15 (w/w enzyme to protein; Promega, V5111). The solutions were diluted in water to 100 µl (2 M urea final concentration). Both samples were digested at 37°C for 18–24 hours and then stored at −80°C until submission to the W. M. Keck Biotechnology Resource Laboratory, Yale University for LC-MS/MS.

For MudPIT, tryptic digests of pooled exudate proteins from *E. scolopes* underwent strong cation exchange (SCX) on an Applied Biosystems Vision Workstation at the W. M. Keck Biotechnology Resource Laboratory at Yale University. During SCX, peptides were separated by charge into fractions, which were then analyzed by LC-MS/MS. MudPIT analyses of separate

pooled exudate samples were run in duplicate. The first analysis used 10 SCX fractions (50 µg, n = 7 squid) and the second used 20 SCX fractions (100 µg, n = 7 squid), allowing greater coverage of lower abundance peptides. The central core sample (40 µg, n = 3 squid) and symbiont exudate 2D-PAGE spots (n = 5 spots) were analyzed by one-dimensional LC-MS/MS (see below).

For SCX, the tryptic digests of pooled exudate proteins were acidified with 2 µl of 1 M phosphoric acid. A 2.1 mm×200 mm PolySULFOETHYL A™ column (PolyLC Inc.) was used to establish a linear gradient for 118 minutes. The gradient was maintained in 10 mM potassium phosphate, 25% acetonitrile (pH 3.0) and the same buffer with the addition of 1 M potassium chloride. Fractions were collected every 2 minutes at a flow rate of 150 µl/min. All fractions were dried, dissolved in 5 µl of 70% formic acid, and diluted to 15 µl in 0.1% trifluoroacetic acid for subsequent LC-MS/MS.

Liquid chromatography tandem mass spectrometry

LC-MS/MS of each exudate SCX fraction, central core peptides, and 2D gel spot peptides was completed at the W. M. Keck Biotechnology Resource Laboratory at Yale University. A LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a Waters nanoAcquity UPLC system operated with a Waters Symmetry® C18 180 µm×20 mm trap column, and a 1.7 µm, 75 µm×250 mm nanoAcquity™ UPLC™ column (35°C) was used for peptide separation. Trapping was performed at 15 µl/min with Buffer A (100% water, 0.1% formic acid) for 1 minute. Peptide separation was performed at 300 nl/min with Buffer A and Buffer B (100% CH3CN, 0.075% formic acid); a 51 minute linear gradient was established

starting with 5% Buffer B, increasing to 50% B at 50 minutes, and finally 85% B at 51 minutes. MS was acquired in the Orbitrap using 1 microscan followed by four data dependent MS/MS acquisitions. Neutral loss scans (MS3) were also obtained for 98.0, 49.0, and 32.7 amu.

Data analysis

All MS/MS spectra were analyzed using the Mascot algorithm for uninterpreted MS/MS spectra [36]. The Mascot Distiller program used the MS/MS spectra to generate Mascot compatible files by combining sequential MS/MS scans from profile data that have the same precursor ion. A charge state of +2 and +3 were preferentially located with a signal to noise ratio of 1.2 or greater and a peak list was created for database searching. The peak list was searched by Mascot using *V. fischeri* amino acid sequence and juvenile *E. scolopes* light organ expressed sequence tag (EST) databases [8,9]. Search parameters included partial methionine oxidation, carboxamidomethylated cysteine, a peptide tolerance of ±20 ppm, MS/MS fragment tolerance of ±0.6 Daltons (Da), and peptide charges of +2 or +3. Normal and decoy database were also searched. Mascot significance scores are based on MOlecular Weight SEarch (MOWSE) scores and rely on multiple matches of more than one peptide to the same protein [37]. The MOWSE based ions score is equal to (−10)*(Log10P), where P is the absolute probability that a match is random. For a match to be significant, the probability of it being a random match should be below 5% (E-value < 0.05) [38]. The protein threshold score depends on the size of the database being searched, therefore, Mascot determined that scores greater than 68 were significant when searching the juvenile light organ EST database and scores greater than 48 were significant when searching the *V. fischeri* ES114 amino acid database. Proteins were considered identified when 2

or more peptides matched the same protein and if the Mascot score was above the respective significance threshold. Proteins with putative identifications contained two or more peptide matches, but had a Mascot score below the threshold for the respective database (E-value > 0.05).

Mascot also calculates the exponentially modified protein abundance index (empai) which estimates the abundance of protein species by using the number of peptides detected in the analysis compared to the number of possible peptides for a particular protein [39,40].

Host proteins identified by Mascot using the juvenile light organ EST database were further analyzed using the Bioinformatics Utility for Data Analysis of Proteomics using ESTs (BUDAPEST) which removed any peptides matching to non-coding reading frames [41]. BLASTx (E-value cutoff 10^{-6}) against the NCBI nr database was used to determine the top protein hit for each EST [42]. In addition, BUDAPEST calculated a peptide score for each protein identified. This score was equal to the number of correct reading frame peptides squared divided by the total number of peptides (all reading frames) identified for that EST. BUDAPEST scores greater than 1 can be considered significant, however, in our study scores of 2 or greater were chosen to represent significant protein identifications.

Results

Exudate samples collected from adult *E. scolopes* light organs were analyzed using a number of proteomic techniques. 1D- and 2D-PAGE revealed that the host soluble fraction of the exudate, derived from host hemocytes and apical surfaces of shed light organ crypt epithelial cells, was comprised of a complex mixture of proteins and peptides, the majority of which are

represented between the isoelectric points of 4 to 7 and a size of 7 to 100 kilodaltons (kD) (Fig. 2.2).

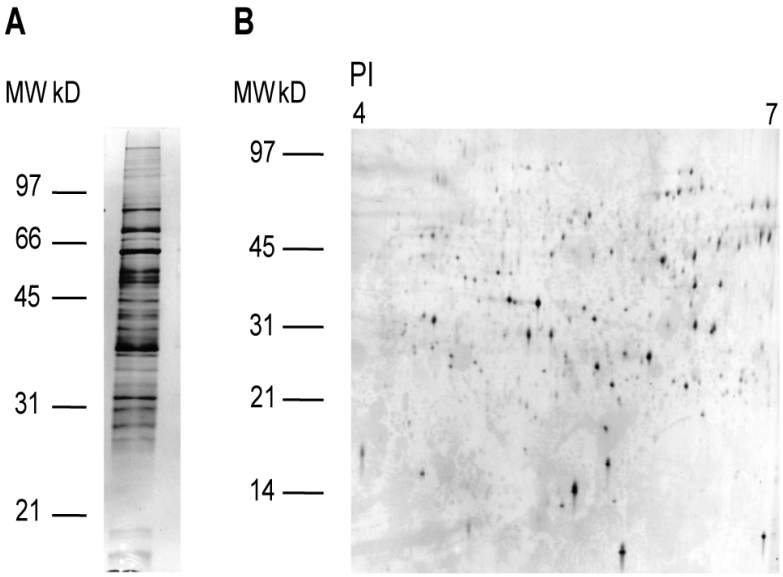


Figure 2. 2: PAGE analysis of the soluble host fraction from light organ exudate.

(A). 1D-PAGE of the host fraction of the exudate on a 12.5% polyacrylamide gel. **(B)**. 2D-PAGE of the host fraction of the exudate on a 12-14% polyacrylamide gel.

Similar analyses of the symbiont fraction of the exudate also revealed a complex protein profile (Fig. 2.3). Comparison of the host and symbiont PAGE gels support previous observations that the exudate appears enriched in bacteria. When comparing proteins expressed by *V. fischeri* in the light organ to proteins expressed by *V. fischeri* in culture, a protein with an isoelectric point of 5.5 and a molecular weight of 10 kD was present in the light organ, but not in solubilized proteins from culture-grown *V. fischeri* (Fig. 2.3C). The protein spot of interest (Fig. 2.3C, spot 2) and four surrounding protein spots (common to both the light organ and culture) were excised and identified by LC-MS/MS (Table 2.1). The unique symbiont light organ protein was determined to be a quorum sensing-regulated protein (QsrP), which has been previously

identified as being expressed by *V. fischeri* in the light organ, but remains functionally uncharacterized [43].

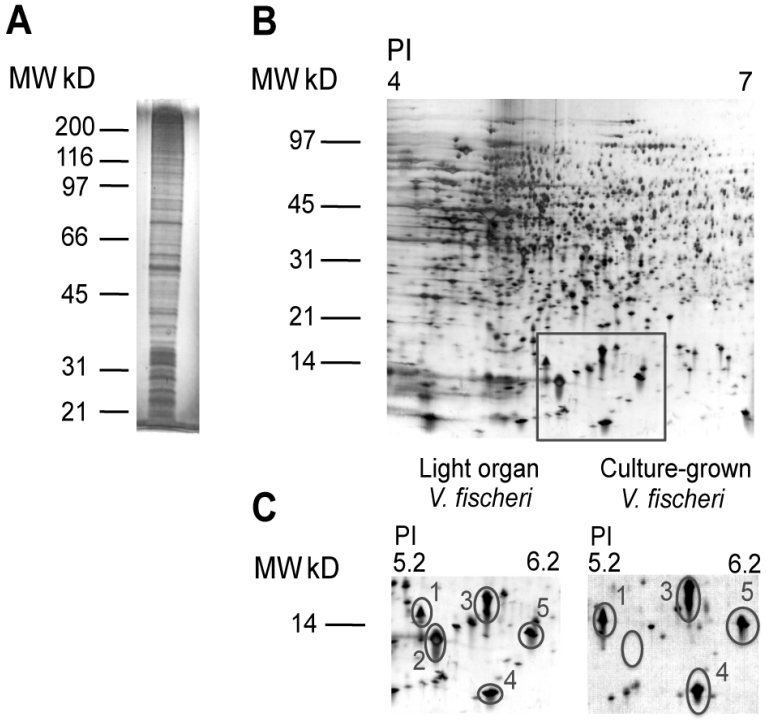


Figure 2. 3: PAGE analysis of the soluble proteins originating from the symbiont fraction of the exudate.

(A). 1D-PAGE of symbiont fraction of the exudate on a 12.5% polyacrylamide gel. **(B)**. 2D-PAGE of the symbiont fraction of the exudate on a 12-14% polyacrylamide gel. Black box highlights the region of the gel compared in C. **(C)**. 2D-PAGE comparison of bacterial soluble proteins from the exudate and culture-grown *V. fischeri*. Numbered protein spots were identified by LC-MS/MS (Table 2.1).

Table 2. 1: Exudate proteins identified by LC-MS/MS from the symbiont 2D-PAGE analysis.

Spot	Mw (kD)/pI[a]	Top Protein Hit to NCBI nr Database	ORF[b]	Score[c]
1	13.3/5.4	50S ribosomal protein L9	VF_2310	1045
2	11.0/5.5	Quorum sensing regulated protein QsrP	VF_A1058	628
3	14.7/5.7	Transcriptional dual regulator H-NS	VF_1631	712
4	8.8/5.7	Cold shock protein	VF_2561	473
5	9.3/6.1	30S ribosomal protein S6	VF_2312	501

a. Predicted molecular weight (Mw) and isoelectric point (pI) for the proteins identified.

b. Open reading frame (ORF) locations of the respective genes on the chromosomes of *V. fischeri*.
c. Scores were assigned by Mascot. Scores greater than 83 were significant (E-value < 0.05) for searches of the NCBI nr database.

In an effort to further characterize the proteins expressed by the host and symbiont we utilized shotgun proteomic techniques (LC-MS/MS and MudPIT). These methods allowed us to putatively identify a combined 1,581 host and symbiont proteins present in the light organ. For MudPIT, light organ exudate samples of 10 or 20 SCX peptide fractions (see Materials and Methods) were analyzed (Table 2.2). In addition, to increase our representation of host proteins we analyzed post-vented central cores by single fraction LC-MS/MS (Table 2.2). A total of 870 unique symbiont proteins were putatively identified by Mascot from all the light organ samples (exudates and central cores; Supplemental file S2.1). 516 of these proteins were above the significance threshold set by Mascot (E-value < 0.05). For the host, we utilized BUDAPEST, a software program developed specifically to identify proteins in the correct open reading frame in cases when only EST databases are available [41]. 676 host proteins with more than 2 peptides matching to the correct reading frame and a BUDAPEST score of greater than or equal to 2 were identified from combining the LC-MS/MS and MudPIT data of the exudate and central core samples (Supplemental file S2.2).

Table 2. 2: Number of host and symbiont proteins identified by shotgun proteomics.[a]

Sample	***Vibrio fischeri***	***Euprymna scolopes*[b]**
Exudate (10 Fractions)	214	163
Exudate (20 Fractions)	708	591
Central Core (1 Fraction)	90	234
Total unique proteins	870	711

a. Protein identifications required 2 or more peptides per protein. Total unique proteins summarize the combined unique proteins from all samples. Numbers include putative identifications (see Materials and Methods).

b. Host protein identifications counts are post-BUDAPEST analysis.

All host and symbiont proteins, including putative identifications, were organized functionally according to the Clusters of Orthologus Groups database (COG and KOG; Supplemental Fig. S2.1; Supplemental files S2.1 and S2.2) [44,45]. In order to achieve a more thorough understanding of the functions represented by the proteins in our data, we first analyzed the relative abundance of each symbiont protein. The 25 most abundant symbiont proteins determined by empai include the protein subunits of luciferase (LuxAB), QsrP, alkyl hydroperoxide reductase C22 (AhpC), and several cold shock proteins (Table 2.3). Our analyses also identified a number of symbiont proteins related to functions involved in stress responses, quorum sensing, motility, and signaling pathways, all of which have been previously implicated as being important in the squid-vibrio association (Table 2.4, Supplemental file S2.3; see discussion). Several of these identified proteins including AhpC and the cold shock proteins have symbiotic roles yet to be characterized.

Host proteins detected in the light organ highlight the innate immune system, oxidative stress, and signaling pathways (Table 2.5). Identified proteins include those involved with the NF-κB signaling pathway and the recognition of microbe-associated molecular patterns (MAMPs) such as peptidoglycan recognition proteins (PGRPs) and a carbohydrate binding protein, galectin (Table 2.5). Proteins related to oxidative stress consist of superoxide dismutase, peroxiredoxins and numerous peroxidases, including the *E. scolopes* halide peroxidase (EsHPO) (Table 2.5; see discussion). Additionally, several host proteins involved in iron-sequestration were detected in the light organ.

Table 2. 3: The 25 most abundant symbiont proteins present in light organ exudates and central cores identified by MudPIT and LC-MS/MS in descending empai order (excluding ribosomal proteins).

#	ORF[a]	Gene	Protein Name	empai	Score[b]
1	VF_A0921	luxA	luciferase alpha chain LuxA	28.74	2840
2	VF_1662	-	DNA-binding protein	21.59	257
3	VF_1739	acpP	acyl carrier protein	19.95	290
4	VF_A1058	qsrP	LuxR-regulated periplasmic protein QsrP	13.61	691
5	VF_1975	ahpC	alkyl hydroperoxide reductase, C22	10.54	994
6	VF_1050	-	hypothetical protein	9.35	179
7	VF_0744	ybeD	hypothetical protein	7.94	194
8	VF_1767	cspD	DNA replication inhibitor	7.66	309
9	VF_A0822	-	hypothetical protein	7.55	101
10	VF_1216	infC	protein chain initiation factor IF-3	6.86	378
11	VF_1237	ihfA	integration host factor subunit alpha	6.41	193
12	VF_2120	arcA	two-component response regulator	6.32	799
13	VF_0625	Ndk	nucleoside diphosphate kinase	5.92	245
14	VF_2151	-	iron(III) ABC transporter	5.3	1109
15	VF_0697	-	putative lipoprotein	5.22	72
16	VF_0442	Pgk	phosphoglycerate kinase	5.08	1747
17	VF_0890	grxA	glutaredoxin 1	4.88	216
18	VF_0262	rpoA	DNA-directed RNA polymerase alpha	4.84	647
19	VF_A0595	-	cold shock protein	4.78	273
20	VF_2096	-	hypothetical protein	4.77	146
21	VF_A1094	cspG	DNA-binding transcriptional regulator	4.64	776
22	VF_1961	Tsf	elongation factor Ts	4.36	794
23	VF_A0920	luxB	luciferase beta chain LuxB	3.93	1434
24	VF_0274	-	immunogenic protein	3.87	442
25	VF_0507	deoD	purine nucleoside phosphorylase	3.86	140

a. Open reading frame (ORF) locations of the respective genes on the chromosomes of *V. fischeri*.
b. Scores were assigned by Mascot. Scores greater than 48 were significant for *V. fischeri*.

Table 2. 4: Symbiont proteins detected in light organ exudates and central cores by MudPIT and LC-MS/MS. [a]

Category	ORF[b]	Protein Name	Score[c]
Quorum Sensing			
	VF_A0921	luciferase alpha chain LuxA	2840
	VF_A0920	luciferase beta chain LuxB	1434
	VF_A1058	LuxR-regulated periplasmic protein QsrP	691
	VF_1349	subtilisin-like serine protease	491
	VF_A0922	acyl transferase LuxD	463
	VF_A0894	putative surface protein	304
	VF_A0923	acyl-CoA reductase LuxC	225
	VF_0545	S-ribosylhomocysteinase LuxS	167
	VF_1725	secretory tripeptidyl aminopeptidase	123
	VF_2177	LitR	78
	VF_A0919	long-chain-fatty-acid ligase LuxE	62
	VF_1978	AcfA-like protein	50
	VF_1165	macrolide ABC transporter	45
	VF_A0925	LuxR	21
Oxidative Stress			
	VF_1975	alkyl hydroperoxide reductase, C22	994
	VF_0902	thioredoxin reductase	238
	VF_A0009	hydroperoxidase HPII(III) KatA	184
	VF_0921	superoxide dismutase, Fe	83
	VF_1920	thioredoxin-dependent thiol peroxidase	54
	VF_2316	nitric oxide dioxygenase	46
	VF_A0890	thioredoxin peroxidase	19
Two-Component Signaling			
	VF_2120	ArcA	799
	VF_A0561	two component response regulator	207
	VF_2342	periplasmic protein CpxP	154
	VF_1570	TorR	111
	VF_0454	transcriptional regulator VpsR	83
	VF_1909	DNA-binding response regulator NarP	63
	VF_1627	response regulator GacA	59
	VF_0114	osmolarity response regulator OmpR	54
	VF_2374	two-component response regulator	46
	VF_A0216	two component response regulator	36
	VF_A1016	two component sensory histidine kinase	32
	VF_1148	response-regulatory protein YehT	23
	VF_1619	hybrid sensory histidine kinase TorS	20
	VF_1401	sigma-54 dependent response regulator	16
	VF_2343	DNA-binding response regulator CpxR	15
Flagellar-related proteins			
	VF_1856	FlrA	107
	VF_1881	flagellar anti-sigma-28 factor FlgM	81
	VF_1864	flagellin	65
	VF_1866	flagellin	62
	VF_1871	flagellar basal body L-ring protein	34
	VF_0715	flagellar motor protein MotB	29
	VF_1870	flagellar basal body P-ring protein	18
	VF_1877	flagellar basal body rod protein FlgB	14

a. For more complete information on MudPIT and LC-MS/MS symbiont protein identifications refer to Supplemental file S2.1 and Supplemental file S2.3.
b. Open reading frame (ORF) locations of the respective genes on the chromosomes of *V. fischeri*.
c. Scores were assigned by Mascot. Scores greater than 48 were significant for *V. fischeri*.

Table 2. 5: Host proteins detected in light organ exudates and central cores by MudPIT and LC-MS/MS. [a]

Category	gi	Top protein hit to NCBI nr database	e-value[b]	Score[c]
Immunity				
	225906399	Galectin [*Pinctada fucata*]	4E-69	10
	63033995	Peptidoglycan recognition protein 2 [*E. scolopes*]	1E-121	7
	223670954	C3 precursor [*Nematosella vectensis*] [d]	4E-12	6
	144952812	Thioester-containing protein [*Chlamys farreri*]	4E-14	6
	113931358	NF-κB repressing factor [*Xenopus tropicalis*]	2E-17	4
	42741753	Importin alpha 3 [*Aplysia californica*]	9E-6	4
	63033997	Peptidoglycan recognition protein 3 [*E. scolopes*]	9E-49	3
	85822201	TEP2 [*Glossina morsitans morsitans*]	3E-16	2
Oxidative stress				
	306451460	Thioredoxin peroxidase [*Cristaria plicata*]	1E-92	13
	110734438	Superoxide dismutase [*Haliotis discus discus*]	7E-61	7
	229366436	Peroxiredoxin-5 [*Anoplopoma fimbria*]	2E-51	5
	67083759	Glutathione-type peroxidase [*Ixodes scapularis*]	8E-50	5
	2239176	Melanogenic peroxidase [*Sepia officinalis*]	3E-59	4.5
	209171295	Peroxiredoxin 4 precursor [*Biomphalaria glabrata*]	1E-101	4
	159008	Halide peroxidase [*Euprymna scolopes*]	1E-141	4
	157136354	Peroxiredoxins, prx-1, prx-2, prx-3 [*Aedes aegypti*]	5E-72	3.2
	77166828	Glutathione peroxidase [*Rhipicephalus microplus*]	4E-62	3
	149688674	Peroxiredoxin [*Chlamys farreri*]	2E-58	2
	126697356	Thioredoxin peroxidase 2 [*Haliotis discus discus*]	1E-54	2
Iron-Binding				
	318067980	Transferrin [*Ictalurus punctatus*]	4E-34	6
	4768842	Ferritin [*Enteroctopus dofleini*]	7E-74	3
	157786780	Melanotransferrin [*Rattus norvegicus*]	6E-21	3

a. For more complete information on MudPIT and LC-MS/MS host protein identifications refer to Supplemental file S2.2.
b. E-value represents the alignment of the light organ EST with the top protein hit in the NCBI nr database.
c. Scores were assigned by BUDAPEST and correlate the number of reading frame peptides matched to the light organ EST to the number of overall peptides. Scores greater than 2 were significant.
d. No alignment with *Euprymna scolopes* C3 (Putatively identified as a thioester-containing protein; see discussion).

Discussion

The daily expulsion of *V. fischeri* from the light organ of *E. scolopes* provides a unique opportunity to characterize the interactions between the host and symbiont in a natural microenvironment. Previous analyses of this exudate have focused on the cellular and biochemical composition of the expelled matrix [13,30]. In this study we characterized the light organ exudate and surrounding epithelial proteome using MudPIT and PAGE. A total of 1,581 unique host and symbiont proteins were putatively identified from the light organ, offering the first proteomic analyses of this symbiotic microenvironment.

Innate immune system

MAMPs and host pattern recognition receptors (PRRs) are two components underlying host-microbe interactions and are significantly involved in the development of this association [46]. MAMPs including lipopolysaccharide (LPS), and peptidoglycan and its derivatives, function in determining the specificity of the squid-vibrio symbiosis as well as initiating morphogenetic changes to the light organ [7,47,48]. We identified several host proteins related to pattern recognition in both the exudate and central core tissues (Table 2.5). *E. scolopes* PGRP 2 and 3 (EsPGRP 2 and EsPGRP 3) are involved in detecting peptidoglycan, a major cell wall component of bacteria [49]. EsPGRP 2 is secreted into the crypts of the light organ where it is thought to degrade tracheal cytotoxin (TCT), a monomer of peptidoglycan [50]. The role of EsPGRP 3 in the symbiosis is currently under investigation, but has been detected in adult and juvenile hemocytes (unpublished data). Certain carbohydrates, such as beta-galactosides, are

another type of MAMP that are recognized by carbohydrate binding proteins known as galectins [51]. A putative galectin was identified in both the exudate and central core tissue (Table 2.5, Supplemental file S2.2) and may have an uncharacterized role in the squid-vibrio symbiosis.

Aside from PRRs and MAMPs, cellular adhesion is often important for host-microbe and cell-to-cell interactions. Outer membrane proteins (OMPs) are localized at the bacterial cell surface and are good candidates for mediating recognition between the partners. OmpU, a symbiont outer membrane protein that we have identified in the light organ (Supplemental file S2.3), was shown to be important in mediating adhesion to adult host hemocytes and during the early stages of colonization [5,52]. Other OMPs identified, such as a hypothetical protein VF_1010, have roles yet to be characterized in binding and adhesion, but may have similar functions (Supplemental file S2.3). Understanding how the symbiont outer membrane proteome varies in the light organ vs. the free-living environment and between symbiosis-competent and incompetent strains may shed light on mechanisms of mediating specificity in this symbiosis.

An immune pathway highlighted by our proteomic data includes NF-κB signaling (Table 2.5). The role of NF-κB signaling during the establishment of the squid-vibrio symbiosis is currently under investigation, however, many important members of the pathway have been identified from juvenile light organ ESTs [49]. We detected a NF-κB repressing factor (see below) and importin alpha 3, a protein involved in shuttling proteins into the nucleus by recognizing nuclear localization signals (Table 2.5) [53]. *In vitro* and *in vivo* studies using cancer cell lines revealed that this protein is a member of the NF-κB signaling pathway and aids in the transport of NF-κB transcription factors into the nucleus [54].

Recently, *E. scolopes* has been shown to have a complement pathway that in other systems is involved with mediating inflammation and opsonization [46,55,56]. The function of

this pathway has yet to be described in the squid-vibrio symbiosis, however, we detected putative components of the complement cascade in both the exudate and the central core (Table 2.5, Supplemental file S2.2). Although one of these identifications was annotated as a complement component C3 precursor (Table 2.5, Supplemental file S2.2), closest to the cnidarian *Nematostella*, further analysis of these peptides using *E. scolopes* transcriptomic data revealed that this protein did not align with the previously described *E. scolopes* C3 (data not shown). Instead, this protein, along with two others, was identified as thioester-containing proteins (TEPs). Among invertebrates, TEPs play an important role in innate immune response as members of the complement system or as protease inhibitors [57,58].

Reactive oxygen and nitrogen stress response

The chemical microenvironment of the light organ crypts likely influences the maintenance of the association and helps to ensure specificity. Although oxygen is critical for the bioluminescence reaction, reactive oxygen species (ROS) and toxic oxygen intermediates have been shown to be abundant in the light organ [59]. Host-derived ROS, such as hypohalous acid, are thought to play key roles in initiation and persistence of the squid-vibrio symbiosis [59]. Hypohalous acid, produced by an abundant light organ peroxidase similar to a halide peroxidase, is believed to help to create an oxidative environment that *V. fischeri* must overcome to colonize the host [60,61]. In addition to the previously described EsHPO, a number of other host peroxidases were present, suggesting that additional ROS may be important to this association (Table 2.5). Peroxiredoxins are antioxidant proteins, which are abundant in the host proteome and have been shown to detoxify reactive molecular species derived from oxygen and nitrogen

[62,63]. Therefore, these ROS mediators may indicate a means by which the host protects its own tissues in the oxidative microenvironment of the light organ.

Another role of host ROS may be maintaining specificity by preventing non-symbiotic bacteria and potential pathogens from infecting the host. The light organ crypts are open to the environment via pores on the surface of the light organ, yet *V. fischeri* is thought to be the sole symbiont of this highly specific association [6]. Proteins expressed by the symbiont reveal functions involved with protecting cells from host ROS (Table 2.4). *V. fischeri* utilizes a periplasmic catalase (KatA) to sequester hydrogen peroxide from the host, which can be used by EsHPO to generate hypohalous acid [64]. We identified, in addition to KatA, the antioxidant enzymes AhpC and thioredoxin-dependent thiol peroxidase (Bcp) (Table 2.4). A *V. fischeri katA* mutant showed no additional catalase activity in culture suggesting that KatA is the major scavenger of H_2O_2 [64]. The additional antioxidant proteins identified in this study may indicate a mechanism by which the symbiont can protect itself from other types of ROS or RNS. AhpC, a peroxiredoxin capable of reducing hydrogen peroxide, organic peroxides, and peroxynitrite, is the most abundant antioxidant symbiont protein present in the light organ (Table 2.3). In *Vibrio vulnificus*, AhpC functions along with another subunit, AhpF, which supplies the reducing equivalents in the form of NADH, to reduce peroxides [65]. However, AhpF is absent from the *V. fischeri* genome, suggesting that another protein is necessary to reduce peroxides by this pathway. Studies involving *Treponema pallidum* show that thioredoxin reductase can substitute for organisms lacking an AhpF homolog [66]. For *V. fischeri*, a thioredoxin reductase FAD/NAD(P)-binding protein (TrxB) was present in our MudPIT data (Table 2.4) and may have the potential of interacting with AhpC. Along with AhpC, proteins implicating that *V. fischeri*

also detoxifies RNS, include nitric oxide dioxygenase (Hmp), and two peptide-methionine (S)-S-oxide reductases (MsrA and VF_A0005; Supplemental file S2.1) [67,68].

Reactive nitrogen species, such as nitric oxide (NO), contribute to signaling and development in the squid-vibrio symbiosis [69]. The role of NO as a toxic product to pathogens has been well studied; however, the function of NO in beneficial associations has only been recently analyzed [69,70]. In juvenile squid the epithelial tissue lining the ducts entering the light organ crypts contain high levels of NO, suggesting that the symbionts must overcome NO in order to colonize the light organ [71]. The detection of nitric oxide dioxygenase (Hmp), recently shown to play a role in NO detoxification, suggests that *V. fischeri* also maintains the ability to manage NO related stress in adult squid (Table 2.4) [68]. Once *V. fischeri* colonizes the light organ, nitric oxide synthase (NOS) is down-regulated and lower levels of NO likely allow the symbiont to grow in the crypt spaces under reduced RNS stress [71]. We identified NF-κB repressing factor, which in addition to other immune functions, has been shown *in vitro* to negatively regulate transcription of NF-κB pathway effectors, including NOS, by directly interacting with promoter region sequences (Table 2.5) [72,73]. The results of this study provide a number of new host and symbiont targets involved in mediating ROS and RNS for further analyses.

The availability of iron has also been shown to be an important factor in squid-vibrio symbioses [68,74,75]. Free iron plays a critical role in host-microbe interactions and under certain circumstances may allow development of pathogenic associations [76,77]. Host proteins involved in sequestering free iron such as ferritin, transferrin, and melanotransferrin were identified (Table 2.5). These iron-binding proteins provide supporting evidence that iron remains limiting in the light organ and suggests a possible role for these proteins in regulating the growth

of *V. fischeri* [74]. In contrast to the host, putative proteins that the symbiont may utilize for acquiring iron include receptors for the siderophores, aerobactin and anguibactin (Supplemental file S2.3). Symbiont proteins involved in utilizing heme, another source of iron, are also present, and include HutZ, HutA, HuvX, and HmuT (Supplemental file S2.3). It is likely that *V. fischeri* employs several different strategies to meet its necessary iron requirements.

Quorum Sensing

First described in *V. fischeri*, quorum sensing regulates bioluminescence, the light from which provides the host with an anti-predatory mechanism known as counter-illumination [11,78,79]. Lux proteins involved in light production were identified and among the most abundant symbiont proteins (Table 2.3). Previous PAGE and transcriptomic analyses first revealed additional quorum sensing-regulated proteins, which were also detected by our characterization of the adult light organ proteome [43,80]. QsrP is one of the most abundant proteins present in the symbiont proteome (Table 2.3), yet this novel protein remains functionally uncharacterized. Another quorum sensing-regulated protein identified in this study is a putative surface protein (VF_A0894) with immunoglobulin-like domains (Table 2.4). This putative surface protein is similar to the Leptospiral immunoglobulin-like proteins (LigA, LigB and LigC) of pathogenic *Leptospira* spp., which are thought to mediate adhesion to host cells [81]. These quorum sensing-regulated proteins may be important to a symbiotic lifestyle. We also detected LuxS, AI-2 synthase, which is involved in a second quorum sensing system in *V. fischeri* and has been implicated in regulating motility in *Vibrio alginolyticus* [82-84]. A link

between LuxS and motility, may implicate a role for quorum sensing and the onset of motility prior to symbiont expulsion from the light organ (see below).

Symbiont Signaling

Two-component signaling pathways are important mechanisms by which bacteria can sense the environment and have been identified in *V. fischeri* [85-88]. The roles in colonization for some of these regulators, which were present in our proteomic data (Table 2.4), such as GacA and ArcA, have been studied in detail, and mutagenesis of these genes has demonstrated that they are important in the association [85,87,88]. Although many regulators have already been characterized with respect to the symbiosis, many proteins involved in two-component signaling have unknown functions in the light organ. For example, CpxP, an abundant symbiont protein (Table 2.4), is a periplasmic component of *Escherichia coli* and *Vibrio cholerae* and involved in modulating the cell envelope stress response through CpxAR signaling, thus providing an appealing target for future studies [89,90].

Other Related Stresses

Although several were identified in this study, cold shock proteins have yet to be described with respect to the light organ symbiosis. Of the top 25 most abundant symbiont proteins present in the light organ, three were cold shock proteins (CspD, CspG, and VF_A0595; Table 2.3). Cold shock proteins often bind nucleic acids and function in general stress responses. Furthermore, they have been shown to play a role in regulating bacterial growth at stationary

phase and may even serve as MAMPs recognized by hosts [91,92]. One cold shock protein identified in the light organ, CspD, prevents replication from occurring in stationary phase *E. coli* cells by binding to single stranded DNA and blocking replication [93]. Prior to expulsion at dawn, the symbiont population is at its most dense during the day/night cycle. Therefore, cold shock proteins may play a role in either maintaining high cell densities in the light organ and/or assisting during the transition between the symbiotic and free-living state.

Motility

Research involving the role of motility in the squid-vibrio symbiosis has focused on the initiation of colonization. Within the light organ *V. fischeri* cells become differentiated with the loss of their flagella [94]. Upon release from the light organ at dawn, *V. fischeri* cells are believed to fully regenerate their flagella within several hours [94]. Our proteomic data show the presence and putative identification of several proteins related to flagellar structure including filamental proteins (FlaA, FlaC), basal body proteins (FlgB, FlgH, FlgI), and a motor protein (MotB; Table 2.4). Proteins related to flagellar regulation (FlrA and FlgM) and chemotaxis (CheW and CheZ) were also detected. A recent study indicated an increase in flagellar gene expression by light organ symbionts in the hours preceding dawn and *V. fischeri* mutants of FlaA and FlrA have been shown to be important for symbiotic competence [10,95,96]. FlrA was also found to be expressed by *V. fischeri* in the light organs of *E. scolopes* and a different squid species, *Euprymna tasmanica*, but not in strains grown in seawater [97]. Together, the data from this present study and others suggests that *V. fischeri* cells are generating flagella prior to expulsion from the light organ and may be preparing for the transition from the symbiotic to the

free-living state. Future studies should focus on signals in the changing microenvironment that may initiate this transition.

Symbiont Metabolism

Within the light organ, *V. fischeri* employs a number of metabolic strategies [10,30,98,99]. The daily rhythm of the light organ symbiont population coincides with fluctuations in symbiont metabolism [10]. Transcriptomics revealed a unique pattern in which during the night the symbiont ferments chitin as a means of obtaining energy. After the majority of the symbiont population is expelled from the light organ, the remaining symbionts anaerobically respire glycerol during the hours in which the light organ becomes replenished with a full symbiont population. The results of this study show abundant symbiont chitin binding proteins and chitinases, thus supporting these previous findings (Supplemental file S2.3). The diel shift in metabolism is one piece of evidence that supports the light organ as being a dynamic microenvironment that is under the regulation of both the host and symbiont [10].

Summary

Proteomic studies of symbioses utilizing high-throughput techniques are becoming more common and have been used for analyses of the pea aphid-*Buchnera* symbiosis, nitrogen fixing symbioses of leguminous plants, human gut microbiota, and in characterizing the function of uncultivable symbionts in hydrothermal vent symbioses [100-104]. Characterization of the light organ proteome with high-throughput techniques allowed for the identification of a large number of host and symbiont proteins using little starting material and demonstrates the value of

proteomic analyses in an effort to understand the relationship of a symbiotic association. The results of this study complement prior transcriptomic data, but have also identified a number of proteins of previously unknown function in the squid-vibrio symbiosis [10]. The high-throughput techniques used here offer new methods for identification of host and symbiont proteins likely important for the maintenance of this and other host-microbe associations.

