Stony Brook University

Molecular characterization of quahog parasite unknown (QPX), a thraustochytird pathogen of the hard clam, *Mercenaria mercenaria*

A Dissertation Presented

by

Ewelina Rubin

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Marine and Atmospheric Sciences

(Concentration – Marine Biology)

Stony Brook University

December, 2014

Stony Brook University
The Graduate School

Ewelina Rubin

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Bassem Allam – Dissertation Advisor
Associate Professor, School of Marine and Atmospheric Sciences

Jackie Collier – Chairperson of Defense
Associate Professor, School of Marine and Atmospheric Sciences

Emmanuelle Pales Espinosa
Adjunct Assistant Professor, School of Marine and Atmospheric Sciences

David Thanassi
Professor, Department of Molecular Genetics and Microbiology

Jerome F. La Peyre
Associate Professor, Department of Veterinary Science, Louisiana State University

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Molecular characterization of quahog parasite unknown (QPX), a thraustochytird pathogen of the hard clam, *Mercenaria mercenaria*

by

Ewelina Rubin

Doctor of Philosophy

in

Marine and Atmospheric Sciences

(Concentration – Marine Biology)

Stony Brook University

2014

The hard clam or northern quahog, *Mercenaria mercenaria*, is one of the most valuable seafood products in the United States. Severe episodes of hard clam mortality have been consistently associated with infections caused by a thraustochytrid parasite called Quahog Parasite Unknown (QPX). Studies in the present dissertation examined the molecular bases of QPX virulence towards clams. Investigation of QPX transcriptome libraries generated a database of QPX putative virulence factors and other molecules potentially involved in interactions with the clam host. Multiple differential gene expression profiles were generated using oligoarray technology for QPX strains isolated from different geographic locations and for parasite cells cultivated under different culture conditions (protein sources, temperature) previously shown to be important influences on parasite physiology and disease development. The transcriptomic study identified genes responsible for QPX survival and infectiveness at temperatures near the extremes of its growth range (10 °C and 27 °C). QPX cells grown in media containing clam factors exhibited a significant number of up-regulated transcripts related to stress response and detoxification. In addition, QPX cells cultivated in that media did not lead to overt disease in challenged clams suggesting that QPX exposure to clam-factors does not enhance its virulence *in*

vivo. Comparative analysis of transcriptomic profiles of five QPX isolates originating from different geographic locations highlighted molecular adaptation of the parasite to local environmental conditions and underlined molecular mechanisms potentially involved in the variability of virulence across different strains. Bioinformatics and proteomic analyses of the molecular composition of QPX secretome revealed an array of extracellular molecules most likely to be involved in QPX infectiveness, including a variety of hydrolytic and proteolytic enzymes. Substrate-incorporated SDS PAGE gels and peptidase inhibitor studies of QPX extracellular proteolytic enzymes showed that QPX secretes mostly serine type peptidases, including subtilisin-like peptidases which digest proteins present in clam plasma, suggesting that they are directly involved in QPX virulence. The studies presented in this dissertation contribute to a better molecular characterization of QPX and could serve as a basis for future targeted investigations of QPX virulence factors.

Dedication Page

I would like to dedicate this work to my little friend – a wonderful cheerful puppy named Ikar Jr. Ikar was only six months old when after a very long and painful fight he finally gave up and died from an unknown infectious disease. Mourning after his loss made me interested in diseases, pathogens and immunity and ultimately in a strange way led me to start PhD research at MADL. During my PhD research, I learned that pursuit of knowledge not only satisfies human curiosity but also eases human suffering. Throughout the entire research, his short but happy life and his incredibly unfair death were always on my mind so I would like to dedicate my work in his memory.

Frontispiece

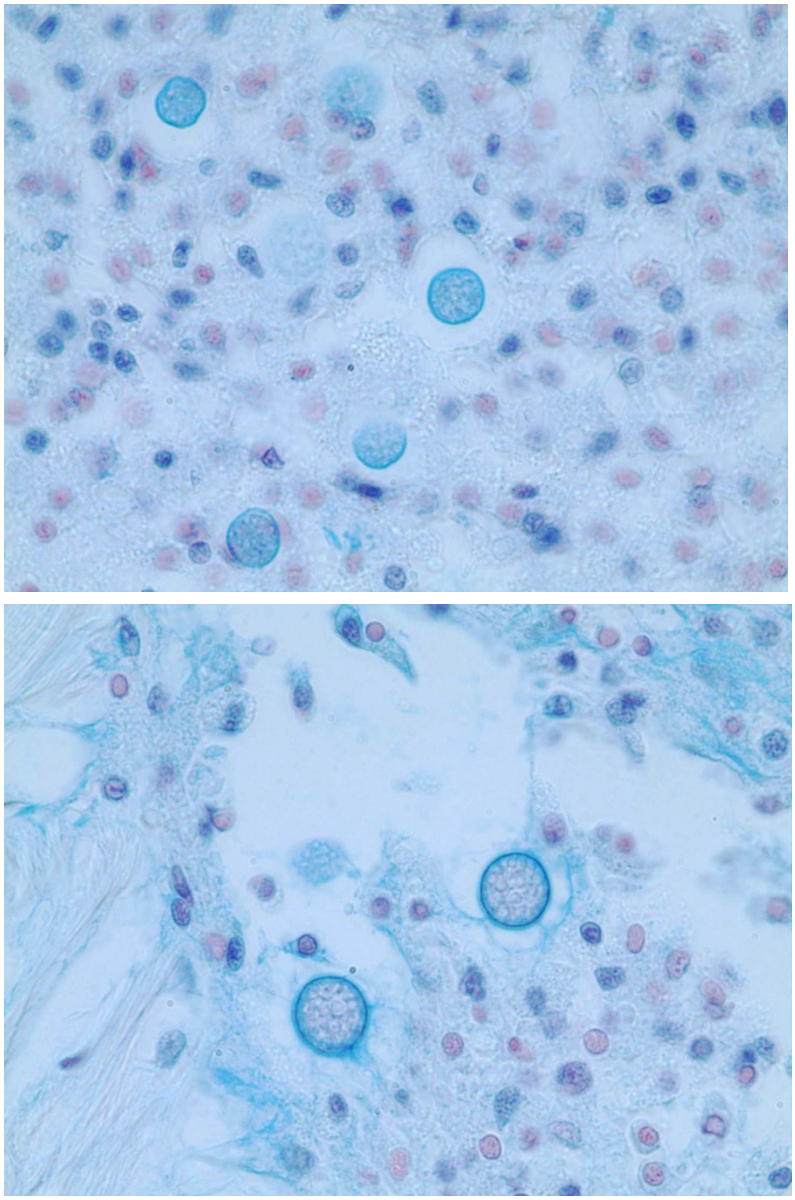


Table of Contents

List of Figures and Tables

different QPX isolates (NY1, NY2, NY3) grown in the same media; MEM – minimal essential medium, YST – yeastolate, FBS – fetal bovine serum, GEL – gelatin, CAMH – clam adductor muscle homogenate, GYP- mix of gelatin hydrolysate, yeastolate, and peptone; protein concentration in the media given in mg/ml

Figure 5.2: Zymograms showing extracellular peptidases produced by QPX (isolate NY20038BC7) grown in: A) minimal essential medium (MEM) supplemented with either yeastolate (YST), fetal bovine serum (FBS), gelatin (GEL) or clam adductor muscle homogenate (CAMH) each to the final protein concentration of 1 $mg \times ml^{-1}$; B) in MEM FBS and CAMH with protein concentration adjusted to 3 $mg \times ml^{-1}$; C) Effect of peptidase inhibitors (10 mM PMSF and 100 μM chymostatin) on extracellular peptidases produced by QPX cells grown in MEM FBS at total protein concentration of about 4 $mg \times ml^{-1}$ (in triplicates); D) Effect of E-64 (cysteine peptidase inhibitor) at 10 uM final concentration on peptidases in QPX ECP samples from part B; -C1, -C2, -C3 and –C4 are media incubated in parallel with the cultures but without QPX cells

Figure 5.3: Extracellular peptidase activity on 0.1% gelatin zymograms of three QPX isolates (NY1 – NY20038BC7, NY2 – NY200320AC6, NY3-NY20073408D) cultivated in minimal essential medium supplemented with gelatin hydrolysate, peptone, and yeastolate (each at 1 mg/ml final concentration); A: fresh ECP without inhibitors; B: ECP frozen for 7 days at -80ºC without inhibitors; C: with 10 mM serine peptidase inhibitor PMSF

Figure 5.4: Differential gene expression of QPX peptidases (papain-like: C1-1 to C1-6 and subtilisin-like: S8-1-S8-5) in QPX cells grown in minimal essential medium (MEM) with addition of one of four different protein supplements (A-B: YST – yeastolate, GEL-gelatin, FBS-fetal bovine serum, CAMH-clam adductor muscle homogenate; at a final concentration of 1 mg/ml) and in QPX cells of three different isolates (C-D: NY1, NY2, and NY3) cultivated in MEM supplemented with gelatin hydrolysate, yeastolate and peptone (1 $mg \times ml^{-1}$ each resulting in 3 $mg \times ml^{-1}$ combined protein concentration). Different letters (a, b, c) designate statistically significant differential gene expression (ANOVA, $p<0.05$, Tukey post hoc analysis)

Figure 5.5: Fractions of QPX ECP purified by ion exchange chromatography (12-19) with protein content revealed on SDS-PAGE gel with silver nitrate stain (A) and peptidase activity revealed on gelatin- and clam plasma- (pool from 3 clams) incorporated gels (B and C, respectively)

.

Acknowledgments

I would like to express my greatest gratitude to my PhD advisor, Dr. Bassem Allam, for the opportunity to pursue research under his supervision. I appreciate all of his help, advice and time which he spent teaching me to be a better research scientist.

I would like to thank Dr. Emmanuelle Pales Espinosa for providing assistance and training on laboratory techniques and her endless advice in experimental planning and design. I also thank her for emotional support in times of self-doubt and discouragement.

I want to thank Joyce Lau for her friendship, cheerfulness and for always offering the right perspective on what's most important in life.

I thank all the other members of the SoMAS Marine Animal Disease Laboratory for their direct help or advice, without which this accomplishment would not be possible: Mickael Perrigault, Sue Pawagi, Joyce Lau, Kailai Wang, Soren Dahl, Chris Del Castillo, and Melissa Ouvrard.

I would like to thank my entire PhD thesis committee: Dr. Jackie Collier, Dr. Jerome La Peyre, Dr. David Thanassi, Dr. Emmanuelle Pales Espinosa and Dr. Bassem Allam, for their time spent on reviewing and critiquing the work presented in this document.

I thank New York Sea Grant, the New York State Department of Environmental Conservation, the Gordon and Betty Moore Foundation and the National Science Foundation for providing resources to conduct the scientific research presented in this thesis.

Special and warmest thanks go to my family, my husband Jon and my doggy Mushroom whose unconditional love makes every moment of my life special and happy.

Background

QPX, the thraustochyrid parasite of hard clams

Quahog Parasite Unknown (QPX, Whyte et al 1994) is a unicellular eukaryote infecting the hard clam, *Mercenaria mercenaria,* which represents one of the most valuable shellfish species in the USA. QPX related mortalities have led to substantial financial losses for the aquaculture industry. QPX disease outbreaks in cultured and wild clams have been reported to occur in Canada (MacCallum & McGladdery 2000, Whyte et al 1994) and the USA (Dove et al 2004, Ragone Calvo et al 1998, Smolowitz et al 1998). QPX disease has not been documented in clam populations living in coastal waters south of Virginia, likely because the parasite does not survive well at temperatures exceeding 27 °C (Perrigault et al 2011). As a result, clam strains originating from southern states, which have not been exposed to QPX in their natural environment, exhibit higher susceptibility to QPX infection (Dahl et al 2008 and 2010, Kraeuter et al 2011, Ragone Calvo et al 2007). QPX is considered to be a facultative opportunistic pathogen that lives in the marine environment outside its clam host (Lyons et al 2007). QPX DNA and in a few cases QPX cells have been detected in marine sediments, marine aggregates, and on shell surfaces of dead marine invertebrates (Gast et al 2008, Liu et al 2008, Lyons et al 2006, Lyons et al 2005). It is hypothesized that QPX is transmitted to clams from organic detritus collected in the mantle cavity as pseudo-feces (Lyons et al 2007). QPX lesions are commonly detected in clam mantle tissue, also suggesting that the mantle is an entry point and the start of the infection. During extensive mortality events, clams are diagnosed with QPX infection spreading to other organs and throughout the visceral mass (Dove et al 2004, Ragone Calvo et al 1998, Smolowitz et al 1998).