Chapter 3: Putative Role of a *Vibrio fischeri* Immunoglobulin-like Protein in the Colonization of the Squid *Euprymna scolopes*

Abstract

The bioluminescent bacterium *Vibrio fischeri* transitions from free-living to forming an association with the Hawaiian bobtail squid *Euprymna scolopes*. Once the host's light organ is colonized, a number of symbiont genes are induced by quorum sensing, including the genes necessary for the production of light. Among the other *V. fischeri* proteins regulated by LuxI-LuxR quorum sensing is a predicted surface protein (Vig) that contains four bacterial immunoglobulin-like (Big) domains. Mostly studied in pathogenic bacteria, Big domains are important for adhesion to host tissues and evasion of the immune response. In order to characterize the function of this *V. fischeri* protein (Vig) in the squid-vibrio association, PCR fusion was used to generate a mutant strain (*vig*$^-$) that lacks the predicted immunoglobulin-like domains. Colonization assays with freshly hatched squid revealed that after a 3 hour incubation with 5,000 cells/ml of *vig*$^-$, the mutant was able to successfully colonize juvenile *E. scolopes*, showing no differences compared to wild type (n=40, 0-96 hrs). However, competition with wild type in a 1:1 initial inoculum revealed that *vig*$^-$ was significantly outcompeted 24 and 48 hours after inoculation (24 hrs- n=49, Relative Competitive Index (RCI) -0.897, p-value < 0.0001 and 48 hrs- n=50, RCI -1.097, p-value < 0.0001). Further analysis of the mutant revealed that although there is no growth defect in culture compared to wild type, *vig*$^-$ was significantly less luminescent in culture with the presence of exogenous autoinducer 3-O-C6-homoserine lactone (HSL) (38% dimmer by comparison of the maximum relative light units (RLUs); n=4 trials). A similar trend was observed in squid colonized with *vig*$^-$ (n= 40, 0-96 hrs). A polyclonal antibody generated to Vig confirmed that the native 65 kD protein is not produced in the mutant. Western

blot analysis of culture-grown wild type *V. fischeri* revealed that Vig is induced using exogenous 3-O-C6-HSL. In addition, Vig was identified in the outer membranes extracted from *V. fischeri*. These data suggest that Vig plays a role in colonization of the host light organ and may influence luminescence production *in situ*. In a broader context, bacterial Ig-like proteins may be important in both pathogenic and beneficial host-microbe associations.

Introduction

Host-microbe relationships are either vertically or horizontally transmitted. For a horizontally transmitted symbiosis to be successfully initiated, the animal host must acquire the correct microorganism(s) from the environment with fidelity each generation [105]. Bacteria involved in these types of associations, must quickly adapt from a free-living lifestyle to a symbiotic one. The transition between these different microenvironments includes changes in bacterial gene expression, which can be triggered by molecular signals such as the acylated homoserine lactones associated with high cell density quorum sensing [106].

The bioluminescent bacterium, *Vibrio fischeri*, makes the transition from that of a free-living marine microbe to forming an association with the Hawaiian bobtail squid, *Euprymna scolopes* [6]. Inside of the light organ, *V. fischeri* utilizes quorum sensing to induce the genes required for light production [27]. The squid in turn uses the bacterial produced light as an anti-predatory behavior called counter-illumination [11]. In addition to bioluminescence, 18 other genes have now been identified that are regulated by the autoinducer 3-O-C6-homoserine lactone (HSL) [80]. One of the uncharacterized members of this LuxI-LuxR regulon is a predicted surface protein with similarity to the bacterial immunoglobulin-like family. Bacterial immunoglobulin-like proteins are associated with adhesion to host tissues and previous research has focused on the role of bacterial Ig-like proteins in pathogenic bacteria such as *Leptospira interrogans*, *Yersinia pestis*, and *Escherichia coli* [107]. The Leptospiral immunoglobulin (Lig) proteins have been characterized in depth to reveal that in addition to increasing adhesion to extracellular matrix proteins, the Lig proteins are also involved with immune evasion and calcium binding [108-110]. For *Leptospira*, these Ig-like proteins have a clear role in pathogenesis. Non-pathogenic strains of *Leptospira* lack the Lig proteins [111]. No vibrio

bacterial Ig-like proteins have been characterized to date. The symbiosis between *V. fischeri* and *E. scolopes* provides the opportunity to study a bacterial Ig-like protein in a beneficial host-microbe association.

In this study, a strain of *V. fischeri* that lacks the Vibrio immunoglobulin-like protein (Vig) was created in order to characterize the role of this protein in colonization of the squid, *E. scolopes*. The *vig*¯ mutant was tested for the ability to colonize freshly hatched juvenile squid alone and in competition with wild type *V. fischeri*. In addition, *in vitro* growth and luminescence were measured. Using an antibody developed to Vig, the cellular localization of this protein was determined by cell fractionation and western blot analyses. These studies of Vig begin to characterize the role of a bacterial Ig-like protein in a beneficial host-microbe relationship. The results from this study suggest that Vig is a quorum sensing regulated protein that may influence the intensity of luminescence produced by *V. fischeri* and thus the ability of the symbiont to colonize the light organ.

Methods

Bacterial strains and growth conditions

All strains of *V. fischeri* were grown in saltwater tryptone (SWT) or Luria-Bertani salt (LBS) at 28°C preferably to an OD 600 under 1.0. Antibiotics were added as necessary. *V. fischeri* pVSV208 was grown in the presence of 2.5 μg/ml chloramphenicol (Sigma-Aldrich). Strains of *Escherichia coli* were grown in LB at 37°C. Growth requirements for *E. coli* included 20 μg/ml chloramphenicol, 1% glucose and 0.3 mM thymidine for *E. coli* π 3813 KV470, and 100 μg/ml kanamycin for *E. coli* pEVS104.

Construction of the mutant

A *V. fischeri* strain lacking the Ig-domain containing surface protein was created as previously described [112]. This method utilized gene specific primers designed to the 5' and 3' ends of the gene in *V. fischeri* pVSV102, resulting in separate 500-600 bp products (5' product- (VFA0894_F1-ATGAATATTATGATTAAAAAACGTCTAAAA, VFA0894_R2-TAGGCGGCCGCACTTAGTATGGCAACGTCTTCGTTTGATGA; 3' product- VFA0894_F2-CATACTAAGTGCGGCCGCCTAAGACGGTCACATCGCTACCT, VFA0894_R1-CTATCTAGAGGGTTTAGGGATTCTACA). PCR was performed with GoTaq Green Mastermix (Promega) and the following conditions: 3 min at 95°C, 35 cycles of 30s at 95°C, 30s at 50°C, and 1.5 min at 72°C, and a final extension at 72°C for 10 minutes. PCR products were separated on a 1% agarose gel supplemented with a 1:10,000 dilution of SYBR safe DNA gel stain (Life Technologies) and imaged using a Molecular Image Gel Doc (Bio-Rad). These products were joined together by an overlap PCR using the same parameters as above (with primers VFA0894_F1 and VFA0894_R1) and the resultant product was cloned using the pGEM T Easy Vector kit (Promega). The construct was subcloned into the suicide plasmid pSW7848 with the restriction enzymes SpeI and EcorI and then transformed into *E. coli* π 3813. The mutation was introduced into *V. fischeri* pVSV102 by triparental conjugation with the suicide plasmid containing *E. coli* π 3813 and *E. coli* pEVS104 on LBS containing 0.3 mM thymidine. *E. coli* π 3813 was removed by growing the ex-conjugates on LBS supplemented with 1% glucose, 1 μ g/ml chloramphenicol and no thymidine. Ex-conjugate colonies were then plated on LBS containing 1% glucose followed by LBS containing 1 μ g/ml chloramphenicol to confirm that

the suicide plasmid had integrated. To select for the loss of the integrated plasmid, colonies were grown on LBS containing 0.2% arabinose, which induced the CcdB toxin (normally repressed by the glucose in the media) and killed the cells still containing the plasmid. This step resulted in *V. fischeri* containing either the wild type gene or the mutated gene, which was confirmed by PCR using gene specific primers designed to the 5' and 3' ends of *vig* (VFA0894_F1 and VFA0894_R1). The colonies were examined for the loss of the antibiotic resistance enabled by the plasmid by plating on LBS containing 1 μ g/ml chloramphenicol.

Luminescence assay

V. fischeri pVSV102 GFP and *vig*$^-$ were grown in 50 ml of SWT for 6 hours. Every 30 minutes the optical density at 600 nm was recorded and luminescence of the culture was determined by a luminometer (Berthold Detection Systems) over a 10 second interval in relative light units per second (RLU/s). When necessary, cultures in mid-log phase (OD 600 ~0.4) were induced with the addition of 10 μM 3-O-C6-homoserine lactone (K3007; Sigma-Aldrich) and were grown to stationary phase. Additionally, cells from 1 ml aliquots of induced or uninduced cultures were harvested by centrifugation (5,000 x g for 10 minutes at room temperature). Cell pellets were then washed twice in 10 mM phosphate buffered saline (PBS), pH 7.4 and then frozen at -80°C until analysis by SDS-PAGE and western blotting.

Colonization

After hatching, aposymbiotic juvenile squid were placed in 0.2 μM filter sterilized artificial seawater (Instant Ocean; FSASW). For colonization, *V. fischeri* was incubated with squid for 3 hours using an initial inoculum of 5,000 cells/ml. After infection, squid were rinsed

twice in FSASW and placed in fresh FSASW in individual scintillation vials. Colonization was monitored with a luminometer (Berthold Detection Systems).

Competition assay

After hatching, juvenile squid were placed in FSASW. An inoculum consisting of equal numbers (5,000 cells/ml each) of wild type *V. fischeri* pVSV208 (RFP) and *V. fischeri vig*$^-$ (GFP) were added to the FSASW for 3 hours. After 3 hours the squid were removed, rinsed 2 times in fresh FSASW and placed in 5 ml of fresh FSASW. Dilutions of the inoculum used for colonization were plated on SWT to confirm the ratio of both strains. Colonization was monitored at 24 and 48 hours with a luminometer (Berthold Detection Systems). Depending on the time point required, either 24 or 48 hours after infection, squid were rinsed twice in FSASW and individually frozen with liquid nitrogen until analysis. The numbers of wild type *V. fischeri* and *vig*$^-$ that colonized the squid light organ were determined as previously described [113]. Individual squid were homogenized in FSASW and serial dilutions were plated on LBS or SWT. The number of RFP (WT) and GFP (*vig*$^-$) CFUs per light organ were determined using a fluorescent microscope. A relative competitive index (RCI) was calculated by comparing the ratio of *vig*$^-$ (GFP) CFUs to the wild type (RFP) CFUs, which was then normalized to the inoculum ratio. Wild type *V. fischeri* dominated when the RCI was less than 1, while the mutant dominated when the RCI was higher than 1. Log transformed RCIs were analyzed using a one sample t-test and considered significant if the p-value was less than 0.05.

Outer membrane preparation

Outer membranes from *V. fischeri* were prepared using a previously published protocol [114]. Briefly, 250 ml SWT cultures of *V. fischeri* ES114 were grown in either the presence or absence of autoinducer to stationary phase (as described above). The cells were harvested by centrifugation (3,000 rpm for 20 minutes at 4°C) and then resuspended in a 200 mM Tris pH 8.0, 500 mM Sucrose, 1mM EDTA (TSE) buffer with the addition 2 mM phenylmethanesulfonyl fluoride (PMSF, Sigma-Aldrich) for 30 minutes on ice. The cells were again pelleted by centrifugation at 16,000 x g for 30 minutes at 4°C (Eppendorf 5415 R). The supernatant contained the cell envelopes that were released from the whole cells by diffusion. Outer membrane proteins were separated from periplasmic proteins by using a micro-ultracentrifuge (RP55-S swinging bucket rotor) at 100,000 x g for 1 hour at 4°C. The resulting pellet of *V. fischeri* outer membranes was resuspended in 1 ml of fresh TSE buffer. Proteins were precipitated with 10% (w/v) trichloroacetic acid as previously described [115]. The resulting protein pellets were resuspended directly in 100-200 µl of Laemmli sample buffer (Bio-Rad) and analyzed by SDS-PAGE and western blot analysis (see below).

Western blot analysis

A rabbit polyclonal antibody was generated (GenScript) using a peptide unique to Vig (GTFDDHTARELDPS). Peptides were chosen based on predicted antigenicity and inspected for homology to any other known *E. scolopes* or *V. fischeri* proteins. *V. fischeri* lysates and outer membrane extractions originating from 3-O-C6-HSL induced or uninduced cultures were analyzed by SDS-PAGE and western blot analysis. Protein concentrations were determined by the *RC DC* protein assay (Bio-Rad). *V. fischeri* proteins (5-20 µg) were separated using 12%

TGX Criterion gels (Bio-Rad) at 200V for 45 minutes. Proteins were transferred onto a .2 μM mini nitrocellulose membrane using a Trans-Blot Turbo transfer system and the Mixed Molecular Weight protocol per the manufacturer's instructions (Bio-Rad). Membranes were blocked in 4% milk, 0.1% Tween at room temperature for one hour. Anti-Vig was diluted in block 1:50,000 and incubated with the blot at room temperature for one hour. After washing the blot 4 times for 5 minutes each in 0.1% Tween, tris buffered saline (TTBS), pH 7.4, a goat anti-rabbit (IgG H+L) HRP conjugate (1:5,000, Bio-Rad) was diluted in fresh block and then incubated with the blot for an additional hour at room temperature. The blot was washed again 6 times in TTBS for 5 minutes each and the antibody was detected using an Amersham Enhanced Chemiluminescence (ECL) Prime Detection kit (GE Healthcare Life Sciences) and imaged using a FluorChem HD2 CCD camera (Cell Biosciences) with an exposure time of 30 seconds.

Transmission electron microscopy

Whole juvenile squid colonized with either wild type *V. fischeri* pVSV102 or *vig*$^-$ were prepared for transmission electron microscopy at 48 hours post infection as previously described with several modifications [116]. Colonization of the light organ by the correct strain was confirmed by plating a subset of the squid on to SWT and checking the resultant colonies with PCR and primers (VFA0894_F1 and VFA0894_R1) designed to 5' and 3' ends of *vig* as described above. Squid were anesthetized on ice and fixed at room temperature in 2.0% paraformaldehyde, 2.5% glutaraldehyde in a 0.1 M sodium cacodylate, 0.395 M NaCl, 1.5 mM $CaCl_2$, 1.5 mM $MgCl_2$, pH 7.4 buffer (Buffer 1). After an initial 15-min fixation, the squid were placed into fresh fixative and stored at 4°C. Following fixation, the squid were washed several times in cold Buffer 1 and stored at 4°C overnight. Whole squid were postfixed in a solution of

1% osmium tetroxide, 0.8% potassium ferricyanide, 0.1 M sodium cacodylate, 0.375 M NaCl, pH 7.4 for 1.5 h at 4°C and then washed in distilled water, dehydrated with ethanol, cleared in 100% acetone, and embedded in a medium formulation epoxy resin (Embed 812, Electron Microscopy Sciences). Thin (80-nm) sections were sliced using a diamond knife on a Leica Ultracut UCT and stained with 2% uranyl acetate and Reynolds' lead citrate. The stained sections were imaged using an FEI Tecnai Biotwin G2 Spirit electron microscope operated at 80 kV.

Phylogenetic analysis

Eleven bacterial immunoglobulin-like proteins predicted to be similar to Vig of *V. fischer*i ES114 by BLASTp against the NCBI nr database (Table 3.1) were compared to the Conserved Domains Database (CDD) using RPS BLAST [117]. Individual bacterial immunoglobulin-like domains were identified using the CDD identifier BID_2- 214752 and then analyzed with SeaView (version 4.4.2) [118]. MUSCLE was used to align the sequences and PhyML was used to create a maximum likelihood tree based on the LG model and 1000 bootstrap replicates [118]. The tree was rooted to the bacterial immunoglobulin-like domain of intimin from *Escherichia coli.*

Table 3. 1: Bacterial immunoglobulin-like protein sequences used for phylogenetic comparisons.

Sequence ID	Protein and Organism
499582929	Surface protein, *Vibrio fischeri* ES114
545472152	Hypothetical protein, *Vibrio azureus*
501528933	Ig domain protein, group 2 domain protein, *Vibrio fischeri* MJ11
550256830	Putative Bacterial Ig-like, group 2, *Vibrio nigripulchritudo*
503623718	Surface protein, *Vibrio anguillarum*
499462021	Surface protein, *Vibrio vulnificus*
487887567	Surface protein Lk90-like protein, *Vibrio cholerae*
492970028	Ig-like repeat domain protein 1, *Vibrio shilonii*
545125133	Bacterial Ig-like domain family protein, *Vibrio parahaemolyticus*
203287111	LigA, *Leptospira interrogans* serovar Kennewicki
2565325	Intimin, *Escherichia coli*

Results

VF_A0894 (*vig*) is a 1,797 bp gene located on the second chromosome of *V. fischeri* ES114 that codes for a 598 amino acid protein with 4 predicted bacterial immunoglobulin-like group 2 domains and a lipid attachment motif (Fig. 3.1). Protein domains were predicted using the Conserved Domains Database (CDD) [117]. Comparison of *V. fischeri* Vig with bacterial proteins present in the NCBI nr protein database revealed that Vig shares homology with other proteins containing bacterial immunoglobulin-like domains (Table 3.1). A maximum likelihood phylogenetic tree was assembled using MUSCLE and PhyML with 61 individual bacterial immunoglobulin-like domains (Fig. 3.2). Immunoglobulin-like domains from various vibrios were compared to the already well-characterized bacterial Ig-like domains from *L. interrogans* (LigA) and *E. coli* (intimin). The 4 predicted immunoglobulin-like domains of Vig do not group together and are scattered throughout the tree (Fig. 3.2). However, bacterial immunoglobulin-like domains from LigA of *L. interrogans* and the Ig-like protein of *V. fischeri* MJ11 generally do cluster together. The bacterial Ig-like domains from Vig appear to be unique and very different from the other available sequences.

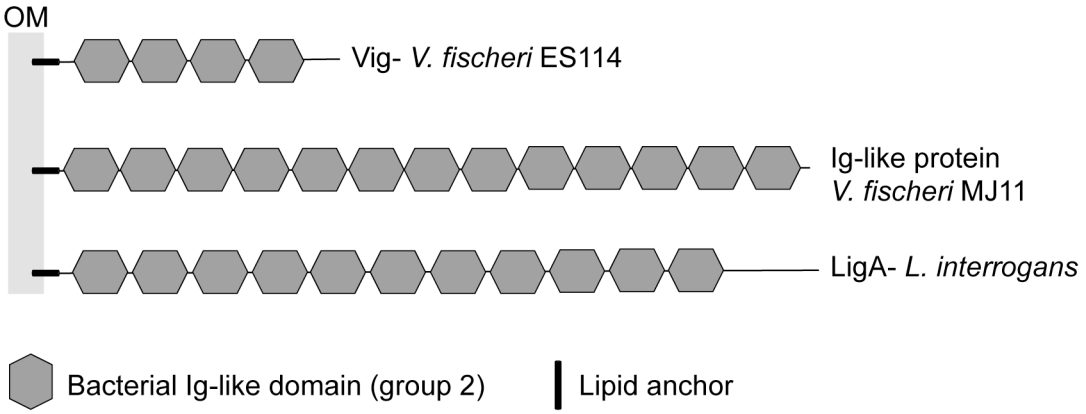


Figure 3. 1: Protein domain organization of immunoglobulin-like proteins from *V. fischeri* ES114, *Vibrio fischeri* MJ11, and *Leptospira interrogans*.

To better understand the function of Vig in the squid-vibrio symbiosis, a *V. fischeri* strain lacking a native Vig was generated. A 633 bp region was deleted from the middle of the gene resulting in a 1,164 bp-truncated product lacking all 5 predicted bacterial Ig-like domains. The mutation introduced into *V. fischeri* generated a stop codon within the first bacterial Ig domain of *vig*, leaving a 185 amino acid protein with a predicted lipid attachment site.

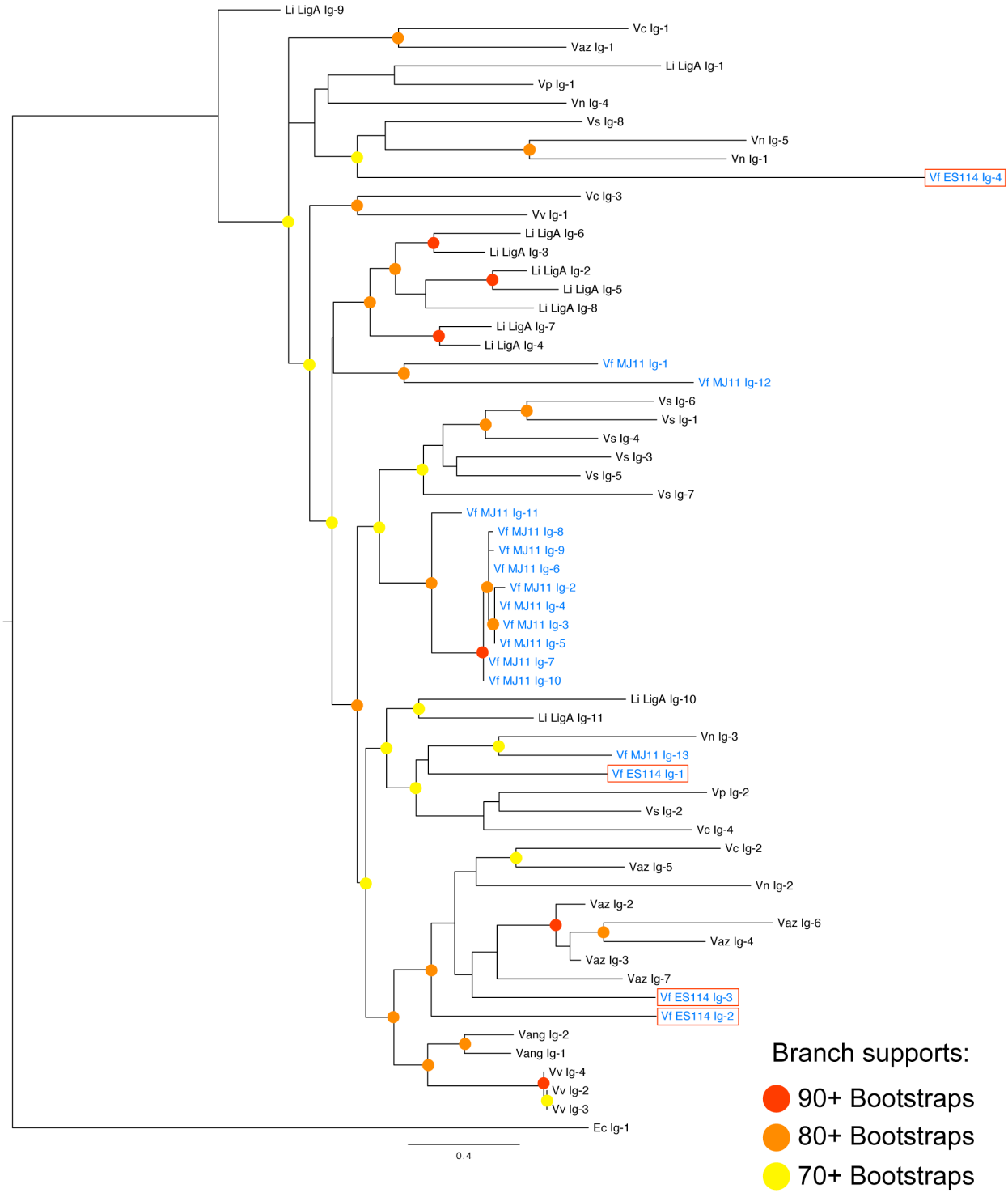


Figure 3. 2: Phylogenetic reconstruction of bacterial immunoglobulin-like domains.

A set of 61 bacterial immunoglobulin-like domains originating from 11 protein sequences was aligned with MUSCLE. A maximum likelihood tree was created with PhyML using the LG model and 1000 bootstrap replicates. The tree is rooted with a bacterial immunoglobulin-like domain from intimin of *E. coli.* Species highlighted in blue are known to participate in symbiotic interactions with animal hosts. Abbreviations: Ig- immunoglobulin-like domain, Vf- *Vibrio fischeri*, Vc- *Vibrio cholerae*, Vs- *Vibrio shilonii*, Vaz- *Vibrio azureus*, Vn- *Vibrio nigripulchritudo*, Vang- *Vibrio anguillarum*, Vv- *Vibrio vulnificus*, Vp- *Vibrio parahaemolyticus*, Li- *Leptospira interrogans*, Ec- *Escherichia coli*

The presence of Vig was confirmed using a rabbit polyclonal antibody developed to Vig. Western blot analysis revealed that wild type cultures of *V. fischeri* did not show detectable levels of Vig, however wild type *V. fischeri* grown in the presence of exogenous autoinducer produced a 65 kD protein corresponding to the predicted molecular weight of Vig (Fig. 3.3). In comparison, the *vig* - did not produce any detectable protein in both the control and induced cultures, supporting the loss of a functional Vig (Fig. 3.3).

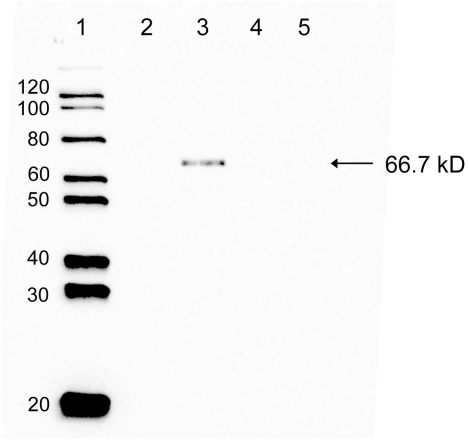


Figure 3. 3: Vig is produced in response to the autoinducer 3-O-C6-homoserine lactone.

Culture grown WT *V. fischeri* ES114 and *vig* - were grown in the presence or absence of autoinducer and harvested in stationary phase. Cell lysates were then analyzed by a western blot analysis with a polyclonal antibody specific to Vig (1-protein standard, kD, 2-WT *V. fischeri*, 3-WT *V. fischeri*+ 3-O-C6-HSL, 4- *vig* -, 5- *vig* - + 3-O-C6-HSL). Vig is indicated by a prominent protein band of 66.7 kD.

Vig is predicted to be a surface protein based on a lipid attachment motif identified at the N-terminal portion of the primary amino acid sequence. To determine the localization of Vig, cell fractionation was used to isolate outer membranes from *V. fischeri*. Outer membranes were separated from *V. fischeri* cells after treatment with a Tris-sucrose-EDTA buffer and ultra centrifugation. The resulting outer membrane pellet was analyzed with SDS-PAGE and a western blot analysis to identify the presence of Vig. Outer membranes from both *V. fischeri*

uninduced and induced with 3-O-C6-HSL contained Vig as indicated by the faint 65 kD bands (Fig. 3.4, Lanes 2 and 3). However, Vig was also detectable in the periplasmic fractions of both samples (Fig. 3.4, Lanes 4 and 5). Both the outer membrane and periplasmic fractions contained lower levels of Vig compared to the induced lysate sample (Fig. 3.4, Lane 7).

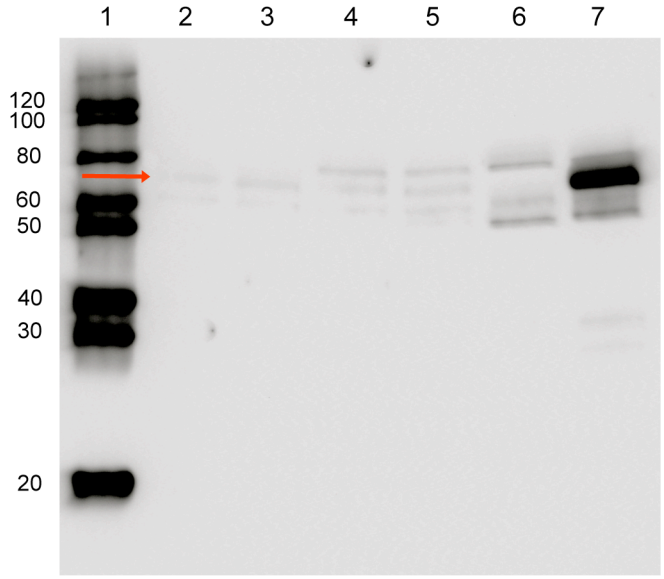


Figure 3. 4: Evidence that Vig can be associated with the outer membrane of *V. fischeri*.

Outer membranes (OM) extracted from WT *V. fischeri* ES114 grown in either the presence or absence of 3-O-C6-HSL were pelleted by ultra centrifugation. The different cellular fractions were analyzed by western blot with a polyclonal antibody specific to Vig (1: protein standard, kD, 2: Outer Membranes, 3: + 3-O-C6-HSL- Outer Membranes, 4: Periplasm, 5: + 3-O-C6-HSL- Periplasm, 6: Lysate, 7: +3-O-C6-HSL Lysate).

The ability of the *vig* ⁻ to colonize freshly hatched juvenile *E. scolopes* was characterized. The mutant was able to colonize the host light organ with no detectable difference from wild type over the first several days (Fig. 3.5A). A separate comparison of the CFUs per light organ for individual squid colonized with either WT or *vig* ⁻ at 48 hours post-inoculation also revealed no significant differences (WT: $5.8x10^4 \pm 1.4x10^4$ CFUs, *vig* ⁻: $7.7x10^4 \pm 1.8x10^4$ CFUs, n=31 squid for each strain; Supplemental file S3.1). However, analysis of the *in vivo* luminescence per animal, as measured by a luminometer, revealed that colonization by the *vig* ⁻ resulted in less

luminescent squid (Fig. 3.5B). This difference in light production was not determined to be significant due to the variation in luminescence per squid (unpaired unequal variance t-test). To further characterize this result, the luminescence of *vig*⁻ was compared to wild type in culture. While there is no detectable difference between the growth rate of wild type and *vig*⁻, in the presence of exogenous autoinducer the mutant never reached the peak luminescence achieved by wild type *V. fischeri* (Fig. 3.6A). The average maximum luminescence of *vig*⁻ was 38% dimmer than wild type (Fig. 3.6B).

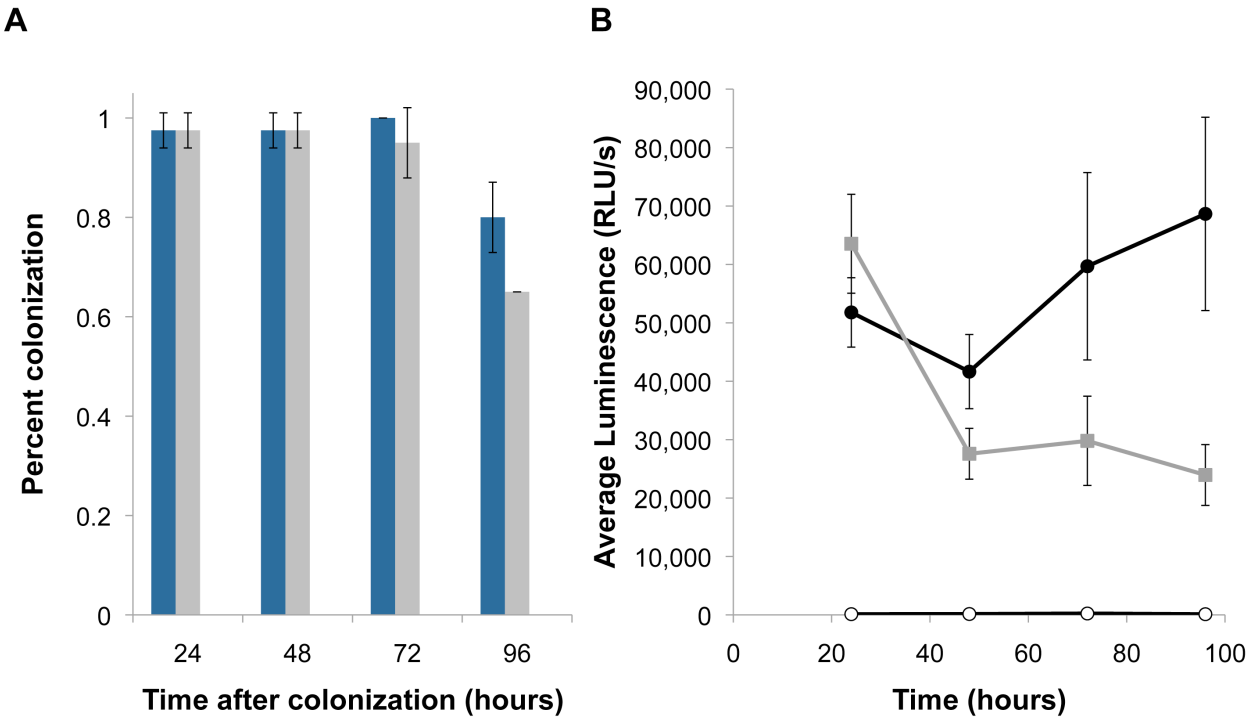


Figure 3. 5: WT *V. fischeri* and *vig*⁻ colonization of the juvenile light organ.

(A). Juvenile squid exposed to WT *V. fischeri* (blue bars) and *vig*⁻ (gray bars) for 3 hours at 5,000 cells/ml were monitored for colonization (determined by luminescence) over a period of 96 hours (2 trials, n=20 squid per strain). Aposymbiotic controls were never colonized during the 96 hour experiment. **(B)**. One representative trial of the average luminescence (relative light units/second, RLU/s) for squid colonized by either WT *V. fischeri* or *vig*⁻ (black circles- WT *V. fischeri*, gray squares- *vig*⁻, open circles- aposymbiotic controls; error bars represent the standard error of the mean for each time point).

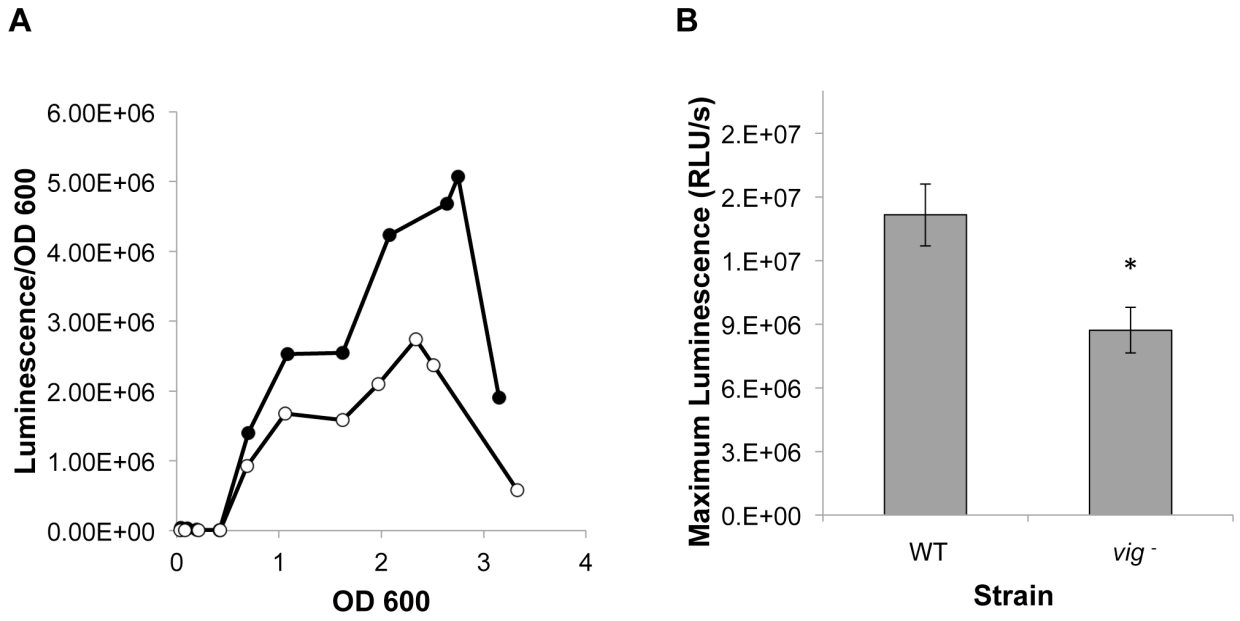


Figure 3. 6: *vig*⁻ grown in culture with exogenous 3-O-C6-HSL produced lower levels of bioluminescence.

(A). Representative experiment for luminescence produced by each strain in culture (Black circles- WT *V. fischeri*, open circles- *vig*⁻). **(B)**. Average maximum luminescence produced by each strain. WT *V. fischeri* produces significantly more luminescence in response to exogenous 3-O-C6-HSL (n= 4 separate experiments, $p < 0.05$, two tailed equal variance t-test, error bars represent SEM).

Although *vig*⁻ was able to colonize the juvenile squid light organ by itself, when the mutant was challenged with wild type in an initial 1:1 inoculum, *vig*⁻ was significantly outcompeted at 24 and 48 hours post-infection. The amount of wild type *V. fischeri* and *vig*⁻ present in the light organ was determined by counting the number of RFP (WT) and GFP (*vig*⁻) colonies that developed on SWT agar plates from homogenized juveniles. The ratio of *vig*⁻ CFUs to WT CFUs (normalized to the initial inoculum) was log transformed and represented a relative competition index (RCI) that was used to reveal if the mutant had a defect in colonizing the light organ in the presence of a competitor strain. Only animals co-colonized by both strains were used to determine the RCI. At 24 hours the mutant was outcompeted by wild type in a majority of the

squid analyzed with an average Log(RCI) of -0.897 (Fig. 3.7A, n=49, p-value < 0.0001). At 48 hours the difference in colonization between wild type and *vig*⁻ was even more prominent as more squid were dominated by wild type *V. fischeri* with an average Log(RCI) of -1.097 (Fig. 3.7B, n=50, p-value < 0.0001). Squid colonized exclusively by either wild type or *vig*⁻ were also recorded. For 5 out of the 8 trials between 24 and 48 hours, a majority of the squid colonized by one strain, were only colonized by wild type (Table 3.2; 35 WT to 6 *vig*⁻). Only one trial (24 hr 4, Table 3.2) resulted in more squid exclusively colonized by *vig*⁻ (Table 3.2; 10 *vig*⁻ to 1 WT).

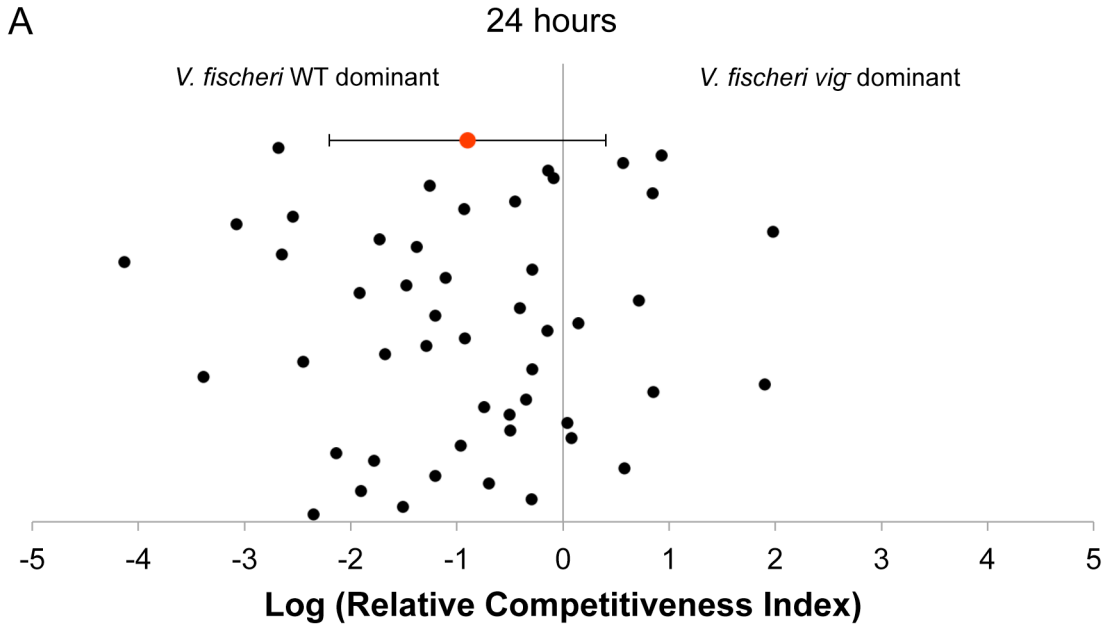
Transmission electron microscopy (TEM) of light organs colonized by either wild type or *vig*⁻ was used to determine if there were any differences in adhesion between the two strains. Both wild type and *vig*⁻ demonstrate intimate contact with the host microvillar brush border of the light organ epithelial tissue (Fig. 3.8). In addition, the bacterial cells themselves appear in close contact with each other for both strains. There appears to be a lower number of *vig*⁻ cells present in the representative light organ TEM, but this may be the result of individual variation (Fig. 3.8B). Previous light organ colony counts for wild type or *vig*⁻ colonized squid have determined that there is no difference between the two strains (Supplemental file S3.1).

Table 3. 2: Summary of *vig*⁻ and wild type co-colonization experiments.

Experiment	Squid	Inoculum [a]	Co-colonized	WT only	*vig*⁻ only	Log(RCI) [b]
24 hr 1	20	0.85	16	4	0	-0.890
24 hr 2	20	1.15	9	1	10	-0.358
24 hr 3	18	1.22	16	2	0	-0.791
24 hr 4	21	1.84	8	9	4	-1.728
48 hr 1	32	1.22	22	5	5	-0.871
48 hr 2	15	2.33	2	11	2	-1.164
48 hr 3	12	0.95	12	0	0	-1.114
48 hr 4	23	0.85	14	9	0	-1.429

a. Ratio of the initial inoculum used to colonize squid (*vig*⁻ to WT)

b. Log(RCI)- relative competitive index, ratio of *vig*$^-$ CFUs to WT CFUs, normalized to the inoculum for each experiment and then log transformed. The average Log(RCI) described in the text for each time point was calculated separately from a complete list of co-colonized squid.



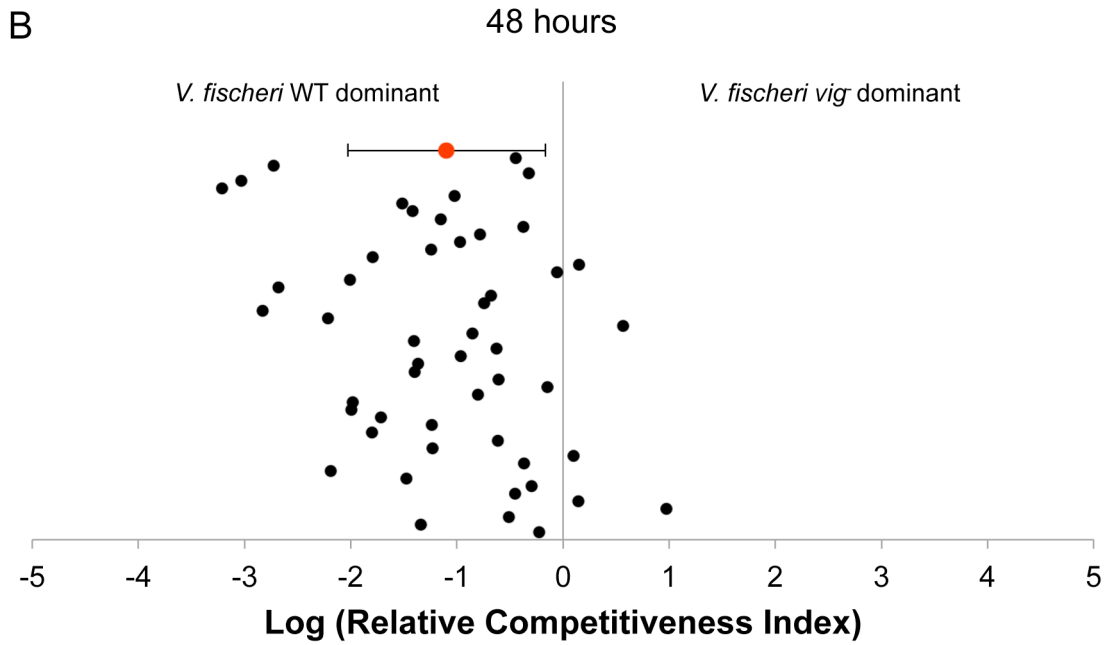


Figure 3. 7: WT *V. fischeri* dominates *vig*$^-$ during co-colonization of the juvenile light organ.

Juvenile squid were exposed to a 1:1 ratio of WT *V. fischeri* RFP and *vig*$^-$ GFP for 3 hours. Fluorescent colonies arising from individual light organs were counted and compared using a relative competition index (RCI, *vig*$^-$ CFUs: WT CFUs normalized to the initial inoculum). **(A)**. Log(RCI)s of co-colonized squid 24 hours post inoculation (n=49, Avg. Log(RCI) -0.897, p-value < 0.0001). **(B)**. Log(RCI)s of co-colonized squid 48 hours post inoculation (n=50, Avg. Log(RCI) of -1.097, p-value < 0.0001). Red circles represent the average Log(RCI). Error bars represent standard deviation.

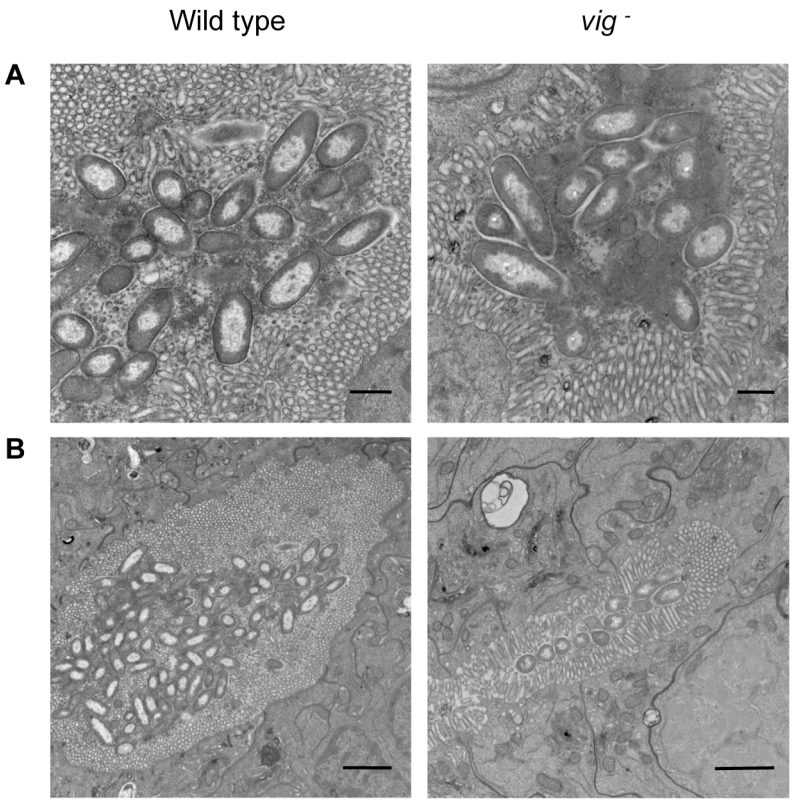


Figure 3. 8: Transmission electron micrographs of 48-hour juvenile light organs colonized with either wild type *V. fischeri* or *vig*$^-$.

V. fischeri lacking Vig remain in contact with the host microvilli brush border (scale- **(A)** 500 nm, **(B)** 2 μM).

Discussion

Within the light organ of the squid *Euprymna scolopes*, the population of *Vibrio fischeri* undergoes a daily change in cell number [6,12]. At night, when the light organ is full, the symbiont can be found tightly associated with the host epithelial tissue and neighboring *V. fischeri* cells [119]. This intimate contact between host-microbe and microbe-microbe provides an opportunity for the exchange of molecular information critical to a functional symbiosis. Quorum sensing is one mechanism of communication used by bacteria in which cells can regulate a set of genes in a cell density dependent manner [106]. In *V. fischeri*, quorum sensing has been extensively studied in regard to bioluminescence, but much less is known about the functions of other proteins influenced by autoinducer [80,120]. In this study a bacterial immunoglobulin-like protein from *V. fischeri* (Vig), previously shown to be up-regulated in the presence of the autoinducer 3-O-C6-homoserine lactone (HSL), was characterized in regards to colonization of the squid light organ. The results provided here suggest that Vig is required to colonize the light organ in the presence of a competitor strain and is associated with the production of normal levels of luminescence. Bacterial immunoglobulin-like proteins now represent a group of proteins that play a role in both pathogenic and beneficial host-microbe interactions.

Colonization of the light organ by *V. fischeri* induces a number of changes to the light organ epithelial tissue. The epithelial cells surrounding the crypt spaces swell in volume and the density of the microvillar brush border increases [121]. In addition, a large portion of the symbiont population (90-95%) is expelled each morning as part of a control mechanism by the host, yet it is unclear how the remaining 5-10% of the population is able to maintain colonization of the light organ [12,13]. The close association between *V. fischeri* and host cells in this

microenvironment indicates that adhesion might contribute to these different events, however the exact mechanisms are not completely understood. In an attempt to address these questions a predicted *V. fischeri* surface protein, Vig, which contains 4 bacterial immunoglobulin-like domains was characterized. Comparison of the individual domains to other Ig-like domains from various vibrios, *E. coli* and *L. interrogans* revealed that the domains from the squid symbiont are distinct from each other and divergent from the other available bacterial Ig-like domains (Fig. 3.2). The Ig-like domains from Vig are not similar to the Ig-like domains from another immunoglobulin-like protein of *V. fischeri* MJ11, the symbiont of the Japanese Pinecone fish light organ, suggesting that these proteins may have different functions even though they are utilized within similar microenvironments.

In addition, Vig is predicted to be associated with the cell surface based on a lipid attachment motif located at the N-terminus of this protein (Fig. 3.1). One hypothesis is that *V. fischeri* uses Vig to adhere to host tissue inside the light organ. Membrane proteins of *V. fischeri* are known to play an important role in this light organ symbiosis. For example, the *V. fischeri* outer membrane protein OmpU is critical for proper symbiont recognition by the hemocytes of *E. scolopes*. Normally, *V. fischeri* adheres to the hemocytes at low levels, but strains lacking OmpU bind to hemocytes at significantly higher numbers [5]. The immunoglobulin-like proteins of pathogens such as *L. interrogans* (Lig), *E. coli* (intimin), and *Yersinia pestis* (invasin) are also important for adhesion to host tissues [107]. In particular, LigA and B of *L. interrogans* enhance adhesion to extracellular matrix proteins such as laminin, collagen, and fibronectin [81,109]. The link between Vig and adhesion remains unknown, but purified outer membranes collected from *V. fischeri* provided evidence for the possible presence of Vig on the cell surface (Fig. 3.4). However, Vig was also identified in the periplasmic fractions and was detected at lower levels in

both the outer membrane and periplasm when compared to the total lysate from an induced culture (Fig. 3.4). It is possible that the outer membrane extraction protocol used in this study was not optimal for *V. fischeri*. Alternatively, Vig could be a predominantly cytoplasmic protein and is rarely or never translocated to the surface. Future work should utilize a different cellular fractionation approach to confirm the localization of Vig.

The Lig proteins of *Leptospira* are only induced in the host [111]. This is similar to what has already been demonstrated for *vig*, as the transcript is only induced in the presence of autoinducer, which occurs with the high cell densities found in the squid light organ [80]. Transmission electron microscopy (TEM) was used in an attempt to identify any anatomical changes in juvenile light organs colonized by either *vig*⁻ or wild type *V. fischeri*, however no detectable differences between the two strains were identified (Fig. 3.8). *V. fischeri* likely has numerous surface proteins involved in adhesion and thus the symbiont may have redundant capabilities to attach to host cells. Future work should characterize if *V. fischeri vig*⁻ has any deficiencies in adhering to extracellular matrix proteins in comparison to wild type.

V. fischeri vig⁻ was able to colonize the light organ with the same efficiency as wild type *V. fischeri* (Fig. 3.5). However, competition of *vig*⁻ with wild type demonstrated that *vig*⁻ was unable to colonize the light organ to the same level. At both 24 and 48 hours *vig*⁻ was significantly outcompeted by wild type *V. fischeri* (Fig. 3.7). It was initially believed that in the presence of wild type, *vig*⁻ might not be able to compete for exclusive binding sites and nutrient access, resulting in lower population numbers. However, analysis of *vig*⁻ also revealed that this strain does not produce the same amount of luminescence in culture and in the light organ of *E. scolopes* as compared to wild type (Fig. 3.5B, 3.6). This represents the first bacterial immunoglobulin-like protein associated with a bioluminescence phenotype. Luminescence is a

critical component for a proper functioning squid-vibrio association. *V. fischeri* strains deficient in the ability to produce light are unable to maintain a symbiosis with the squid [122]. Proteins responsible for the production of light (luciferase, LuxA and B) and the regulation of quorum sensing (LuxR and LuxI) are essential for *V. fischeri* to achieve normal levels of colonization in the light organ [122]. Previous studies have also demonstrated that the presence of luminescent cells does not rescue non-luminescent cells in the light organ [122]. This is similar to what has been reported in this study with *vig*$^-$/WT colonized light organs. It is possible that when colonizing the light organ alone *vig*$^-$ produces enough luminescence to remain a colonizer, but in competition with wild type, the dimmer *vig*$^-$ is removed by an unknown mechanism. The host appears to be able to differentiate between luminescent cells and dim or dark cells. The light organ epithelial cells express the genes necessary to respond to light, which may represent one mechanism the host utilizes to select for proper light production by the resident symbiont [123]. In addition, it was recently discovered that the luminescence of *V. fischeri* influences the transcription of a cryptochrome, a gene that is involved in maintaining circadian rhythms [124]. However, the molecular pathways regulated by light, visual transduction, and cryptochromes in the light organ remain to be characterized.

One possible mechanism by which Vig could influence luminescence is through the binding of cations such as Mg^{2+} or Ca^{2+}. Bacterial Ig-domains from LigA and LigB of *L. interrogans* have a demonstrated an affinity for calcium [110,125]. Furthermore, calcium appears to influence LigB adhesion to fibronectin [110]. Cations, specifically magnesium, have already been shown to effect *V. fischeri* in respect to regulation of flagellation and bioluminescence [126,127]. The PhoQ-PhoP two component regulatory system represses bioluminescence in the presence of low magnesium, however there appears to be an independent, and unexplained

mechanism utilized by *V. fischeri* to increase luminescence in the absence of *phoQ* [126]. Less is known about the impact of calcium levels on *V. fischeri*. It is possible that Vig enhances the affinity of *V. fischeri* to magnesium and/or other cations, resulting in increased luminescence of the symbiont. Future work will involve determining the affinity and influence of cations on Vig and luminescence.

In conclusion, the function of bacterial Ig-like proteins likely depends on the context of the relationship with the host. This study, along with previously published data, demonstrates that bacterial Ig-like proteins are induced in both pathogenic and beneficial associations. As more bacterial-immunoglobulin-like proteins are characterized in depth, it is increasingly clear that this group of proteins is not only important for adhesion, but also immune evasion, cation binding, bioluminescence (this study) and a wide variety of other cellular processes [107-109]. Here we show that Vig plays a role in colonization of the host light organ and may influence luminescence production *in situ*, but the exact mechanism of how Vig functions in the light organ remains unknown. This initial description of Vig provides the opportunity to further characterize the functions of an immunoglobulin-like protein, particularly in regards to quorum sensing, bioluminescence, and adhesion in a beneficial host-microbe association.

Chapter 4: Colonization state influences the hemocyte proteome in a beneficial squid-vibrio symbiosis

Abstract

The squid, *Euprymna scolopes*, and the luminescent bacterium, *Vibrio fischeri*, form a highly specific beneficial light organ symbiosis. Not only does the host have to select *V. fischeri* from the environment, but it must also prevent subsequent colonization by non-symbiotic microorganisms. Host macrophage-like hemocytes are believed to play a role in mediating the symbiosis with *V. fischeri*. Previous studies have shown that the colonization state of the light organ influences the host's hemocyte response to the symbiont. To further understand the molecular mechanisms behind this process, two quantitative mass spectrometry-based proteomic techniques, isobaric tags for relative and absolute quantification (iTRAQ) and label-free spectral counting, were used to compare and quantify the adult hemocyte proteomes from colonized (sym) and uncolonized (antibiotic-treated/cured) squid. Overall, iTRAQ allowed for the quantification of 1,024 proteins with 2 or more peptides. Thirty-seven unique proteins were determined to be significantly different between sym and cured hemocytes (p-value < 0.05), with 20 more abundant proteins and 17 less abundant in sym hemocytes. The label-free approach resulted in 1,241 proteins that were identified in all replicates. Of 185 unique proteins present at significantly different amounts in sym hemocytes (as determined by spectral counting), 92 were more abundant and 93 were less abundant. Comparisons between iTRAQ and spectral counting revealed that 30 of the 37 proteins quantified by iTRAQ exhibited similar trends identified by the label-free method. Both proteomic techniques mutually identified 16 proteins that were significantly different between the two groups of hemocytes (p-value < 0.05). The presence of *V.*

fischeri in the host light organ influenced the abundance of proteins associated with the cytoskeleton, adhesion, lysosomes, proteolysis, and the innate immune response. These data provide evidence that colonization by *V. fischeri* alters the hemocyte proteome and reveals proteins that may be important for maintaining host-symbiont specificity.

Introduction

Animals must have mechanisms to differentiate between their normal microbiota and pathogenic microorganisms. In order to recognize microorganisms, vertebrates utilize adaptive immunity, which is mediated by the ability to produce antibodies to specific antigens. However, invertebrates lack this adaptive component and immunological memory is poorly understood in these animals [128]. Instead, invertebrates rely solely on components of the innate immune system to interact with microorganisms. Long recognized for its ability to remove pathogens, the innate immune system has more recently been studied for interactions that foster benign and beneficial symbioses [1]. Factors such as reactive oxygen and nitrogen species, antimicrobial peptides, and complement-like proteins have a function in maintaining populations of beneficial bacteria in a number of animal hosts [1]. In addition, invertebrates have a cellular component of their innate immune system, specifically macrophage-like-cells, which can bind and phagocytose bacteria [5,129,130]. These cells utilize pattern recognition receptors (PRRs), which can detect microbe-associated molecular patterns (MAMPs) such as peptidoglycan and lipopolysaccharide (LPS), to activate an immune response [1]. A growing body of evidence from several invertebrates suggests that these phagocytic cells are capable of responses to specific bacteria [5,131,132].

The Hawaiian bobtail squid, *Euprymna scolopes*, is used as a model organism to study the role of the innate immune response in a beneficial symbiosis [1,46]. Upon hatching, the host establishes a highly-specific binary association with the bacterium *Vibrio fischeri*, which colonizes a specialized light organ [6]. Inside this organ *V. fischeri* is extracellular and maintained within epithelium-lined crypt spaces. Although the light organ is open to the environment, non-symbiotic bacteria are excluded, and *V. fischeri* remains the sole symbiont for

the duration of the animal's life. While the mechanisms behind this specificity are not completely understood, studies suggest that macrophage-like hemocytes play a role [1,5,15,46].

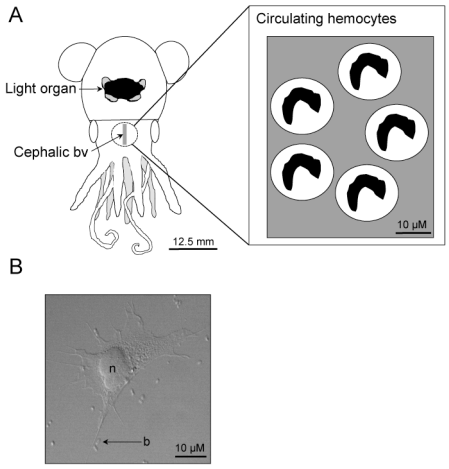


Figure 4. 1: *Euprymna scolopes* has one type of blood cell, the macrophage-like hemocyte.

(A). The dorsal view of *E. scolopes* reveals the location of the light organ and the cephalic blood vessel (bv). The hemocytes circulate through the vasculature of the squid (scale 12.5 mm). **(B)**. A DIC image of an adult *E. scolopes* hemocyte interacting with *Vibrio fischeri* cells (b- bacteria, n- nucleus, scale 10 μM).

Hemocytes are the only type of blood cell found in *E. scolopes* (Fig. 4.1). Upon initiation of the symbiosis, these hemocytes traffic to the juvenile light organ in response to *V. fischeri* [14]. Within the colonized juvenile light organ crypt spaces, hemocytes with internalized bacteria have been observed [13]. It is unclear whether these bacteria are *V. fischeri* or non-symbiotic interlopers that have entered the organ from the environment. However, after the symbiosis has had time to mature, the hemocytes of an adult squid have never been visualized with engulfed bacteria in the light organ, even amongst the dense population of *V. fischeri* in the crypt spaces. A previous binding study revealed that adult hemocytes adhere to the symbiont significantly less than related non-symbiotic bacteria [5]. Removal of the symbiont from the light organ with antibiotics leads to hemocytes that adhere to *V. fischeri* at significantly higher numbers, suggesting colonization induces host immune tolerance of the symbiont [5].

Although the molecular mechanisms behind this change in hemocyte response to *V. fischeri,* remain poorly characterized, a transcriptome and proteome identified hemocyte genes and proteins that may be important to the symbiosis [15]. Quantitative PCR revealed that nitric oxide synthase (NOS), complement component C3 (C3), and peptidoglycan recognition protein 5 (PGRP 5) were differentially expressed between the hemocytes from colonized (sym) and cured squid [15]. In order to further understand the effect of light organ colonization on hemocyte function, two quantitative MS-based proteomic techniques, isobaric tags for relative and absolute quantification (iTRAQ), and label-free spectral counting, were used to compare the proteomes of sym and cured hemocytes. Proteins that showed significant differences in abundance in sym hemocytes included those involved with the immune response, adhesion, the cytoskeleton, and lysosomal processes. These data, along with the results from previous studies, suggest that colonization by *V. fischeri* alters the host's immune response by influencing the hemocyte proteome to favor tolerance of the symbiont.

Methods

Animal collection and maintenance

Adult *Euprymna scolopes* were caught within the shallow sand flats off of Oahu, HI by dip net. Hemocyte samples not collected immediately in Hawaii were obtained from animals maintained at the University of Connecticut in artificial seawater (ASW, Instant Ocean) at 23°C on an approximate 12hr light / 12hr dark cycle [15,115]. All animals were acclimated at least 48 hours under laboratory conditions prior to sample collection.

Curing experiments (symbiont removal)

Vibrio fischeri was cured from the light organ using a cocktail of chloramphenicol (20 μg /m1; Sigma Aldrich, St. Louis, MI) and gentamicin (20 μg /m1; Affymetrix/USB, Cleveland, Ohio) for 5 consecutive days [5,15]. Curing was confirmed by homogenizing the symbiont-containing central core of the light organ, and plating the homogenate on seawater tryptone agar [5,15]. Plates with central core homogenates from cured hosts contained no detectable *V. fischeri* colonies after an overnight incubation at 28°C. Hemocytes from cured hosts were collected and prepared as described below.

Hemocyte collection and protein extraction

Squid hemocytes were collected from adult *E. scolopes* as previously described [5,15,16]. Approximately 50-100 μl of hemolymph (~5,000 hemocytes/ μl; [5]) was extracted from the cephalic blood vessel using a sterile 1-ml syringe with a 28-gauge needle. Freshly collected hemocytes were washed in 100-200 μl of Squid Ringer's solution (530 mM NaCl, 10 mM KCl, 25 mM $MgCl_2$, 10 mM $CaCl_2$ and 10 mM HEPES buffer, pH 7.5) 2 times to remove excess hemolymph. Hemocytes collected from untreated/colonized squid were designated as "sym" and hemocytes from antibiotic treated/cured squid were referred to as "cured".

Sym hemocyte samples (n=8) and cured hemocyte samples (n=8) were resuspended in 400 μl of RIPA (Radio Immunoprecipitation Assay) buffer (50 mM Tris, 150 mM NaCl, 1% Triton, 0.5% sodium deoxycholate, 0.1% SDS, pH 7.5), pooled, and homogenized with a pestle, resulting in two samples, sym and cured. The lysates were centrifuged (Eppendorf 5810 R, 14,000 rpm, 4°C, 30 minutes) and the protein concentration of the supernatants was quantified

spectrophotometrically using the *RC DC* protein assay (Bio-Rad, Hercules, CA). Approximately 300 μ g of protein were obtained from sym hemocytes and 500 μ g from cured hemocytes.

Immunocytochemistry

Hemocytes were prepared for immunocytochemistry as previously described with minor modifications [133]. Approximately 50 µl of hemolymph per squid was adjusted to a total volume of 300 µl with Squid Ringer's (SR) and then evenly distributed among 3 wells (containing glass coverslips submerged in 1 ml of SR in a 12-well tissue culture plate (~$8.3x10^4$ hemocytes/slide). Hemocytes were allowed to adhere to glass coverslips for 30 minutes (all steps were carried out at room temperature). After attachment, the hemocytes were rinsed two times in SR to remove excess tissue or cellular debris and then fixed with 4% paraformaldehyde in SR for 30 minutes. After fixation, the hemocytes were washed 4 times with marine phosphate buffered saline, pH 7.4 (mPBS; 50 mM sodium phosphate, 0.45 M NaCl) for 10 minutes each. Fixed hemocytes were permeabolized with 1% Triton X-100, mPBS, pH 7.4 for 30 minutes. The cells were blocked (mPBS, pH 7.4, 1% Triton X-100, 1% goat serum, 0.5% BSA) for 1 hour. Hemocytes were then exposed to anti-cathepsin L2 (donated by M. McFall-Ngai, University of Wisconsin-Madison) at a 1:1,000 dilution in block overnight. The primary antibody was then removed and the hemocytes were washed four times in 1% Triton X-100, mPBS pH 7.4 for 5 minutes each and then blocked for an additional hour. The hemocytes were then incubated with an Alexa Fluor 488 goat anti-rabbit secondary antibody (Life Technologies, Grand Island, NY) at a 1:1,000 dilution in fresh block for 2 hours in the dark. Hemocytes were washed four times in 1% Triton X-100, mPBS pH 7.4 for 5 minutes each and then counterstained using a nuclear stain, DRAQ5 (2.5 µM; Thermo Scientific, Rockford, IL), in 1% Triton X-100, mPBS, pH 7.4

overnight in the dark. Finally, the hemocytes were mounted onto glass slides with Vectashield (Vector Laboratories, Burlingame, CA), sealed with nail polish, and imaged using an A1R confocal microscope equipped with NIS elements software (v. 4.13). Z-sections (0.225 μm each) were collected for more than 10 hemocytes per squid combined from greater than 2 microscopic fields (3 sym squid and 2 cured squid). Maximum intensity projections and fluorescence measurements were made using FIJI as previously described [134,135].

iTRAQ Quantitation Methodology

Protein digestion and iTRAQ labeling

The pooled protein stocks collected from sym and cured hemocytes were split into two samples as experimental replicates (Fig. 4.2). Proteins were precipitated using a methanol and chloroform protocol as previously described [136]. The resulting pellet was resuspended in 0.5 M triethylammonium bicarbonate, 2% SDS. The proteins were reduced using 50 mM tris-(2-carboxyethyl) phosphine (TCEP; Thermo Scientific, Rockford, IL) at 60°C for 1 hour and alkylated with 200 mM methyl methanethiosulfonate (MMTS; Thermo Scientific, Rockford, IL) at room temperature for 10 minutes. Samples were digested with trypsin (Promega, Madison, WI) at a 1:10 ratio (enzyme: protein) at 37°C overnight (18-20 hours). ITRAQ labeling was performed on 100 μg of protein from each replicate per condition using the reagents from a 4-plex iTRAQ kit (AB SCIEX, Framingham, MA). The labeled samples were then pooled.

Strong cation exchange

The labeled hemocyte protein sample was acidified with 1M phosphoric acid to a pH less than 3.0 and then separated on a Hewlett Packard 1090 HPLC system fitted with a polySulfoethylA column (The Nest Group, Southborough, MA). At 0.5 ml/min of Buffer A (10 mM KH_2PO_4, pH 3.0, 25% acetonitrile), a gradient of 0 to 100% Buffer B (10 mM KH_2PO_4, 1 M NaCl, pH 3.0, 25% acetonitrile), was established over 120 minutes. Fractions were collected at 1-minute intervals. The broad, unresolved A214 peak was pooled into 10 fractions according to absorbance. An additional cleanup step was performed for each fraction with a C18 MacroSpin column according to the manufacturer's protocol (The Nest Group, Southborough, MA).

LC-MS/MS

ITRAQ samples were analyzed by LC-MS/MS at Yale University's W.M. Keck Biotechnology Resource Laboratory. Three micrograms of each strong cation exchange (SCX) fraction (n=10) were separated and analyzed on a Waters nanoAcquity UPLC system equipped with a 5600TripleTOF (AB SCIEX, Framingham, MA); fitted with a Nanospray III source (AB SCIEX, Framingham, MA) and a pulled quartz tip as the emitter (New Objectives, Woburn, MA) as previously described with minor modifications [15]. For trapping, a flow rate of 5 µl/min with 99% Buffer A (100% water, 0.1% formic acid) was maintained for 1 minute using a Waters Symmetry® C18 180 µm x 20 mm trap column. A 1.7 µm, 75 µm x 150 mm nanoAcquity™ UPLC™ column (at 45ºC) was used for peptide separation. At a flow rate of 500 nl/min, a 161-minute linear gradient was maintained with Buffer A and Buffer B (100% CH_3CN, 0.075% formic acid). Initial conditions consisted of 95% Buffer A and 5% Buffer B. At 160 minutes conditions reached 60% A, 40% B and 15% A, 85% B at 161 minutes. Mass spectrometer

settings included an ion spray voltage of 2.2 kV, a curtain gas of 20 PSI, and a sweeping collision energy setting of 35 ± 15 eV for collision-induced dissociation. During information dependent acquisition (IDA) mode, survey scans were acquired for 250 ms for a mass range of 400-1250 Da. For ions with a charge state of +2 to +5 and which exceeded 125 counts per second, up to 20 product ion scans were collected. Four time bins were summed for each scan at a pulser frequency value of 15.420 kHz through monitoring of the 40 GHz multichannel TDC detector with four-anode/channel detection. The total cycle time was fixed to 1.3 seconds.

Data analysis

The combined raw MS/MS files (*.wiff) from Analyst TF 1.5.1 were analyzed with the Paragon^{TM} search algorithm of ProteinPilot (version 4.0; [137]). Data were searched against an *E. scolopes* protein sequence database originating from published transcriptomic sequences (34, 684 sequences; [8,15,21]). The transcriptomic sequences were translated and the longest open reading frame was annotated using BlastP and the NCBI nr database with an e-value cut-off of $1E^{-3}$ (due to the low representation of cephalopod protein sequences in the nr database). The ProteinPilot software determined the mass tolerance of precursor and fragment ions during the calibration of the data. Searching parameters included the 4-plex iTRAQ reagents, trypsin digestion, MMTS cysteine alkylation, bias correction, and background correction, in addition to selecting no special factors, biological modifications or amino acid substitutions. The false discovery analysis conducted by the ProteinPilot software utilized a reversed sequence decoy database to determine the false discovery rate (Supplemental file S4.1). Peptides identified by ProteinPilot were filtered using the "auto" setting to only include unique peptides, no missed cleavages, and at least 2 or more iTRAQ ions per peptide. To minimize iTRAQ ratios from low intensity ions, ProteinPilot also required peptides to maintain a signal to noise >9 from the

combined intensities of the contributing iTRAQ ions. Additionally, each protein quantified by ProteinPilot required two or more peptides. To compensate for any technical variation between iTRAQ labels, the remaining peptides were subject to a Cyclic Loess normalization using a script designed at Yale University's W.M. Keck MS and Proteomics Resource Laboratory [138]. Pairwise comparisons of each iTRAQ reporter during normalization revealed that the experimental variation arising from the iTRAQ reagents and methods was around 1.75 fold, which was used as the fold change cut-off (Supplemental file S4.2). The iTRAQ ion areas for labels 114 and 115 (sym) and for labels 116 and 117 (cured) were averaged for each peptide of each protein. These average iTRAQ values were then used to compare the protein levels between sym and cured hemocytes. Proteins were determined to be significantly more or less abundant if they had a p-value less than 0.05 (established by a t-test of the iTRAQ ion areas from contributing peptides) and a Log2 fold change greater than 0.80 or less than -0.80 (fold change of 1.75). Only proteins with a ProteinPilot unused score of >2.0 (99% confidence level), a significant p-value, and those meeting the Log2 fold change cut-offs were considered for further analysis. The after-normalization ProteinPilot summary and iTRAQ fold change analysis can be found in supplemental files S4.3 and S4.4 respectively.