Two life stages of QPX have been observed in clam tissue and *in vitro* cultures. These include vegetative cells (thalli) which are between 2 to 15 μm in diameter, and sporangia (sori) which are about 20 to 45 μm in diameter (Ragone Calvo et al 1998, Smolowitz et al 1998). Each sporangium contains between 30 to 40 immature thalli (endospores) which are released by the break of the sporangial cell wall (Figure B-1)

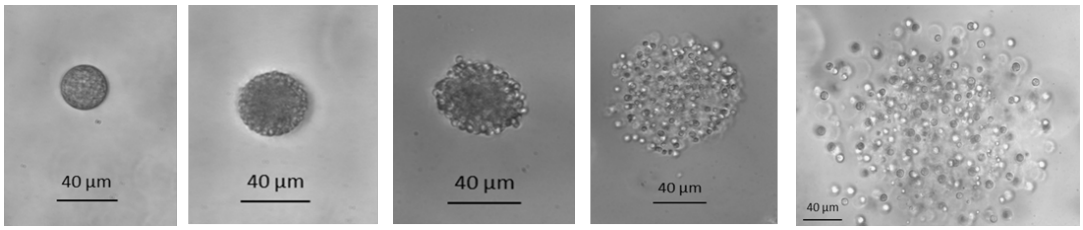


Figure B-1 Life cycle of Quahog Parasite Unknown (QPX) strain NY0313808bc7 grown in minimal essential medium supplemented with salts. Life stages include A. Mature vegetative cells (thalli) B. Developing sporangia (sori) C. mature sporangium, D. release of immature vegetative cells (endospores) via cell wall dissolution E. growth and expansion of young thalli into the media

QPX cells and sporangia inside clam tissue are surrounded by muco-polysaccharide secretion which protects QPX cells from antimicrobial compounds produced by the clam (Anderson et al 2003). QPX life stages, specifically the reproducing sporangia, have not been detected in environmental samples. In culture, QPX does not produce a motile stage (zoospores), which is usually present in the life cycle of other thraustochytrids. However, the presence of QPX zoospores in the environment cannot be excluded. It is also unknown how long QPX cells can persist in the environment without entering the clam host. No sexual stages have been reported for any thraustochytrid (Porter 1990). Indirect evidence that QPX possesses only asexual reproduction can be speculated from the very high intragenomic nucleotide variation of the internal transcribed spacer (ITS1 and ITS2) regions (Qian et al 2007).

QPX is a taxonomically unnamed species which, based on the nucleotide sequence of its small ribosomal subunit, belongs to the thraustochytrid branch within the *Labyrinthulomycetes* phylogenic group (Maas et al 1999, Ragan et al 2000). *Labyrinthulomycetes* (aka slime nets or labys) are composed of ecologically important but greatly understudied marine protists most closely related to water molds (oomycetes), diatoms, golden and brown algae, all of which collectively form a monophyletic cluster named stramenopiles, aka heterokonts (Tsui et al 2009). Slime nets have many different life strategies. Many are free-living saprophytes associated with algal or animal detritus, but some are parasitic, commensal and mutualistic species (Raghukumar 2002). In the past, several Labrinthulomycete organisms have been reported to cause diseases in other mollusks, including abalone, squids and nudibranchs (Bower 1987, Jones & O'Dor 1983, McLean & Porter 1982), but these diseases are no longer studied, so the data on the Labyrinthulomycete pathogens affecting animals is very scarce.

One labyrinthulid pathogen, *Labyrinthula zostera*, causes devastation of marine seagrasses worldwide (Bergmann et al 2011). However, no molecular information about the pathogen has been gathered to date. In recent years, other Labyrinthulomycete species have been emerging as parasites in marine invertebrates. For example, an undescribed thraustochytrid species is infecting eggs of a mud crab in Australia (Kvingedal et al 2006), and another was found to be parasitic to a free-living flatworm (Schärer et al 2007). Another recently described species, *Labyrinthula terrestris*, is the first labyrinthulid to affect terrestrial grasses (Bigelow et al 2005). Several free-living species of thraustochytrids have been more extensively investigated because of their ability to produce large amount of polyunsaturated fatty acids (e.g. omega-3-oil) (Gupta et al 2012, Raghukumar 2008). The increasing interest in their biotechnological application resulted in the first sequenced transcriptome of a free-living thraustochytrid, *Schizochytrium* sp. (Huang et al 2008). Most of the parasitic labyrinthulids lack any significant genomic or transcriptomic sequence information. It is also unknown how the labyrinthulids evolved their pathogenic or parasitic abilities and what are their cellular and molecular virulence mechanisms that enable them to invade and cause diseases in marine invertebrates and plants.

Virulence factors of infectious microorganisms

Pathogenic organisms produce a variety of biomolecules which facilitate their invasion and propagation within host tissues and fluids. Molecules which are essential for a successful infection and colonization of host tissue are frequently referred to as pathogenicity factors (Latijnhouwers et al 2003). In addition, many pathogenic agents produce biochemical molecules that are not essential but contribute to the efficiency of an infection process (aggressiveness factors) and/or evoke defense responses from their hosts (elicitors or effectors) (Kale & Tyler 2011, Latijnhouwers et al 2003). In the current thesis, virulence factors will be considered as any molecules, mechanisms or characteristics which contribute to pathogen infectivity. These could be any characteristics of the pathogen which provide protection against the immune defense of the host or directly attack (destroy or manipulate) host immune cells and molecules.

Common infection mechanisms are shared by some evolutionarily divergent lines of eukaryotic pathogens. For example, some pathogens have evolved similar machinery of secretion and injection of virulence molecules directly into host tissues and cells. Conserved amino acid motifs in proteins that are directly secreted into the host cells (known as host-targeting sequence) can be found in the malaria pathogen, *Plasmodium falciparum* and the oomycete potato

pathogen, *Phytophthora infestans* (Haldar et al 2006). Other pathogens release virulence proteins into the extracellular space, alter host receptor and consequently enter the host cell via endocytosis. Some of the major and better studied groups of virulence factors which are common to various pathogens from distant phylogenetic lineages include hydrolytic and proteolytic enzymes and their inhibitors, antioxidants, surface and extracellular molecules used for attachment, recognition and protection (Haldar et al., 2006; Meng et al, 2009)

Hydrolyzing enzymes are produced by pathogenic organisms to destroy host derived polymers (proteins, polysaccharides and lipids). The major function of these enzymes is nutrient acquisition, but they may play a role in direct invasion and killing of host tissue and immune cells, or neutralization of defense proteins (Sajid & McKerrow 2002). For example, phospholipases are important enzymes hydrolyzing the cell membranes of host cells (Stehr et al., 2003). Proteolytic enzymes (also known as peptidases) degrade host protein material such as collagen, fibrin and actin in the connective tissue of the extracellular matrix and also are cable of destroying immunoglobulins (Klemba & Goldberg 2002, McKerrow 1993, Olivos-Garcia et al 2009, Sajid & McKerrow 2002).

The molecular host-parasite ware fares have also led to the evolution of antioxidant molecules in pathogens (Bogdan 2007). For instance, an important defense mechanism of metazoan host cells is the production of reactive oxygen species (ROS). It is a result of the respiratory burst in response to contact with invading microorganisms. To neutralize these toxic ROS, parasites co-evolved a counter-defense mechanism allowing them to survive within host cells and extracellular matrices (Faust and Guillen, 2012, Osorio et al., 2012). The cellular defense cells, the hemocytes of bivalves, including *M. mercenraria,* produce ROS when in contact with pathogens (Bugge et al 2007).

The diversity and nature of molecules covering the surface of pathogenic microorganisms have been shown to be involved in their infectivity. A key step in the establishment of the infection site is the ability of the pathogen to recognize and to attach to the host (Calderone and Fonzi, 2001). Thus, cell adhesion molecules and cell surface receptors are commonly investigated. . For example, mucins and mucin related molecules produced by pathogens have an essential role in their ability to establish infection (Jain et al 2001) via adhesion, host surface recognition, and protection. Other important surface glycoproteins involved in host surface recognition are lectins. Lectins have been implicated in virulence of animal pathogens such as

Entamoeba spp. (Gilchrist & Petri 1999, Mann 2002, Stanley & Reed 2001) and *Trypanosoma cruzi* (Jacobs et al 2010), and also in plant oomycete pathogens such as *Phytophthora parastica* (Gaulin et al 2002).

The above multitude of molecules enables microorganisms to become infectious agents by allowing them to recognize and enter host tissue or cells, attach or anchor themselves to host extracellular matrices or cells, survive in the hostile host environment, protect themselves against host immune responses and eventually damage or destroy host tissues. The research presented in the current dissertation was designed to identify and describe molecules produced by QPX which might play a role as virulence factors for its invasion, host tissue degradation and manipulation of clam immune responses, thus allowing for QPX disease development. The approaches taken to reach the overall aim were:

A. To sequence and annotate the QPX transcriptome and provide a database of putative QPX virulence factors (Chapter 1)
B. To evaluate transcriptional changes of QPX cell exposed to temperatures known to control the disease development (Chapter 1)
C. To measure the effect of different media on the expression of putative QPX virulence factors (Chapter 2)
D. To study gene expression profiles of different QPX isolates (Chapter 3)
E. To characterize the QPX secretome (Chapter 4)
F. To functionally characterize QPX extracellular peptidases (Chapter 5)

Chapter 1: Characterization of the transcriptome and temperature-induced differential gene expression in QPX, the thraustochytrid parasite of hard clams

Manuscript published in BMC Genomics 2014, 15:245

Abstract

The hard clam or northern quahog, *Mercenaria mercenaria*, is one of the most profitable shellfish species in the United States representing the first marine resource in some Northeastern states. Severe episodes of hard clam mortality have been consistently associated with infections caused by a thraustochytrid parasite called Quahog Parasite Unknown (QPX). QPX is considered as a cold/temperate water organism since the disease occurs only in the coastal waters of the northwestern Atlantic Ocean from Maritime Canada to Virginia. High disease development at cold temperatures was also confirmed in laboratory studies and is thought to be caused predominantly by immunosuppression of the clam host even though the effect of temperature on QPX virulence has not been fully investigated. In this study, the QPX transcriptome was sequenced using Roche 454 technology to better characterize this microbe and initiate research on the molecular basis of QPX virulence towards hard clams. Close to 18,000 transcriptomic sequences were generated and functionally annotated. Results revealed a wide array of QPX putative virulence factors including a variety of peptidases, antioxidant enzymes, and proteins involved in extracellular mucus production and other secretory proteins potentially involved in interactions with the clam host. Based on these sequences, a 15 K oligonucleotide array was constructed and used to investigate the effect of temperature on QPX fitness and virulence factors. Results identified a set of molecular chaperones that could explain QPX acclimation to cold temperatures. Finally, several putative virulence-related factors were up-regulated at low temperature providing molecular targets for further investigations of increased QPX pathogenicity in cold water conditions. This is one of the first studies to characterize the transcriptome of a parasitic labyrinthulid, offering new insights into the molecular bases of the pathogenicity of members of this group. Results from the oligoarray study demonstrated the ability of QPX to cope with a wide range of environmental temperatures, including those considered to be suboptimal for clam immunity (low temperature) providing a mechanistic scenario for disease distribution in the field and for high disease prevalence and intensity at low

temperature. These results will serve as the basis for studies aimed at a better characterization of specific putative virulence factors.

Introduction

QPX (Quahog Parasite Unknown) is a unicellular eukaryote that infects hard clam, *Mercenaria mercenaria*, populations along the northeastern coast of North America (Dove et al 2004, Ragone Calvo et al 1998, Smolowitz et al 1998, Whyte et al 1994). QPX is thought to be an opportunistic pathogen occurring at low abundances in coastal waters, sediments and organic debris (Gast et al 2008, Haas et al 2009, Lyons et al 2005), that occasionally becomes invasive to clams exhibiting lower resistance (Dahl et al 2008, Dahl et al 2010, Ragone Calvo et al 2007), suppressed immunity and/or during unfavorable environmental conditions (Dahl et al 2011, Ford et al 2002, Perrigault et al 2011). QPX is an undescribed species belonging to thraustochytrids in the phylogenetic group *Labyrinthulomycetes* (Maas et al 1999, Ragan et al 2000, Stokes et al 2002). Nothing is known about the evolution of pathogenicity in this group of marine protists. The closest relatives of the Labyrinthulomycetes are the Oomycetes, also known as water molds. The group contains many species which are pathogenic to important terrestrial crops, vegetables, or ornamental plants and trees. Because of their economic importance they have been extensively studied. For example, the potato blight pathogen, *Phytophthora infestans*, which is a model species for the oomycetes, has had its genome, transcriptome, and secretome sequenced and described (Avrova et al 2007, Gavino & Fry 2002, Haas et al 2009, Raffaele et al 2010). Similarly, genomes or at least transcriptomes of many other pathogenic oomycetes have been explored, providing a comprehensive view of their pathogenicity (Cheung et al 2008, Gaulin et al 2008, Mao & Tyler 1996, Panabieres & Le Berre 1999). The extensive nucleotide and protein sequence information of these pathogens has led to major advancements in the understanding of their virulence characteristics and the molecular mechanisms of the infection process. The present study was directed to provide a first look into transcriptomic sequences of a pathogenic member of the Labyrinthulomycetes. .

As is often the case for opportunistic infections, QPX disease occurrence and development is dependent upon external environmental conditions affecting both the invader and the host. QPX's natural geographic range of distribution shows that QPX is a temperate/cold climate adapted species, as it has not been reported in relatively warmer waters south of Virginia. Interestingly, QPX optimal *in vitro* growth temperature is between 20 °C and 23 °C, and

suppression of its growth occurs at both colder (below 8 °C) and elevated (above 27 °C) temperatures (Perrigault et al 2010). The major impact of temperature on disease development is reflected in the natural disease patterns in the temperate waters for which QPX prevalence in clams shows seasonal fluctuations, with peaks ranging from fall in MA to early summer and spring in NY and VA, respectively (Dove et al 2004, Ragone Calvo et al 1998, Smolowitz et al 1998 , Whyte et al 1994). On the other hand, experimental transmission studies reveal maximal disease development at low water temperatures, which has been attributed to the suppression of clam immunity (Dahl et al 2011, Perrigault et al 2011) and it is hypothesized that immunosuppression is a major factor controlling QPX disease development. The effects of cold temperature on some of the clam immune parameters (Perrigault et al 2011) support the hypothesis of cold water influencing the disease. On the contrary, very little is known about the effect of temperature on QPX physiology, fitness or its virulence characteristics and ability to cause disease.