Label-Free Spectral Counting Methodology

Protein digestion

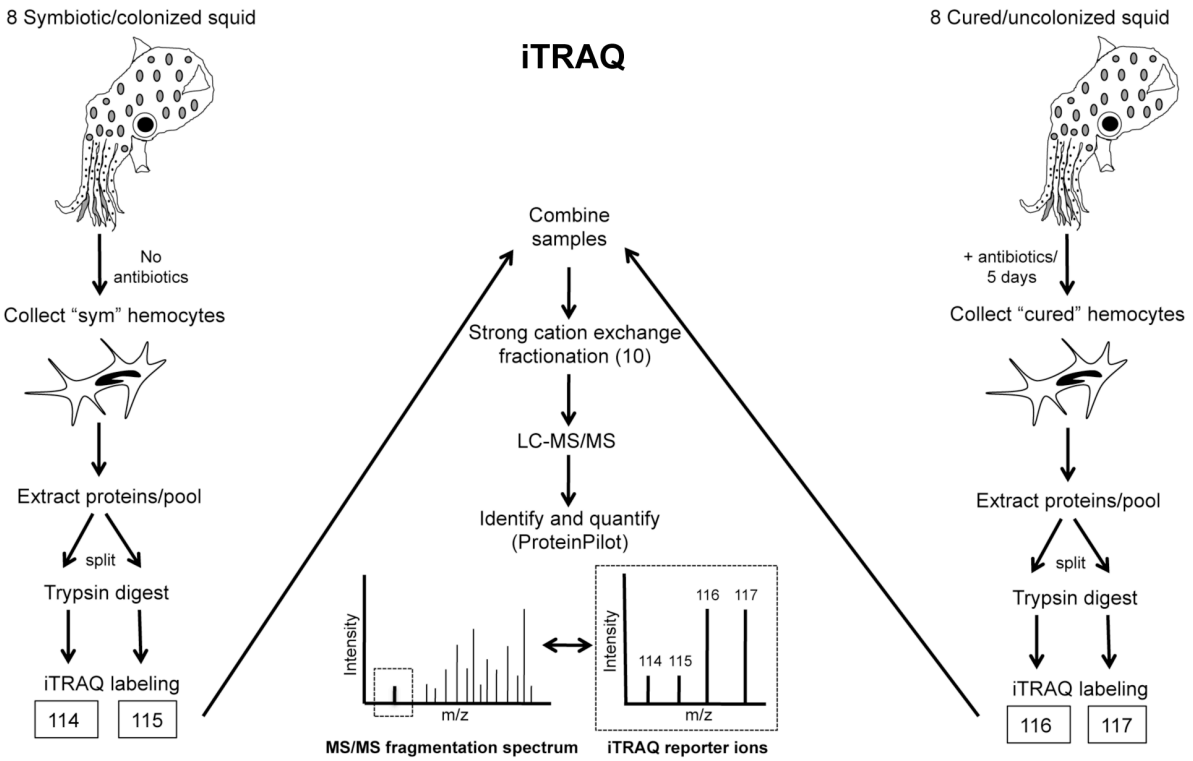
Proteome samples were digested using a modified filter assisted sample preparation technique (FASP; Expedeon, United Kingdom). Briefly ~100 μ g of each proteome sample were diluted in 1% SDS and heated at 60°C for 5 minutes. Samples were allowed to cool and then further diluted with 1% SDS, 8 M urea, 10 mM DTT (made in 50 mM Tris buffer pH 8.0, all

buffers above and below were made in this solution) and rocked at room temperature for 45 minutes. Samples were then transferred onto Expedeon FASP filters and centrifuged at 14,000 x g for 15 minutes. Fresh 8 M urea was added to the filters and centrifuged at 14,000 x g for 15 minutes. Ten µL of iodoacetamide solution (provided in the FASP kit) and 90 µL of urea solution (no DTT, no SDS) were added to each filter preparation. Samples were vortexed for 1 minute and then incubated without mixing for 20 minutes in the dark followed by centrifugation at 14,000 x g for 15 minutes. One hundred µL of 8 M urea solution was then added to the filters (no DTT, no SDS) followed by centrifugation at 14,000 x g for 15 minutes. This step was repeated twice followed by the addition of 100 µL of a 50 mM ammonium bicarbonate solution (provided with the FASP kit) and centrifuged at 14,000 x g for 15 minutes. The filter was then transferred to a new collection tube for the proteolytic digestion step. Seventy-five µL of digestion solution was added to the filter and incubated at 37°C for 18 hours with no rocking. The digestion solution contained 75 uL of ammonium bicarbonate solution with the addition of 10 µg of trypsin (Trypsin Ultra, New England Biolabs, Ipswich, MA). The next day peptides were eluted into the clean collection tube via three steps of centrifugation at 14,000 x g for 15 minutes as follows: Step 1: 40 µL 50 mM ammonium bicarbonate solution; Step 2: 50 µL 0.5 M sodium chloride solution (provided with the FASP kit); Step 3: 170 µL H_2O with formic acid. The final solution was split into three aliquots per sample and frozen at -80°C until analysis by 2d-LC-MS/MS. The eluted solution was ready to load onto a 2d-nano LC column without further purification or de-salting.

2d-LC-MS/MS Analyses

Tryptic digests of sym and cured unlabeled protein were analyzed in technical duplicates via a 2d nano-LC-MS/MS system with a split-phase nano column (RP-SCX-RP; [139]) on a

A


8 Symbiotic/colonized squid
iTRAQ
8 Cured/uncolonized squid
No antibiotics
Collect "sym" hemocytes
Extract proteins/pool
split
Trypsin digest
iTRAQ labeling
114
115
Combine samples
Strong cation exchange fractionation (10)
LC-MS/MS
Identify and quantify (ProteinPilot)
Intensity
m/z
116 117
114 115
MS/MS fragmentation spectrum
iTRAQ reporter ions
+ antibiotics/ 5 days
Collect "cured" hemocytes
Extract proteins/pool
split
Trypsin digest
iTRAQ labeling
116
117

B

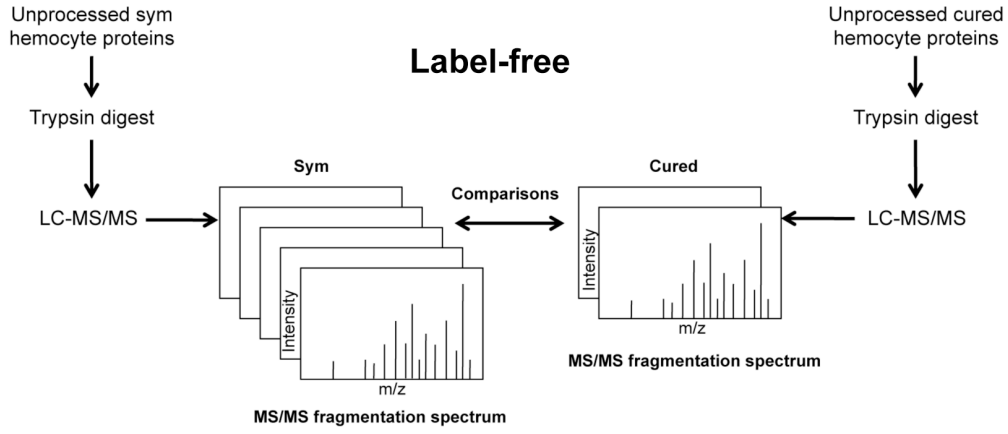

Unprocessed sym hemocyte proteins
Label-free
Unprocessed cured hemocyte proteins
Trypsin digest
Trypsin digest
Sym
Comparisons
Cured
LC-MS/MS
LC-MS/MS
Intensity
m/z
MS/MS fragmentation spectrum
Intensity
m/z
MS/MS fragmentation spectrum

Figure 4. 2: A visual diagram of the methods for collecting and quantifying hemocyte proteomes from symbiotic and cured squid using iTRAQ and label-free proteomics.

(A). Hemocytes were collected from the cephalic blood vessel of healthy adult squid. For cured hemocytes, hemocytes were collected from squid treated with an antibiotic cocktail for five days. Hemocytes from eight squid from each condition were pooled and the proteins were extracted. The protein pool from sym and cured hemocytes was then split in half resulting in experimental replicates. After digestion with trypsin, the peptides from the two sym replicates were labeled with the iTRAQ reagents 114 and 115 and the two cured replicates were labeled with 116 and 117 (these reagents are isobaric mass tags which bind to the free amine groups of tryptic peptides). The four samples were combined and then fractionated into ten fractions using strong cation exchange. LC-MS/MS was performed on each individual fraction. The spectral data from each fraction was combined and searched against an *E. scolopes* protein database using ProteinPilot. After peptide identification, the different intensities of iTRAQ reagent ions (found in the low *m/z* region of the MS/MS spectrum) allowed the comparison of sym and cured hemocyte proteomes. The example in this figure suggests that the identified peptide from cured hemocytes (116 and 117, iTRAQ reporter ions) may be more abundant than the peptide found in sym hemocytes (114 and 115). (B). For label-free proteomics, unprocessed sym or cured hemocyte proteins from the iTRAQ experiment were digested with trypsin, subjected to LC-MS/MS, and analyzed using SEQUEST. For quantification the normalized spectral abundance factor (NSAF) was determined for all proteins in each sample per replicate. The average NSAFs between sym and cured hemocyte proteins were compared to identify differences.

QExactive mass spectrometer (Thermo Scientific, Rockford, IL) with 22 hour runs per sample (LC as described earlier; [104,140,141]). Briefly peptides were loaded onto the RP-SCX nano back column and de-salted with a water to organic to water gradient over 15 minutes using 95% H_2O, 5% acetonitrile, 0.1% formic acid and 30% H_2O, 70% acetonitrile, 0.1% formic acid. The back column was then connected to a 15 cm RP resolving nano-column positioned on a nanospray source (Proxeon; Thermo Scientific, Rockford, IL) directly connected to the QExactive. Peptides were eluted from the SCX column with 11 increasing salt pulses (0 to 500 mM ammonium acetate), followed by water-organic gradients (using 95% H_2O, 5% acetonitrile, 0.1% formic acid and 30% H_2O, 70% acetonitrile, 0.1% formic acid) for 2 hours to resolve the peptides and subsequently ionize via nanospray into the QExactive. For all 2d-LC-MS/MS analyses, the QExactive was operated in data-dependent mode with the top ten ions selected for isolation and HCD fragmentation (higher-energy collisional dissociation) from the most

abundant peptides from the survey scan (400-1600 m/z). The QExactive settings were as follows: the normalized collision energy for HCD was 28 eV, a full scan resolution of 70,000K, a HCD MS/MS resolution of 17,500, and the dynamic exclusion was set at 15 seconds. Peptides were not excluded based on charge state.

Proteome informatics

All MS/MS spectra were searched with the SEQUEST algorithm (v.27; [142]) and filtered with DTASelect/Contrast (v1.9; [143]) at the peptide level (Xcorrs of at least 1.8 (+1), 2.5 (+2), 3.5 (+3)), with a minimum DeltCN of 0.08. SEQUEST parameters included a fixed modification for carboxyamidomethylated cysteines, a variable modification for urea carbamylation of arginine and lysine residues, trypsin digestion, up to 4 missed cleavages, a precursor mass tolerance of 3.0 Da and a fragment mass tolerance of 0.8 Da. Only proteins identified with two fully tryptic peptides from a 22 hour run were considered for further biological study. Tandem MS/MS spectra were searched against the same *E. scolopes* protein database used for iTRAQ, but with several modifications. In addition to the squid protein sequences originating from the light organ and hemocytes, *V. fischeri* ES114 protein sequences and common contaminants (such as trypsin, keratins and protein lab standards) were included in this combined database (22, 733 protein sequences). To remove redundancy in protein sequences from the original database used during iTRAQ (34,684 sequences), protein sequences originating from hemocyte singleton transcripts were removed from the combined database. Spectral counts were extracted for all proteins and compared across all runs. False discovery rates were determined by selecting one technical run from each sym and cured proteome analysis and searching those data against a reversed decoy database created from the squid-vibrio database used during the label-free analyses as described previously [140,144]. All false positive rates

were below 5% and below 0.5% when using parent peptides with a mass accuracy between -10< PPM <10 (Table 4.1). Protein spectral counts were normalized and compared between sym and cured hemocytes using the normalized spectral abundance factor method (NSAF; [145,146]). Significant differences between NSAF values for proteins of sym and cured hemocytes were determined by an unpaired equal variance t-test (p-value < 0.05). For each significantly different protein, the ratio of the average NSAF for sym relative to the average NSAF for cured was Log2 transformed and filtered for proteins meeting a cut-off of a 1.5 fold change.

Results

ITRAQ and label-free spectral counting were utilized to compare hemocyte proteomes from hosts with colonized or uncolonized light organs (Fig. 4.2).

Table 4. 1: Summary of iTRAQ and label-free spectral counting.

Method	**iTRAQ**	**Label-free**			
Sample	Sym and cured	Sym 1	Sym 2	Cured 1	Cured 2
Spectra	40,771	69,007	69,503	35,926	40,805
Peptides	15,415	23,339	19,215	12,444	10,337
Proteins	2,001	3,313	2,923	1,713	1,506
PPM	N/A	83.47	81.33	80.95	81.59
Proteins quantified	1,024	1,241			
Significant proteins	37	185			
More abundant	20	92			
Less abundant	17	93			
FDR (FDRppm[a])	1%	4.63% (0.48%)		2.96% (0.33%)	

a. FDR in parentheses represents the FDR when using parent peptides with a mass accuracy between -10< PPM <10. Applied to the label-free approach only.

Overall iTRAQ identified 40,771 spectra, 15,415 peptides and 2,001 proteins at a global false discovery rate (FDR) of 1% (Table 4.1; Supplemental files S4.1 and S4.3). Of the 1,024 proteins quantified with 2 or more peptides, 983 had an unused protein score of greater than or

equal to 2.0 (99% confidence; Supplemental file S4.4). General housekeeping proteins, which were not expected to change between sym and cured hemocytes (e.g. elongation factor 1- α (-0.04 Log2 fold change), heat shock protein 70 (-0.09 Log2 fold change), lactate dehydrogenase (0.10 Log2

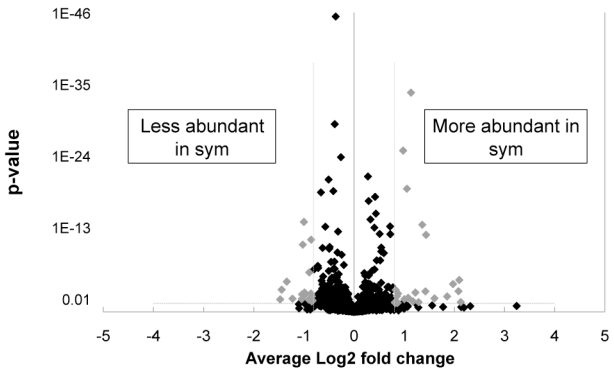


Figure 4. 3: Volcano plot of hemocyte iTRAQ data.

The average Log2 fold change and respective p-value for each protein identified using iTRAQ with more than 2 peptides was plotted. Proteins identified as significantly more abundant or less abundant in sym hemocytes are highlighted in gray (p-value < 0.05). Gray dotted lines highlight the Log2 fold change cutoffs (± .80) and a p-value cutoff of 0.05.

fold change), and glyceraldehyde-3-phosphate dehydrogenase (0.06 Log2 fold change)), maintained a Log2 fold change near zero (Supplemental file S4.4). Proteins with a Log2 fold change greater than .80 or less than -.80 (corresponding to a fold change of 1.75) were considered significant after comparison of the experimental variation arising from reagent labeling and/or digestion (see Methods). From the 1,024 proteins, 37 unique proteins displayed significantly different abundances between the samples (p-value < 0.05; Fig. 4.3). These proteins were evenly distributed between being more and less abundant. Twenty proteins were more abundant and 17 were less abundant in sym hemocytes (Table 4.2). The protein with the greatest increase in abundance was a ganglioside GM2 activator precursor (2.12 Log2 fold change; Table 4.2) while a hypothetical protein revealed the largest reduction (-1.47 Log2 fold change; Table

4.2). These differentially produced proteins were categorized into several functional groups including cell adhesion, cytoskeleton, innate immunity, and lysosomal (Fig. 4.4).

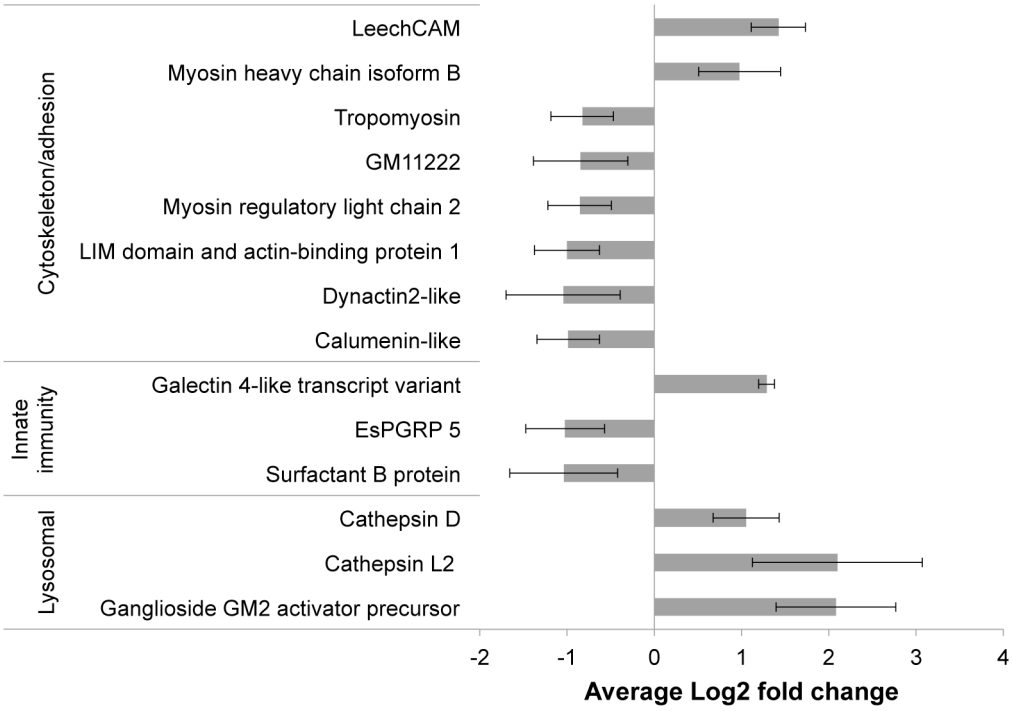


Figure 4. 4: Proteins detected at significantly different amounts between sym and cured hemocytes by iTRAQ.

Proteins with the most peptides and the largest fold change were selected to represent a group of similar proteins (Table 4.2).

To further validate the iTRAQ data, the same unprocessed starting samples from iTRAQ were analyzed by 2d-LC-MS/MS and label-free spectral counting. Each sample had a FDR below 5% (4.63% for sym and 2.96% for cured; Table 4.1). The FDR's were lower than 0.5% when using peptides with a mass accuracy of -10 < PPM <10 (Table 4.1). Normalized spectral abundance factors (NSAFs) were used to determine the relative abundance of each protein identified. A total of 1,241 proteins were identified in all technical runs (Supplemental file S4.5). From these analyses there were 185 unique proteins that were significantly more or less abundant in sym hemocytes (p-value < 0.05; Supplemental file S4.5). The distribution of significantly

different sym and cured proteins showed a similar trend to what was found with iTRAQ, although more proteins were identified with the label-free technique (Supplemental file S4.5). Comparison of these proteins to the iTRAQ data revealed 33 of the 37 unique iTRAQ proteins were also identified using the label-free method (Table 4.3). Thirty of these 33 proteins followed the same trend identified by iTRAQ (Table 4.3). Only 3 of the significant iTRAQ proteins were identified by the label-free method as having an opposite trend (Table 4.3). All 3 of these proteins (s-formylglutathione hydrolase, rootletin, and mps one binder kinase activator), generally had a lower number of peptides contributing to the iTRAQ ratio, which may explain the opposite trends resulting between the two techniques. While there was a large overlap between all of the iTRAQ proteins and all of the label-free proteins identified (795 in common), only 16 proteins were significantly different between sym and cured hemocytes using both methods (p-value < 0.05, Table 4.3; Fig. 4.5).

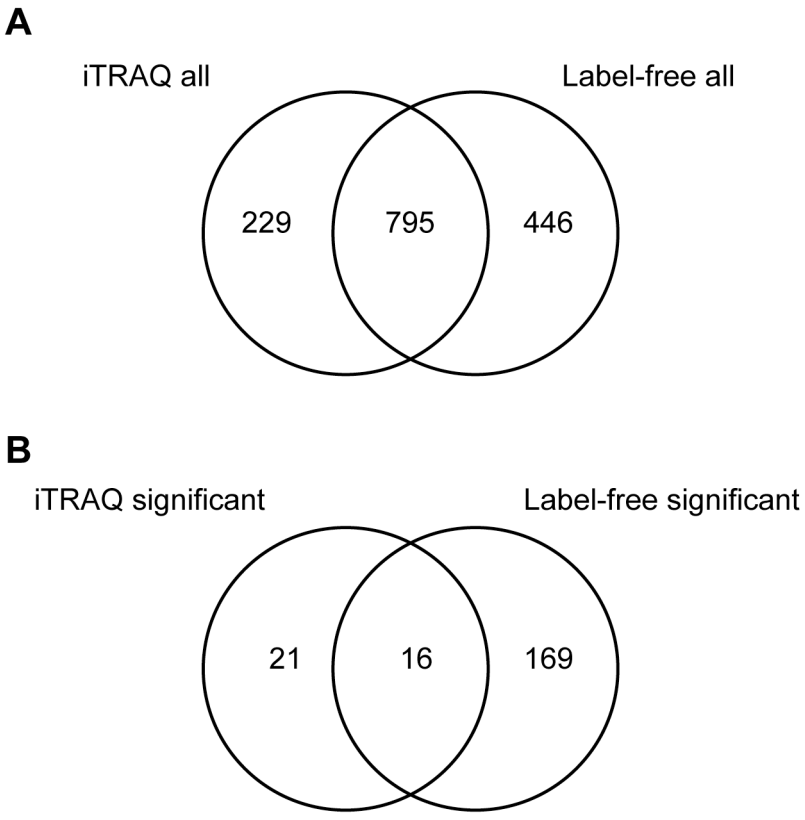


Figure 4. 5: Comparison of iTRAQ and label-free proteomic analyses of sym and cured hemocytes.

(A). Venn diagram showing the overlap of all iTRAQ proteins identified (>2 peptides) and all label-free proteins identified (>2 peptides, identified in both samples and in both replicates). **(B)**. Venn diagram of the differentially produced proteins identified using iTRAQ and label-free methods (16 are significant using both iTRAQ and label-free).

In comparison to the iTRAQ data, the label-free analysis identified an additional 169 unique proteins (Figure 4.5). Among the significantly more abundant proteins identified by both methods was cathepsin L2 and a ganglioside GM2 activator precursor (Table 4.3). The less abundant proteins mutually identified by iTRAQ and spectral counting included a myosin regulatory light chain 2, a LIM domain and actin binding protein, and a surfactant B-like protein (Table 4.3). One common limitation to iTRAQ is ratio suppression, which may lead to smaller fold changes and therefore a lower number of significantly different proteins detected [147,148]. However, analysis of the significant proteins identified by label-free spectral counting revealed categories of proteins similar to the proteins identified by iTRAQ (Fig. 4.6). Proteins associated with the cytoskeleton and cell adhesion were generally less abundant (16 of 20), while proteases were mostly more abundant (4 of 5) in sym hemocytes. Other differences were also detected for proteins associated with innate immunity, lysosomes, and stress responses (Fig. 4.6).

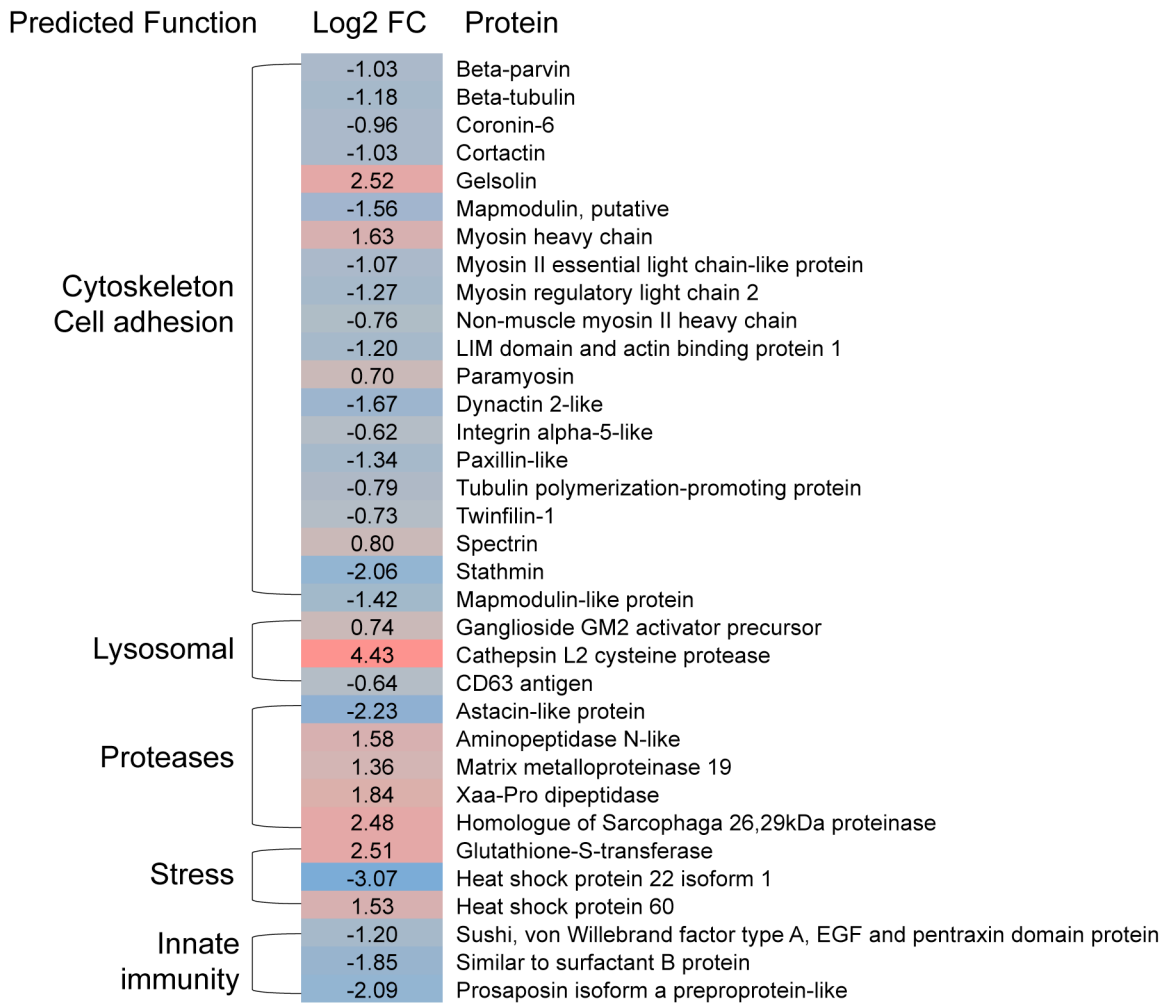


Figure 4. 6: Heat map of proteins significantly more or less abundant in sym hemocytes identified by label-free spectral counting.

For each protein, an average NSAF ratio of sym relative to cured hemocytes was Log2 transformed. Proteins that are blue are less abundant and proteins that are red are more abundant. The scale ranges from a -3.1 Log2 fold change (blue) to a 4.5 Log2 fold change (red).

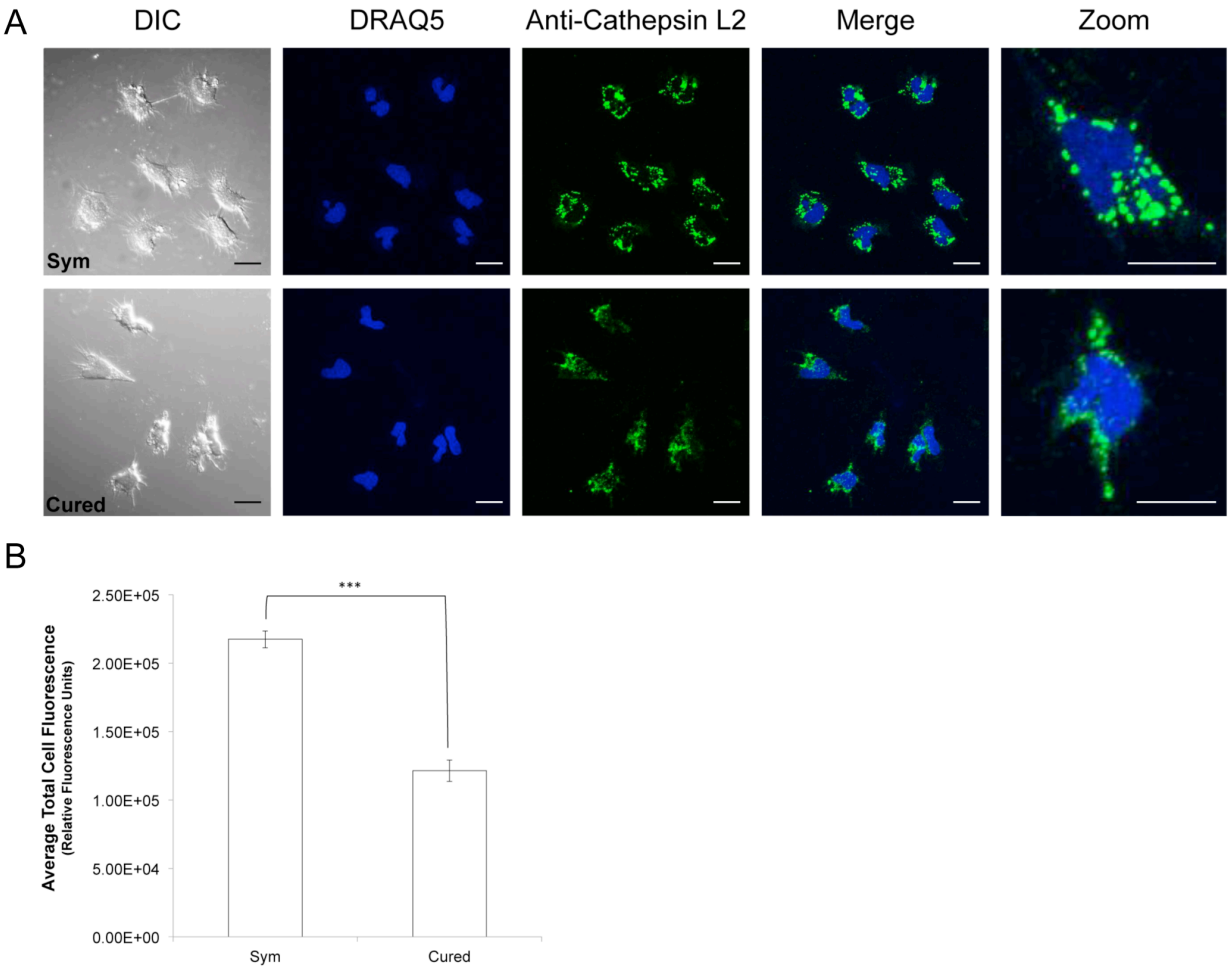


Figure 4. 7: Protein localization of cathepsin L2 in sym and cured hemocytes.

(A). Representative maximum projection composite confocal images of sym and cured hemocytes. Anti-cathepsin L2 was detected with an Alexa Fluor 488 goat anti-rabbit secondary antibody. The hemocytes were counterstained with the nuclear stain DRAQ5. A higher magnification view of additional representative hemocytes revealed cathepsin L2 was abundant and localized in punctate granules consistent with lysosomes in sym hemocytes. In hemocytes from cured animals, cathepsin L2 was less intense and had a more diffuse staining pattern (DIC- differential interference contrast image, scale bar measurements- 10 μM). Hemocyte negative controls (secondary antibody only) can be found in supplemental file S4.6. (B). Fluorescence measurements from maximum projections of individual hemocytes were utilized to compare the abundance of cathepsin L2 in sym (n=3 squid, 89 hemocytes total) and cured hemocytes (n=2 squid, 28 hemocytes total). Normalized cell fluorescence was determined with the software FIJI by measuring the integrated density of the cell and then subtracting the mean background fluorescence multiplied by the area of the cell measured. Cathepsin L2 was significantly more abundant in sym hemocytes (1.79 fold, ***- two-tailed unequal variance t-test, p-value < 0.0001; error bars represent standard error of the mean (SEM)).

An antibody generated to recognize a peptide specific to cathepsin L2, a protein which demonstrated one of the more significant increases in sym hemocytes by both iTRAQ and spectral counting, was used to compare the localization and abundance of this cathepsin in sym and cured hemocytes by immunocytochemistry. Fluorescence measurements of composite confocal images of individual hemocytes revealed that cathepsin L2 was significantly more abundant in sym hemocytes (+1.79 fold) providing further confirmation of the mass spectrometry-based proteomic results (Fig. 4.7). In addition, the localization of cathepsin L2 in sym hemocytes appeared to be organized in punctate granules consistent with lysosomes (Fig. 4.7A; negative controls-Supplemental file S4.6). In contrast, the localization of cathepsin L2 in cured hemocytes appeared to be more diffuse and present both inside and outside of granules (Fig. 4.7A).

Discussion

The *E. scolopes* and *V. fischeri* association presents a unique opportunity to study the influence of the innate immune system on a beneficial symbiosis [1,46]. Because there is a high degree of specificity such that only *V. fischeri* can colonize the light organ, mechanisms must be in place to safeguard the association. Host hemocytes are believed to play a significant role in this process and a previous study revealed that these cells are able to differentiate between *V. fischeri* and non-symbiotic bacteria, even in the absence of an adaptive immune system [5]. The molecular mechanisms that allow these hemocytes to make this distinction are poorly understood, but some target genes and proteins have been identified [15]. In this study, the hemocyte

proteome from *E. scolopes* was characterized based on colonization of the light organ in order to identify host proteins that may be involved with symbiont recognition. Quantitative proteomics is becoming a reliable technique for characterizing host-microbe interactions during symbiosis as label or label-free methods have been applied to beneficial associations involving plants, insects, corals, and humans [103,104,149-152]. This study represents, to the best of our knowledge, the first use of two complimentary quantitative proteomic approaches to further understand the influence of a bacterial symbiont on the immune response of an animal host.

Cytoskeletal dynamics and cell adhesion

Many of the hemocyte proteins influenced by *V. fischeri* colonization were related to the cytoskeleton. Both iTRAQ and spectral counting supported this observation, however the label-free approach resulted in the identification of a greater number of differentially produced cytoskeletal and adhesion proteins (iTRAQ- 10 of 37; label-free- 20 of 185; Table 4.2; Fig. 4.6; Supplemental file S4.5). Proteins associated with integrin signaling were also identified by both methods (Table 4.2; Fig. 4.6). Although involved in a variety of cellular pathways, integrin signaling is a well-characterized component of cell adhesion and cytoskeletal rearrangements in other types of macrophages [153,154]. The less abundant cytoskeletal/adhesion proteins identified in this study may represent a decrease in the ability of the hemocytes to adhere and/or phagocytose *V. fischeri*. Previous results have shown that *V. fischeri* binds significantly less to hemocytes from a colonized host compared to one that has been cured [5]. Bacteria can influence the host cell cytoskeleton in a number of ways, especially as an effective strategy for immune evasion [155,156]. In the context of the squid-vibrio symbiosis, it has been shown that *V. fischeri* induces changes to the host's actin cytoskeleton in the tissues surrounding the light organ [157].

The mechanism by which *V. fischeri* may deliver a signal that alters the hemocyte cytoskeleton remains unknown. However, an outer membrane protein OmpU appears to be involved with mediating adhesion to host hemocytes in both the squid-vibrio symbiosis and in the pathogenic association between oysters and *Vibrio splendidus* [5,52,158]. The exact mechanism by which OmpU or another bacterial factor(s) may mediate the hemocyte response remains to be characterized, but the results of this study suggest that colonization by *V. fischeri* does alter the hemocyte cytoskeleton.

Other hemocyte proteins putatively associated with cell adhesion and innate immunity were a leech CAM (cell adhesion molecule), surfactant-like proteins and a galectin 4-like protein (Tables 4.2 and 4.3; Fig. 4.6). In leech neurons, CAM is a transmembrane protein that has been implicated in playing a role in cell attachment and migration [159-162]. Two putative saposin/surfactant-like proteins were identified as less abundant by both methods. In other systems these proteins have the ability to bind microorganisms and enhance phagocytosis [163,164]. Future studies will focus on characterizing whether these surfactant-like proteins in *E. scolopes* are involved with the observed increased binding of *V. fischeri* to cured hemocytes. Additionally, iTRAQ and spectral counting identified one galectin (a galectin 4-like transcript variant) as more abundant in sym hemocytes. Galectins are carbohydrate-binding proteins that are believed to recognize the surface-exposed carbohydrates of microorganisms [165,166]. Many galectins are now implicated as important factors of the invertebrate innate immune response as these proteins can directly bind bacteria and enhance phagocytosis [167,168]. Recently in the lancelet, *Branchiostoma belcheri tsingtauense*, it was discovered that a galectin can bind to *Vibrio vulnificus,* but not to *Vibrio parahaemolyticus* [169]. The apparent selectivity of galectins

may also contribute to the ability of *E. scolopes* to differentiate between closely related *Vibrio* species, but this remains to be characterized.

Taken together, the data from this and previous studies suggest that a specific response to *V. fischeri* may require coordination between host proteins involved with microbial recognition, adhesion, and engulfment. It is unclear how a putative decrease in cytoskeletal function may be involved with a specific response to *V. fischeri* as sym and cured hemocytes still have the ability to adhere and phagocytose non-symbiotic bacteria [5]. As an alternative hypothesis, the change in abundance of cell adhesion and cytoskeletal proteins may influence hemocyte migration. The light organ is highly vascularized and hemocytes are known to migrate into the light organ crypt spaces where the symbiont is found at a high density [5,13,14]. Future studies should determine whether hemocytes from sym and cured hosts migrate to the light organ at the same frequency and numbers.

Lysosomal proteins

Some of the more significantly abundant proteins are predicted to be involved with lysosome function. Similar to other macrophage-like cells, hemocytes of *E. scolopes* have lysosomes that play an important role in the immune response of phagocytic cells by degrading engulfed bacteria [13,170]. Cathepsins and a ganglioside GM2 activator were more abundant in sym hemocytes (Tables 4.2 and 4.3). Although linked to other cellular pathways, cathepsins are proteases activated by a low lysosomal pH and the ganglioside GM2 activator precursor is a lipid transfer protein also associated with lysosomes [171,172]. As part of normal lysosomal function, these proteins may be important for processing MAMPs from phagocytosed bacteria. The differences found in lysosomal and other proteins identified in this study may not be restricted to

hemocytes. For example, a recent study analyzed the influence of *V. fischeri* on the host's light organ transcriptome during the initiation of the association and found that both cathepsin D and L2, along with the GM2 activator precursor protein and a triacylglycerol lipase were also significantly more abundant in symbiotic light organs [21]. Further examination using an antibody generated to cathepsin L2 corroborated that colonization by *V. fischeri* increases the abundance of this protease in hemocytes (Fig. 4.7). Cathepsin L-like proteins have been described in other invertebrate hemocytes [173,174] and are modulated by colonization in both pathogenic [175,176] and beneficial associations ([177], this study).