Even though significant clam mortality has been associated with QPX disease since the early 1950's (Whyte et al 1994), many aspects of the disease remain unexplored and understudied. Little progress has been made to understand QPX basic cell biology, including its life stages, nutritional requirements, and factors mediating its virulence toward clams. The cellular mechanisms of the infection process are only partially described from histopathologic examinations. Very recently, the partial sequencing of the QPX genome and two transcriptomes of QPX initiated the work on QPX molecular characteristics and the effect of temperature on its physiology (Garcia-Vedrenne et al 2013). The present study was also conducted to gather a better understanding of QPX cell biology and to explore its potential virulence factors. For that purpose, transcriptome sequences of QPX, were generated, assembled and annotated and then used to build a custom designed 15000 oligonucleotide array that was subsequently used to investigate QPX molecular response to temperatures known to regulate disease development (Dahl et al 2011, Perrigault et al 2011). Findings are discussed with a focus on the characterization of possible links between temperature and the expression profiles of virulence-related factors.

Materials and Methods

Library construction and sequencing

QPX isolate NY0313808BC7 (Qian et al 2007) initially isolated from a diseased clam collected from Raritan Bay, New York, was used for RNA extraction and library construction. To enhance library coverage, QPX was separately grown in Minimum Essential Medium (MEM, Sigma) supplemented with one of the following protein sources: clam muscle homogenates (500, 1000 or 3000 μg ml^{-1} (Perrigault et al 2009a), fetal bovine serum (0%, 2.5%, 5% or 10%) or gelatin (0.1%, 0.2% or 0.3%). Additional cultures were grown in MEM supplemented with yeastolate (0.1%, 0.2% or 0.3%) or with clam muscle homogenate made in sterile seawater without MEM. Cultures were incubated at 23 °C and parasite cells were separately harvested on Day 8 and Day 14. Parasite biovolume in each sample was assessed using a fluorometric technique (Bugge & Allam 2005) before samples were combined using the same parasite biovolume from each culture condition. Total RNA was extracted from pooled samples using Trizol (Molecular Research Center, Inc.). The quality of total RNA was verified using Agilent 2100 bioanalyser and quantified by assessing the A260/280 and A260/230 ratios using a Nanodrop (ND 1000) spectrophotometer. Poly(A+) RNA was isolated from 300 μg of total RNA using the PolyATract® mRNA Isolation system (Promega) following the manufacturer's instructions and used for cDNA synthesis. Preparation and sequencing of the cDNA library were performed at the McGill University and Génome Québec Innovation Centre (Canada) following the manufacturer's protocol (Roche-454 Life Sciences, Brandford, CT, USA). The SMART adaptor sequences were removed from reads using a custom script. Reads less than 50 bp and low-quality reads were filtered out and remaining high-quality reads were assembled using MIRA assembler (version 3.0.5) (Chevreux et al 2004).

Transcriptome annotation

The annotation of the QPX transcriptome was completed using the online bioinformatics tool Blast2GO (Gotz et al 2008). First, all sequences were subjected to blastx (basic local alignment tool) searches against the National Center for Bioinformatics (NCBI) sequence database with E-value cut off of 10^{3}. Next the sequences proceeded through mapping to Gene Ontology (GO) Consortium database of standardized phrases describing functional information of known gene products, and finally GO functional annotation was completed using cutoff value 10^{-3}. Using the same tool, the sequences in all six translation frames were subjected to the InterProScan to find conserved protein domain matches in the Integrated Protein database of the European Bioinformatics Institute (Hunter et al 2009).

In silico identification of QPX putative virulence factors

The selection of potential virulence factors was based on the above annotated QPX transcriptome library and similarity of QPX transcripts and protein sequences to the virulence factors of other protistan parasites and pathogens. The QPX sequence library was screened for sequences encoding for a variety of peptidases, hydrolytic enzymes, antioxidants, cell surface receptor and adhesion molecules. Manual curation of automated annotation as well as alignments and translations into amino acid sequences were accomplished using Geneious software (Drummond et al 2011). The translated amino acid sequences were manually checked for the correct protein signatures and conserved protein domain using MEROPS, the peptidase database at http://merops.sanger.ac.uk/ (Rawlings et al 2012) and in the InterPro database from the European Bioinformatics Institute, http://www.ebi.ac.uk/nterProScan (Hunter et al 2009, Zdobnov & Apweiler 2001) .

Transcriptomic changes in QPX in response to temperature

For the temperature treatment, 500 µl of an exponentially growing QPX (isolate NY0313808BC7) culture was inoculated into twelve 25-ml culture flasks (Falcon) filled with 5 ml of MEM supplemented with 10% fetal bovine serum (FBS, Sigma). The flasks were incubated at four different temperatures: 27 °C, 23 °C, 13 °C and 10 °C, in triplicate (except 10 °C which was done in duplicate) for each treatment. After seven days of incubation, the cultures were diluted with equal volume of filter sterilized artificial seawater and passed several times through a syringe to facilitate liquefaction of QPX mucus secretion. The mixtures were then transferred into 15-ml conical tubes and centrifuged at 3000g for 40 minutes at 4 °C. The supernatant was discarded and cell pellets collected and kept on ice for immediate RNA extraction. Trizol reagent (Molecular Research Center, Inc.) was used to isolate RNA from all samples following manufacturer's protocol. RNA quality and quantity were estimated spectophotometrically using a Nandrop spectrophotometer.

Oligoarray design and hybridization

A subset of contigs produced by MIRA was used for the production of a 60-mer oligonucleotide array using the Agilent eArray application (https://earray.chem.agilent.com/earray/). These included 6,781 curated annotated sequences and 8,297 non-annotated sequences (minimal size = 215 b) to emphasize gene discovery. One probe

was produced for each submitted sequence. Probes were synthesized in situ along with positive and negative controls using 8x15K-feature Agilent format slides. Labeled (Cy3 or Cy5) complementary RNA (cRNA) was synthesized from 150 ng of RNA purified from cultures grown at different temperatures using Trizol (Invitrogen) and the Two-Color Microarray-Based Gene Expression Analysis Protocol (Quick Amp Labeling) following the manufacturer's protocol. Labeled cRNA was purified using Illustra CyScribe GFX Purification Kit (GE Healthcare). cRNA quantity and quality (including dye incorporation) were determined by spectrophotometry (Nanodrop). Samples were considered satisfactory if cRNA concentration and incorporation efficiency exceeded 300 ng/µl and 8 pmol Cy/µg cRNA, respectively. All arrays were hybridized with the same amount of cRNA (300 ng of each Cy3- and Cy5-labeled cRNA). Arrayhybridization and washes were conducted according to the kit protocol and the arrays were scanned with a GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA, USA) using the suggested Agilent scan settings.

Oligoarray data normalization and analysis

Spot fluorescence intensities were extracted using GenePix software. LIMMA package in R software was used to normalize the intensity data and to remove within-array (method: global lowess) and between-array (method: quantile) non-biological variation (Smyth & Speed 2003). After normalization, the intensities in separate color channels were exported into an excel spreadsheet for further data quality control and trimming. The intensities which were less than twofold background intensity were eliminated from further analysis. The relative expression of each transcript at each temperature treatment was calculated as the ratio of the mean intensity (n=3 for 27 °C, 23 °C, 13 °C, and n=2 for 10 °C) for each temperature treatment and the mean intensity of that transcript in all temperature treatments (n=11). Hierarchical clustering of all samples and genes (Pearson correlation) and the determination of statistically significant differentially expressed genes (one way ANOVA, p-value < 0.01) were completed in the TM4-suite using MeV program (Saeed et al 2006, Saeed et al 2003). The final criteria for differential gene expression were the significance by ANOVA analysis and a one and a half fold increase from the mean (up-regulation) or a one and a half fold decrease from the mean (down-regulation) in at least one of the experimental treatments. The transcript sequences in the few categories of interest were investigated further by amino acid sequence alignment analysis to identify the

number of unique QPX genes responsive to temperature. All translations and alignments were accomplished using Geneious software.

Real- time quantitative PCR and oligo array validation

Real-time quantitative PCR was performed on selected transcripts of interest which were shown to be differentially and no-differentially expressed by oligoarray methodology. Total RNA (2.5 ug) from each sample was used to synthesize cDNA using the MMLV reverse transcription kit (Promega) and oligo dT primers following the manufacturer's protocol. Relative quantification was carried out in 10-ul reactions with Brilliant II SYBR green qPCR master mix (Agilent), 100 nM final primer concentration and 5 ng of RNA-equivalent cDNA. The PCR reactions were performed using a Mastercycler ep realplex PCR machine (Eppendorf). Primers with a melting temperature of 60° C were designed using the PrimerQuest program (Integrated DNA Technologies, IDT) within the open reading frames encoding six different QPX peptidases and actin, which was used as a reference gene for mRNA normalization. The amplification products (135 – 148 bp) were confirmed using gel electrophoresis. The relative transcript levels were calculated using the delta delta Ct method (Livak & Schmittgen 2001).

Results

Characterization of the QPX transcriptome

A total of 490,726 sequence reads (average length of 215 bp) were generated by the 454 GS-FLX Titanium platform. Quality and length filtering yielded 223,652 high-quality reads that were assembled using MIRA. The resulting QPX transcriptome library consists of 17,934 transcript sequences including contigs (continuous sequences, 14,636), singletons (single copy sequences, 1636), and long repetitive contigs (lrc, 1662) with sequence length ranging from 40 bp to over 2700 bp. Sequences longer than 200 bp have been deposited at DDBJ/EMBL/GenBank under the TSA (Transcript Shotgun Assembly) accession number: GALJ00000000. A total of 7,093 transcripts were found to have positive blastx similarity matches to other sequences in the NCBI non-redundant (nr) protein database ($E < 10^{-3}$), and 4,644 sequences were identified to have a match to at least one conserved protein domain in the InterPro database. In addition, 5,764 sequences were assigned at least one functional annotation term based on the Gene Ontology vocabulary database (**Figure 1.1**). By default, the GO annotation phrases are grouped in three major categories: biological process, molecular function,

and cellular component. Each main category is subdivided into hierarchical subcategories (levels), with each subsequently higher level containing more detailed functional description of a hypothetical gene product. The level-2 functional annotation of the QPX transcriptome returned 5,768 and 5,459 gene ontology phrases associated with the biological process (**Figure 1.1A**) and molecular function (**Figure 1.1B**) categories, respectively. The combined level 2 and 3 functional annotation of QPX transcripts resulted in 3,081 cellular component gene ontologies (**Figure 1.1C**). Within the biological processes, metabolic (GO: 0008152) and cellular (GO: 0009987) processes were represented by the highest number of QPX transcripts, 2,177 and 1,849 respectively (**Figure 1.1A**).

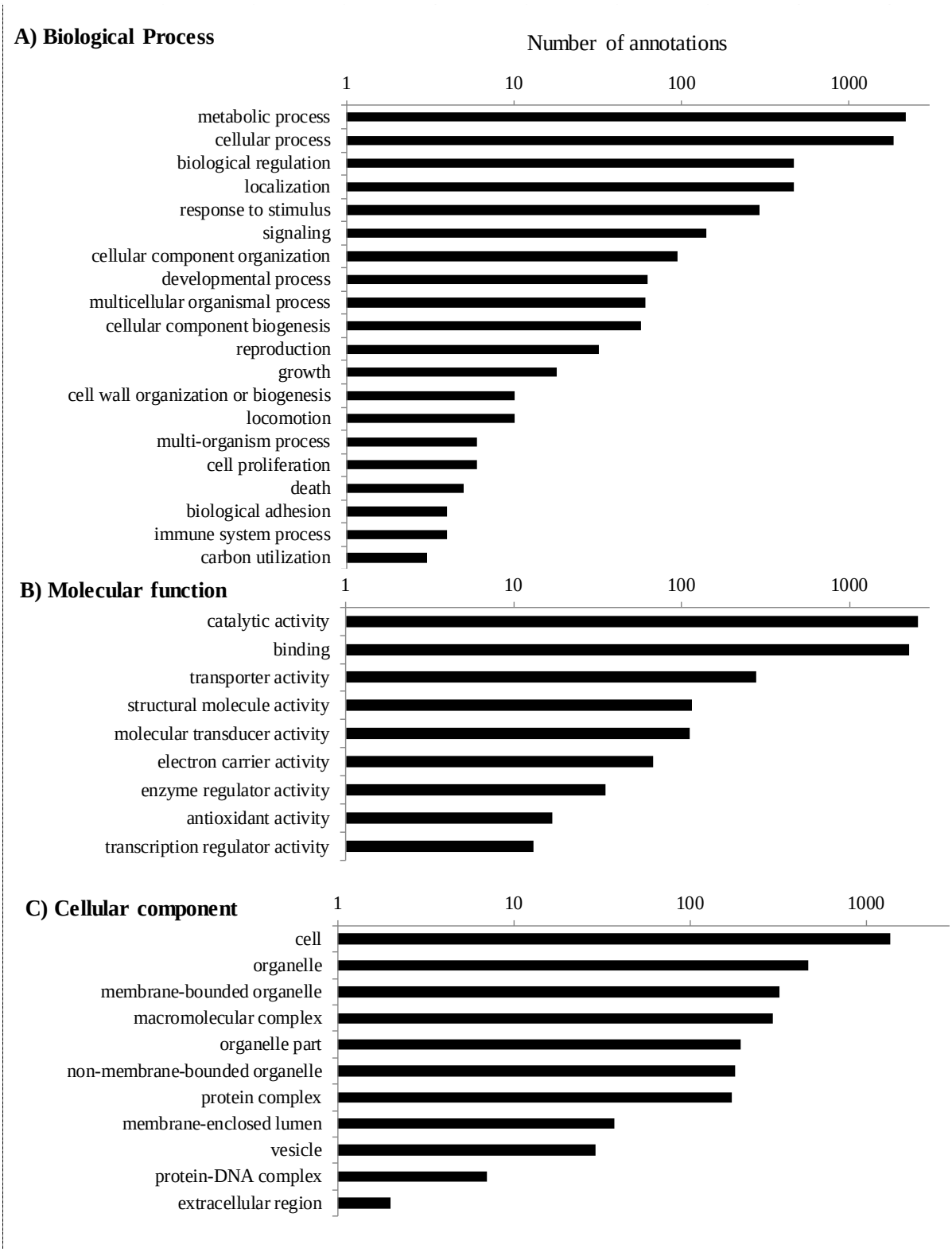


Figure 1.1: Gene Ontology (GO) annotations identified by the Blast2GO search tool in QPX transcriptome sequence library. A) a total of 5768 level-2 annotations within the biological process ontology; B) 5459 molecular function level-2 annotations C) combined level-2 and level-3 annotations (3081) within the cellular component ontology

The QPX transcriptome library contained a variety of functional annotations commonly used to predict biological processes involved in pathogenicity and host-pathogen interactions **(Table 1.1**). A total of 777 QPX transcripts were annotated to encode hydrolytic enzymes (GO:0016787), and some of them could possibly be implicated in QPX virulence (**Table 1.1**). For example, within the GO biological process category, 203 transcripts were annotated with the molecular function defined as peptidase activity (GO:0008233). These included transcripts encoding protein digesting enzymes in the five major peptidase categories: serine (45), cysteine (35), metallo (47), aspartate (10), and threonine peptidases (7) (**Table 1.1).** In addition, transcripts annotated as lipases (GO:0016298), phospholipases (GO:0004623, GO:0004620) and glycosidases (GO:0016160, GO:0004568, GO:0015925, GO:0015926) were also identified.

Table 1.1: Number of QPX transcripts annotated with gene ontology phrases commonly associated with virulence characteristics. F: molecular functional, P: biological process, C: cellular component

Gene Ontology Term	GO Category	GO Level	GO ID	#Seqs
Degradation of biopolymers				
hydrolase activity	F	3	GO:0016787	777
peptidase activity	F	4	GO:0008233	203
peptidase activity, acting on L-amino acid peptides	F	5	GO:0070011	167
proteolysis	P	5	GO:0006508	158
endopeptidase activity	F	6	GO:0004175	88
metallopeptidase activity	F	6	GO:0008237	47
serine-type peptidase activity	F	5	GO:0008236	45
exopeptidase activity	F	6	GO:0008238	38
cysteine-type peptidase activity	F	6	GO:0008234	35
metalloendopeptidase activity	F	7	GO:0004222	30
nuclease activity	F	5	GO:0004518	29
phosphatase activity	F	6	GO:0016791	27
serine-type endopeptidase activity	F	6	GO:0004252	26
cysteine-type endopeptidase activity	F	7	GO:0004197	17
exonuclease activity	F	6	GO:0004527	14
aminopeptidase activity	F	7	GO:0004177	11
aspartic-type endopeptidase activity	F	7	GO:0004190	10
endonuclease activity	F	6	GO:0004519	9
endopeptidase inhibitor activity	F	5	GO:0004866	9
lipase activity	F	5	GO:0016298	9
peptidase inhibitor activity	F	4	GO:0030414	9
serine-type carboxypeptidase activity	F	8	GO:0004185	7
serine-type exopeptidase activity	F	6	GO:0070008	7
threonine-type peptidase activity	F	6	GO:0070003	7
metalloexopeptidase activity	F	7	GO:0008235	6
cysteine-type endopeptidase inhibitor activity	F	6	GO:0004869	4
phospholipase activity	F	6	GO:0004620	4
urease activity	F	6	GO:0009039	4
chitinase activity	F	6	GO:0004568	3
serine-type endopeptidase inhibitor activity	F	6	GO:0004867	3
acid phosphatase activity	F	7	GO:0003993	2
glucosidase activity	F	6	GO:0015926	2
amylase activity	F	6	GO:0016160	1
galactosidase activity	F	6	GO:0015925	1
Cell homostasis and protection against oxidative stress				
antioxidant activity	F	2	GO:0016209	17
response to oxidative stress	P	4	GO:0006979	11
peroxidase activity	F	3	GO:0004601	10
response to reactive oxygen species	P	5	GO:0000302	6
response to hydrogen peroxide	P	6	GO:0042542	6
superoxide metabolic process	P	5	GO:0006801	5

Table 1.1: continued. Number of QPX transcripts annotated with gene ontology phrases commonly associated with virulence characteristics. F: molecular functional, P: biological process, C: cellular component

Gene Ontology Term	GO Category	GO Level	GO ID	#Seqs
cellular response to oxidative stress	P	5	GO:0034599	4
catalase activity	F	4	GO:0004096	4
cellular response to reactive oxygen species	P	6	GO:0034614	3
superoxide dismutase activity	F	3	GO:0004784	2
thioredoxin-disulfide reductase activity	F	3	GO:0004791	2
peroxiredoxin activity	F	5	GO:0051920	2
thioredoxin peroxidase activity	F	4	GO:0008379	1
glutathione peroxidase activity	F	4	GO:0004602	1
Secretion and polysaccharide production				
envelope	C	4	GO:0031975	70
vesicle-mediated transport	P	5	GO:0016192	38
external encapsulating structure	C	4	GO:0030312	31
cell envelope	C	5	GO:0030313	26
external encapsulating structure part	C	4	GO:0044462	25
cell outer membrane	C	5	GO:0009279	22
polysaccharide metabolic process	P	4	GO:0005976	21
polysaccharide biosynthetic process	P	6	GO:0000271	13
cellular polysaccharide metabolic process	P	6	GO:0044264	10
external encapsulating structure organization	P	3	GO:0045229	8
secretion	P	5	GO:0046903	8
secretion by cell	P	5	GO:0032940	8
cellular polysaccharide biosynthetic process	P	6	GO:0033692	8
polysaccharide catabolic process	P	6	GO:0000272	7
cell wall	C	5	GO:0005618	6
flagellum	C	5	GO:0019861	5
exocytosis	P	6	GO:0006887	4
cell surface	C	4	GO:0009986	4
extracellular structure organization	P	3	GO:0043062	3
vesicle docking	P	4	GO:0048278	3
vesicle docking involved in exocytosis	P	5	GO:0006904	3
lipopolysaccharide biosynthetic process	P	6	GO:0009103	3
lipopolysaccharide metabolic process	P	6	GO:0008653	3
extracellular matrix organization	P	4	GO:0030198	2
extracellular polysaccharide metabolic process	P	6	GO:0046379	2
Adhesion and recognition				
receptor activity	F	4	GO:0004872	34
cell surface receptor linked signaling pathway	P	4	GO:0007166	20
G-protein coupled receptor protein signaling pathway	P	5	GO:0007186	18
cell communication	P	3	GO:0007154	13
cell adhesion	P	3	GO:0007155	4

Table 1.1: continued. Number of QPX transcripts annotated with gene ontology phrases commonly associated with virulence characteristics. F: molecular functional, P: biological process, C: cellular component

Gene Ontology Term	GO Category	GO Level	GO ID	#Seqs
cellular response to external stimulus	P	4	GO:0071496	4
cellular response to extracellular stimulus	P	4	GO:0031668	4
GPI anchor metabolic process	P	7	GO:0006505	3
GPI anchor biosynthetic process	P	8	GO:0006506	3
receptor binding	F	4	GO:0005102	3
cell-cell adhesion	P	4	GO:0016337	2
cellular response to endogenous stimulus	P	4	GO:0071495	2
cell-cell signaling	P	3	GO:0007267	1

Molecules commonly involved in protection against host immune response include enzymes that shield microbes against oxidative damage. Seventeen QPX sequences were identified to be involved in antioxidant activity (GO:0016209, **Figure 1.1B**), oxidative stress response and protection against oxidative stress (GO:0006979, GO:0034599) including molecules annotated with catalase activity (GO:0004096), superoxide dismutase activity (GO:0004784), and peroxidase activity (GO:0004601, 10 sequences), one sequence with thioredoxin peroxidase activity (GO:0008379), and two sequences with glutathione peroxidase activity (GO:0004602) (**Table 1.1**).

During infection, molecules that play the most important roles are the ones on the interface between the host cell or its extracellular matrix and the pathogen cell. In the QPX cDNA library, several transcripts were identified to be associated with the biological processes of secretion (GO:0046903) and exocytosis (GO:0006887) (**Table 1.1**). In addition, cellular component annotations revealed transcripts localized in an envelope (GO:0031975, 70 transcripts), external encapsulating structure (GO:0031975, 31) and cell surface (GO:0009986, 4) (**Table 1.1**). Eight sequences involved in cellular polysaccharide biosynthetic process (GO:0033692) and ten sequences corresponding to polysaccharide metabolic process (GO:0000271) were also identified (**Table 1.1**) and may be involved in the biosynthesis of the mucopolysaccharide secretion produced by QPX.

Other host-parasite surface interacting molecules of interest are those involved in cell recognition, cell adhesion and cell communication, and the library allowed the identification of 20 sequences related to cell surface receptor linked signaling pathways GO:0007166), 18 linked

to G-protein coupled receptor protein signaling pathways (GO:0007186), 13 transcripts involved in cell communication (GO:0007154), and 4 transcripts related to cell adhesion (GO:0007155) (**Table 1.1**).

The sequenced transcriptome also allowed the identification of a number of transcripts that may provide new insight into QPX cell biology and add to the debate on the possible existence of a sexual life stage of the parasite, including 32 sequences related to reproductive process (GO:0000003, **Figure 1.1 A**) within which 25 sequences were annotated to be involved in cell cycle (GO:0007049), 14 in cell division (GO:0051301), six in sexual reproduction (GO:0019953), two in meiosis (GO:0051321), and five annotated to the cellular component "flagellum" (GO:0019861) (**Table 1.1**).

QPX oligoarray gene expression profiles in response to temperature

The gene expression profiles of QPX cells cultivated at four different temperatures (27 °C, 23 °C, 13 °C, 10 °C) were investigated using 15K 60-mer oligonucleotide arrays. A total of 1580 transcripts were differentially expressed (DE, at least 1.5 fold in conjunction to Anova $p<0.01$) in response to temperature changes (**Figure 1.2**).

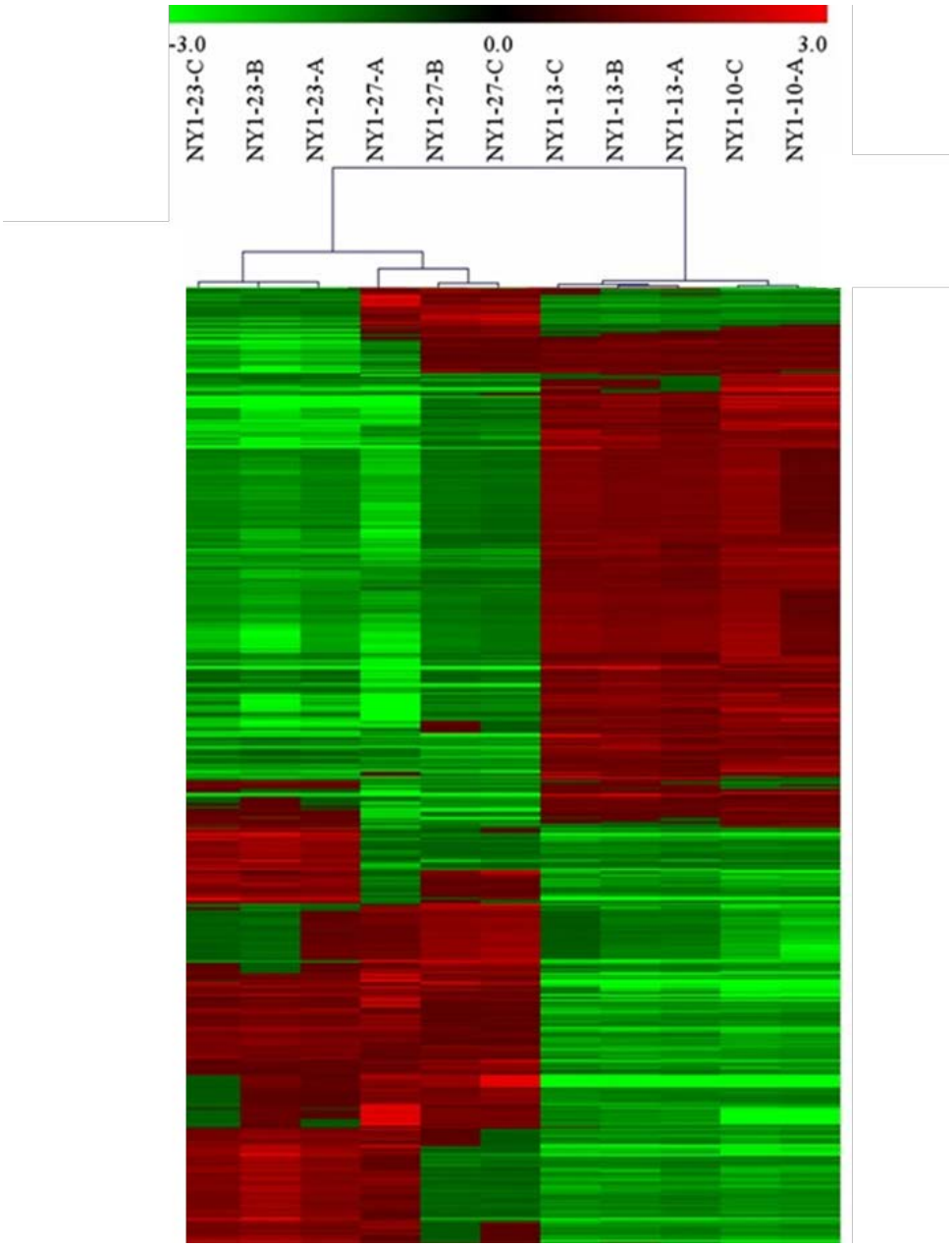


Figure 1.2: Hierarchical clustering of 1580 quahog parasite (QPX) transcripts (Pearson's correlation centered) identified to be differentially expressed at four temperature conditions by oligoarray methodology. The clustering of samples within each treatment (27 ℃, 23 ℃, 13 ℃, n=3 and 10 ℃, n=2) highlights the reproducibility of the assay