Additional innate immunity proteins

A transcriptome and proteome of circulating hemocytes from symbiotic hosts identified a number of genes and proteins related to innate immunity, including members of the NF- κ B pathway, PRRs (2 PGRPs and a galectin) and complement-like proteins [15]. Building on this previous research, iTRAQ revealed that the EsPGRP 5 protein is less abundant in sym hemocytes, although transcript levels of this gene were up-regulated (Table 4.2; Fig. 4.4; [15]). The retention time and turnover of EsPGRP 5 is unknown and may explain this discrepancy. EsPGRP 5 also has a putative signal peptide and secretion of this protein may be influenced by the presence or absence of *V. fischeri* [15]. The cellular localizations of two other PGRPs from *E. scolopes* (EsPGRP 1 and EsPGRP 2) are also altered in response to colonization of the host light organ [15,50,178]. Because iTRAQ does not quantify extracellular proteins lost during sample preparation, colonization may lead to an increased release of EsPGRP 5 and therefore a loss of the cellular protein, although this remains to be tested.

A matrix metalloproteinase (MMP), a protein also associated with innate immunity, was significantly more abundant in sym hemocytes analyzed by label-free spectral counting (Fig. 4.6). This MMP with homology to a complement-related protein, vitronectin, was previously identified by transcriptomics and shotgun proteomics [15]. Unlike EsPGRP 5, an increase in protein abundance for MMP did correlate with previously identified increases in gene expression [15]. Recent in-depth comparisons of mRNA and protein levels in yeast, mice, and humans have demonstrated that mRNA abundance does not always correlate with protein abundance. In these studies approximately 40% of mRNA transcripts predicted protein abundance [179,180]. The variation between the rest of the transcripts and proteins is likely due to the complex dynamics associated with regulating post-transcriptional, post-translational, and protein degradation processes. These pathways are poorly understood in the squid-vibrio symbiosis, but a global transcriptomic comparison between sym and cured hemocytes is currently underway and will be useful for future analyses.

Summary

The mechanisms by which hemocytes of *E. scolopes* differentiate between the light organ symbiont *V. fischeri* and non-symbiotic bacteria are still under investigation, but it is clear that colonization influences the cellular response, gene expression, and proteome of this cell type. Previous proteomic analyses of the squid-vibrio symbiosis have been focused on 2-Dimensional gel electrophoresis or individual protein identification and characterization [15,25,115]. This study represents the first application of quantitative proteomics to the squid-vibrio symbiosis. In addition, this study represents the first combination of iTRAQ and a label-free proteomic analysis to enhance our understanding of a beneficial symbiosis. Proteomic and gene expression

studies in other organisms including oysters and pigs have revealed that similar genes and proteins (cathepsins, GM2AP, heat shock protein 22, and cytoskeletal proteins) are altered by both pathogenic and beneficial bacterial colonization [181,182]. Few studies have combined both iTRAQ and label-free proteomics; however, applying a label-free method has been shown to identify a larger number of significantly different proteins than a labeling approach, similar to what was observed in this study [183,184]. Employing both methods, however, may offer a means to identify those proteins or pathways displaying the most significant changes.

In this study, after curing, both quantitative proteomic techniques (iTRAQ and spectral counting) revealed significant differences between hemocytes from colonized and uncolonized squid, bolstering the hypothesis that *V. fischeri* influences host hemocyte function. Although the mechanisms are still unclear, the proteins identified independently by both techniques suggest that a number of cellular processes (e.g. cytoskeletal, lysosomal and immune recognition) may help mediate host tolerance of *V. fischeri*. The mechanisms by which invertebrate innate immune systems target and recognize specific microorganisms are poorly understood and the results from this study may have broader implications for understanding these processes in other taxa. This study acts as a foundation for understanding how the hemocyte proteome changes in response to colonization and future studies will focus on characterizing hemocyte proteins with respect to their role in the *E. scolopes*-*V. fischeri* symbiosis. In addition, directly challenging sym and cured hemocytes with symbiotic and non-symbiotic bacteria will likely help identify dynamics involved with host recognition of bacteria.

Table 4. 2: Proteins identified by iTRAQ as significantly more or less abundant in sym hemocytes.

Trend [a]	Database ID	Protein	Pep[b]	Log2 fold change[c]	p-value
More abundant					
	Meta990274633	cathepsin D	39	1.054	1.20E-19
	Meta990279877	cathepsin L2	12	2.097	1.29E-05
	Meta990255245	galectin 4-like protein transcript variant	2	1.287	3.21E-02
	Meta990274730	ganglioside GM2 activator precursor	4	1.222	1.08E-03
	Meta990237061	ganglioside GM2 activator precursor	6	2.082	6.87E-04
	Meta990280234	ganglioside GM2 activator precursor	4	2.128	3.18E-02
	Meta990084014	glucosamine-fructose-6-phosphate aminotransferase	2	0.820	4.50E-02
	Meta990256250	histone H4-I	56	1.135	1.99E-34
	Meta990272817	26S proteasome non-ATPase subunit 12	2	0.825	4.26E-02
	Meta990275393	hypothetical protein BRAFLDRAFT_185672	5	0.829	4.33E-04
	Meta990279786	pancreatic triacylglycerol lipase-like	4	1.856	4.30E-03
	Meta990297914	amino acid transport protein rBAT	8	0.877	1.58E-03
	Meta990261223	pyruvate carboxylase	2	0.830	3.02E-02
	Meta990234931	histone H3.2-like	24	1.434	1.46E-12
	Meta990123568	intermediate filament protein	5	0.853	3.93E-03
	Meta990236123	intermediate filament protein	12	1.972	5.32E-05
	Meta990244043	l-3-hydroxyacyl-coenzyme a dehydrogenase	4	0.862	4.43E-02
	Meta990233579	leechCAM	5	1.422	7.15E-04
	Meta990279781	myosin heavy chain	4	0.869	3.98E-02
	Meta990231669	myosin heavy chain isoform B	66	0.978	1.64E-25
	Meta990231856	aconitate hydratase, mitochondrial-like	4	1.078	6.88E-03
	Meta990275031	s-formylglutathione hydrolase-like	5	0.812	3.57E-02
	Meta990210359	WW domain-containing oxidoreductase-like	20	1.359	4.13E-14
	Meta990236159	WW domain-containing oxidoreductase-like	4	1.603	8.54E-03
	Meta990272315	Rab3	3	1.176	2.51E-02
Less abundant					
	Meta990253072	GM11222	5	-0.844	2.56E-02
	Meta990275900	heat shock protein 22 isoform 1	7	-0.975	2.13E-03
	Meta990300071	hypothetical protein	3	-0.858	1.51E-03
	Meta990081459	hypothetical protein	8	-1.341	2.30E-05
	Meta990232278	LIM domain and actin-binding protein 1	29	-1.000	1.52E-14
	Meta990297867	Myosin regulatory light chain 2	26	-0.856	8.14E-12
	Meta990275793	EsPGRP 5	25	-1.023	4.71E-11
	Meta990299707	calumenin-like	3	-1.227	9.64E-03
	Meta990236998	calumenin-like	6	-0.988	1.08E-03
	Meta990232620	dynactin 2-like	6	-1.045	1.13E-02
	Meta990220758	hypothetical protein LOC100115922	3	-1.474	1.23E-02
	Meta990232946	mps one binder kinase activator-like 2B	3	-0.807	2.46E-02
	Meta990231811	rootletin-like	2	-0.904	1.35E-02
	Meta990232518	protein kinase, AMP-activated, beta 1	5	-0.839	4.23E-03
	Meta990274714	similar to surfactant B protein	8	-1.037	2.10E-03
	Meta990255307	THO complex 4-like	3	-1.444	3.91E-04
	Meta990299300	translationally-controlled tumor protein	18	-0.894	8.92E-07

	Meta990139263	tropomyosin	4	-0.826	1.89E-02

a. Protein trend based on iTRAQ data.
b. Pep: number of peptides contributing to the iTRAQ ratio for that protein.
c. Log2 fold change represents a protein difference in sym relative to cured.

Table 4. 3: Comparison of the proteins identified by iTRAQ with NSAF values from the label-free analysis.

iTRAQ trend[a]	Database ID	Protein	NSAF Sym[b]	NSAF Cured[b]	p-value[c]	LF trend[d]
More abundant						
	Meta990274633	cathepsin D	1.42E-03	1.24E-03	0.388	↑
	Meta990279877	**cathepsin L2**	**1.83E-03**	**8.50E-05**	**0.016**	↑
	Meta990255245	galectin 4-like protein transcript variant	2.20E-04	4.00E-05	-	↑
	Meta990274730	**ganglioside GM2 activator precursor**	**4.50E-04**	**2.70E-04**	**0.038**	↑
	Meta990237061	ganglioside GM2 activator precursor	2.80E-04	N/A	-	↑
	Meta990280234	ganglioside GM2 activator precursor	7.50E-04	N/A	-	↑
	Meta990084014	glucosamine-fructose-6-phosphate aminotransferase	N/A	N/A	-	N/A
	Meta990256250	histone H4-I	2.70E-03	2.52E-03	0.683	↑
	Meta990272817	26S proteasome non-ATPase subunit 12	3.85E-04	1.65E-04	0.174	↑
	Meta990275393	**hypothetical protein BRAFLDRAFT_185672**	**6.45E-04**	**2.05E-04**	**0.011**	↑
	Meta990279786	**pancreatic triacylglycerol lipase-like**	**2.37E-03**	**7.30E-04**	**0.006**	↑
	Meta990297914	**amino acid transport protein rBAT**	**6.35E-04**	**1.45E-04**	**0.020**	↑
	Meta990261223	pyruvate carboxylase	2.40E-04	N/A	-	↑
	Meta990234931	histone H3.2-like	1.57E-03	7.05E-04	0.202	↑
	Meta990123568	intermediate filament protein	N/A	N/A	-	N/A
	Meta990236123	**intermediate filament protein**	**8.80E-04**	**7.00E-05**	**0.010**	↑
	Meta990244043	l-3-hydroxyacyl-coenzyme a dehydrogenase	5.80E-04	1.85E-04	0.074	↑
	Meta990233579	leechCAM	1.80E-04	4.50E-05	0.123	↑
	Meta990279781	**myosin heavy chain**	**3.40E-04**	**1.10E-04**	**0.044**	↑
	Meta990231669	myosin heavy chain isoform B	7.05E-04	4.20E-04	0.062	↑
	Meta990231856	**aconitate hydratase, mitochondrial-like**	**3.15E-04**	**1.15E-04**	**0.016**	↑
	Meta990275031	s-formylglutathione	2.60E-04	3.15E-04	0.409	↓

		hydrolase-like				
	Meta990210359	WW domain-containing oxidoreductase-like	N/A	N/A	-	N/A
	Meta990236159	**WW domain-containing oxidoreductase-like**	**1.08E-03**	**3.25E-04**	**0.017**	**↑**
	Meta990272315	Rab3	5.95E-04	3.10E-04	0.224	↑
Less abundant						
	Meta990253072	GM11222	5.45E-04	7.35E-04	0.127	↓
	Meta990275900	**heat shock protein 22 isoform 1**	**4.00E-05**	**3.35E-04**	**0.033**	**↓**
	Meta990300071	hypothetical protein	N/A	N/A	-	N/A
	Meta990081459	hypothetical protein	N/A	N/A	-	N/A
	Meta990232278	**LIM domain and actin-binding protein 1**	**7.50E-04**	**1.73E-03**	**0.012**	**↓**
	Meta990297867	**Myosin regulatory light chain 2**	**2.12E-03**	**5.12E-03**	**0.007**	**↓**
	Meta990275793	EsPGRP 5	1.85E-04	3.25E-04	0.067	↓
	Meta990299707	calumenin-like	2.15E-04	8.75E-04	0.158	↓
	Meta990236998	calumenin-like	4.00E-04	1.12E-03	0.235	↓
	Meta990232620	**dynactin 2-like**	**1.90E-04**	**6.05E-04**	**0.010**	**↓**
	Meta990220758	**hypothetical protein LOC100115922**	**2.95E-04**	**7.20E-04**	**0.010**	**↓**
	Meta990232946	mps one binder kinase activator-like 2B-like	4.00E-05	1.00E-04	-	↑
	Meta990231811	rootletin-like	5.00E-05	3.00E-05	-	↑
	Meta990232518	protein kinase, AMP-activated, beta 1	3.83E-03	5.97E-03	0.161	↓
	Meta990274714	**similar to surfactant B protein**	**9.00E-05**	**3.25E-04**	**0.002**	**↓**
	Meta990255307	**THO complex 4-like**	**1.00E-04**	**1.80E-04**	**0.015**	**↓**
	Meta990299300	translationally-controlled tumor protein	1.55E-03	3.36E-03	0.079	↓
	Meta990139263	tropomyosin	N/A	N/A	-	N/A

a. Trend represents sym hemocytes relative to cured hemocytes.
b. Normalized spectral abundance factor (NSAF). NSAFs were calculated as an average of two technical replicates per hemocyte condition.
c. Proteins needed to be identified in both replicates of sym and cured hemocytes to calculate a p-value. Proteins with "N/A" were not identified using the label-free approach. Trends were still determined if a p-value was not calculated. Proteins highlighted in bold were identified as significant by both iTRAQ and spectral counting.
d. LF: Label-free trend as determined by comparing the average NSAF value of the specific sym hemocyte protein relative to the same cured hemocyte protein. Up arrows indicate proteins that were more abundant in sym. Down arrows indicate proteins that were less abundant in sym.

Chapter 5: Investigating the role of putative squid galectins in a light organ symbiosis

Abstract

Only *Vibrio fischeri* from the seawater environment can colonize the light organ of the squid *Euprymna scolopes*. *E. scolopes* is able to recognize *V. fischeri* and remove possible non-symbiotic bacteria, but the exact mechanism that allows this distinction is not completely understood. Two putative carbohydrate-binding proteins (galectins) have been identified in hemocytes from *E. scolopes* using mass spectrometry-based proteomics. Galectins in other molluscs have been shown to adhere to a variety of microorganisms. Therefore the galectins of *E. scolopes* may represent uncharacterized innate immunity components important for the recognition of bacteria. Both galectins of *E. scolopes* were cloned from the light organ central core and shown to differ in the protein structure. EsGalectin 1 contains four carbohydrate recognition domains (CRDs) while EsGalectin 2 contains two CRDs. The type of CRD organization found in EsGalectin 1 is unique to molluscs and has not yet been discovered in mammals. Reverse transcription PCR revealed that both EsGalectin 1 and 2 are found ubiquitously in tissues from adult *E. scolopes* including gills, hemocytes, central core, hindgut, accessory nidamental gland, mantle, heart, and brain. Polyclonal antibodies generated to each squid galectin were used to determine the localization of each galectin in squid hemocytes with immunocytochemistry. EsGalectin 1 revealed several distinct hemocyte localizations including the filopodia, the nucleus, and the cytoplasm. EsGalectin 2 was predominantly a nuclear protein, but was lost from the nucleus in cells containing membrane vesicles. This study represents the

first analysis of galectins from *E. scolopes* and indicates that these proteins may contribute to the high degree of host-symbiont specificity observed in this squid-vibrio association.

Introduction

Euprymna scolopes maintains a lifelong association with the bioluminescent bacterium *Vibrio fischeri* [4,6,26]. One aspect of this squid-vibrio symbiosis that remains uncharacterized is the ability of the squid to differentiate between *V. fischeri* and non-symbiotic bacteria. This phenomenon is believed to contribute to the high degree of specificity seen in this relationship. Previous studies have demonstrated that the hemocytes of *E. scolopes* appear to develop a tolerance to *V. fischeri* in response to colonization of the light organ [5]. Quantitative proteomics comparing hemocytes from squid with colonized (sym) and uncolonized (cured) light organs did not find an overwhelming absence or presence of one protein that could be responsible for this response to *V. fischeri* (Chapter 4). Rather it appears that the difference between the high-binding phenotype of cured hemocytes and the low-binding phenotype of sym hemocytes may involve a number of complex protein interactions.

A group of proteins identified as candidates for potentially playing a role in the recognition of *V. fischeri* from other non-symbiotic bacteria are galectins. Galectins are carbohydrate-binding proteins that specifically bind to beta-galactoside residues [165]. The surfaces of bacteria, including *V. fischeri*, are decorated with a variety of carbohydrate structures, which provide an opportunity for interactions with galectins or other lectin-like proteins [185,186]. While galectins are known to play a role in a diverse set of cellular processes, both mammals and invertebrates have galectins that have been shown to function in association with the immune response [166,187]. In several invertebrates, galectins are up-regulated in response to bacterial challenges [188-191]. In addition, galectins from *Crassostrea virginica* and *Branchiostoma belcheri* have the ability to bind to microorganisms [168,169]. Transcriptomic and proteomic analyses of hemocytes from *E. scolopes* have identified two galectin-like proteins

so far ([15,115]; Chapter 4). Comparison of sym and cured hemocytes revealed that one of these galectins, EsGalectin 2, is more abundant in hemocytes when the light organ is colonized by *V. fischeri* (Chapter 4). However, it is unclear the exact role of the galectins from *E. scolopes* in the squid-vibrio association.

To date no cephalopod galectins have been characterized. In addition, there is very little knowledge about the function of galectins in beneficial host-microbe symbioses. To help address these issues two previously identified galectin-like sequences were cloned from the light organ central core of *E. scolopes*. A combination of sequence analyses, reverse transcription (RT) PCR, and immunocytochemistry (ICC) were used to analyze EsGalectin 1 and EsGalectin 2. These initial studies set the foundation for further characterization of galectins in this squid-vibrio symbiosis.

Methods

Animal care and maintenance

Adult and juvenile *E. scolopes* squid were collected and maintained in the laboratory as previously described [15,115]. Juvenile squid were maintained as aposymbiotic in filter sterilized artificial seawater (FSASW) or symbiotic with the addition of *V. fischeri* to the FSASW at concentrations ranging from 5,000 -10,000 cells/ml. After incubation with the symbiont for 3 hours to overnight, colonization was monitored the next day with a luminometer (Berthold Detection Systems). *V. fischeri* ES114 was grown in either salt water tryptone (SWT) or Luria-Bertani Salt (LBS) at 28°C to an OD 600 between 0.6 and 1.0.

RNA extraction and cDNA synthesis

RNA was collected from *E. scolopes* tissues (hemocytes, central core, gills, mantle, brachial heart, systemic heart, brain, accessory nidamental gland, hindgut) using TRIzol reagent per the manufacturer's instructions (Life Technologies). Extracted RNA was treated with Turbo DNase as described in the manufacturer's methods (Life Technologies) and cDNA was synthesized using an iScript kit (Bio-Rad) from 200 ng of starting template and random primers.

Cloning and phylogenetic analysis

E. scolopes galectin protein sequences were identified from a database consisting of combined published and unpublished *E. scolopes* light organ and hemocyte transcriptomic sequences [8,15,21]. EsGalectin 1 and EsGalectin 2 sequences originating from this database were used for primer design, alignments, and other sequence analyses. Gene specific primers (Galectin1: 430GalectinF-ATGTCTTACAACTACACTAAATTTGATACTCC, 430GalectinR-TAAGATTTCAGGTTCAAAGAAATGG; Galectin2_20_F- ATGCTCGCTGAATCTGGAC, Galectin2_1160_R- ATGGCCAACAAGAATAAAATGG) were used to clone full length ORFs corresponding to both EsGalectin 1 and EsGalectin 2 from cDNA prepared from central core tissue. Both galectin sequences were ligated and cloned using a PGEM-T Easy kit and *Escherichia coli* JM109 cells (Promega). In order to verify the cloned galectin sequences with those originating from transcriptomic libraries, the plasmids containing the galectin sequences were sequenced using BigDye v1.1 according to the manufacturer's protocol (Applied

Biosystems). To get a better understanding of the relationship between the galectins from *E. scolopes* and other galectins from related invertebrates, a phylogenetic analysis of the individual galectin domains originating from 21 molluscan galectin sequences was performed (Table 5.1). Full-length galectin sequences were compared against the NCBI Conserved Domains Database (CDD) with RPS-BLAST [117]. Individual galectin domains were detected within these sequences with the GLECT CDD identifier 214596. In SeaView (version 4.4.2), MUSCLE was used to align the individual galectin domains and PhyML was used to create a maximum likelihood tree based on the LG model and 1000 bootstrap replicates [118]. The tree was rooted to a single galectin domain from the sponge, *Geodia cydonium*.

Table 5. 1: Galectin sequences originating from different molluscs used for phylogenetic comparisons.

Sequence ID	Protein and Organism
225906399	Galectin *Pinctada fucata*
110559485	Galectin *Crassostrea virginica*
316936680	Tandem-repeat galectin *Pinctada fucata*
241913786	Galectin 2 *Argopecten irradians*
241913784	Galectin 1 *Argopecten irradians*
295136551	Galectin *Ostrea edulis*
363902082	Galectin *Solen grandis*
460002040	Galectin *Tegillarca granosa*
167888935	Tandem repeat galectin *Ruditapes philippinarum*
146199199	Tandem-repeat galectin *Biomphalaria glabrata*
125743200	Galectin 4-like protein *Haliotis discus hannai*
524886489	Uncharacterized protein LOC101856665 *Aplysia californica*
405966479	Galectin-6 *Crassostrea gigas*
524891029	Galectin-5-like *Aplysia californica*
405972434	Galectin-4 *Crassostrea gigas*
405975976	Galectin-9 *Crassostrea gigas*
405974002	Galectin-8 *Crassostrea gigas*
524890997	Galectin-4-like isoform X1 *Aplysia californica*
524893908	Uncharacterized protein LOC101859875 *Aplysia californica*
N/A	Galectin 1 *Euprymna scolopes*
N/A	Galectin 2 *Euprymna scolopes*
23306810	Galectin *Geodia cydonium*

Western blot analysis

Rabbit polyclonal antibodies were generated (GenScript) using a peptide from EsGalectin 1 (DFMKEIPSGLQPKD) and EsGalectin 2 (WGSEVRSKNLPLRK). Peptides were chosen based on predicted antigenicity and inspected for homology to any other known *E. scolopes* or *V. fischeri* proteins.

Hemocyte and central core proteins were prepared as previously described [15,115]. Protein concentrations were determined by the *RC DC* protein assay (Bio-Rad). Proteins (50-60 μg) were separated using a 12% TGX Criterion gel (Bio-Rad) at 200V for 45 minutes. Proteins were transferred onto a .2 μM mini nitrocellulose membrane using a Trans-Blot Turbo transfer system and the Mixed Molecular Weight protocol per the manufacturer's instructions (Bio-Rad). Membranes were blocked in 4% milk, 0.1% Tween at room temperature for one hour. Primary antibody was diluted in block (anti-EsGalectin 1- 1:10,00 or anti-EsGalectin 2- 1:1,000) and incubated at room temperature for one hour. After washing the blot 4 times for 5 minutes each in 0.1% Tween, tris buffered saline (TTBS), pH 7.4, a goat anti-rabbit (IgG H+L) HRP conjugate (1:5,000, Bio-Rad) was diluted in fresh block and then incubated with the blot for an additional one hour at room temperature. The blot was washed again 6 times in TTBS for 5 minutes each and the antibody was detected using the Amersham Enhanced Chemiluminescence (ECL) Prime Detection kit (GE Healthcare Life Sciences) and imaged using a FluorChem HD2 CCD camera (Cell Biosciences) with an exposure time of 30 seconds.

Immunocytochemistry

Hemocytes were collected from adult and juvenile squid as previously described [15,16,133]. To collect juvenile hemocytes, 10 juvenile squid were anesthetized in 2% ethanol, FSASW and homogenized in 1 ml Squid Ringer's (50 mM $MgCl_2$, 10 mM $CaCl_2$, 10 mM KCl, 530 mM NaCl, 10 mM HEPES). The homogenate was applied to a glass coverslip. Both adult and juvenile hemocytes were allowed to adhere to coverslips in 12-well tissue culture plates for 30 minutes. After attachment, the hemocytes were rinsed two times in Squid Ringer's to remove excess tissue or cellular debris and then fixed with 4% paraformaldehyde in Squid Ringer's for 30 minutes at room temperature. After fixation the hemocytes were washed 4 times with marine phosphate buffered saline, pH 7.4 (mPBS; 50 mM sodium phosphate, 0.45 M NaCl) for 5 minutes each. Fixed hemocytes were permeabolized with 1% Triton X-100, mPBS, pH 7.4 for 30 minutes at room temperature. The cells were blocked (mPBS, pH 7.4, 1% Triton X-100, 1% goat serum, 0.5% BSA) at room temperature for 1 hour. Hemocytes were then exposed to anti-EsGalectin 1 or anti-EsGalectin 2 at a 1:400 dilution in block overnight. The primary antibody was then removed and the hemocytes were washed four times in 1% Triton X-100, mPBS pH 7.4 for 5 minutes each and then blocked for an additional hour. The hemocytes were then incubated with an Alexa Fluor 488 goat anti-rabbit secondary antibody (Life Technologies) at a 1:3,000 dilution in fresh block for 4 hours in the dark. Finally, hemocytes were washed four times in 1% Triton X-100, mPBS pH 7.4 for 5 minutes each and then counterstained using a lectin stain, ConA-TritC (0.2 μg/ml), and a nuclear stain, DRAQ5 (2.5 μM), in 1% Triton X-100, mPBS, pH 7.4 overnight in the dark. Finally, the hemocytes were mounted onto glass slides with Vectashield, sealed with nail polish, and imaged using an A1R confocal microscope.

End point RT-PCR

PCR was performed using gene specific primers (Galectin1_F TCCGCAAAGAAAACTCCATT, Galectin1_R GGTTGTTGTCCGCAATTCTT; Galectin2_F TGTAAGAATGCCTGCACGAG, Galectin2_R CAGTGGGACAAATCACGTTG; EsL21F GCCTTGGCTTGAGCCTTCAACTTT, EsL21R GGTGATCGTCAACAAACGCGTGAA), GoTaq Green Mastermix (Promega) and the following conditions: 3 min at 95°C, 35 cycles of 30s at 95°C, 30s at 55°C, and 1 min at 72°C, and a final extension at 72°C for 10 minutes. PCR products were separated on a 1% agarose gel supplemented with a 1:10,000 dilution of SYBR safe DNA gel stain (Life Technologies) and imaged using a Molecular Image Gel Doc (Bio-Rad).

Results

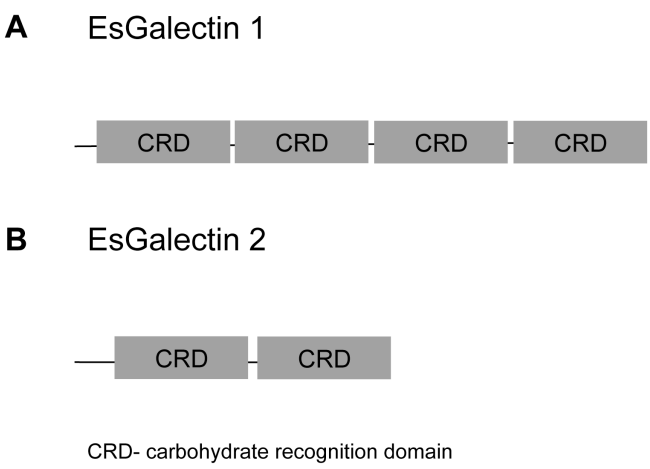


Figure 5. 1: Protein domain organization of galectins identified from *E. scolopes*.

Two galectin-like proteins have previously been identified by transcriptomics and proteomics from hemocytes and the light organ tissue of *E. scolopes* [15]. Using gene specific primers, both galectins were cloned from light organ central core cDNA. Analysis of the primary

amino acid sequence of both putative galectins revealed that EsGalectin 1 has 4 carbohydrate recognition domains (CRDs, Fig. 5.1A) and the other, EsGalectin 2 had 2 CRDs (Fig. 5.1B).

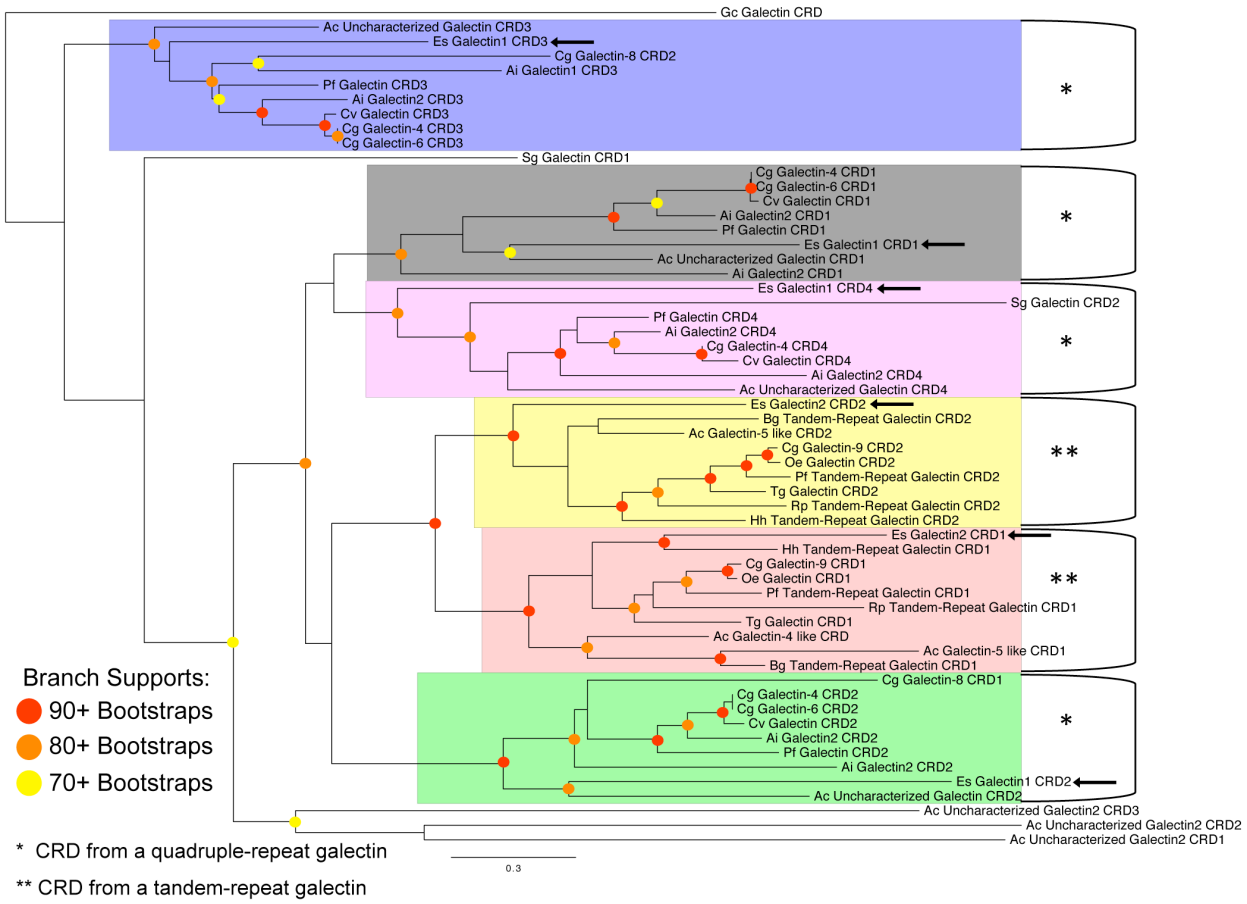


Figure 5. 2: Phylogenetic reconstruction of individual galectin domains originating from different species of molluscs.

Individual galectin domains originating from a set of 21 molluscan galectin sequences were identified with the Conserved Domains Database (CDD) and aligned with MUSCLE. A maximum likelihood tree was created with PhyML using the LG model and 1000 bootstrap replicates. Similar types of carbohydrate recognition domains, based on position within a full-length sequence, cluster together and are highlighted by different colors. Abbreviations: CRD- carbohydrate recognition domain; Es- *Euprymna scolopes*; Cg- *Crassostrea gigas*; Cv- *Crassostrea virginica*; Pf- *Pinctada fucata*; Ac- *Aplysia californica*; Bg- *Biomphalaria glabrata*; Rp- *Ruditapes philippinarum*; Hh- *Haliotis discus hannai*; Tg- *Tegillarca granosa*; Sg- *Solen grandis*; Oe- *Ostrea edulis*; Ai- *Argopecten irradians*; Gc- *Geodia cydonium*

A phylogenetic tree composed of 58 CRDs from 21 molluscan galectin sequences (Table 5.1) demonstrated that the CRDs from EsGalectin 1 cluster with the CRDs from other quadruple-repeat galectins (Fig. 5.2: *) and the CRDs from EsGalectin 2 group with CRDs from other tandem repeat-like galectins consisting of 2 CRDs, with only a few exceptions (Fig. 5.2: **). The tree also highlights clustering specific to the various positions of the CRDs within a particular galectin. Both galectins from *E. scolopes*, as revealed by the CRDs, are more divergent compared to the CRDs from the other organisms in each group.

To better characterize the presence of native galectin proteins in squid tissues, rabbit polyclonal antibodies were generated to unique peptides originating from each galectin. A western blot analysis with anti-EsGalectin 1 revealed prominent bands corresponding to a 65 kD protein in both the central core and adult sym hemocyte samples (Fig. 5.3A). For anti-EsGalectin 2 a faint band of approximately 40.5 kD was detected in adult sym hemocytes (Fig. 5.3B). The predicted molecular weights of EsGalectin 1 and 2 from the primary amino acid sequences are 63.8 kD and 35.6 kD, respectively.

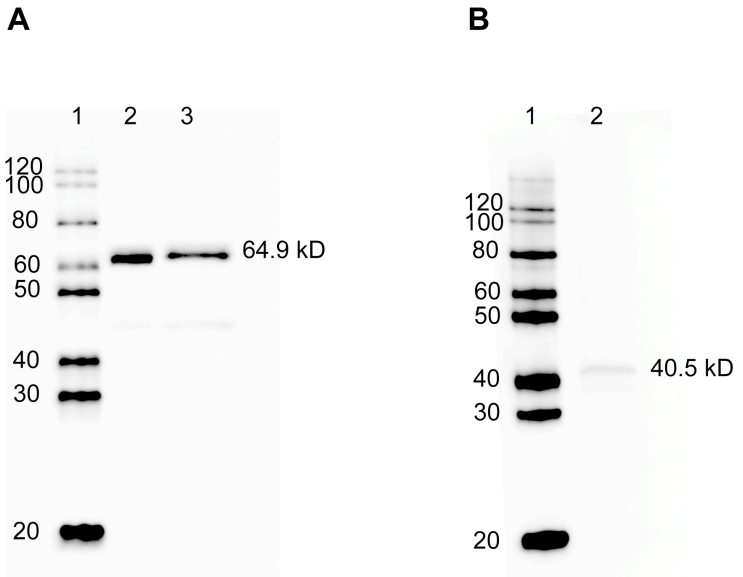


Figure 5. 3: Galectin western blot analysis.

(A). EsGalectin 1: Lane 1-protein standard, kD, Lane 2- hemocytes, Lane 3- central core **(B)**. EsGalectin 2: Lane 1- -protein standard, kD, Lane 2- hemocytes. EsGalectin 1 is indicated by a prominent 64.9 kD protein band present in both the hemocyte and central core tissue. EsGalectin 2 is represented by a faint band of approximately 40.5 kD.

Using reverse transcribed RNA, the distribution of galectin transcripts in adult tissues from *E. scolopes* was characterized by end point PCR. Both EsGalectin 1 and EsGalectin 2 were found in each adult tissue analyzed (Fig. 5.4). Overall this widespread pattern of distribution did not reveal differences between tissues and between galectins. However, there was one exception, transcripts of EsGalectin 1 appear to be more abundant than EsGalectin 2 in the accessory nidamental gland (ANG, Fig. 5.4).

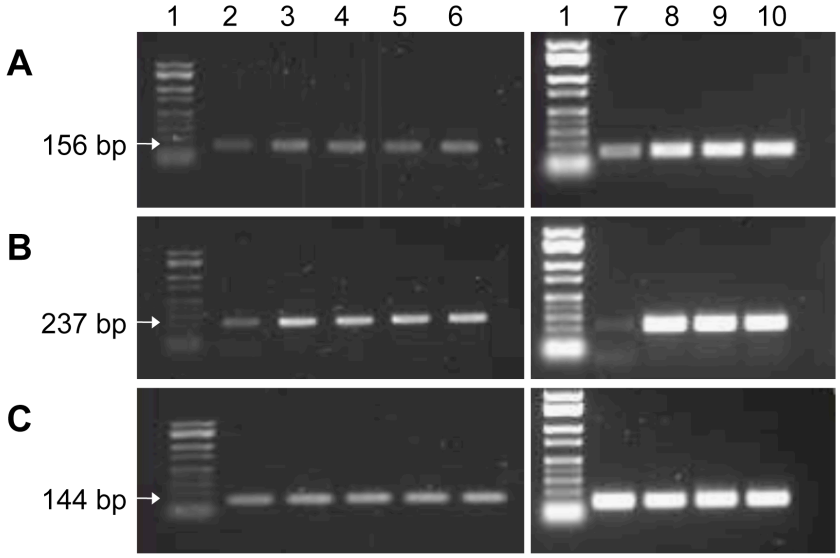


Figure 5. 4: Both galectin transcripts are ubiquitously expressed in *E. scolopes*.

End point RT-PCR of RNA collected from adult squid tissues was used to determine the presence of **(A)**. EsGalectin 1 (156 bp), **(B)**. EsGalectin 2 (237 bp) and **(C)**. L21 ribosomal subunit (144 bp) transcripts using gene specific primers (1-100 bp DNA standard, 2-hemocytes, 3-gills, 4-central core, 5-mantle, 6-hindgut, 7-accessory nidamental gland, 8-brachial heart, 9-systemic heart, 10-brain).

Immunocytochemistry with anti-EsGalectin 1 and anti-EsGalectin 2 was used to analyze the localization of each galectin in adult and juvenile hemocytes (Table 5.2). EsGalectin 1 was identified in several cellular locations including the filopodia, cytoplasmic vacuole-like compartments, and the nucleus (Fig. 5.5). These localizations for EsGalectin 1 have been identified in both juvenile and adult hemocytes with no detectable differences between the two developmental stages (Table 2). EsGalectin 2 demonstrated a more consistent pattern of localization as this protein was predominantly detected in the nucleus of both juvenile and adult hemocytes (Fig. 5.6). While EsGalectin 2 is mostly nuclear there appears to be some diffuse staining present in the cytoplasm as well (Fig. 5.6). An exception to the nuclear localization of EsGalectin 2 was found in several cells imaged from hemocytes of 24 hour symbiotic juveniles that were challenged with or without *V. fischeri* (Fig. 5.7). EsGalectin 2 switched from a nuclear to a perinuclear localization (Fig. 5.7). These cells had noticeable vesicle-like formations present on the filopodia (Fig. 5.7C,D).

Table 5. 2: Maturity and colonization state of representative hemocyte images.

Figure	Hemocyte Sample
Figure 4	
A	Adult sym
B	Juvenile 24-48 hr apo
C	Juvenile 24 hr apo
D	Adult sym
Figure 5	
A	Adult sym
B	Juvenile 24 hr sym
C	Juvenile 24-48 hr apo
Figure 6	
A	Juvenile 24 hr sym + *V. fischeri*
B	Juvenile 24 hr sym + *V. fischeri*
C	Juvenile 24 hr sym + *V. fischeri*
D	Juvenile 24 hr sym + *V. fischeri*

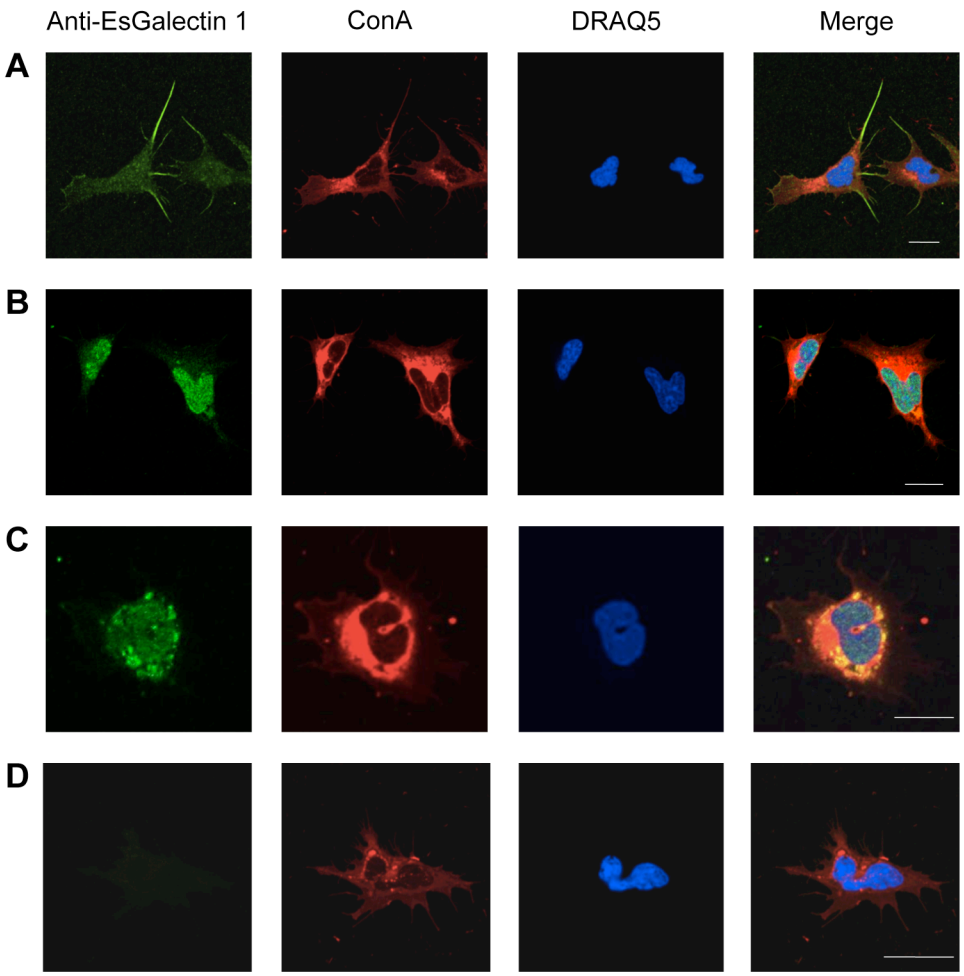


Figure 5. 5: The localization of EsGalectin 1 in hemocytes.

EsGalectin 1 can be found in the **(A)**. filopodia **(B)**. nucleus and **(C)**. cytoplasm of hemocytes (**D)**. Secondary control, scale-10 µM. Anti-EsGalectin 1 is stained with a goat anti-rabbit Alexa fluor 488 antibody. The hemocytes are counterstained with ConA-TritC (red, a lectin stain) and DRAQ5 (blue, a nuclear stain).

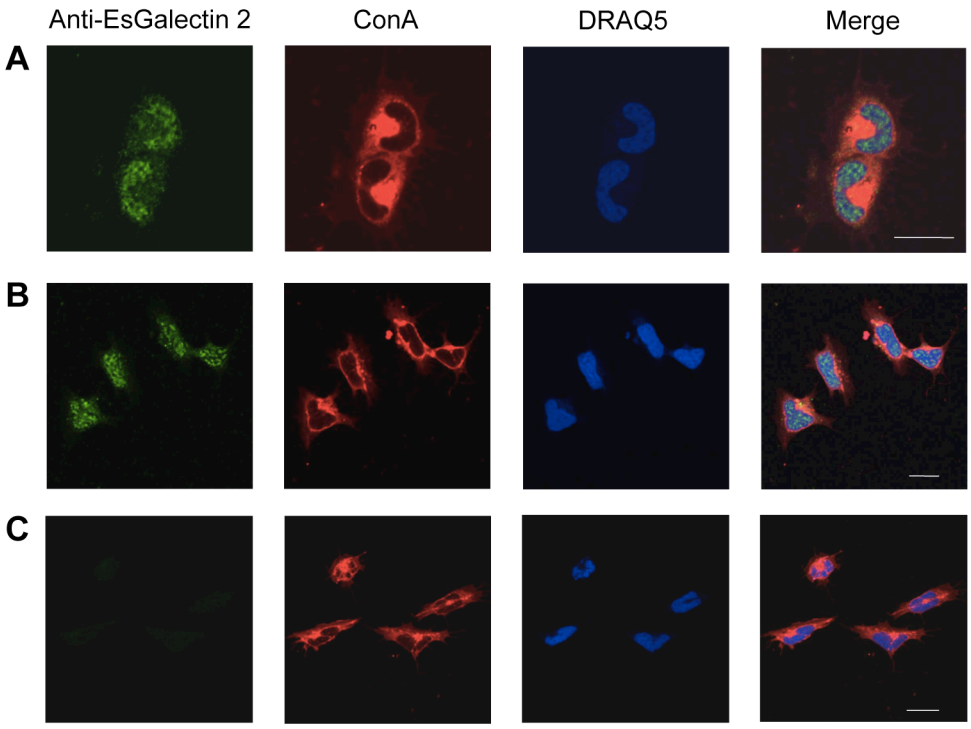


Figure 5. 6: The localization of EsGalectin 2 in hemocytes.

EsGalectin 2 can be found in the nucleus of **(A)**. adult **(B)**. juvenile hemocytes **(C)**. secondary control, scale-10 μM. Anti-EsGalectin 2 is stained with a goat anti-rabbit Alexa fluor 488 antibody. The hemocytes are counterstained with ConA-TritC (red, a lectin stain) and DRAQ5 (blue, a nuclear stain).

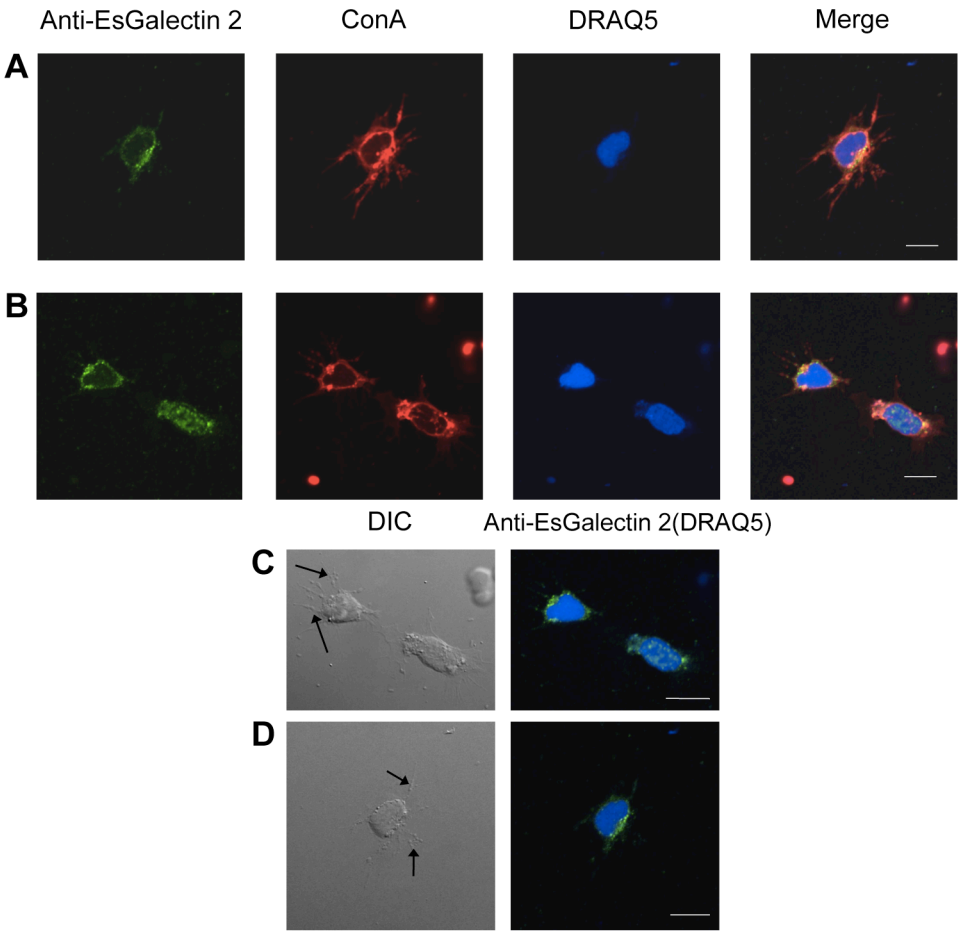


Figure 5. 7: EsGalectin 2 can become perinuclear.

Hemocytes in **(A)**, **(B)** show an alternative cellular localization from Figure 5.6. **(C)**, **(D)**. DIC images and Anti-EsGalectin 2/Nuclear images of hemocytes shown in **(A)**, **(B)**. Black arrows indicate possible membrane vesicles associated with this EsGalectin 2 localization.

Discussion

The mechanisms by which *E. scolopes* recognizes *V. fischeri* are partially mediated by the innate immune response of the host [1,46]. However the proteins underlying this process remain uncharacterized. Recent quantitative proteomic comparisons of sym and cured hemocytes suggested that one protein is not responsible for the host tolerance to *V. fischeri*, but the coordination between many proteins may be necessary (Chapter 4). Among the proteins that were significantly more abundant in sym hemocytes was EsGalectin 2. Galectins are proteins

that bind carbohydrate residues (a type of microbe-associated molecular pattern (MAMP)) and are utilized as pattern recognition receptors in a number of animal hosts to identify microorganisms and influence the immune response. Given the distribution of sugars on the surface of microorganisms, including *V. fischeri*, the galectins from *E. scolopes* represent a possible mechanism by which the host might be able to differentiate the symbiont from other related non-symbiotic bacteria [185]. Cephalopod galectins have yet to be described, but other galectins from related molluscs have already been recognized for their ability to interact with microbes [167,168,188-192]. In this study two putative galectins from *E. scolopes* were cloned and observed with fluorescence confocal microscopy to further understand how these proteins might contribute to the host-symbiont specificity.

Analysis of *E. scolopes* galectin sequences revealed that these squid proteins contain the canonical carbohydrate recognition domains (CRDs) found in galectins of other species. EsGalectin 1 contains 4 CRDs and EsGalectin 2 has 2 CRDs (Fig. 5.1). The organization of CRDs is used to classify galectins into different groups. Currently there are four described CRD organizations that include proto (1 CRD that forms a dimer), chimera (1 CRD with a N-terminal polypeptide), tandem-repeat (2 CRDs linked together by a peptide), and quadruple-repeat (4 sequential CRDs) [165]. Galectins identified with 4 CRDs have only been described in molluscs [168,189,190]. EsGalectin 1 represents a new quadruple-repeat galectin and is the first representative galectin from the cephalopods. EsGalectin 2 is similar to the tandem-repeat galectins, which demonstrate a much broader distribution among taxonomic groups. Comparison of the CRDs from the *E. scolopes* galectins to the CRDs from other available molluscan galectins revealed that, generally, individual CRDs group separately from one another and cluster specifically by CRD position and type of galectin sequence, similar to what has been described

by other galectin studies (Fig. 5.2; [168,188]). The squid galectins represent the first cephalopod sequences analyzed and are very divergent relative to the other molluscan galectins already described. However, this hypothesis is limited by the lack of other cephalopod sequences in public databases and further analyses will be necessary to clarify the evolutionary history of molluscan galectins. In addition, it is likely that more galectins from *E. scolopes* will be discovered once a genome is made available.

Both EsGalectin 1 and 2 transcripts were identified in all adult squid tissues tested (Fig. 5.4). The ubiquitous distribution of both galectins is not surprising. The widespread presence of galectins has also been reported in oysters and scallops [168,188,190]. It is possible that these end point RT-PCR experiments may represent basal levels of galectin expression and any changes in expression after a bacterial infection may offer a better understanding of any immunological role for each galectin in these different tissues. Additionally, galectins are known to be multifunctional proteins and may be important to cellular processes outside of the immune response. Of all of the tissues analyzed only the accessory nidamental gland (ANG) appeared to show a difference in galectin expression with a higher expression of EsGalectin 1 transcripts (Fig. 5.4). The ANG is a female reproductive organ that contains a dense population of bacteria, but the benefit of this symbiosis remains unknown [116]. EsGalectin 2 represents a unique target to study the role of this differentially expressed transcript in two separate symbiotic tissues, the light organ and the ANG.

In order to get a better understanding of how galectins might contribute to the immune response of the host, immunocytochemistry was used to identify the cellular localization of each galectin in the hemocytes of the squid (Fig. 5.5-5.7). No differences in galectin localization were identified in hemocytes depending on colonization state of the light organ or maturity of the

squid. EsGalectin 1 was identified in several cellular locations including the filopodia, the nucleus and the cytoplasm (Fig. 5.5). EsGalectin 2 was predominantly found in the nucleus (Fig. 5.6). However, hemocytes imaged from symbiotic juvenile squid revealed that EsGalectin 2 could also be perinuclear (Fig. 5.7). In addition, hemocytes with a perinuclear localization of EsGalectin 2 had putative membrane vesicles located on the filopodia (Fig. 5.7C,D). It is not uncommon for galectins to exhibit multiple localizations. Galectins in mammals are known to shuttle between the nucleus and the membrane [193]. In addition, even with the absence of a canonical secretion signal some galectins can be secreted. In the oyster, *Crassostrea virginica*, a quadruple-repeat galectin (CvGal) similar to EsGalectin 1 translocates to the membrane upon hemocyte attachment and is then secreted by an unknown mechanism [168]. It is possible that the both *E. scolopes* galectins can be secreted given the filopodia localization of EsGalectin 1 and the loss of EsGalectin 2 from the nucleus in cells with membrane vesicles, however this hypothesis remains to be characterized.

This study represents the initial characterization of two galectins from *E. scolopes*. Not only are these the first cephalopod galectins to be analyzed, but these proteins also provide the opportunity to study the role of galectins in a beneficial host-microbe relationship. The results presented here demonstrated that the galectins of *E. scolopes* are widely distributed in adult squid tissues and can be detected in multiple cellular localizations in one cell type (hemocytes), regardless of the developmental stage or colonization state. In addition, this study confirms the presence of these galectins in the hemocytes and further suggests that these proteins may function in regulating the host immune response. The galectins of *E. scolopes* share a number of characteristics with the galectins identified in other molluscs so far, but there are still many questions to be answered. Future work should involve determining the ability of each galectin to

bind bacteria. In *Branchiostoma belcheri*, a tandem-repeat galectin was shown to bind *Vibrio vulnificus*, but not *Vibrio parahaemolyticus* or *Staphylococcus aureus* [169]. The selective binding of this galectin from *Branchiostoma* to one vibrio species and not another is promising given the desire to understand symbiont recognition by *E. scolopes*. Comparing the ability of each *E. scolopes* galectin to bind *V. fischeri* and other non-symbiotic bacteria will provide important information on how the squid can achieve the host-symbiont specificity required by this light organ association.

Appendix I: Comparison of the hemocyte proteome and transcriptome

Background

Increasing evidence supports the hypothesis that macrophage-like hemocytes of the squid, *Euprymna scolopes*, play an important, but poorly characterized role in the maintenance of the light organ symbiosis with *Vibrio fischeri* [5,14]. To date there have been very few in depth "omic" analyses involving invertebrate hemocytes, especially in regards to cephalopods. In order to get a better understanding of these squid blood cells, both protein and RNA were extracted from hemocytes collected directly from the cephalic blood vessel of *E. scolopes*. These biological samples were used to generate a representative transcriptome and proteome of the hemocytes from adult squid colonized with *V. fischeri*.

The methods for the mass spectrometry (MS) based proteomics and analysis of the identified hemocyte proteins presented in Collins et al., 2012 are explained below [15].

Methods

Protein extraction

Hemocytes collected from an adult symbiotic squid were homogenized in a 64 mM Tris buffer, pH 8.0 with the addition of 1% SDS, and a 1× protease inhibitor cocktail (P2714, Sigma Aldrich). After the homogenate was centrifuged (Eppendorf 5810 R, 14,000 rpm, 4°C, 30 min), the protein concentration of the supernatant was quantified spectrophotometrically using the *RC DC* protein assay (Bio-Rad). Alternatively, hemocyte proteins from adult symbiotic squid were extracted using the Ready Prep Protein Extraction Kit (Membrane I) as described by the

manufacturer's protocol (Bio-Rad). Hemocyte proteins from two adult squid were separated into either a hydrophilic or a hydrophobic layer based on temperature dependent phase partitioning. The protein concentration of each layer was determined with the *RC DC* protein assay as described above (Bio-Rad).

Protein digestion

Proteins were digested as previously described [115]. Hemocyte proteins were precipitated with 10% (w/v) trichloroacetic acid (Fisher Scientific) at 4°C overnight. The precipitates were collected by centrifugation (Eppendorf 5810 R, 11,000 × g, 30 min, 4°C) and washed twice with ice-cold acetone. Protein pellets were briefly air-dried and then resuspended in 25 µl of 8 M urea, 0.4 M ammonium bicarbonate, pH 8.0. Samples were reduced and alkylated with 5 µl of 45 mM dithiothreitol (DTT; Acros Organics) at 37°C for 20 min and 5 µl of 100 mM iodoacetamide (Acros Organics) at room temperature in the dark for 20 additional minutes. Sequencing grade trypsin was added 1:15 (w/w enzyme to protein; Promega, V5111). The solutions were diluted in water to 100 µl (2 M urea final concentration). Both samples were digested at 37°C for 18–24 h and then stored at −80°C until further analysis.

Protein identification using nanoLC-MS/MS

Protein identifications were completed at the W.M. Keck Biotechnology Resource Laboratory at Yale University as previously described [115]. Briefly, 5 µl (0.375 µg) of the hydrophilic and hydrophobic fraction of hemocyte protein extraction from the Ready Prep Protein Extraction Kit I were analyzed on a LTQ Orbitrap equipped with a Waters nanoAcquity UPLC system. Samples were concentrated at 15 µl/min with 99% Buffer A (100% water, 0.1%

formic acid) for 1 min on a trap column (Waters Symmetry® C18 180 µm × 20 mm). Peptides were separated at 300 nl/min with Buffer A and Buffer B (100% CH3CN, 0.075% formic acid) using a 1.7-µm, 75 µm × 250 mm nanoAcquity™ UPLC™ column (35°C). A 51-min linear gradient was established using increasing amounts of Buffer B (95% Buffer A, 5% Buffer B at initial conditions, 50% A, 50% B, at 50 min, and 15% A, 85% B at 51 min). MS was acquired in the Orbitrap using one microscan and an inject time of 900 followed by four data dependent MS/MS acquisitions in the ion trap.

Hemocyte proteins from an adult symbiotic squid were also identified using a 5600 TripleTOF (AB SCIEX) equipped with a Waters nanoAcquity UPLC system. Briefly, 5 µl (3 µg) of sample were loaded onto a Waters Symmetry® C18 180 µm × 20 mm trap column at 15 µl/min with 99% Buffer A for 1 min. Peptides were separated with a 1.7-µm, 75-µm × 150-mm nanoAcquity™ UPLC™ column (45°C) using a 161-min linear gradient at 500 nl/min. The gradient included 95% Buffer A, 5% Buffer B at initial conditions, 60% A, 40% B at 160 min, and 15% A, 85% B at 161 min. MS was acquired from 400 to 1250 Da for 0.25 s. While in high sensitivity mode, 20 data dependent MS/MS were acquired for 0.05 s each.

Proteomic data analysis

MS/MS spectra were converted to Mascot compatible files using either the Mascot Distiller (Orbitrap MS/MS) or the AB SCIEX MGF Converter (5600 TripleTOF MS/MS) and searched with the Mascot algorithm (Version 2.3.01; [36,38]) using an *E. scolopes* protein sequence database (25,745 sequences) generated from available transcriptomic data: juvenile light organ ESTs [8] and hemocyte transcripts (this study). Searching parameters included peptide charge states of +2 or +3, partial methionine oxidation, carboxamidomethylated cysteine,

a peptide tolerance of ±20 ppm, and MS/MS fragment tolerance of ±0.6 Da. A decoy database search was also performed by Mascot.

Scaffold (version Scaffold_3.2.0, Proteome Software Inc.) was used to validate the MS/MS based peptide and protein identifications made by Mascot. In addition, Scaffold applied X! Tandem [194] [The GPM, thegpm.org; version CYCLONE (2010.12.01.1)] as a supplementary search to further confirm protein and peptide identifications. X! Tandem applied the same search parameters as Mascot. Peptide and protein identifications were accepted at greater than 95.0% probabilities each. Only proteins that met the necessary criteria and contained two or more peptides were considered as positive identifications. For positive identifications, the false discovery rates (FDRs) were 0% at both the protein and peptide level as calculated by Scaffold. Putative protein identifications included proteins with only one matching peptide (protein FDR, 0.4%).

Results

Mass spectrometry (MS) based proteomics was used to identify proteins from circulating hemocytes collected directly from the cephalic blood vessel of adult *E. scolopes* colonized by *V. fischeri.* Two protein extraction methods were used to isolate proteins from the hemocytes. One involved the total soluble protein collected after homogenization of the hemocytes from 1 adult squid and the other involved a hydrophobic and hydrophilic phase separation of proteins from the hemocytes pooled from 2 adult squid. The spectral data from the hemocytes of these 3 adult squid were combined in Scaffold and compared against either a translated version of the hemocyte transcriptome or a translated combination of the hemocyte and a previously published juvenile light organ transcriptome. The combined translated *E. scolopes* database allowed for the

detection of 702 unique proteins at a false discovery rate (FDR) of 0.4% (Table A1.1, Supplemental file A1.1). Of the 702 proteins, 630 were identified with 2 or more peptides while 72 were putatively identified from 1 peptide (Table A1.1, Supplemental file A1.1).

Table A1. 1: MS-based proteomic analysis of hemocytes from *E. scolopes*.

Number of Squid	3
Number of MS/MS Spectra	107,523
Number of Spectra Identified	8,147
Number of Proteins Identified	630
Number of Proteins Putatively Identified (1 peptide)	72
Peptide False Discovery Rate	0.0%
Protein False Discovery Rate	0.4%

Both the hemocyte proteome and the transcriptome were annotated using gene ontologies (GO) [195]. Comparison of these datasets revealed that both the transcriptome and proteome yielded similar proportions of genes and proteins involved with biological processes, molecular function, and cell components (Fig. A1.1). The proteome was enriched in sequences associated with structural molecule activity and macromolecular processes as opposed to the transcriptome, which had more sequences representing metabolic processes and catalytic activity (Fig. A1.1). Direct comparison of the protein sequences identified with the hemocyte transcriptome revealed 596 genes/proteins in common between the two datasets and 145 proteins that were only detected by proteomics (Fig. A1.2, Supplemental file A1.2). Among the proteins unique to the proteome were two histones, three 26S proteasome subunits, two thioester proteins (TEPs), an alpha macroglobulin, and two transmembrane proteins with emp24 domains (Table A1.2, Supplemental file A1.2).

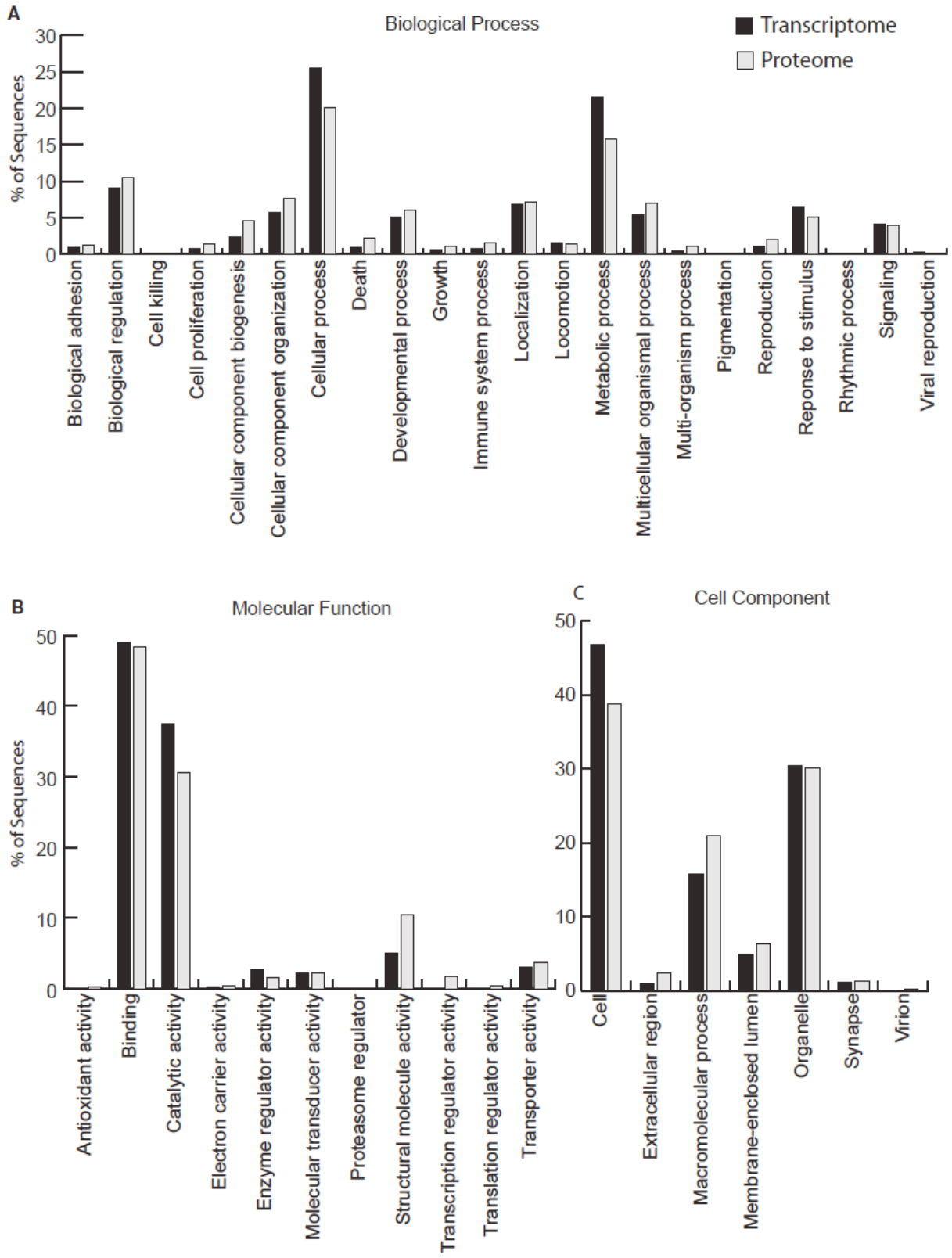

A
Biological Process
Transcriptome
Proteome
% of Sequences
0
5
10
15
20
25
30
Biological adhesion
Biological regulation
Cell killing
Cell proliferation
Cellular component biogenesis
Cellular component organization
Cellular process
Death
Developmental process
Growth
Immune system process
Localization
Locomotion
Metabolic process
Multicellular organismal process
Multi-organism process
Pigmentation
Reproduction
Reponse to stimulus
Rhythmic process
Signaling
Viral reproduction
B
Molecular Function
% of Sequences
0
10
20
30
40
50
Antioxidant activity
Binding
Catalytic activity
Electron carrier activity
Enzyme regulator activity
Molecular transducer activity
Proteasome regulator
Structural molecule activity
Transcription regulator activity
Translation regulator activity
Transporter activity
C
Cell Component
0
10
20
30
40
50
Cell
Extracellular region
Macromolecular process
Membrane-enclosed lumen
Organelle
Synapse
Virion

Figure A1. 1: Gene ontology (GO) analysis of the hemocyte transcriptome and proteome.

Annotation by Blast2GO revealed similar proportions of level 2 GO terms in both the transcriptome and proteome for (A) biological process, (B) molecular function and (C) cell component. Black bars represent transcriptome sequences and gray bars represent protein sequences.

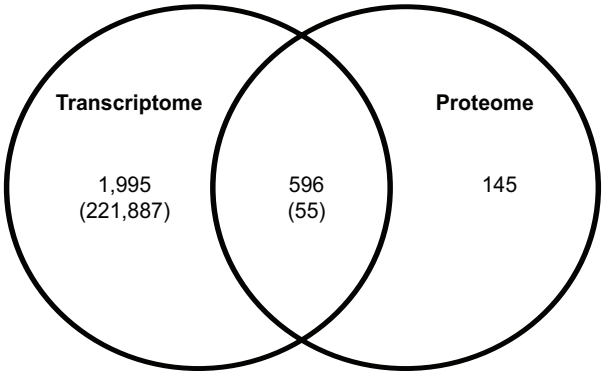


Figure A1. 2: Comparison of the hemocyte transcriptome and proteome.

Protein sequences identified by hemocyte peptides were compared to the hemocyte transcriptome using tblastn. 596 sequences were found to be in common between the transcriptome and proteome while 145 protein sequences were identified solely using proteomics. Numbers in parentheses represent the total number of annotated singletons (transcriptome) and the proteome matches to the singletons (shared).

Several proteins with similarity to members of the complement cascade were identified from the hemocyte proteome, including the central complement component EsC3, two thioester proteins (TEPs), an alpha macroglobulin, and a matrix metalloproteinase similar to vitronectin (Table A1.3). A 398 aa residue protein with six sushi repeat domains and a protein with similarity to the complement protein C1qBP were also detected. As mentioned above, some of these complement-related proteins, the TEPs and the alpha macroglobulin, were unique to the proteome. In addition, two toxin-related proteins, cephalotoxin and plancitoxin, were discovered by both the proteome and the transcriptome (Table A1.4). Other proteins identified in this study related to the innate immune response include pattern recognition receptors (PRRs) with the ability to respond to microbe-associated molecular patterns (MAMPs) such as peptidoglycan and

carbohydrates. Two peptidoglycan recognition proteins (PGRPs; EsPGRP 4 and EsPGRP 5) and one galectin were detected by both analyses (Table A1.4).

Table A1. 2: Proteins not identified by the hemocyte transcriptome.

ID	Protein	Total SpC	Unique Pep
UI-S-GU1-aed-n-06-0-UI.contig	26S proteasome non-ATPase regulatory subunit 14	1	1
UI-S-GB1-aar-p-22-0-UI.contig	26S proteasome non-ATPase subunit 13-like	3	3
UI-S-GU0-ado-m-18-18-UI.r23	26S proteasome non-ATPase subunit 3-like	2	2
UI-S-GG1-abg-f-07-18-UI.r23	Alpha macroglobulin	4	2
UI-S-GG1-abc-k-03-0-UI.contig	Histone H2A	8	2
UI-S-GB1-aah-c-18-0-UI.s1	Histone H3.3	37	5
UI-S-HH0-adf-c-09-0-UI.s1	Myophilin	3	2
UI-S-GS1-acj-a-19-0-UI.s1	TEP2 protein	1	1
UI-S-GU0-acx-i-20-18-UI.r23	Thioester-containing protein	5	2
UI-S-GG1-abh-b-09-0-UI.s1	Transmembrane emp24 domain containing 3-like	1	1
UI-S-GB1-aak-f-14-18-UI.r23	Transmembrane emp24 domain-containing protein 9	4	4

Abbreviations: SpC- Total number of spectra identified; Pep- Total number of unique peptides identified.

Table A1. 3: Complement pathway and other innate immunity-related proteins identified in the hemocytes.

Protein	Mascot score[a]	Scaffold[b]	Pep[c]	SpC	Annotation	e-value
EsC3	24	60.2	1	1	*E. scolopes*	0
Cephalotoxin	39	99.8	1	1	*S. esculenta*	2e-118
TEP 1	218	99.8	2	5	*C. farreri*	2e-11
TEP 2	40	98.9	1	1	*G. morsitans morsitans*	1e-13
Alpha macroglobulin	83	99.8	2	4	*C. farreri*	4e-7
MMP/ vitronectin	45	99.8	2	2	*S. kowalevskii*	7e-7
Sushi repeat protein	403	100	9	17	*X. tropicalis*	5e-31
Hypothetical protein (similar to C1qBP)	25	nv	1	1	*B. floridae*	6e-37

Abbreviations: *E. scolopes*: *Euprymna scolopes*, *B. floridae*: *Branchiostoma floridae*, *C. farreri*: *Chlamys farreri*, *G. morsitans morsitans*: *Glossina morsitans morsitans*, *X. tropicalis*, *Xenopus tropicalis* *S. kowalevskii*: *Saccoglossus kowalevskii*, S. *esculenta*; *Sepia esculenta*. TEP: thioester protein, MMP: Matrix metalloproteinase, nv: not validated.

a. Protein scores were derived from peptide ion scores and were determined by Mascot. Peptide scores greater than 20 were considered significant (p-value < 0.05).
b. Mascot identifications were validated by Scaffold. Scaffold utilized the Mascot scores to calculate a protein identification probability.
c. Unique peptides matched to protein sequences generated from translated light organ and hemocyte ESTs in addition to published *E. scolopes* protein sequences present in the NCBI nr database.

Table A1. 4: Additional innate immunity-related transcripts and proteins identified in the hemocytes.