Among these, 617 DE transcripts were functionally annotated and grouped by biological and molecular function categories (**Figure 1.3**). The biological group with the highest number of differentially expressed transcripts (78 up and down regulated) was related to transmembrane transport and molecule trafficking (**Figure 1.3**). The next two groups of biological processes responsive to temperature were amino acid metabolism (74) and replication/transcription/translation (73), followed by carbohydrate metabolism (61), protein metabolism (53), response to temperature stress (49), signal transduction (44), fatty acid metabolism (36), antioxidant pathways (20) and lipid metabolism (17) categories (**Figure 1.3**).

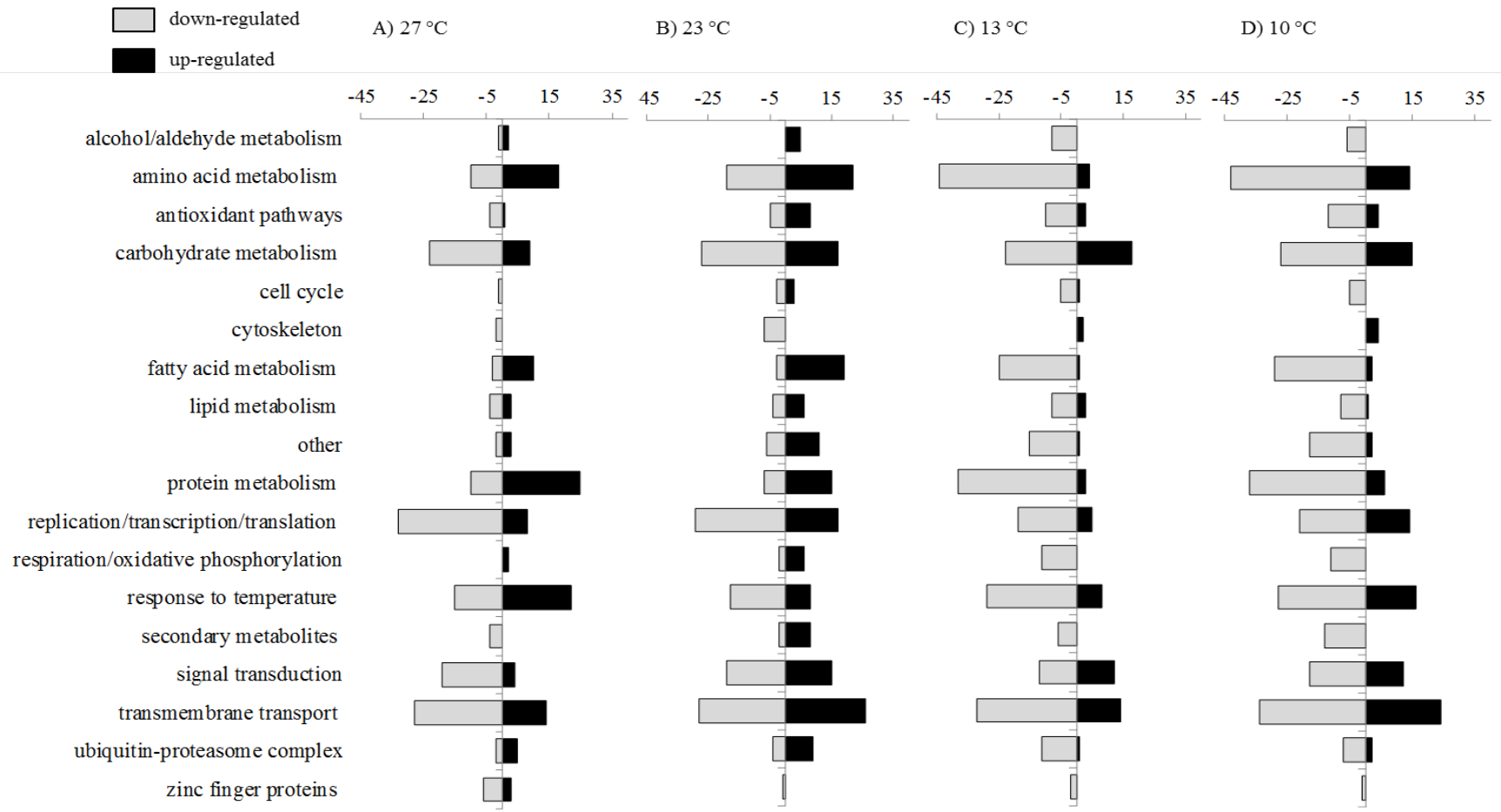


Figure 1.3: Number of annotated QPX transcripts differentially expressed (DE) in response to four temperature treatments. Transcripts are grouped by their putative biological process and molecular function into 17 operational categories

The 49 transcripts in the “response to temperature” category were aligned together for the identification of unique mRNA sequences coding for heat and cold shock related proteins. The alignments of QPX transcripts responsive to temperature changes revealed unique sequences coding for three heat shock proteins (hsp): 18, 16, and 15 kDa (**Figure 1.4**). The other “heat” shock proteins were up-regulated in the two studied cold temperatures (13 °C and 10 °C) and could be referred to as cold shock proteins. These included: three hsp of predicted molecular weight (MW) 70 kDa, three DNAj–family proteins with predicted MW of 61.5, 64.7 and 74.7 kDa, and also one hsp of around 40 kDa, one of 90 kDa and one of 101 kDa (**Figure 1.4**).

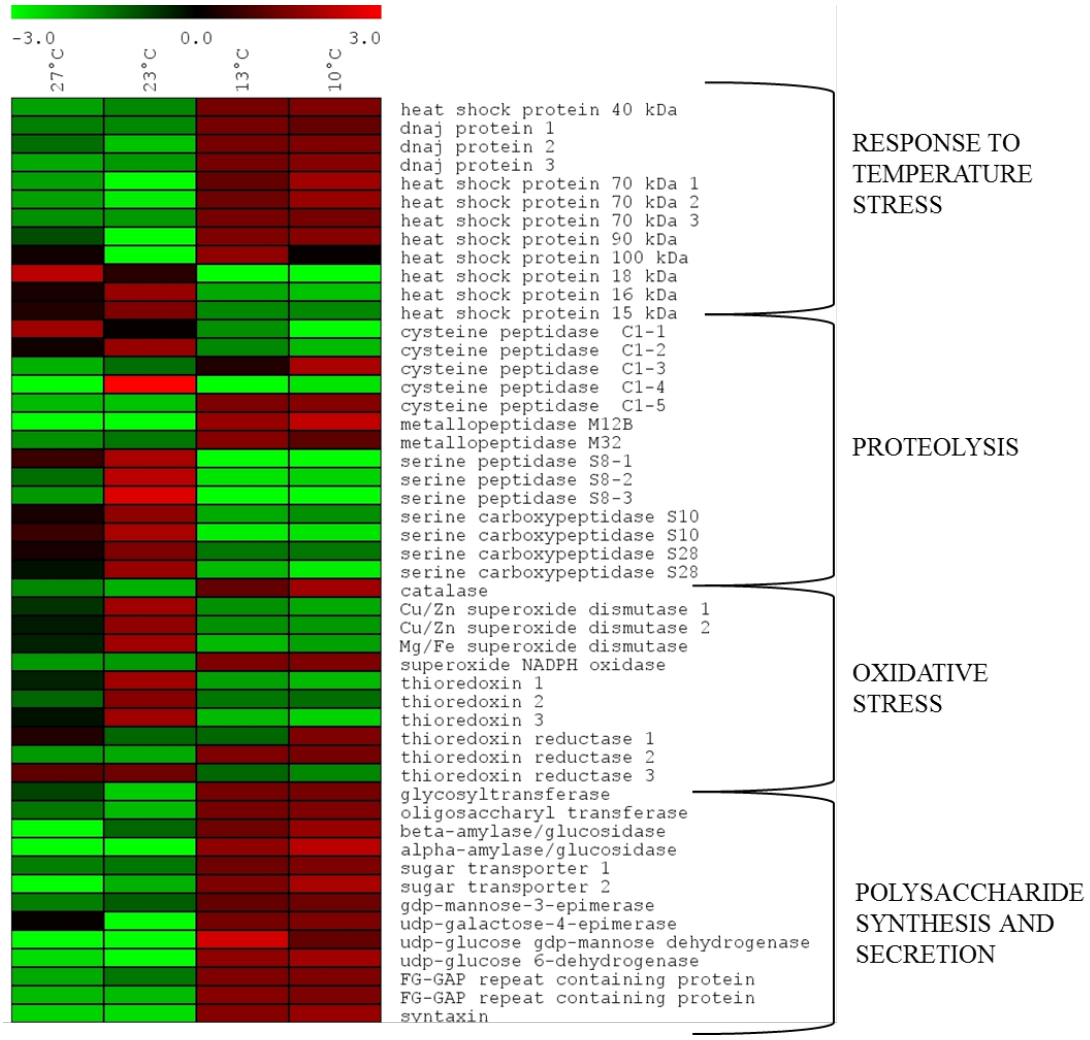


Figure 1.4: Heat map showing changes in selected transcripts levels in QPX cultures exposed to different temperatures. Each column represents the average fold change of 3 biological replicates

Within the protein metabolism category (**Figure 1.3**), only transcripts coding for protein degrading enzymes (peptidases) were investigated in detail. 14 different peptidases belonging to different families (MEROPS nomenclature) were found to be regulated with temperature changes. Seven of these peptidases were predicted to be extracellular enzymes: three subtilisin-like peptidases (S8), two serine carboxypeptidases (S10) and two papain-like cysteine peptidases (C1) (**Figure 1.4**). Two metallopeptidases (M12B and M32) were up-regulated at lower temperature (10 and 13 °C) as compared to 23 and 27 °C (**Figure 1.4**). Molecules related to oxidative stress were also shown to be responsive to temperature, including three different thioredoxins, three thioredoxin reductases, three superoxide dismutases, a catalase, and a superoxide generating oxidase (**Figure 1.4**).

In the remaining biological process categories (**Figure 1.4**), only transcripts which were up-regulated at either 13 or 10 ℃ were further investigated. A total of 76 transcripts were up-regulated at 13 ℃ (**Figure 1.4**) and 115 transcripts were up-regulated at 10 ℃. From these, several transcripts of particular interest for potential involvement in QPX pathogenesis are presented in **Figure 1.4**. These were sequences coding for: two transmembrane sugar transporters, two glycotransferases, two glycosidases, one mannose epimerase, one galactose epimerase, two UDP-glucose-GDP-mannose-dehydrogenases, syntaxin-related protein and two integrin-related proteins (**Figure 1.4**).

Real time PCR and oligo-array validation

To confirm gene expression patterns obtained by the oligoarray analysis, transcription levels of six different QPX proteases at four different temperatures were examined using quantitative real time PCR. The comparison of the relative transcript levels determined by both methods, oligoarray and qPCR, revealed very similar expression patterns at different temperature conditions (**Figure 1.5**) with a strong statistically significant correlation between the values determined by both methods (Pearson, $r=0.83$). The mRNA expression patterns of four of the six examined peptidases altered by the temperature conditions were statistically significant (ANOVA, $p<0.001$; **Figure 1.5**). The highest expression levels of four peptidases (three serine peptidases: S8-1, S8-2, S8-3 and one cysteine peptidase: C1A-2) were measured in cultures incubated at 23 ℃.