Putative ID	Accession	Organism	e-value	Putative function	Pep[a]
NF-κB cascade					
TRAF6	AAY27978.1	*E. scolopes*	4e-48	Nf-κB signaling	nd
TRAF4	ABN04153.1	*B. belcheri*	9e-50	Nf-κB signaling	nd
REL/Nf-κB	AAY27981.1	*E. scolopes*	9e-65	Nf-κB signaling	nd
Inhibitor of Nf-κB	AAY27980.1	*E. scolopes*	4e-4	Nf-κB signaling	nd
Nf-κB activating protein	CAX69433.1	*S. japonicum*	1e-36	Nf-κB signaling	nd
Calcineurin subunit A	ABO26624.1	*H. discus*	5e-54	Suppressor of Nf-κB	nd
Calcineurin subunit B	ACI96107.1	*P. fucata*	2e-56	Suppressor of Nf-κB	nd
Regulator of calcineurin	NP_001011215.1	*X. tropicalis*	4e-11	Regulator of calcineurin	nd
MyD88	BAC76897.1	*C. porcellus*	1e-20	Toll signaling protein	nd
MAMP binding					
Galectin	ACS72241.1	*A. irradians*	3e-43	Glycan binding	3
EsPGRP 5	ACV67267.1	*B. manjavacas*	5e-38	PGN binding	4
EsPGRP 4	AAY27976.1	*E. scolopes*	1e-52	PGN binding	4[b]
Other					
IK cytokine	ABV24915.1	*C. gigas*	7e-47	Anti-interferon cytokine	nd
CD63 antigen	ABY87409.1	*H. diversicolor*	3e-25	Phagosome protein	3
Plancitoxin	Q75WF2.1	*A. planci*	1e-68	DNA-degrading	2
MMP/ vitronectin	BAC66058.1	*S. kowalevskii*	7e-7	Immune homeostasis	1
Cephalotoxin	B2DCR8.1	*S. esculenta*	2e-118	Toxin	2

Abbreviations: *E. scolopes*: *Euprymna scolopes*, *B. belcheri*: *Branchiostoma belcheri*, *S. japonicum*: *Schistosoma japonicum*, *H. discus*: *Haliotis discus*, *P. fucata*: *Pinctada fucata*, *X. tropicalis*, *Xenopus tropicalis*, *C. porcellus*: *Cavia porcellus*, *A. irradians*: *Argopecten irradians*, *B. manjavacas*: *Brachionus manjavacas*, *C. gigas*: *Crassostrea gigas*, *H. diversicolor*: *Haliotis diversicolor*, *A. planci*: *Acanthaster planci, S. kowalevskii*: *Saccoglossus kowalevskii*, S. *esculenta*; *Sepia esculenta*. PGN: peptidoglycan, MMP: Matrix metalloproteinase, nd: not detected.

a. Mascot and Scaffold validated peptide matches using translated hemocytes ESTs
b. PGRP 4 peptide matches were identified in an additional search using a combination of light organ and hemocyte translated transcripts as well as published full-length *E. scolopes* protein sequences

Discussion

This study represents the first transcriptomic and proteomic analyses of hemocytes from adult *E. scolopes* colonized by *V. fischeri*. The data presented here also revealed a number of genes and proteins that have been previously characterized in respect to this squid-vibrio association. In addition, this study allowed the identification of proteins which have not been described in regards to this symbiosis including a 5th peptidoglycan recognition protein (EsPGRP 5) of *E. scolopes* and two putative toxins (cephalotoxin and plancitoxin; Table A1.4). Both of these toxins have not been previously identified in invertebrate hemocytes. Quantitative PCR also revealed that several transcripts (EsPGRP 5, EsC3, cephalotoxin, and a MMP) of the hemocytes are differentially expressed in response to the colonization state of the light organ (described in detail in [15]). This study established the foundation for exploring the molecular mechanisms of host-symbiont specificity in the squid-vibrio symbiosis in respect to the hemocytes.

Several putative pattern recognition receptors (PRRs) including a galectin and two PGRPs (EsPGRP 4 and EsPGRP 5) were detected in both the hemocyte proteome and transcriptome (Table A1.4). Galectins are carbohydrate-binding proteins that have been implicated as PRRs during the immune response [51,166,187]. The role of galectins in this

symbiosis have yet to be fully characterized, but efforts are underway to compare gene expression and protein levels in hemocytes of colonized and uncolonized hosts (Chapters 4 and 5). The other PRRs identified were two peptidoglycan recognition proteins, EsPGRP 4 and EsPGRP 5 (Table A1.4). A total of five PGRPs have now been described from *E. scolopes* ([49,50,178]; this study). Previous work has demonstrated that *V. fischeri* influences the localization and expression of two PGRPs in the juvenile light organ (EsPGRP 1 and 2; [50,178]). Proteomics also contributed to the identification of the prominent PGRPs of the adult light organ, EsPGRP 2 and EsPGRP 3 [115]. It appears that the light organ PGRPs differ from the hemocyte PGRPs identified in this study (EsPGRP 4 and EsPGRP 5). The role of both hemocyte PGRPs in host-symbiont specificity are currently being investigated.

Recent analyses of the squid-vibrio association have identified members of the complement pathway [55,115]. C3 is involved in opsonization and phagocytosis and is considered to be the central protein to all three pathways of the complement system [196]. A C3 ortholog was recently discovered in *E. scolopes* and localized to the apical surfaces of light organ crypt epithelial cells [55]. In this study, C3 was found in the hemocyte proteome (Table A1.3). In general, the complement pathway remains poorly characterized in respect to invertebrates. Many of the traditional vertebrate homologs have been difficult to identify. Even though some proteins have been identified, there are inconsistent patterns of distribution across species [197,198]. It is useful to search for proteins that share sequence characteristics to the known vertebrate complement components. Two thioester proteins (TEPs) were identified in the hemocyte proteome (Table A1.3). TEPs are similar to complement-like proteins and alpha macroglobulins, which function as proteases and protease inhibitors respectively [58]. Among invertebrates, TEPs are interesting targets because they play an important role in the innate immune response

as members of complement-like pathways [57,58]. A recent analysis of the light organ proteome also identified several TEPs [115]. Another complement-like protein detected in the hemocytes is an alpha macroglobulin (Table A1.3). This group of proteins can function as protease inhibitors and have been shown to be an important component of the innate immune response of arthropods [199].

Among the other putative complement associated proteins identified was a matrix metalloproteinase (MMP) similar to vitronectin (Table A1.4). Vitronectin has been described as a serum glycoprotein involved in regulating innate immunity homeostasis [200]. In humans, vitronectin inhibits the membrane attack complex of the complement pathway. Bacteria, including *Haemophilus influenzae* and *Neisseria meningitidis* can interact with vitronectin and evade the complement response [201]. The putative MMP/vitronectin in this study contains a hemopexin-like domain (data not shown), but whether it serves a role similar to vitronectin or an MMP in the squid host remains to be determined. Another complement-like protein identified was a partial ORF containing six sushi repeat domains (Table A1.3). Sushi domains are commonly found among complement receptors (e.g., CR1, CR2, C4b binding protein, and factor H) as well as other proteins involved with innate immunity and self/non-self recognition [202,203]. This protein has yet to be characterized in the squid-vibrio association but offers an enticing future target for analysis. A putative peptide was also detected from the hemocytes that has similarity to the complement protein C1qBP, but this will have to be validated further (Table A1.3).

One of the most unique discoveries by both the transcriptome and proteome was a protein with high similarity to cephalotoxin (Table A1.4). In octopus and cuttlefish, cephalotoxins are normally expressed in the salivary glands and are involved with immobilizing prey [204]. This is

the first report of this protein in another cell type. A search of the UniProt database identified a thrombospondin domain similar to ones described in properdin, a proenzyme in the complement cascade [205,206]. Whether cephalotoxin functions as a toxin, a protease, or in another capacity in the hemocytes of *E. scolopes* remains to be determined. A second toxin was identified with similarity to the plancitoxins found in the crown-of-thorns starfish, *Acanthaster planci* (Table A1.4)[207]. These proteins are predicted to be deoxyribonucleases and future studies should focus on characterizing any potential DNAse activity in squid hemocytes.

In conclusion, MS-based proteomics of the hemocytes from *E. scolopes* enabled the initial characterization of the molecular components of this cell type. In addition, the complementary analysis of the hemocyte transcriptome provided a more comprehensive understanding of the squid's circulating macrophages. One advantage of combining both "omic" techniques is that one approach was able to detect genes/proteins not identified by the other. For example, TEPs and an alpha macroglobulin were found in the proteome, but not in the transcriptome (Table A1.2). On the other hand, the transcriptome identified several transcripts related to the NF-κB signaling pathway (Table A1.4). These differences between the two techniques may be explained by the limit of detection for proteins and transcripts of lower abundance. Another possibility is that some genes may not be transcribed as often or under certain physiological conditions. New data resulting from quantitative proteomics (Chapter 4) has provided greater coverage of the hemocyte proteome, allowing the detection of low abundant proteins including REL/NF-κB and a putative Toll-like receptor. As new technology involving mass spectrometry and data visualization emerges, the complex information provided by the hemocyte proteome will contribute to a better understanding of how these macrophages are involved in this squid-vibrio symbiosis.

Appendix II: Histone H4 of *Euprymna scolopes*: a putative antimicrobial protein

Background

Quantitative proteomics identified histone H4 as one of the more abundant proteins in hemocytes in response to colonization of the light organ by *V. fischeri* (Chapter 4). In addition, a transcript of histone H4 is up-regulated in the light organ prior to the expulsion of a majority of the symbiont population at dawn [10]. Histones are normally found within the nucleus where they play a critical role in chromatin organization and structure [208]. However, recent research suggests that histones, including histone H4, can serve as antimicrobial molecules [209-212]. One hypothesis may be that histone H4 serves as a host antimicrobial protein that is used to maintain the specificity of the light organ microenvironment. In this study a polyclonal antibody was used to confirm the presence of histone H4 in hemocytes by western blot and immunocytochemistry analyses. The possible antimicrobial properties of histone H4 were also investigated by testing the susceptibility of several marine microbes, including *V. fischeri*, to a recombinant human histone H4.

Methods

Histone H4

A rabbit polyclonal antibody to human histone H4 (ab10158) was purchased from Abcam. Recombinant human histone H4 (10264; 100 μg) was purchased from Cayman Chemical.

Western blot analysis

Hemocyte protein was prepared as previously described [15]. The protein concentration was determined by the *RC DC* protein assay (Bio-Rad). Total hemocyte protein (20 μg) and purified recombinant human histone H4 (2 μg) were separated using a 12% TGX Criterion gel (Bio-Rad) at 200V for 45 minutes. Proteins were transferred onto a .2 μM mini nitrocellulose membrane using a Trans-Blot Turbo transfer system and the Low Molecular Weight protocol as described by the manufacturer's instructions (Bio-Rad). Membranes were blocked in 4% milk, 0.1% Tween at room temperature for one hour. Anti-histone H4 was diluted in block 1:1,000 and incubated with the blot at room temperature for one hour. After washing the blot 4 times for 5 minutes each in 0.1% Tween, tris buffered saline (TTBS), pH 7.4, a goat anti-rabbit (IgG H+L) HRP conjugate (1:5,000, Bio-Rad) was diluted in fresh block and then incubated with the membrane for an additional one hour at room temperature. The blot was then washed again in TTBS 6 times for 5 minutes each. The antibody was detected using the Amersham Enhanced Chemiluminescence (ECL) Prime Detection kit (GE Healthcare Life Sciences) and imaged using a FluorChem HD2 CCD camera (Cell Biosciences) with an exposure time of 15 to 30 seconds.

Immunocytochemistry

Hemocytes were collected from adult squid as previously described [15,16]. Adult hemocytes were allowed to adhere to glass coverslips in 12-well tissue culture plates for 30 minutes. After attachment, the hemocytes were rinsed two times with Squid Ringer's (50 mM $MgCl_2$, 10 mM $CaCl_2$, 10 mM KCl, 530 mM NaCl, 10 mM HEPES) to remove excess tissue or cellular debris. The adherent cells were fixed with 4% paraformaldehyde in Squid Ringer's for 30 minutes at room temperature. After fixation the hemocytes were washed 4 times with marine

phosphate buffered saline, pH 7.4 (mPBS; 50 mM sodium phosphate, 0.45 M NaCl) for 5 minutes each. Fixed hemocytes were permeabolized with 1% Triton X-100, mPBS, pH 7.4 for 30 minutes at room temperature. The cells were blocked (mPBS, pH 7.4, 1% Triton X-100, 2% BSA) at room temperature for 1 hour. Hemocytes were then exposed to anti-histone H4 at a 1:250 dilution in block overnight. The primary antibody was removed and the hemocytes were washed in 1% Triton X-100, mPBS pH 7.4 four times for 5 minutes each and blocked for an additional hour. The hemocytes were then incubated with an Alexa Fluor 488 goat anti-rabbit secondary antibody (Life Technologies) at a 1:3,000 dilution in fresh block for 4 hours in the dark. The hemocytes were then washed in 1% Triton X-100, mPBS pH 7.4 four times for 5 minutes each and then counterstained using a lectin stain, ConA-TritC (0.2 μg/ml), and a nuclear stain, DRAQ5 (2.5 μM), in 1% Triton X-100, mPBS, pH 7.4 overnight in the dark. Hemocytes were mounted onto glass slides with Vectashield, sealed with nail polish and imaged using an A1R confocal microscope.

Antibacterial activity assay

To determine any possible antibacterial activity of recombinant human histone H4, a minimum inhibitory concentration (MIC) assay (performed by René Augustin). 96-well round bottom microtiter plates were coated with 200 μl of 0.1% BSA for 30 min. After the BSA solution was removed, the first wells were filled with 125 μg/mL of recombinant histone H4 in 90 μl of a 10 mM PIPES based buffer (50 mM $MgSO_4$, 10 mM $CaCl_2$, 300 mM NaCl, 10 mM KCl, 10 mM PIPES, pH 6.3). The concentration of recombinant histone H4 was then serially diluted 2 fold in the additional wells. Finally, each well was inoculated with 100 CFU of either *V. fischeri* ES114, *Vibrio parahaemolyticus*, *Photobacterium leiognathi*, *Exiguabacteria aestuori*

CNJ 771, *Bacillus megaterium* CNJ 778, or *Bacillus algicola* CNJ 803 for a final volume of 100 µL per well. The microtiter plates were incubated overnight at 28°C and the MICs were determined for each strain by the absence of a bacterial cell pellet. Three biological replicates were completed.

Results

The histone H4 protein sequence from *Euprymna scolopes* is highly conserved and 100% identical to the human histone H4. A polyclonal antibody designed to recognize human histone H4 cross-reacted with histone H4 from the squid hemocytes. A western blot analysis confirmed the presence of histone H4 in adult hemocytes from squid with colonized light organs (Fig. A2.1). A prominent protein band of approximately 14.7 kD was detected by anti-histone H4, which was similar to the corresponding band from the human recombinant histone H4 (Fig. A2.1). In addition, immunocytochemistry with anti-histone H4 revealed punctate staining within the nuclear region of adult hemocytes (Fig. A2.2).

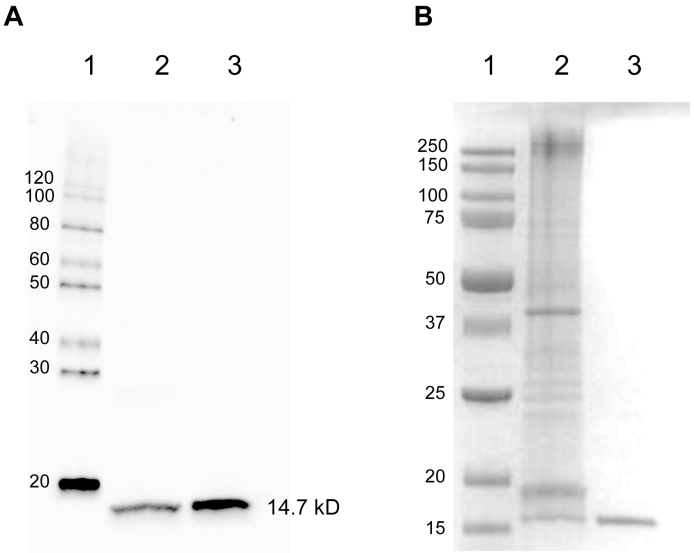


Figure A2. 1: Histone H4 western blot analysis.

A human polyclonal antibody that recognizes histone H4 cross-reacted with histone H4 from *Euprymna scolopes*. **(A)**. A western blot analysis with anti-histone H4 identified both histone H4 in the hemocytes (Lane 2) and the recombinant human histone H4 (Lane 3) at 14.7 kD (protein standard, kD -Lane 1) **(B)**. Total soluble hemocyte protein (Lane 2, 60 µg) and purified recombinant human histone H4 (Lane 3, 3 µg) were separated on a 12% TGX Criterion gel and stained with Coomassie R-250. A prominent protein band is present in the hemocyte protein sample in the region where recombinant histone H4 also appears.

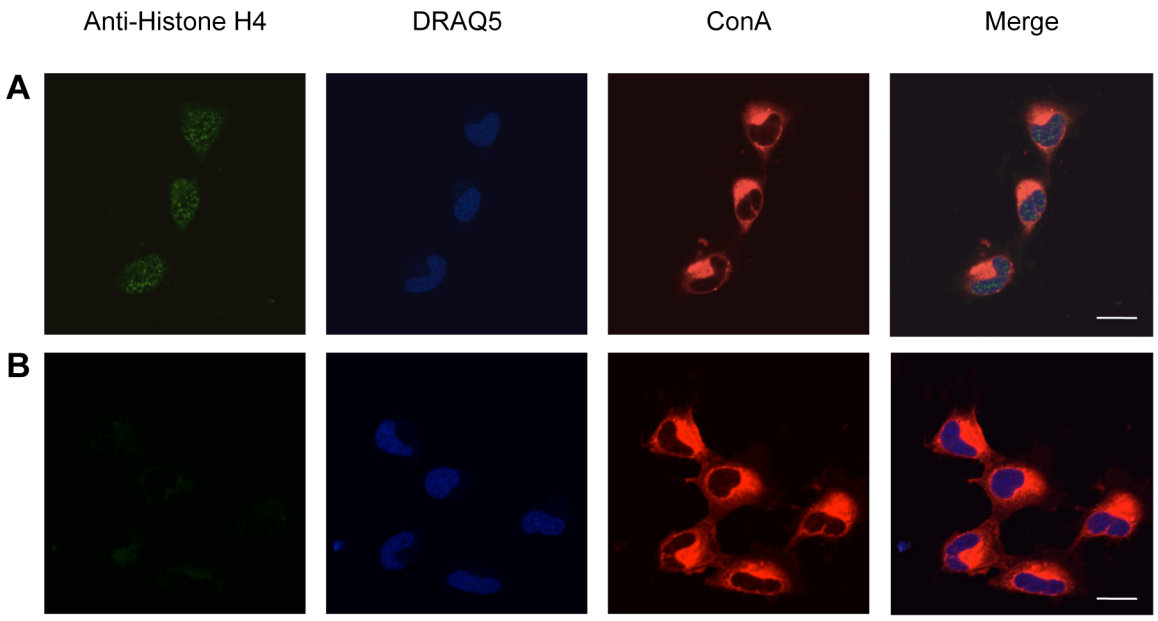


Figure A2. 2: The localization of histone H4 in hemocytes.

(A). Histone H4 can be found in the nucleus of adult hemocytes. **(B)**. Secondary antibody control. Hemocytes treated with anti-human histone H4 were stained with a goat anti-rabbit Alexa fluor 488 antibody. The hemocytes are counterstained with ConA-TritC (red, a lectin stain) and DRAQ5 (blue, a nuclear stain). Scale-10 µM

Analyses of hemocyte lysates revealed that histones were components of protein fractions demonstrating antimicrobial activity (R. Augustin). To determine the possible contribution of histone H4 to the antimicrobial activity of the hemocytes, a minimum inhibitory concentration

(MIC) assay was performed by René Augustin (Table A2.1). With this assay several marine Gram-negative and Gram-positive bacteria were tested for sensitivity to purified recombinant human histone H4. All 3 Gram-positive marine bacteria analyzed (*E. aestuori* CNJ 771, *B.megaterium* CNJ 778, *B. algicola* CNJ 778) were extremely sensitive to recombinant histone H4 (Table A2.1). Of the Gram-negative microbes tested, only *P. leiognathi* was affected by histone H4. *V. fischeri* and *V. parahaemolyticus* were not sensitive to histone H4 as demonstrated by this particular assay (Table A2.1).

Table A2. 1: The antimicrobial activity of recombinant human histone H4.

Bacterial stain	**MIC (µM)**	**MBC (µM)**
Vibrio fischeri ES114	>11	>11
Vibrio parahaemolyticus	>11	>11
Photobacterium leiognathi	2.8	5.5
Exiguabacteria aestuori CNJ 771	0.17-0.34	0.17-0.34
Bacillus megaterium CNJ 778	0.7-1.4	1.4
Bacillus algicola CNJ 778	0.7-1.4	0.7-1.4

Abbreviations: MIC- minimum inhibitory concentration, MBC- minimum bactericidal concentration
* Data provided by R. Augustin, University of Wisconsin-Madison

Discussion

The mechanisms by which hemocytes from *E. scolopes* maintain the high level of specificity found in the light organ are uncharacterized. Quantitative proteomics identified histone H4 as more abundant in hemocytes in response to colonization, but the basis for this difference is not known (Chapter 4). It is possible that this protein could function as an antimicrobial molecule important for preventing unwanted microorganisms from colonizing the light organ. In this study an antibody designed to recognize human histone H4 confirmed the

presence of histone H4 as a component of the hemocytes from *E. scolopes* (Fig. A2.1). As expected, immunocytochemistry demonstrated that histone H4 localizes to the nucleus of this cell type (Fig. A2.2). It is unclear how this protein may translocate out of the nucleus or be released from the hemocytes to take part in an antimicrobial response. There are a couple of possibilities for finding histones outside of the nucleus. In *Drosophila*, antimicrobial histones can be attached and released from lipid droplets [213]. In addition, human neutrophils can release antimicrobial neutrophil extracellular traps (NETs), which contain nuclear material and histones [214]. This study also revealed that marine Gram-positive bacteria are very sensitive to the antimicrobial activity of recombinant human histone H4 (Table A2.1). *V. fischeri* was not sensitive to histone H4 (Table A2.1). The differences between the cell envelopes of Gram-positive and Gram-negative bacteria may be behind the resistance to histone H4. The current mechanism of action for histones involves their ability to translocate bacterial membranes and interact with nucleic acids [215-217]. Potentially, *E. scolopes* uses the highly conserved histone H4 alone or in combination with other histones as potent antimicrobial molecules important for selecting *V. fischeri* from the environment and maintaining the specificity of the light organ [6]. This study establishes histone H4 as putative antimicrobial effector used in a beneficial host-microbe relationship.

Appendix III: Detection of peptidoglycan recognition protein 5 from *Euprymna scolopes* in the hemolymph

Background

One of the most abundant proteins detected in the hemocytes of *E. scolopes* is an uncharacterized peptidoglycan recognition protein (PGRP), EsPGRP 5 [15]. There have been 4 previously described PGRPs from *Euprymna scolopes* and only two, EsPGRP 1 and EsPGRP 2, have been studied in depth [49,50,178]. EsPGRP 5 was first described from hemocyte transcriptomic and proteomic analyses [15]. This 253 amino acid protein contains the conserved residues for amidase activity, which highlight the potential ability of this PGRP to bind and degrade peptidoglycan. Another feature of EsPGRP 5 is a predicted signal peptide suggesting that this PGRP may be secreted from the hemocytes [15]. In this study cell free hemolymph was surveyed for the presence of EsPGRP 5 using a polyclonal antibody and a western blot analysis. Detection of EsPGRP 5 in the hemolymph would provide evidence that this protein can be secreted from hemocytes.

Methods

Hemolymph protein preparation

Hemocyte proteins were prepared as previously described [15]. For hemolymph proteins, hemocytes were first removed from hemolymph by centrifugation (10 minutes, 2000 rpm, 4°C). Cell-free hemolymph was then transferred into a fresh 1.5 ml tube and frozen at -80°C until analysis. Hemocyanin is a major component of hemolymph from *Euprymna scolopes* and

increases the difficulty in finding plasma proteins of lower abundance. To address this, hemolymph was centrifuged at maximum speed (21,000 x g) for 4 hours at 4°C. This long centrifugation pelleted some of the hemocyanin, leaving a clear layer, which was collected for protein analysis. To minimize protein degradation a 1× protease inhibitor cocktail (P2714, Sigma Aldrich) made in Squid Ringer's buffer was added to the hemolymph prior to centrifugation. Protein concentration was determined by the *RC DC* protein assay (Bio-Rad).

Western blot analysis

Total hemocyte protein (20 μg) and cell-free hemolymph (150 μg) were separated using a 12% TGX Criterion gel (Bio-Rad) at 200V for 45 minutes. Proteins were transferred onto a .2 μM mini nitrocellulose membrane using a Trans-Blot Turbo transfer system and the Mixed Molecular Weight protocol as described by the manufacturer's instructions (Bio-Rad). Membranes were blocked in 4% milk, 0.1% Tween at room temperature for one hour. Anti-EsPGRP 5 was diluted in block 1:1,000 and incubated at room temperature for one hour. After washing the blot 4 times for 5 minutes each in 0.1% Tween, tris buffered saline (TTBS), pH 7.4, a goat anti-rabbit (IgG H+L) HRP conjugate was diluted in fresh block (1:5,000, Bio-Rad) and incubated with the blot for an additional hour at room temperature. The blot was then washed again 6 times in TTBS for 5 minutes each. The antibody was detected using the Amersham Enhanced Chemiluminescence (ECL) Prime Detection kit (GE Healthcare Life Sciences) and imaged using a FluorChem HD2 CCD camera (Cell Biosciences) with an exposure time of 30 seconds.

Results

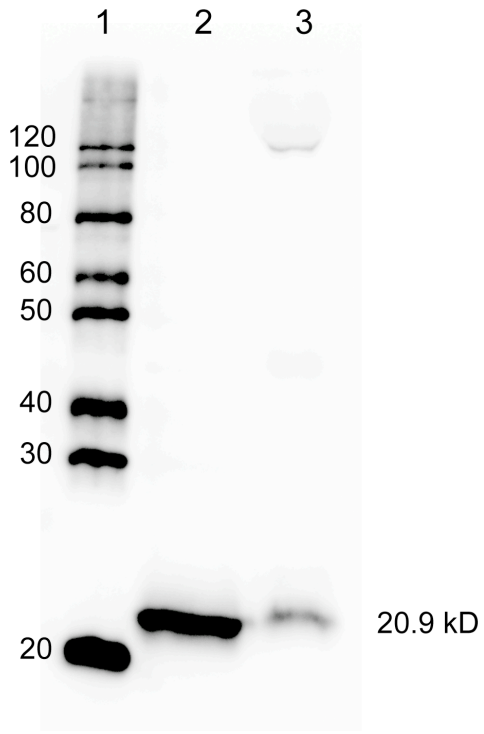


Figure A3. 1: EsPGRP 5 is present in the hemolymph.

Representative anti-EsPGRP 5 western blot of hemocytes (Lane 2) and hemolymph (lane 3). EsPGRP 5 is indicated by a 20.9 kD band (Lane 1- protein standard, kD).

To determine if EsPGRP 5 is secreted, the hemolymph from individual squid (n=3) were separately analyzed using a polyclonal antibody and a western blot analysis (Fig. A3.1). To control for the possibility that residual hemocytes may be responsible for the presence of EsPGRP 5 in the hemolymph, EsGalectin 1 was used as a control. EsGalectin 1 is known to be an abundant protein in hemocytes from *E. scolopes* and research involving similar galectins in other molluscs demonstrated that they are not secreted into the hemolymph [168,192]. EsPGRP 5 was detected in the hemolymph (Fig. A3.1), while EsGalectin 1 was not identified (not shown).

Discussion

This study represents the first attempt to determine if EsPGRP 5 is secreted. Analysis of cell-free hemolymph from three individual animals revealed that EsPGRP 5 can be found in the plasma of *Euprymna scolopes* (Fig. A3.1). The presence of EsPGRP 5 in the hemolymph and the absence of EsGalectin 1 suggested that EsPGRP 5 is secreted from hemocytes. Future work should validate these findings, determine how EsPGRP 5 is secreted and under what physiological conditions secretion occurs. One of the other PGRPs from *E. scolopes*, which has already been characterized, is secreted too. EsPGRP 2 is secreted into the light organ crypt spaces in response to colonization by *V. fischeri* [50]. In the light organ EsPGRP 2 utilizes amidase activity to degrade tracheal cytotoxin (TCT), a toxic byproduct of the peptidoglycan released by *V. fischeri*. The role of EsPGRP 5 in the squid-vibrio symbiosis and the function of this protein remain unclear at this point. This abundant hemocyte protein is a promising target for future characterization of host-symbiont specificity in this association.

Summary

Chapter 2

- This study was the first to use high throughput mass spectrometry-based proteomics to identify proteins from the symbiotic light organ of the squid, *Euprymna scolopes*. The proteomic data generated from this study gave us a better understanding of which proteins are present in the light organ at an important transitional period during the diel rhythm.

- The combination of spectral data from light organ central core and exudate samples resulted in the identification of 1,581 proteins including 870 proteins from *V. fischeri* and 711 from *E. scolopes*.

- Proteins identified were categorized by function using the Cluster of Orthologus Groups database for prokaryotes and eukaryotes (COG/KOG). Host and symbiont proteins were identified that emphasize how each organism responds to the oxidative environment found in the light organ. Symbiont proteins also highlighted proteins regulated by quorum sensing. In addition, host proteins related to innate immunity were detected. Many of the proteins identified represent new targets to characterize in respect to beneficial host-microbe interactions.

Chapter 3

- The *Vibrio fischeri* immunoglobulin-like surface protein (Vig), a quorum sensing-regulated protein identified from the proteomic analyses of the light organ, was selected for characterization.

- A *V. fischeri vig*$^-$ mutant was significantly outcompeted in the light organ in the presence of wild type *V. fischeri* at 24 and 48 hours post inoculation. In addition, *vig*$^-$ demonstrated reduced light production in culture and in the light organ when compared to wild type.

- Alone *vig*$^-$ can colonize the light organ efficiently, but in competition the host may select against the dimmer *vig*$^-$ strain. This study provides further evidence that wild type luminescence does not rescue dim or dark strains of *V. fischeri*. In addition, this study provides the first characterization of a bacterial immunoglobulin-like protein with a role in a beneficial host-microbe relationship.

Chapter 4

- This study was the first to apply two different quantitative proteomics techniques to characterize the influence of colonization on the hemocytes of *E. scolopes*.

- Isobaric tags for relative and absolute quantification (iTRAQ) quantified 1,024 proteins of which 37 were determined to be significantly different between sym and cured hemocytes. The label-free approach resulted in 1,241 proteins identified with 185

proteins that were present at significantly different levels. The combination of these two techniques provided a more comprehensive analysis and further validated the functional groups of proteins identified between the two methods.

- This study provided more evidence that colonization alters the response of the squid hemocytes in favor of tolerance of the symbiont *V. fischeri*. Quantitative proteomics did not identify an overwhelming absence or presence of one specific protein, but rather a complex adjustment of proteins may be involved in symbiont recognition and accommodation. The proteins altered between sym and cured hemocytes are associated with the cytoskeleton, cell adhesion, innate immunity, and lysosomes.

Chapter 5

- *E. scolopes* has two galectins as identified from the light organ and hemocytes by MS-based proteomics. Galectins, due to their inherent carbohydrate binding abilities, represent possible pattern recognition receptors, which might be involved in host-symbiont recognition.

- EsGalectin 1 has 4 carbohydrate recognition domains (CRDs) and EsGalectin 2 has 2 CRDs. Transcripts for both galectins were ubiquitously distributed throughout adult squid tissues. One noticeable difference from the comparison of both transcripts by end point RT-PCR was that EsGalectin 1 appeared to be more abundant in the accessory nidamental gland (ANG). Also, iTRAQ previously identified that the EsGalectin 2 protein was more abundant in sym hemocytes.

- Antibodies revealed the localization of both galectins in adult and juvenile hemocytes. No differences were discovered based on the maturity or the colonization state of the squid. EsGalectin 1 can be found in the filopodia, nucleus, and cytoplasm. EsGalectin 2 was detected in the nucleus and in the cytoplasm. On rare occasions EsGalectin 2 was perinuclear in cells with membrane vesicles.

- The exact role of these galectins in the squid-vibrio symbiosis remains unknown. Future work should involve characterizing the ability of each galectin to bind different bacteria.

Appendix I

- The composition of the hemocyte proteome of *E. scolopes* was characterized with MS-based proteomics. This represented the first time a hemocyte proteome from a squid was analyzed.

- Proteins identified from the hemocytes included complement-like proteins, pattern recognition receptors, proteases, cytoskeletal proteins, and two putative toxins.

- These proteomic analyses established the foundation for understanding what proteins might help regulate the symbiosis with *V. fischeri* from the perspective of the hemocytes.

Appendix II

- Histone H4 was determined by iTRAQ to be more abundant in hemocytes in response to colonization by *V. fischeri*. The amino acid sequence of histone H4 from *E. scolopes* is also highly conserved.

- A polyclonal antibody to human histone H4 cross-reacted with histone H4 from squid hemocytes by western blot and immunocytochemistry. Histone H4 appears to be nuclear in the hemocytes. A mechanism for the release of histone H4 from the nucleus remains to be determined.

- Recombinant human histone H4 demonstrated strong antibacterial activity to marine Gram-positive bacteria. *V. fischeri* is not sensitive to recombinant histone H4.

- Histone H4 may represent a novel antimicrobial mechanism used by the squid to maintain the specificity found in the light organ.

Appendix III

- An abundant peptidoglycan recognition protein of the hemocytes from *E. scolopes*, EsPGRP 5, has a predicted signal sequence. In this study the possible secretion of this PGRP was characterized.

- An antibody specific to EsPGRP 5 detected EsPGRP 5 in the hemolymph of adult squid. The lack of another prominent hemocyte protein, EsGalectin 1, suggested that EsPGRP 5

is secreted from the hemocytes into the hemolymph. The mechanism and role of this secretion remains uncharacterized.

References

1. Nyholm SV, Graf J (2012) Knowing your friends: invertebrate innate immunity fosters beneficial bacterial symbioses. Nat Rev Microbiol 10: 815-827.

2. McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Loso T, et al. (2013) Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A 110: 3229-3236.

3. Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI (2005) Host-bacterial mutualism in the human intestine. Science 307: 1915-1920.

4. McFall-Ngai M (2014) Divining the essence of symbiosis: insights from the squid-vibrio model. PLoS Biol 12: e1001783.

5. Nyholm SV, Stewart JJ, Ruby EG, McFall-Ngai MJ (2009) Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. Environ Microbiol 11: 483-493.

6. Nyholm SV, McFall-Ngai MJ (2004) The winnowing: establishing the squid-vibrio symbiosis. Nat Rev Microbiol 2: 632-642.

7. Koropatnick TA, Engle JT, Apicella MA, Stabb EV, Goldman WE, et al. (2004) Microbial factor-mediated development in a host-bacterial mutualism. Science 306: 1186-1188.

8. Chun CK, Scheetz TE, Bonaldo Mde F, Brown B, Clemens A, et al. (2006) An annotated cDNA library of juvenile *Euprymna scolopes* with and without colonization by the symbiont *Vibrio fischeri*. BMC Genomics 7: 154.

9. Ruby EG, Urbanowski M, Campbell J, Dunn A, Faini M, et al. (2005) Complete genome sequence of *Vibrio fischeri*: a symbiotic bacterium with pathogenic congeners. Proc Natl Acad Sci U S A 102: 3004-3009.

10. Wier AM, Nyholm SV, Mandel MJ, Massengo-Tiasse RP, Schaefer AL, et al. (2010) Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. Proc Natl Acad Sci U S A 107: 2259-2264.

11. Jones BW, Nishiguchi MK (2004) Counterillumination in the hawaiian bobtail squid, *Euprymna scolopes* Berry (Mollusca : Cephalopoda). Marine Biology 144: 1151-1155.

12. Boettcher KJ, Ruby EG, McFallNgai MJ (1996) Bioluminescence in the symbiotic squid *Euprymna scolopes* is controlled by a daily biological rhythm. Journal of Comparative Physiology a-Sensory Neural and Behavioral Physiology 179: 65-73.

13. Nyholm SV, McFall-Ngai MJ (1998) Sampling the light-organ microenvironment of *Euprymna scolopes*: description of a population of host cells in association with the bacterial symbiont *Vibrio fischeri*. Biol Bull 195: 89-97.

14. Koropatnick TA, Kimbell JR, McFall-Ngai MJ (2007) Responses of host hemocytes during the initiation of the squid-Vibrio symbiosis. Biol Bull 212: 29-39.

15. Collins AJ, Schleicher TR, Rader BA, Nyholm SV (2012) Understanding the role of host hemocytes in a squid/vibrio symbiosis using transcriptomics and proteomics. Front Immunol 3: 91 doi: 10.3389/fimmu.2012.00091

16. Collins AJ, Nyholm SV (2010) Obtaining hemocytes from the Hawaiian bobtail squid *Euprymna scolopes* and observing their adherence to symbiotic and non-symbiotic bacteria. J Vis Exp. 36: doi:10.3791/1714

17. Rader BA, Nyholm SV (2012) Host/microbe interactions revealed through "omics" in the symbiosis between the Hawaiian bobtail squid *Euprymna scolopes* and the bioluminescent bacterium *Vibrio fischeri*. Biol Bull 223: 103-111.

18. Meyer E, Weis VM (2012) Study of cnidarian-algal symbiosis in the "omics" age. Biol Bull 223: 44-65.

19. Chaston J, Douglas AE (2012) Making the most of "omics" for symbiosis research. Biol Bull 223: 21-29.

20. Chun CK, Troll JV, Koroleva I, Brown B, Manzella L, et al. (2008) Effects of colonization, luminescence, and autoinducer on host transcription during development of the squid-vibrio association. Proc Natl Acad Sci U S A 105: 11323-11328.

21. Kremer N, Philipp EE, Carpentier MC, Brennan CA, Kraemer L, et al. (2013) Initial symbiont contact orchestrates host-organ-wide transcriptional changes that prime tissue colonization. Cell Host Microbe 14: 183-194.

22. Bomar L, Graf J (2012) Investigation into the physiologies of *Aeromonas veronii in vitro* and inside the digestive tract of the medicinal leech using RNA-seq. Biol Bull 223: 155-166.

23. Yates JR, Ruse CI, Nakorchevsky A (2009) Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng 11: 49-79.

24. Hegde PS, White IR, Debouck C (2003) Interplay of transcriptomics and proteomics. Curr Opin Biotechnol 14: 647-651.

25. Doino Lemus J, McFall-Ngai MJ (2000) Alterations in the proteome of the *Euprymna scolopes* light organ in response to symbiotic *Vibrio fischeri*. Appl Environ Microbiol 66: 4091-4097.

26. McFall-Ngai M (2008) Host-microbe symbiosis: the squid-Vibrio association--a naturally occurring, experimental model of animal/bacterial partnerships. Adv Exp Med Biol 635: 102-112.

27. Visick KL, Ruby EG (2006) *Vibrio fischeri* and its host: it takes two to tango. Curr Opin Microbiol 9: 632-638.

28. Mandel MJ (2010) Models and approaches to dissect host-symbiont specificity. Trends Microbiol 18: 504-511.

29. Lee KH, Ruby EG (1994) Effect of the Squid Host on the Abundance and Distribution of Symbiotic *Vibrio fischeri* in Nature. Appl Environ Microbiol 60: 1565-1571.

30. Graf J, Ruby EG (1998) Host-derived amino acids support the proliferation of symbiotic bacteria. Proc Natl Acad Sci U S A 95: 1818-1822.

31. Washburn MP, Wolters D, Yates JR, 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19: 242-247.

32. Boettcher KJ, Ruby EG (1990) Depressed light emission by symbiotic *Vibrio fischeri* of the sepiolid squid *Euprymna scolopes*. J Bacteriol 172: 3701-3706.

33. Ho YP, Hsu PH (2002) Investigating the effects of protein patterns on microorganism identification by high-performance liquid chromatography-mass spectrometry and protein database searches. J Chromatogr A 976: 103-111.

34. Whitaker JR, Granum PE (1980) An absolute method for protein determination based on difference in absorbance at 235 and 280 nm. Anal Biochem 109: 156-159.

35. Heukeshoven J, Dernick R (1985) Simplified Method for Silver Staining of Proteins in Polyacrylamide Gels and the Mechanism of Silver Staining. Electrophoresis 6: 103-112.

36. Hirosawa M, Hoshida M, Ishikawa M, Toya T (1993) MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming. Comput Appl Biosci 9: 161-167.

37. Pappin DJ, Hojrup P, Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol 3: 327-332.

38. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20: 3551-3567.

39. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, et al. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics 4: 1265-1272.

40. Shinoda K, Tomita M, Ishihama Y (2010) emPAI Calc--for the estimation of protein abundance from large-scale identification data by liquid chromatography-tandem mass spectrometry. Bioinformatics 26: 576-577.

41. Jones BM, Edwards RJ, Skipp PJ, O'Connor CD, Iglesias-Rodriguez MD (2011) Shotgun Proteomic Analysis of *Emiliania huxleyi*, a Marine Phytoplankton Species of Major Biogeochemical Importance. Marine Biotechnology 13: 496-504.

42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.

43. Callahan SM, Dunlap PV (2000) LuxR- and acyl-homoserine-lactone-controlled non-lux genes define a quorum-sensing regulon in *Vibrio fischeri*. J Bacteriol 182: 2811-2822.

44. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, et al. (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4: 41.

45. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28: 33-36.

46. McFall-Ngai M, Nyholm SV, Castillo MG (2010) The role of the immune system in the initiation and persistence of the *Euprymna scolopes*--*Vibrio fischeri* symbiosis. Semin Immunol 22: 48-53.

47. Foster JS, Apicella MA, McFall-Ngai MJ (2000) *Vibrio fischeri* lipopolysaccharide induces developmental apoptosis, but not complete morphogenesis, of the *Euprymna scolopes* symbiotic light organ. Dev Biol 226: 242-254.

48. Nyholm SV, Deplancke B, Gaskins HR, Apicella MA, McFall-Ngai MJ (2002) Roles of *Vibrio fischeri* and nonsymbiotic bacteria in the dynamics of mucus secretion during symbiont colonization of the *Euprymna scolopes* light organ. Appl Environ Microbiol 68: 5113-5122.

49. Goodson MS, Kojadinovic M, Troll JV, Scheetz TE, Casavant TL, et al. (2005) Identifying components of the NF-kappaB pathway in the beneficial *Euprymna scolopes*-*Vibrio fischeri* light organ symbiosis. Appl Environ Microbiol 71: 6934-6946.

50. Troll JV, Bent EH, Pacquette N, Wier AM, Goldman WE, et al. (2010) Taming the symbiont for coexistence: a host PGRP neutralizes a bacterial symbiont toxin. Environ Microbiol 12: 2190-2203.

51. Cerliani JP, Stowell SR, Mascanfroni ID, Arthur CM, Cummings RD, et al. (2011) Expanding the universe of cytokines and pattern recognition receptors: galectins and glycans in innate immunity. J Clin Immunol 31: 10-21.

52. Aeckersberg F, Lupp C, Feliciano B, Ruby EG (2001) *Vibrio fischeri* outer membrane protein OmpU plays a role in normal symbiotic colonization. J Bacteriol 183: 6590-6597.

53. Goldfarb DS, Corbett AH, Mason DA, Harreman MT, Adam SA (2004) Importin alpha: a multipurpose nuclear-transport receptor. Trends Cell Biol 14: 505-514.

54. Fagerlund R, Kinnunen L, Kohler M, Julkunen I, Melen K (2005) NF-{kappa}B is transported into the nucleus by importin {alpha}3 and importin {alpha}4. J Biol Chem 280: 15942-15951.

55. Castillo MG, Goodson MS, McFall-Ngai M (2009) Identification and molecular characterization of a complement C3 molecule in a lophotrochozoan, the Hawaiian bobtail squid *Euprymna scolopes*. Dev Comp Immunol 33: 69-76.

56. Rus H, Cudrici C, Niculescu F (2005) The role of the complement system in innate immunity. Immunol Res 33: 103-112.

57. Blandin S, Levashina EA (2004) Thioester-containing proteins and insect immunity. Mol Immunol 40: 903-908.

58. Bou Aoun R, Hetru C, Troxler L, Doucet D, Ferrandon D, et al. (2011) Analysis of thioester-containing proteins during the innate immune response of *Drosophila melanogaster*. J Innate Immun 3: 52-64.

59. Ruby EG, McFall-Ngai MJ (1999) Oxygen-utilizing reactions and symbiotic colonization of the squid light organ by *Vibrio fischeri*. Trends Microbiol 7: 414-420.

60. Small AL, McFall-Ngai MJ (1999) Halide peroxidase in tissues that interact with bacteria in the host squid *Euprymna scolopes*. J Cell Biochem 72: 445-457.

61. Tomarev SI, Zinovieva RD, Weis VM, Chepelinsky AB, Piatigorsky J, et al. (1993) Abundant mRNAs in the squid light organ encode proteins with a high similarity to mammalian peroxidases. Gene 132: 219-226.

62. Abbas K, Breton J, Drapier JC (2008) The interplay between nitric oxide and peroxiredoxins. Immunobiology 213: 815-822.

63. Robinson MW, Hutchinson AT, Dalton JP, Donnelly S (2010) Peroxiredoxin: a central player in immune modulation. Parasite Immunol 32: 305-313.

64. Visick KL, Ruby EG (1998) The periplasmic, group III catalase of *Vibrio fischeri* is required for normal symbiotic competence and is induced both by oxidative stress and by approach to stationary phase. J Bacteriol 180: 2087-2092.

65. Baek WK, Lee HS, Oh MH, Koh MJ, Kim KS, et al. (2009) Identification of the *Vibrio vulnificus* ahpCl gene and its influence on survival under oxidative stress and virulence. J Microbiol 47: 624-632.

66. Parsonage D, Desrosiers DC, Hazlett KR, Sun Y, Nelson KJ, et al. (2010) Broad specificity AhpC-like peroxiredoxin and its thioredoxin reductant in the sparse antioxidant defense system of *Treponema pallidum*. Proc Natl Acad Sci U S A 107: 6240-6245.

67. Barth KR, Isabella VM, Wright LF, Clark VL (2009) Resistance to peroxynitrite in *Neisseria gonorrhoeae*. Microbiology 155: 2532-2545.

68. Wang Y, Dunn AK, Wilneff J, McFall-Ngai MJ, Spiro S, et al. (2010) *Vibrio fischeri* flavohaemoglobin protects against nitric oxide during initiation of the squid-Vibrio symbiosis. Mol Microbiol 78: 903-915.

69. Wang Y, Ruby EG (2011) The roles of NO in microbial symbioses. Cell Microbiol 13: 518-526.

70. Fang FC (2004) Antimicrobial reactive oxygen and nitrogen species: concepts and controversies. Nat Rev Microbiol 2: 820-832.

71. Davidson SK, Koropatnick TA, Kossmehl R, Sycuro L, McFall-Ngai MJ (2004) NO means 'yes' in the squid-vibrio symbiosis: nitric oxide (NO) during the initial stages of a beneficial association. Cell Microbiol 6: 1139-1151.

72. Feng X, Guo Z, Nourbakhsh M, Hauser H, Ganster R, et al. (2002) Identification of a negative response element in the human inducible nitric-oxide synthase (hiNOS) promoter: The role of NF-kappa B-repressing factor (NRF) in basal repression of the hiNOS gene. Proc Natl Acad Sci U S A 99: 14212-14217.

73. Froese N, Schwarzer M, Niedick I, Frischmann U, Koster M, et al. (2006) Innate immune responses in NF-kappaB-repressing factor-deficient mice. Mol Cell Biol 26: 293-302.

74. Graf J, Ruby EG (2000) Novel effects of a transposon insertion in the *Vibrio fischeri* glnD gene: defects in iron uptake and symbiotic persistence in addition to nitrogen utilization. Mol Microbiol 37: 168-179.

75. Guerrero-Ferreira RC, Nishiguchi MK (2010) Differential gene expression in bacterial symbionts from loliginid squids demonstrates variation between mutualistic and environmental niches. Environ Microbiol Rep 2: 514-523.

76. Faraldo-Gomez JD, Sansom MS (2003) Acquisition of siderophores in gram-negative bacteria. Nat Rev Mol Cell Biol 4: 105-116.

77. Ratledge C, Dover LG (2000) Iron metabolism in pathogenic bacteria. Annu Rev Microbiol 54: 881-941.

78. Engebrecht J, Silverman M (1984) Identification of genes and gene products necessary for bacterial bioluminescence. Proc Natl Acad Sci U S A 81: 4154-4158.

79. Nealson KH (1977) Autoinduction of bacterial luciferase. Occurrence, mechanism and significance. Arch Microbiol 112: 73-79.

80. Antunes LC, Schaefer AL, Ferreira RB, Qin N, Stevens AM, et al. (2007) Transcriptome analysis of the *Vibrio fischeri* LuxR-LuxI regulon. J Bacteriol 189: 8387-8391.

81. Lin YP, McDonough SP, Sharma Y, Chang YF (2010) The terminal immunoglobulin-like repeats of LigA and LigB of *Leptospira* enhance their binding to gelatin binding domain of fibronectin and host cells. PLoS One 5: e11301.

82. Lupp C, Ruby EG (2004) *Vibrio fischeri* LuxS and AinS: comparative study of two signal synthases. J Bacteriol 186: 3873-3881.

83. Lupp C, Urbanowski M, Greenberg EP, Ruby EG (2003) The *Vibrio fischeri* quorum-sensing systems ain and lux sequentially induce luminescence gene expression and are important for persistence in the squid host. Mol Microbiol 50: 319-331.

84. Tian Y, Wang Q, Liu Q, Ma Y, Cao X, et al. (2008) Involvement of LuxS in the regulation of motility and flagella biogenesis in *Vibrio alginolyticus*. Biosci Biotechnol Biochem 72: 1063-1071.

85. Bose JL, Kim U, Bartkowski W, Gunsalus RP, Overley AM, et al. (2007) Bioluminescence in *Vibrio fischeri* is controlled by the redox-responsive regulator ArcA. Mol Microbiol 65: 538-553.

86. Hussa EA, O'Shea TM, Darnell CL, Ruby EG, Visick KL (2007) Two-component response regulators of *Vibrio fischeri*: identification, mutagenesis, and characterization. J Bacteriol 189: 5825-5838.

87. Whistler CA, Koropatnick TA, Pollack A, McFall-Ngai MJ, Ruby EG (2007) The GacA global regulator of *Vibrio fischeri* is required for normal host tissue responses that limit subsequent bacterial colonization. Cell Microbiol 9: 766-778.

88. Whistler CA, Ruby EG (2003) GacA regulates symbiotic colonization traits of *Vibrio fischeri* and facilitates a beneficial association with an animal host. J Bacteriol 185: 7202-7212.

89. Nguyen DT, Ngo TC, Tran HH, Nguyen TP, Nguyen BM, et al. (2009) Two different mechanisms of ampicillin resistance operating in strains of *Vibrio cholerae* O1 independent of resistance genes. FEMS Microbiol Lett 298: 37-43.

90. Raivio TL, Popkin DL, Silhavy TJ (1999) The Cpx envelope stress response is controlled by amplification and feedback inhibition. J Bacteriol 181: 5263-5272.

91. Balhesteros H, Mazzon RR, da Silva CA, Lang EA, Marques MV (2010) CspC and CspD are essential for *Caulobacter crescentus* stationary phase survival. Arch Microbiol 192: 747-758.

92. Felix G, Boller T (2003) Molecular sensing of bacteria in plants. The highly conserved RNA-binding motif RNP-1 of bacterial cold shock proteins is recognized as an elicitor signal in tobacco. J Biol Chem 278: 6201-6208.

93. Yamanaka K, Zheng W, Crooke E, Wang YH, Inouye M (2001) CspD, a novel DNA replication inhibitor induced during the stationary phase in *Escherichia coli*. Mol Microbiol 39: 1572-1584.

94. Ruby EG, Asato LM (1993) Growth and flagellation of *Vibrio fischeri* during initiation of the sepiolid squid light organ symbiosis. Arch Microbiol 159: 160-167.

95. Millikan DS, Ruby EG (2003) FlrA, a sigma54-dependent transcriptional activator in *Vibrio fischeri*, is required for motility and symbiotic light-organ colonization. J Bacteriol 185: 3547-3557.

96. Millikan DS, Ruby EG (2004) *Vibrio fischeri* flagellin A is essential for normal motility and for symbiotic competence during initial squid light organ colonization. J Bacteriol 186: 4315-4325.

97. Jones BW, Nishiguchi MK (2006) Differentially expressed genes reveal adaptations between free-living and symbiotic niches of *Vibrio fischeri* in a fully established mutualism. Can J Microbiol 52: 1218-1227.

98. Dunn AK, Stabb EV (2008) Genetic analysis of trimethylamine N-oxide reductases in the light organ symbiont *Vibrio fischeri* ES114. J Bacteriol 190: 5814-5823.

99. Studer SV, Mandel MJ, Ruby EG (2008) AinS quorum sensing regulates the *Vibrio fischeri* acetate switch. J Bacteriol 190: 5915-5923.

100. Delmotte N, Ahrens CH, Knief C, Qeli E, Koch M, et al. (2010) An integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium japonicum* bacteroid metabolism in soybean root nodules. Proteomics 10: 1391-1400.

101. Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, et al. (2007) Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. Science 315: 247-250.

102. Mastronunzio JE, Benson DR (2010) Wild nodules can be broken: proteomics of *Frankia* in field-collected root nodules. Symbiosis 50: 13-26.

103. Poliakov A, Russell CW, Ponnala L, Hoops HJ, Sun Q, et al. (2011) Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. Mol Cell Proteomics 10: M110 007039.

104. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, et al. (2009) Shotgun metaproteomics of the human distal gut microbiota. ISME J 3: 179-189.

105. Bright M, Bulgheresi S (2010) A complex journey: transmission of microbial symbionts. Nat Rev Microbiol 8: 218-230.

106. Waters CM, Bassler BL (2005) Quorum sensing: cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol 21: 319-346.

107. Bodelon G, Palomino C, Fernandez LA (2013) Immunoglobulin domains in *Escherichia coli* and other enterobacteria: from pathogenesis to applications in antibody technologies. Fems Microbiology Reviews 37: 204-250.

108. Castiblanco-Valencia MM, Fraga TR, Silva LB, Monaris D, Abreu PA, et al. (2012) Leptospiral immunoglobulin-like proteins interact with human complement regulators factor H, FHL-1, FHR-1, and C4BP. J Infect Dis 205: 995-1004.

109. Choy HA, Kelley MM, Chen TL, Moller AK, Matsunaga J, et al. (2007) Physiological osmotic induction of *Leptospira interrogans* adhesion: LigA and LigB bind extracellular matrix proteins and fibrinogen. Infect Immun 75: 2441-2450.

110. Lin YP, Raman R, Sharma Y, Chang YF (2008) Calcium binds to leptospiral immunoglobulin-like protein, LigB, and modulates fibronectin binding. J Biol Chem 283: 25140-25149.

111. Matsunaga J, Barocchi MA, Croda J, Young TA, Sanchez Y, et al. (2003) Pathogenic *Leptospira* species express surface-exposed proteins belonging to the bacterial immunoglobulin superfamily. Mol Microbiol 49: 929-945.

112. Le Roux F, Binesse J, Saulnier D, Mazel D (2007) Construction of a *Vibrio splendidus* mutant lacking the metalloprotease gene vsm by use of a novel counterselectable suicide vector. Appl Environ Microbiol 73: 777-784.

113. Dunn AK, Millikan DS, Adin DM, Bose JL, Stabb EV (2006) New rfp- and pES213-derived tools for analyzing symbiotic *Vibrio fischeri* reveal patterns of infection and lux expression *in situ*. Appl Environ Microbiol 72: 802-810.

114. Quan S, Hiniker A, Collet JF, Bardwell JC (2013) Isolation of bacteria envelope proteins. Methods Mol Biol 966: 359-366.

115. Schleicher TR, Nyholm SV (2011) Characterizing the host and symbiont proteomes in the association between the Bobtail squid, *Euprymna scolopes*, and the bacterium, *Vibrio fischeri*. PLoS One 6: e25649.

116. Collins AJ, LaBarre BA, Won BS, Shah MV, Heng S, et al. (2012) Diversity and partitioning of bacterial populations within the accessory nidamental gland of the squid *Euprymna scolopes*. Appl Environ Microbiol 78: 4200-4208.

117. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, et al. (2013) CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res 41: D348-352.

118. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27: 221-224.

119. Ruby EG (1996) Lessons from a cooperative, bacterial-animal association: the *Vibrio fischeri-Euprymna scolopes* light organ symbiosis. Annu Rev Microbiol 50: 591-624.

120. Miyashiro T, Ruby EG (2012) Shedding light on bioluminescence regulation in *Vibrio fischeri*. Mol Microbiol 84: 795-806.

121. Lamarcq LH, McFall-Ngai MJ (1998) Induction of a gradual, reversible morphogenesis of its host's epithelial brush border by *Vibrio fischeri*. Infect Immun 66: 777-785.

122. Visick KL, Foster J, Doino J, McFall-Ngai M, Ruby EG (2000) *Vibrio fischeri* lux genes play an important role in colonization and development of the host light organ. J Bacteriol 182: 4578-4586.

123. Tong D, Rozas NS, Oakley TH, Mitchell J, Colley NJ, et al. (2009) Evidence for light perception in a bioluminescent organ. Proc Natl Acad Sci U S A 106: 9836-9841.

124. Heath-Heckman EA, Peyer SM, Whistler CA, Apicella MA, Goldman WE, et al. (2013) Bacterial bioluminescence regulates expression of a host cryptochrome gene in the squid-Vibrio symbiosis. MBio 4.

125. Raman R, Rajanikanth V, Palaniappan RUM, Lin YP, He HX, et al. (2010) Big Domains Are Novel Ca2+-Binding Modules: Evidences from Big Domains of *Leptospira* Immunoglobulin-Like (Lig) Proteins. Plos One 5(12) e14377

126. Lyell NL, Dunn AK, Bose JL, Stabb EV (2010) Bright Mutants of *Vibrio fischeri* ES114 Reveal Conditions and Regulators That Control Bioluminescence and Expression of the lux Operon. Journal of Bacteriology 192: 5103-5114.

127. O'shea TM, DeLoney-Marino CR, Shibata S, Aizawa SI, Wolfe AJ, et al. (2005) Magnesium promotes flagellation of *Vibrio fischeri*. Journal of Bacteriology 187: 2058-2065.

128. McFall-Ngai M (2007) Adaptive immunity: care for the community. Nature 445: 153.

129. Canesi L, Gallo G, Gavioli M, Pruzzo C (2002) Bacteria-hemocyte interactions and phagocytosis in marine bivalves. Microsc Res Tech 57: 469-476.

130. Marmaras VJ, Lampropoulou M (2009) Regulators and signalling in insect haemocyte immunity. Cell Signal 21: 186-195.

131. Pham LN, Dionne MS, Shirasu-Hiza M, Schneider DS (2007) A specific primed immune response in Drosophila is dependent on phagocytes. PLoS Pathog 3: e26.

132. Rodrigues J, Brayner FA, Alves LC, Dixit R, Barillas-Mury C (2010) Hemocyte differentiation mediates innate immune memory in *Anopheles gambiae* mosquitoes. Science 329: 1353-1355.

133. Heath-Heckman EA, McFall-Ngai MJ (2011) The occurrence of chitin in the hemocytes of invertebrates. Zoology (Jena) 114: 191-198.

134. Burgess A, Vigneron S, Brioudes E, Labbe JC, Lorca T, et al. (2010) Loss of human Greatwall results in G2 arrest and multiple mitotic defects due to deregulation of the cyclin B-Cdc2/PP2A balance. Proc Natl Acad Sci U S A 107: 12564-12569.

135. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, et al. (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9: 676-682.

136. Wessel D, Flugge UI (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal Biochem 138: 141-143.

137. Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, et al. (2007) The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. Mol Cell Proteomics 6: 1638-1655.

138. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19: 185-193.

139. McDonald WH, Ohi R, Miyamoto DT, Mitchison TJ, Yates Iii JR (2002) Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. International Journal of Mass Spectrometry 219: 245-251.

140. Lo I, Denef VJ, Verberkmoes NC, Shah MB, Goltsman D, et al. (2007) Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. Nature 446: 537-541.

141. Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, et al. (2005) Community proteomics of a natural microbial biofilm. Science 308: 1915-1920.

142. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5: 976-989.

143. Tabb DL, McDonald WH, Yates JR, 3rd (2002) DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res 1: 21-26.

144. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res 2: 43-50.

145. Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, et al. (2006) Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. Methods 40: 303-311.

146. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, et al. (2006) Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. Journal of Proteome Research 5: 2339-2347.

147. Karp NA, Huber W, Sadowski PG, Charles PD, Hester SV, et al. (2010) Addressing accuracy and precision issues in iTRAQ quantitation. Mol Cell Proteomics 9: 1885-1897.

148. Ow SY, Salim M, Noirel J, Evans C, Rehman I, et al. (2009) iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". J Proteome Res 8: 5347-5355.

149. Fan Y, Thompson JW, Dubois LG, Moseley MA, Wernegreen JJ (2013) Proteomic analysis of an unculturable bacterial endosymbiont (*Blochmannia*) reveals high abundance of chaperonins and biosynthetic enzymes. J Proteome Res 12: 704-718.

150. Li X, LeBlanc J, Truong A, Vuthoori R, Chen SS, et al. (2011) A metaproteomic approach to study human-microbial ecosystems at the mucosal luminal interface. PLoS One 6: e26542.

151. Tolin S, Arrigoni G, Moscatiello R, Masi A, Navazio L, et al. (2013) Quantitative analysis of the naringenin-inducible proteome in *Rhizobium leguminosarum* by isobaric tagging and mass spectrometry. Proteomics 13: 1961-1972.

152. Weston AJ, Dunlap WC, Shick JM, Klueter A, Iglic K, et al. (2012) A profile of an endosymbiont-enriched fraction of the coral *Stylophora pistillata* reveals proteins relevant to microbial-host interactions. Mol Cell Proteomics 11: M111 015487.

153. Berton G, Lowell CA (1999) Integrin signalling in neutrophils and macrophages. Cell Signal 11: 621-635.

154. Dupuy AG, Caron E (2008) Integrin-dependent phagocytosis: spreading from microadhesion to new concepts. J Cell Sci 121: 1773-1783.

155. Baxt LA, Garza-Mayers AC, Goldberg MB (2013) Bacterial subversion of host innate immune pathways. Science 340: 697-701.

156. Ham H, Sreelatha A, Orth K (2011) Manipulation of host membranes by bacterial effectors. Nat Rev Microbiol 9: 635-646.

157. Kimbell JR, McFall-Ngai MJ (2004) Symbiont-induced changes in host actin during the onset of a beneficial animal-bacterial association. Appl Environ Microbiol 70: 1434-1441.

158. Duperthuy M, Schmitt P, Garzon E, Caro A, Rosa RD, et al. (2011) Use of OmpU porins for attachment and invasion of *Crassostrea gigas* immune cells by the oyster pathogen *Vibrio splendidus*. Proc Natl Acad Sci U S A 108: 2993-2998.

159. Huang Y, Jellies J, Johansen KM, Johansen J (1997) Differential glycosylation of tractin and LeechCAM, two novel Ig superfamily members, regulates neurite extension and fascicle formation. J Cell Biol 138: 143-157.

160. Jie C, Zipser B, Jellies J, Johansen KM, Johansen J (1999) Differential glycosylation and proteolytical processing of LeechCAM in central and peripheral leech neurons. Biochim Biophys Acta 1452: 161-171.

161. Johansson MW (1999) Cell adhesion molecules in invertebrate immunity. Dev Comp Immunol 23: 303-315.

162. Springer TA (1994) Traffic signals for lymphocyte recirculation and leukocyte emigration: the multistep paradigm. Cell 76: 301-314.

163. Chroneos ZC, Sever-Chroneos Z, Shepherd VL (2010) Pulmonary surfactant: an immunological perspective. Cell Physiol Biochem 25: 13-26.

164. Yang L, Johansson J, Ridsdale R, Willander H, Fitzen M, et al. (2010) Surfactant protein B propeptide contains a saposin-like protein domain with antimicrobial activity at low pH. J Immunol 184: 975-983.

165. Vasta GR (2009) Roles of galectins in infection. Nat Rev Microbiol 7: 424-438.

166. Vasta GR (2012) Galectins as pattern recognition receptors: structure, function, and evolution. Adv Exp Med Biol 946: 21-36.

167. Shi XZ, Wang L, Xu S, Zhang XW, Zhao XF, et al. (2014) A galectin from the kuruma shrimp (*Marsupenaeus japonicus*) functions as an opsonin and promotes bacterial clearance from hemolymph. PLoS One 9: e91794.

168. Tasumi S, Vasta GR (2007) A galectin of unique domain organization from hemocytes of the Eastern oyster (*Crassostrea virginica*) is a receptor for the protistan parasite Perkinsus marinus. J Immunol 179: 3086-3098.

169. Yu Y, Yuan S, Yu Y, Huang H, Feng K, et al. (2007) Molecular and biochemical characterization of galectin from amphioxus: primitive galectin of chordates participated in the infection processes. Glycobiology 17: 774-783.

170. Saftig P, Klumperman J (2009) Lysosome biogenesis and lysosomal membrane proteins: trafficking meets function. Nat Rev Mol Cell Biol 10: 623-635.

171. Darmoise A, Maschmeyer P, Winau F (2010) The immunological functions of saposins. Adv Immunol 105: 25-62.

172. Zavasnik-Bergant T, Turk B (2006) Cysteine cathepsins in the immune response. Tissue Antigens 67: 349-355.

173. Lefebvre C, Vandenbulcke F, Bocquet B, Tasiemski A, Desmons A, et al. (2008) Cathepsin L and cystatin B gene expression discriminates immune coelomic cells in the leech *Theromyzon tessulatum*. Developmental and Comparative Immunology 32: 795-807.

174. Tryselius Y, Hultmark D (1997) Cysteine proteinase 1 (CP1), a cathepsin L-like enzyme expressed in the *Drosophila melanogaster* haemocyte cell line mbn-2. Insect Molecular Biology 6: 173-181.

175. Nepal RM, Mampe S, Shaffer B, Erickson AH, Bryant P (2006) Cathepsin L maturation and activity is impaired in macrophages harboring *M. avium* and *M. tuberculosis*. International Immunology 18: 931-939.

176. Thomas V, Samanta S, Fikrig E (2008) *Anaplasma phagocytophilum* Increases Cathepsin L Activity, Thereby Globally Influencing Neutrophil Function. Infection and Immunity 76: 4905-4912.

177. Nishikori K, Morioka K, Kubo T, Morioka M (2009) Age- and morph-dependent activation of the lysosomal system and *Buchnera* degradation in aphid endosymbiosis. Journal of Insect Physiology 55: 351-357.

178. Troll JV, Adin DM, Wier AM, Paquette N, Silverman N, et al. (2009) Peptidoglycan induces loss of a nuclear peptidoglycan recognition protein during host tissue development in a beneficial animal-bacterial symbiosis. Cell Microbiol 11: 1114-1127.

179. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, et al. (2011) Global quantification of mammalian gene expression control. Nature 473: 337-342.

180. Vogel C, Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nature Reviews Genetics 13: 227-232.

181. Danielsen M, Hornshoj H, Siggers RH, Jensen BB, van Kessel AG, et al. (2007) Effects of bacterial colonization on the porcine intestinal proteome. J Proteome Res 6: 2596-2604.

182. de Lorgeril J, Zenagui R, Rosa RD, Piquemal D, Bachere E (2011) Whole Transcriptome Profiling of Successful Immune Response to Vibrio Infections in the Oyster *Crassostrea gigas* by Digital Gene Expression Analysis. Plos One 6(8): e23142

183. Neilson KA, Mariani M, Haynes PA (2011) Quantitative proteomic analysis of cold-responsive proteins in rice. Proteomics 11: 1696-1706.

184. Wang H, Alvarez S, Hicks LM (2012) Comprehensive comparison of iTRAQ and label-free LC-based quantitative proteomics approaches using two *Chlamydomonas reinhardtii* strains of interest for biofuels engineering. J Proteome Res 11: 487-501.

185. Comstock LE, Kasper DL (2006) Bacterial glycans: key mediators of diverse host immune responses. Cell 126: 847-850.

186. Post DM, Yu L, Krasity BC, Choudhury B, Mandel MJ, et al. (2012) O-antigen and core carbohydrate of *Vibrio fischeri* lipopolysaccharide: composition and analysis of their role in *Euprymna scolopes* light organ colonization. J Biol Chem 287: 8515-8530.

187. Vasta GR, Ahmed H, Nita-Lazar M, Banerjee A, Pasek M, et al. (2012) Galectins as self/non-self recognition receptors in innate and adaptive immunity: an unresolved paradox. Front Immunol 3:199 doi: 10.3389/fimmu.2012.00199

188. Zhang DC, Hu YT, Guo HY, Cui SG, Su TF, et al. (2011) cDNA cloning and mRNA expression of a tandem-repeat galectin (PoGal2) from the pearl oyster, *Pinctada fucata*. Genet Mol Res 10: 1963-1974.

189. Zhang D, Jiang S, Hu Y, Cui S, Guo H, et al. (2011) A multidomain galectin involved in innate immune response of pearl oyster *Pinctada fucata*. Dev Comp Immunol 35: 1-6.

190. Song X, Zhang H, Wang L, Zhao J, Mu C, et al. (2011) A galectin with quadruple-domain from bay scallop *Argopecten irradians* is involved in innate immune response. Dev Comp Immunol 35: 592-602.

191. Song X, Zhang H, Zhao J, Wang L, Qiu L, et al. (2010) An immune responsive multidomain galectin from bay scallop *Argopectens irradians*. Fish Shellfish Immunol 28: 326-332.

192. Yoshino TP, Dinguirard N, Kunert J, Hokke CH (2008) Molecular and functional characterization of a tandem-repeat galectin from the freshwater snail *Biomphalaria glabrata*, intermediate host of the human blood fluke *Schistosoma mansoni*. Gene 411: 46-58.

193. Haudek KC, Spronk KJ, Voss PG, Patterson RJ, Wang JL, et al. (2010) Dynamics of galectin-3 in the nucleus and cytoplasm. Biochim Biophys Acta 1800: 181-189.

194. Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20: 1466-1467.

195. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674-3676.

196. Dunkelberger JR, Song WC (2010) Complement and its role in innate and adaptive immune responses. Cell Res 20: 34-50.

197. Pinto MR, Melillo D, Giacomelli S, Sfyroera G, Lambris JD (2007) Ancient origin of the complement system: emerging invertebrate models. Adv Exp Med Biol 598: 372-388.

198. Smith LC, Azumi K, Nonaka M (1999) Complement systems in invertebrates. The ancient alternative and lectin pathways. Immunopharmacology 42: 107-120.

199. Sottrup-Jensen L (1989) Alpha-macroglobulins: structure, shape, and mechanism of proteinase complex formation. J Biol Chem 264: 11539-11542.

200. Schvartz I, Seger D, Shaltiel S (1999) Vitronectin. Int J Biochem Cell Biol 31: 539-544.

201. Singh B, Su YC, Riesbeck K (2010) Vitronectin in bacterial pathogenesis: a host protein used in complement escape and cellular invasion. Mol Microbiol 78: 545-560.

202. Kirkitadze MD, Barlow PN (2001) Structure and flexibility of the multiple domain proteins that regulate complement activation. Immunol Rev 180: 146-161.

203. Nyholm SV, Passegue E, Ludington WB, Voskoboynik A, Mitchel K, et al. (2006) fester, A candidate allorecognition receptor from a primitive chordate. Immunity 25: 163-173.

204. Ueda A, Nagai H, Ishida M, Nagashima Y, Shiomi K (2008) Purification and molecular cloning of SE-cephalotoxin, a novel proteinaceous toxin from the posterior salivary gland of cuttlefish *Sepia esculenta*. Toxicon 52: 574-581.

205. Kemper C, Atkinson JP, Hourcade DE (2010) Properdin: emerging roles of a pattern-recognition molecule. Annu Rev Immunol 28: 131-155.

206. Kemper C, Hourcade DE (2008) Properdin: New roles in pattern recognition and target clearance. Mol Immunol 45: 4048-4056.

207. Shiomi K, Midorikawa S, Ishida M, Nagashima Y, Nagai H (2004) Plancitoxins, lethal factors from the crown-of-thorns starfish *Acanthaster planci*, are deoxyribonucleases II. Toxicon 44: 499-506.

208. Talbert PB, Henikoff S (2010) Histone variants--ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol 11: 264-275.

209. Dorrington T, Villamil L, Gomez-chiarri M (2011) Upregulation in response to infection and antibacterial activity of oyster histone H4. Fish Shellfish Immunol 30: 94-101.

210. Lee DY, Huang CM, Nakatsuji T, Thiboutot D, Kang SA, et al. (2009) Histone H4 is a major component of the antimicrobial action of human sebocytes. J Invest Dermatol 129: 2489-2496.

211. Parseghian MH, Luhrs KA (2006) Beyond the walls of the nucleus: the role of histones in cellular signaling and innate immunity. Biochem Cell Biol 84: 589-604.

212. Patat SA, Carnegie RB, Kingsbury C, Gross PS, Chapman R, et al. (2004) Antimicrobial activity of histones from hemocytes of the Pacific white shrimp. Eur J Biochem 271: 4825-4833.

213. Anand P, Cermelli S, Li Z, Kassan A, Bosch M, et al. (2012) A novel role for lipid droplets in the organismal antibacterial response. Elife 1: e00003.

214. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, et al. (2004) Neutrophil extracellular traps kill bacteria. Science 303: 1532-1535.

215. Park CB, Kim HS, Kim SC (1998) Mechanism of action of the antimicrobial peptide buforin II: buforin II kills microorganisms by penetrating the cell membrane and inhibiting cellular functions. Biochem Biophys Res Commun 244: 253-257.

216. Tagai C, Morita S, Shiraishi T, Miyaji K, Iwamuro S (2011) Antimicrobial properties of arginine- and lysine-rich histones and involvement of bacterial outer membrane protease T in their differential mode of actions. Peptides 32: 2003-2009.

217. Uyterhoeven ET, Butler CH, Ko D, Elmore DE (2008) Investigating the nucleic acid interactions and antimicrobial mechanism of buforin II. FEBS Lett 582: 1715-1718.