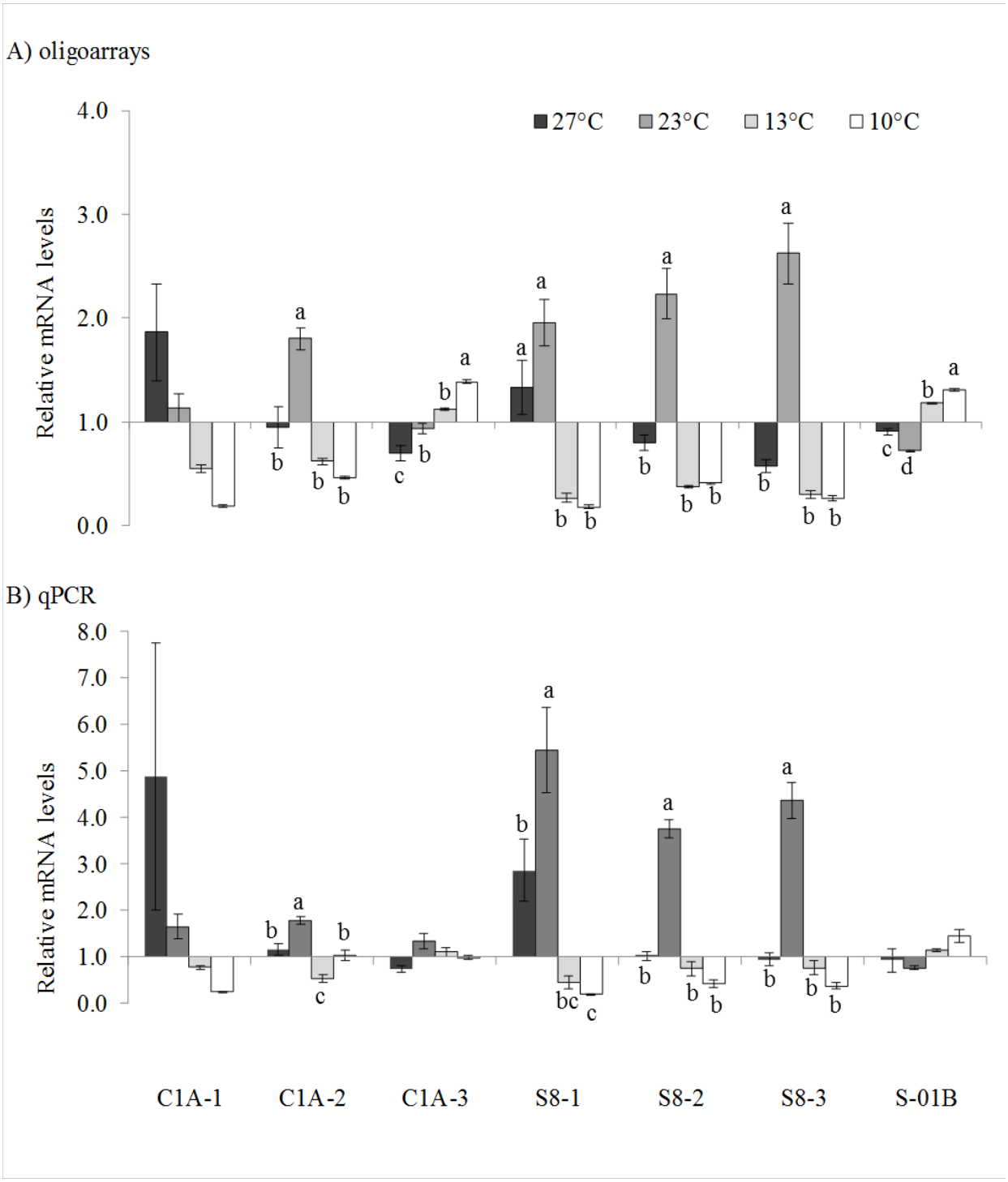


Figure 1.5: Relative mRNA expression levels (mean ± std error, n=3) of six QPX peptidases in response to four different growth-temperature conditions as determined byoligoarrays (A) and quantitative PCR (A)Significant differences (ANOVA, $p<0.01$, Tukey post-hoc test) between treatments for each gene are indicated by letter labels

Discussion

Characterization of QPX transcriptome

In the present study, we sequenced, assembled and annotated the transcriptome of quahog parasite unknown, the thraustochytrid protist which causes disease and mortality in an important aquaculture species, the hard clam *Mercenaria mercenaria*. The cDNA sequence library was generated from QPX cells (NY, Raritan Bay isolate) grown in a variety of culture media to obtain satisfactory coverage for transcripts that may have different expression profiles under various culture conditions. The gene ontology annotation of the sequences provides are source for future investigations of QPX virulence characteristics. Over 7,000 QPX transcripts matched other sequences in publicly available databases. The top four species showing the most similarity to QPX sequences were the oomycete *Phytophthora infestans*, the brown alga *Ectocarpus siliculosus*, and the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*, all of which are of the stramenopilan species, thus confirming QPX affiliation to this phylogenetic group (Tsui et al 2009).

Approximately 5,000 QPX cDNA sequences were assigned information regarding their molecular and cellular functions. The gene ontology annotation and *in silico* data mining process are initial yet critical steps in the discovery of genes involved in the molecular process of infection. Previous bioinformatics-based studies have proven to be very successful in predicting putative virulence factors, and in particular certain GO terms have been used as predictors of virulence proteins in bacteria (Chia-Ta et al 2009) and infection effector proteins in oomycetes (Torto et al 2003). In this study, we used similar approaches to search conserved GO terms and classify QPX sequences as candidate virulence factors and discuss the possible infection process inside the host clam.

Infectious microorganisms produce virulence factors that allow them to invade and colonize their hosts (Casadevall & Pirofski 1999). Likely the most destructive virulence molecules are hydrolytic enzymes produced by pathogenic organisms to destroy host-derived biopolymers (proteins, polysaccharides and lipids). QPX possesses sequences encoding peptidases, glycosidases and phospholipases, which are commonly predicted to be involved in pathogenicity (Joseph et al 2010, Torto-Alalibo et al 2005). Glycosyl hydrolases are enyzmes which digest various carbohydrates. These enzymes, such as pectinase cellulases, xylanases and glucanases, are long-studied and important virulence factors of fungal and oomycete pathogens of plants

because they can break down plant cell walls (Latijnhouwers et al 2003). Recently, chitinases from the oomycete pathogen *Aphanomyces astaci* of freshwater crayfish have been investigated as potential virulence factors (Hochwimmer et al 2009). The QPX cDNA library contains a sequence showing similarity to bacterial chitinases. Since the molluskan shell is partially composed of chitin (Addadi et al 2006, Furuhashi et al 2009) and QPX cells have been previously found on empty molluskan shells (Gast et al 2008), QPX's ability to digest the shell material cannot be excluded. Phospholipases are important enzymes utilized by many pathogenic organisms, including bacteria, fungi and protozoans, to hydrolyze cell membranes of host cells (Belaunzaran et al 2013, Calderone & Fonzi 2001, Cox et al 2001, Ghannoum 2000, Stehr et al 2003). QPX possesses sequences coding for b-like (IPR007000) and d-like (IPR001736) phospholipases which are required for virulence in some pathogens (Lev et al 2013, Li et al 2012).

Peptidases can degrade host protein material such as collagen, fibrin and actin in the connective tissue of the extracellular matrix, and are also capable of destroying immunoglobulins (Klemba & Goldberg 2002, McKerrow 1993, Olivos-Garcia et al 2009a, Sajid & McKerrow 2002b). In this study, we found over 200 transcripts annotated to encode peptidases belonging to five major peptidase types which can be distinguished following the MEROPS peptidase database (Rawlings et al 2012). Within the dataset of all peptidases, QPX was found to possess peptidases related to families commonly associated with pathogenicity. These include papain-type cysteine peptidases (family, C1A: IPR000169; IPR000668; IPR013128) which are represented by the highest number of transcripts. Papain-like peptidases, also known as cathepsins, are the dominant peptidases in pathogenic protozoans (Atkinson et al 2009). Further, QPX possesses serine peptidases in the S8 family known as subtilisins or subtilases (IPR000209; IPR015500), which are universal virulence factors found in pathogenic bacteria (Bonifait et al 2011, Hu et al 2010), protozoa (Hernandez-Romano et al 2010, Yeoh et al 2007), and fungi. QPX also possesses sequences encoding peptidases found to be virulence factors in other pathogens, including aspartic peptidases (family A1), serine peptidases (families S1, S9 and S10) and metallopeptidases (families M16, M28 and M35) (Coetzer et al 2008, Dubovenko et al 2010, Gvozdeva et al 2004, Klemba & Goldberg 2002, Monod et al 2002). Based on the sequence information, QPX is expected to use hydrolytic enzymes, including peptidases, to break down extracellular material of clam tissue to be able to propagate throughout the host body. *In situ*

observation of degraded clam tissues surrounding QPX cells (Dove et al 2004, Ragon Calvo et al 1998)supports the involvement of peptidases in the pathogenesis of this parasite.

Histopathologic examinations of infected clam tissue document an inflammatory response of clam hemocytes aggregating around QPX cells. It is also known that the hemocytes of bivalves, including *M. mercenaria*, produce reactive oxygen species (ROS) to combat invading organisms (Bugge et al 2007). To survive the clam defense, QPX would be expected to produce antioxidant molecules which can neutralize the toxic effect of ROS. Based on the transcriptome sequences investigated in this study, QPX contains five superoxide dismutases (IPR001424, IPR001189) and a catalase (IPR002226) for the detoxification of superoxide ions and hydrogen peroxide, respectively. QPX's antioxidant defense consists also of six thioredoxin-like proteins (IPR012336, IPR005746), three glutathione-dependent peroxidases (IPR000889, IPR002109), and two plant-like ascorbate peroxidases (IPR010255 IPR002207). All of these antioxidant molecules have been shown to be effective against endogenous (metabolism related) and exogenous (from host defense) ROS species and are essential for resistance to oxidative stress in many pathogenic protozoans during disease development (Olivos-Garcia et al 2009b, Osorio et al 2012). QPX antioxidant armor shows that the pathogen is prepared to overcome *M. mercenaria* hemocyte-derived ROS, and indirectly facilitate the establishment of an infection.

Another key step in the establishment of the infection site is the ability of the pathogen to recognize and to attach to the host cells or extracellular matrices. The diversity of molecules covering the surface of pathogens has been shown to be involved in their infectivity. For example, cell adhesion molecules and cell surface receptors are commonly found to participate in infection processes. Adhesive and ligand-binding molecules of plant and animal pathogens are usually glycoproteins and lipoproteins related to thrombospondin (Robold & Hardham 2005, Yuda et al 1999), integrin (Harper et al 2004, Yuda et al 1999) and lectins (Gaulin et al 2002, Gilchrist & Petri 1999, Jacobs et al 2010, Mann 2002, Stanley & Reed 2001, Vines et al 1998). The QPX transcriptome reveals a number of possible adhesive proteins, ligands and cell recognition proteins including fasciclin (IPR000782), fibronectin (IPR003961) and integrin (IPR013519) related proteins, G-protein coupled receptors (IPR001019, IPR017978), and lectins (IPR000772). These molecules are targets for future research studies on the capabilities of QPX to anchor itself to the extracellular matrix within clam tissue.

QPX putative virulence factors regulated by temperature

Because previous studies indicate that temperature is one of the environmental factors controlling QPX disease development in hard clams, we further investigated how putative virulence transcripts respond to temperature changes. The data from an ongoing 10-year New York State monitoring program of QPX prevalence and intensity in the Raritan Bay (New York) hard clam population show clear seasonality of the disease that peaks during summer (unpublished data – Marine Animal Disease Laboratory). In contrast, experimental QPX transmission studies show that low temperature conditions favor disease development which usually peaks four to five months after clams are challenged with the parasite (Dahl et al 2011, Perrigault et al 2011). It could then be hypothesized that QPX disease develops most effectively during winter and spring seasons as cooler temperatures suppress the clam's immune system (Perrigault et al 2011), allowing the infection process to occur followed by severe disease outbreaks during summer months. It should be noted that different virulence molecules may play different roles during the different stages of the infection process and disease progression. For example, the mucus polysaccharide secretion could be more important during the establishment of the infection foci and provide protection against clam immune response, while the extracellular peptidase can be highly expressed during later stages of infection and may be involved in extensive host tissue degradation.

The current study reveals a number of putative QPX virulence molecules regulated in response to temperature. Cysteine peptidases belonging to the cathepsin-like subfamily (C1A) have been shown to be involved in many protozoan and metazoan diseases and are some of the most extensively studied virulence factors (Atkinson et al 2009, Klemba & Goldberg 2002, Sajid & McKerrow 2002). They play a crucial role in parasite biology, including growth, development and replication. They are also implicated in host tissue degradation, including digestion of host extracellular matrix proteins and destruction of host immune-related proteins such as immunoglobulins (Lecaille et al 2002). One QPX cathepsin-like peptidase was up-regulated at 27 °C, two were up-regulated at 23 °C, and two were up-regulated at 13 °C and 10 °C, suggesting different roles for these genes in basic metabolism and/or pathogenesis of the parasite. Further specific experimental studies are needed to characterize the biochemical function and role of each cathepsin in QPX biology and virulence abilities. Other QPX peptidases for which expression was shown to be regulated by temperature were subtilases. The amino acid sequence alignments of QPX transcripts annotated as belonging to the S8 family revealed that QPX

possesses at least 6 different subtilisin-like peptidases. Only three subtilases were shown to be differentially expressed in response to temperature, with the highest expression measured at 23 °C, which corresponds to QPX's optimal growth temperature. QPX is a saprophytic microorganism which needs to degrade nutrient proteins extra-cellularly and transport oligopeptides and amino acids into the cell via transporter assisted diffusion. QPX maximal growth at 23 °C could then explain the need for an efficient extracellular protein degradation and nutrient acquisition system associated with high expression of its digestive enzymes. Subtilases, however are also universally occurring secreted enzymes found in many medically important pathogenic microorganisms, including bacteria e.g. *Mycobacterium*, *Streptococcus* (Bonifait et al 2011, Brown et al 2000, Coutte et al 2001, Hu et al 2010, Kennan et al 2010), protozoa e.g. *Plasmodium*, *Leishmania* and *Toxoplasma* (Lagal et al 2010, Liao et al 2010, Swenerton et al 2010, Yeoh et al 2007), and fungi e.g. *Asperigillus* spp. (Muszewska et al 2011)). Therefore, the role of subtilisins as QPX virulence factors might also be a possibility, but these enzymes may be important for host tissue digestion and disease progression during advanced and late stages of infection when overcoming the host immune system is less important and degradation of the tissue and pathogen proliferation takes over.

Four different QPX metallopeptidases, classified in the MEROPS peptidase families M12B, M16, M32 and M41, were up-regulated at 13 °C and 10 °C and also require special attention. The M12B peptidases (also known as ADAMs - A Dis-integrin And Metalloprotease or adamalysin family) are zinc-dependent endopeptidases (IPR001590). The same QPX peptidase was also found to be regulated with temperature in another study (Garcia-Vedrenne et al 2013). In animals, ADAMs are membrane-anchored glycosylated enzymes capable of degrading proteins from the surface of cells, thus playing important roles in cell adhesion, signaling, cell-cell fusion, and cell-cell interactions (Seals & Courtneidge 2003, van Goor et al 2009). They possess two different protein domains in addition to the metalloprotease domain, including an integrin domain (with adhesion and receptor activities) and a cytosolic domain which provides an attachment for various signal transduction proteins (Seals & Courtneidge 2003). Cell-surface glycol-metallo-proteases play an important role in the pathogenesis of trypanosomatids, as they allow these parasites to adhere and move through connective tissue of their hosts (Masini d'Avila-Levy et al., 2006). The specific role, if any, of adamysin-like protease in the virulence of QPX requires additional studies.

The QPX M32 enzyme is a carboxy-Taq-metallopeptidase (IPR001333). The first zinc-containing thermostable metallopeptidase was originally discovered and purified from *Thermus aquaticus* and had an optimal enzymatic activity at 80 °C (Lee et al 1992). Most bacterial and archaebacterial species are known to possess M32 carobxypeptidases, but they are considered absent in most eukaryotic genomes except for a few protozoan species (Niemirowicz et al 2007). Their biological function in these organisms remains unknown. However, the basic characteristics of two recombinant M32-peptidases of *Trypanosoma cruzi* have been investigated (Niemirowicz et al 2007). They were shown not to behave as thermostable enzymes, with their activity significantly decreasing with increased temperature. The current study is the first report of a M32 enzyme in a thraustochytrid species for which activity increases at lower temperatures. This possibly cold resistant QPX's metallopeptidase is a very interesting target for investigation of its biological function and role in the infection process.

QPX cells release a gel-like muco-filamentous secretion which appears to protect QPX from clam-hemocyte phagocytosis and encapsulation as seen during histo-pathologic observation of infected clam tissues (Dahl & Allam 2007, Dahl et al 2008, Dahl et al 2010, Smolowitz et al 1998a). In addition, QPX mucus causes necrotic response in clam hemocytes *in vitro* (Perrigault & Allam 2009) and provides protection against clam humoral defense (Anderson et al 2003a). The consistency of QPX mucus secretion changes with different temperatures, from a gel with low viscosity at 27 °C to a very rigid gel at 10 °C (personal observation), thus it is surmised that cold temperatures would benefit the parasite by providing stronger protection during the infection process. Thicker mucus secretion at lower temperature was associated with up-regulation of GDP-mannose dehydrogenase in QPX cells from cold temperatures. This enzyme essentially catalyzes irreversible conversion of GDP-mannose to GDP-mannuronic acid and ultimately leads to the biosynthesis of alginates in *Pseudomonas aeruginosa* (Snook et al 2003) and in the brown alga *Ectocarpus siliculosus* (Tenhaken et al 2011). In addition, the expression of two enzymes involved in the production of mannose (gdp-mannose-3- epimerase) and galactose (udp-galactose-4-epimerase) were also up-regulated at the two colder water temperatures. Mannose is the most important monosaccharide required for the production of mannuronic acid, which is the major component of alginates (Rehm 2009). Interestingly, the mRNA levels of several different transcripts coding for UDP-glucose/GDP-mannose dehydrogenase in QPX were up-regulated at the two cold temperatures.

Further, there were also a few other molecules in the carbohydrate metabolism and vesicular secretion pathways that were induced at 10 and 13 °C. These include two glycosyltransferases (IPR007577, IPR0036 or enzymes that transfer and add a range of different sugars to other sugars, phosphates and proteins, thus participating in glycosylation of proteins and in synthesis of polysaccharides and glycoconjugates (Breton & Imberty 1999). If the up-regulated glycosyltransferases are involved in the glycosylation process of QPX extracellular mucopolysaccharides, it can provide additional support for the hypothesis of considerable changes to the QPX mucus structure at cold temperature. In addition, two sugar transporters or sugar permeases that are carriers of dissolved simple sugars across membranes in response to chemosmotic gradient (Pao et al 1998) (IPR003663, PTHR19432), and a syntaxin-type protein (IPR006011) that is a membrane protein participating in vesicular membrane fusion during exocytosis (Dacks & Doolittle 2004) , were over-expressed at the colder temperatures. Again, it can be speculated that these molecules participate in mucus polymer production and/or rearrangements. Even though the composition and structure of QPX mucus is still to be determined, these findings identify specific candidates that may be involved in mucus biosynthesis and provide a plausible scenario for the regulation of mucus consistency at different temperatures. The present hypothesis that QPX mucus production is increased in cold conditions can also be supported by the up-regulation of several molecules in carbohydrate synthesis and secretion pathways.

QPX acclimation to temperature changes

One of the major biochemical mechanism by which organisms have evolved to adapt to stress environmental conditions, including temperature shifts , is maintenance of protein stability and refolding, which is accomplished by a set of molecules referred to as heat shock proteins (hsp). Temperature changes lead to changes in protein configuration and their potential malfunction. Hsp are molecular chaperones that control aggregation of denatured protein in the cell, and are responsible for their reconfiguration, translocation across membranes and degradation. QPX optimal growth temperature is between 20 °C and 23 °C, and growth decreases significantly below and above that range (Perrigault et al 2010). In the current study, the QPX transcriptome expression profiles under to different temperature conditions revealed molecular changes occurring at suboptimal temperatures. QPX was found to possess a number of molecular

chaperones which were up-regulated at different temperatures, both below and above its optimal growth temperature, reflecting heat and cold temperature stress.

The amino acid sequence alignments of theses transcripts show that QPX has at least four different heat shock proteins which can be classified into a group of molecular chaperones of low monomeric molecular mass, ranging from 12 to 43 kDa (de Jong et al., 1998). These low molecular weight chaperones were up-regulated at the elevated temperature of 27 °C. This is contrary to the results of a recent study on QPX gene expression changes with temperature which found these molecules to be overexpressed at the low temperature of 10 °C (Garcia-Vedrenne et al., 2013). A possible explanation for this discrepancy is that the studies had different experimental designs to measure gene expression changes due to temperature shifts. While the study of Garcia-Vedrenne et al. shows gene expression changes in QPX cells after short exposure to different temperatures, the current study was aimed at understanding QPX physiological changes and expression of putative virulence factors due to different temperatures during QPX growth. This design was chosen to reflect seasonal changes experienced by QPX cells in the environment and to better mirror the effect of temperature on this chronic disease. The transcription expression levels of these small molecular weight chaperones have been commonly investigated in pathogenic protozoans affecting humans. Their mRNA levels are usually up-regulated during heat stress related to fever. For example, the mRNA expression of heat shock protein 16 from *Trypanosoma cruzi*, the human pathogen of Chagas disease, is significantly up-regulated at 42ºC (fever conditions) as compared to 37° C (control). The recombinant hsp16 of *T. cruzi* prevents aggregation of denatured proteins (Perez-Morales et al 2009). Similarly, *Toxoplasma gondii* which possesses five different small heat shock proteins (Hsp20, Hsp21, Hsp28, Hsp29 and Hsp30) shows increased expression of these proteins upon heat stress conditions (de Miguel et al 2009).

Similar to the study of Garcia-Vedrenne et al (2013), we found a number of QPX molecular chaperones for which mRNA levels increased at cold temperatures. The expression of three different hsp70, three dnaj family hsp, one hsp40, one hsp90 and one hsp101 were up-regulated at both 13 °C and 10 °C. Heat shock proteins 70 and 90 are commonly investigated as genetic markers of environmental stress conditions, including elevated temperatures, exposure to toxic chemicals and pollutants. They are less commonly investigated as cold temperature shock biological markers; however some studies do indicate they function during cold stress. For

example, one of several hsp70 in *Escherichia coli* functions as a cold shock molecule (Lelivelt & Kawula 1995). The parasitic nematode larvae *Trichinella sp.* increases the expression of hsp70 during temperature shift from 37 °C to 42 °C (Martinez et al 2001). Similarly, the levels of hsp70 mRNA from an oomycete, *Bremia lactucae*, increase in response to both heat and cold treatment (Judelson & Michelmore 1989). The dnaj proteins are co-chaperones of hsp70 and contain a conserved amino acid J domain which binds to hsp70 and stabilizes the interaction between the substrate and hsp70 (Qiu et al 2006). The up-regulation of the above hsps in QPX cells at low temperature highlights a concerted response to cold stress and underlines the ability of the parasite to cope with suboptimal temperatures.

The response of ectothermic organisms to temperature changes is usually reflected in adjustment of their metabolic rates such that their metabolism slows down during cold temperatures and during wintering (Blagojevic et al 2011). The slower metabolic rate reduces oxygen consumption and subsequently causes production of reactive oxygen species as there is an oxygen surplus from respiration (Blagojevic et al 2011). To cope with the oxidative stress ectothermic organisms have evolved cold-survival mechanisms involving many antioxidant molecules. The differential expression of QPX antioxidant molecules in response to temperature could be explained by the adjustment of the metabolic rate of the parasite. At 13 °C and 10 °C, a total of 159 and 163 transcripts, respectively, related to metabolism had suppressed expression, and in response, the mitochondrial membrane bound enzyme, superoxide NADH generating oxidase and the cytosolic catalase (hydrogen peroxide catalyzing enzyme) were also up-regulated. Further QPX's molecules of the thioredoxin system, which are responsible for maintaining reduced conditions inside the cells, were also up regulated during the cold exposure.

The gene expression profile of QPX cells at 23 °C suggests that the fast growing parasite cells also suffer from oxidative stress. Slight oxidative stress, characterized by accumulation of reactive oxygen species (ROS), happens in every cell as a result of imbalance between oxygen availability and oxygen consumption from respiration. If a small amount of ROS, specifically superoxide anion, is produced during the normal respiration process as a simple by-product, an increased respiration rate would cause increased production of ROS. Seven QPX transcripts related to respiration and oxidative phosphorylation were also induced at 23 °C, suggesting an increased respiration rate. QPX possesses six different sequences encoding superoxide dismutase enzymes (SODs, data not shown), including four copper/zinc related SODs which are usually

found in the cytosol of a cell, and two iron/manganese SODs which are usually found in mitochondria. The expression of only three QPX SODs (two Cu/Zn SODs and one Fe/Mn SOD) was up-regulated at 23 °C, suggesting that these three SODs are the main enzymes expressed in QPX cells to cope with the accumulation of endogenous ROS during the normal respiration process. Any potential role of the six QPX SOD enzymes in dismutation of exogenous superoxide such as those produced by clam hemocytes would have to be investigated in a different study. But this study shows that QPX possesses molecular protection machinery to stabilize its proteins from oxidation at the range of temperatures at which the infection process starts, develops and progresses. In other words, this parasite seems to be better equipped to cope with environmental stress than the clam host, allowing it to cause infection and disease under conditions (low temperature) that may be suboptimal for both the host (Perrigault & Allam 2012, Perrigault et al 2012, Perrigault et al 2011) and the parasite (Perrigault et al 2010, this study).

Chapter 2: Effect of culture conditions on QPX virulence

Manuscript in preparation for Journal of Invertebrate Pathology

Abstract

Previous investigations showed increased of QPX cytotoxicity following its growth in clam tissue homogenates. The current study was designed to evaluate the effect of exposure to clam-factors on the transcription level of putative virulence genes. Transcriptome profiles were generated (oligoarray technology) for QPX cells cultivated in either the minimum essential media supplemented with fetal bovine serum or clam adductor muscle homogenate. A total of 1,623 transcripts were differentially expressed between the two culture media. Among these, 207 functionally annotated genes were found to be up-regulated in parasite cells cultivated in the clam adductor muscle homogenate. Most of the up-regulated transcripts were related to stress response and cell detoxification revealing QPX acclimation to a hostile environment created by host factors. Genes commonly associated with virulence such as subtilisin-like peptidases, papain-like peptidases, and phospholipases were down regulated, suggesting that QPX exposed to clam-factors did not enhance its virulence. To further test the effect of clam factors on QPX virulence, parasite cells cultivated in the two media were injected into clams in an in vivo challenge experiment. Three months after QPX injection QPX was detected in clam tissues by quantitative PCR (qPCR) even though there was no clam mortality or signs of disease by histopathology, which suggests that QPX cells might be able to survive inside clam tissues without causing overt disease. There was no difference in QPX cell counts between clams exposed to QPX cells derived from standard media or parasite cells grown in adductor muscle homogenate. In parallel to the media exposure experiments, the virulence of two QPX isolates, one from 2003 and another from 2012, was also compared. The 2012 QPX isolate was detected at higher intensities than the 2003 isolate, suggesting that the prolonged subculture conditions reduce the virulence of the parasite.

Introduction

The production of virulence factors in pathogens is generally stimulated by contact with host cells or extracellular matrix. Contact of pathogens with host tissue components (host factors) triggers large scale transcriptomic changes which can be used to identify genes involved in

virulence and other host-pathogen interactions (Gobert et al 2005). Transcriptome-wide expression changes have been investigated in some medically important pathogens. For example, microarray studies of gene expression in *Entamoeba histolytica* during its *in vitro* interactions with human epithelial cell lines, human mucin and collagen revealed the transcriptional regulation of molecules which could explain its pathogenicity (Debnath et al 2004, Debnath et al 2007). Similarly, an *in vitro* time series study of gene expression profiles of *Giardia duodenalis* cells after contact with human intestinal epithelium cells revealed genes which might be important for its infectivity (Ma'ayeh & Brook-Carter 2012, Ringqvist et al 2011).

It is also well documented that the virulence of pathogenic organisms diminishes during long subculture conditions. For example, obligate pathogens of humans and other mammals such as *Leishmania donovani* or *E. histolytica* have to be re-injected on a regular basis into laboratory animals and re-collected, or the media for their sub-culturing must be supplemented with host derived tissue or serum in order to preserve their pathogenicity (Giannini 1974, Lushbaugh et al 1978). Similar effects of virulence attenuation during prolonged subculture have been documented for facultative environmental opportunistic pathogens such as entomogenous fungi, and their virulence can be restored by their passage through an appropriate host species (Shah et al 2007). A few studies on the restoration of virulence by *in vivo* passages have been conducted on shellfish pathogens (e.g. the oyster pathogen *Perkinsus marinus*). Ford et al (2002) showed that freshly isolated *P. marinus* cells are much more virulent than those propagated in culture. It has been also shown that supplementation of *P. marinus* cultures with host plasma, tissue homogenates (Earnhart et al 2004) and pallial mucus (Pales Espinosa et al 2013) increases its infectivity. Interestingly, a substantial number of virulence-related transcripts are differentially expressed when *P. marinus* cells are cultivated in media supplemented with oyster pallial mucus (Pales Espinosa et al 2014). On the contrary, the ciliate facultative parasite of lobsters *Anophryoides haemophila* shows no change in expression of common virulence genes when it is cultivated in media containing lobster serum (Acorn et al 2011). The latter study measured expression of a few selected genes so it might be omitting other transcriptional changes. Acorn et al. 2011suggested that the expression of virulence factors in response to host media might be dependent on the obligate versus facultative parasitic life strategy.

In this study, it was hypothesized that host factors modulate virulence characteristics of QPX (quahog parasite unknown), the opportunistic extracellular parasite of the hard clam,

Mercenaria mercenaria. Perrigault et al. (2009) showed that homogenates of clam adductor muscle enhance QPX growth as compared to media supplemented with bovine serum albumin at the same protein concentration. In addition, the QPX extracellular products (ECP) derived from cultures supplemented with clam homogenates had stronger cytotoxic effects on clam hemocytes than the QPX ECP derived from standard media (Perrigault et al 2009). In this study, we used QPX cells propagated in standard culture media and in clam tissue homogenate to generate gene expression profiles of the parasite after contact with host factors to examine potential regulation of QPX putative virulence factors. Experimental transmission experiments were conducted to determine if the virulence of QPX *in vivo* increases with exposure to clam-adductor muscle homogenate. In addition, transmission experiments were conducted to compare virulence between freshly isolated QPX cells and QPX cells kept in subculture conditions for approximately 5 years.

Materials and Methods

QPX cultures and RNA extraction

QPX isolate NY0313808BC7 (Qian et al 2007, NY1-03 in this study) was used in this experiment. This strain was nitially isolated from a diseased clam in 2003, cryopreserved and continuously subcultured since 2007 using minimal essential medium (MEM, Sigma) supplemented with salts and 10% of fetal bovine serum (MEM-FBS or "standard media") (Kleinschuster et al 1998). For the experiment, parasite cells were washed with filtered artificial seawater (FASW) then subcultured in two different media: 1) MEM-FBS, and 2) clam adductor muscle homogenate adjusted to 2 mg/ml of protein with FASW (Perrigault et al 2009). The 7-ml cultures were prepared in 25-ml culture flasks (Falcon) in triplicate for each treatment and grown at 23°C for seven days. After seven days, the cultures were diluted with an equal volume of FASW and passed several times through a syringe to facilitate liquefaction of QPX mucus secretion. The mixtures were transferred into 15-ml conical tubes and centrifuged at 3000 g for 40 minutes at 4 °C. The supernatant was discarded and cell pellets were collected and kept on ice for immediate RNA extraction. Trizol reagent (Molecular Research Center, Inc.) and the manufacturer’s protocol were used to isolate RNA from all samples and RNA quality and quantity were estimated spectrophotometrically by Nandrop and Agilent Bioanalyzer 2100.

Oligoarray hybridization and data analysis

The oligoarray platform and protocols used in this study were described earlier (Rubin et al., 2014; see Chapter 1). Briefly, previously generated transcriptomic data were used for the production of a 8×15 k 60-mer oligonucleotide array using the Agilent eArray application (https://earray.chem.agilent.com/earray/). These included 6,781 curated annotated sequences and 8,297 non-annotated sequences (Rubin et al., 2014). Probes (1 probe/sequence) were synthesized in situ along with positive and negative controls using 8x15K-feature Agilent format slides. Labeled (Cy3 or Cy5) complementary RNA (cRNA) was synthesized from 150 ng of RNA purified from cultures grown in the 2 experimental media using the Two-Color Microarray-Based Gene Expression Analysis Protocol (Quick Amp Labeling) following the manufacturer's protocol. Labeled cRNA was purified using Illustra CyScribe GFX Purification Kit (GE Healthcare) and used for array hybridization (300 ng of each Cy3- and Cy5-labeled cRNA). Arrays hybridization and scanning were conducted as previously described (Rubin et al., 2014; Chapter 1). Fluorescence data were normalized using LIMMA package in R software to remove within-array (method: global lowess) and between-array (method: quantile) non-biological variation (Smyth & Speed 2003). The intensities which were less than twofold of the background intensities were eliminated from further analysis. Hierarchical clustering of samples and genes (Pearson correlation) and the determination of statistically significant differentially expressed genes (t-test, p-value < 0.01) were completed in the TM4-suite using the MeV program (Saeed et al 2003, Saeed et al 2006). The final criteria for differential gene expression were the significance by t-test analysis and a one and one-half fold increase (up-regulation) or one and one-half fold decrease (down-regulation) from the mean as previously described (Rubin et al., 2014). The differentially expressed transcripts (DE) were functionally annotated using the online bioinformatics tool Blast2GO (Gotz, 2008).

QPX infection experiments

Two different QPX isolates were used in the experiment. Both QPX isolates originated from diseased clams from Raritan Bay, New York, however one QPX isolate had been initially isolated in 2003 (NY0313808BC7,Qian et al., 2007) and continuously sub-cultured in MEM-FBS since 2007 (see above; here designated NY1-03) and the other was freshly isolated in 2012 (NY-12) and continuously grown in either MEM FBS or clam adductor muscle homogenate both

adjusted to a total protein concentration of 2 mg/ml. For the infection experiments, both isolates were propagated in MEM-FBS and clam adductor muscle homogenate (2 mg protein/ml). In addition, NY-12 initially grown in standard media was subcultured into standard media supplemented with 0.3 mg clam pallial mucus proteins/ml (mucus covering gills and mantle) because these secretions were shown to enhance virulence of another parasite of bivalves (Pales Espinosa et al 2013). All cultures were established in 50-ml culture flask (353108, BD Falcon) containing 10 ml of sterile media plus with QPX inoculla of 300 µl of a parasite suspension established in seawater at a concentration of 2×10^6 cell×ml^{-1}. After seven days of growth, QPX cell concentration in each flask was estimated using a hemocytometer, and adjusted to 5×10^6 cell×ml^{-1} using appropriate media to. Clams, *M. mercenaria* var. *notata*, approximately 50 mm in length, were obtained from the Senator George Kirkpatrick Marine Lab in Cedar Key, FL. Clams were kept in 150-L re-circulating seawater (28 -30 ppt) in tanks and gradually (about 2°C per day) acclimated from 20°C down to 13°C. The low temperature condition was selected for the experiment because it has been previously shown to favor QPX disease development (Dahl et al 2011). The previously established method of injection of QPX cell suspension (here 100 µl per clam) into the pericardial cavity was used to induce QPX disease in clams (Dahl & Allam 2007). QPX cells from the five cultures described above and three negative controls (corresponding to sterile media) were injected into 26 clams per treatment. Clams were distributed into ten 25-L tanks resulting in 20 or 21 clams per tank in a common garden experimental design. After 3 months of no recorded clam mortality, 12 clams from the five treatments were sampled for histopathology and qPCR analysis was conducted using procedures previously described (Dahl & Allam 2007, Liu et al 2009).

Results

Trancriptomic changes

The current study used microarray gene expression technology to investigate transcriptional changes in QPX cells grown in clam-adductor muscle homogenate (CAMH) in comparison to QPX cells cultivated in minimal essential medium supplemented with fetal bovine serum (MEM FBS). A total of 1623 transcripts were found to be differentially expressed (DE) using t-test ($p<0.01$) in conjunction with one and one-half fold increase or decrease in expression. A total of 625 (38%) transcripts were up-regulated and 998 (61%) were down-

regulated in QPX cells derived from CAMH as compared to cells propagated in MEM FBS. A total of 774 functionally annotated transcripts were further analyzed and distributed into biological process and molecular function categories based on gene ontology (GO) annotation (**Figure 2.1**). The annotated DE transcripts were disproportionally distributed between the up-regulated and down-regulated groups. Only 209 (27%) DE and annotated transcripts belonged to the up regulated group while 565 (73%) DE and annotated transcripts belonged to the down regulated group. The GO classification of DE transcripts revealed that the highest number of DE transcripts were associated with basic cell processes such as metabolic process, replication, transcription, and translation (a total of 240 DE transcripts, **Figure 2.1**). A total of 70 DE transcripts were related to the transport of molecules (transmembrane transport and vesicle-mediated transport in **Figure 2.1**). An additional 63 DE transcripts were related to cell communication and response to external environment (signal transduction, response to the external stimuli, and membrane associated proteins in **Figure 2.1**). Another 60 DE transcripts were associated with biosynthetic process carried out by the cell (**Figure 2.1**). A significant number (37 in total) of transcripts coding for ubiquitins and other proteins which constitute proteasomes were also differentially expressed (**Figure 2.1**). A total of 35 DE sequences were coding for proteins containing "zinc finger domain" (**Figure 2.1**). A small number of DE transcripts, 14 in total, encoded constituents of cytoskeleton and proteins involved in cytoskeleton rearrangement such as regulation of actin filament polymerization (**Figure 2.1)**. Transcripts which were not annotated with a biological process category were grouped by molecular function terminology, including peptidase activity (42 DE transcripts), hydrolase activity (29 DE transcripts), transferase activity (45 DE transcripts), oxidoreductase activity (43 DE transcripts), kinase activity (7 DE transcript), and protein folding (10 DE transcripts) (**Figure 2.1**). Finally, all other DE transcripts with GO annotations that did belong to any of the above functional categories were classified as "other" (79 DE, **Figure 2.1**).

Further, all DE transcripts were screened for their similarity to virulence factors of other pathogenic organisms. A few candidate virulence factors were up-regulated in QPX cells from CAMH. These included: two phosphatases, three papain-like cysteine peptidases (C1-1, C1-2, C1-3), a glycosyl hydrolase, and three lipases (**Table 2.1**). However, other putative virulence-related genes such as extracellular subtilisins (S8-1, S8-2), three different papain-like peptidases (C1-4, C1-5, C1-6) and two phospholipases were down regulated in QPX cells from CAMH

(**Table 2.1**). **Table 2.1** also includes selected transcripts coding for proteins known to be associated with stress response (four heat shock proteins, four zinc finger domain containing proteins), oxidative stress response (glutathione peroxidase and iron-cluster protein) and detoxification (cytochrome P450, cytochrome b5). Proteins related to signal transduction , cell recognition, and transport across the cell membrane which were up-regulated in CAMH were also included in Table 2.1 as a possible list of QPX cell surface molecules possibly involved in host-parasite interactions.

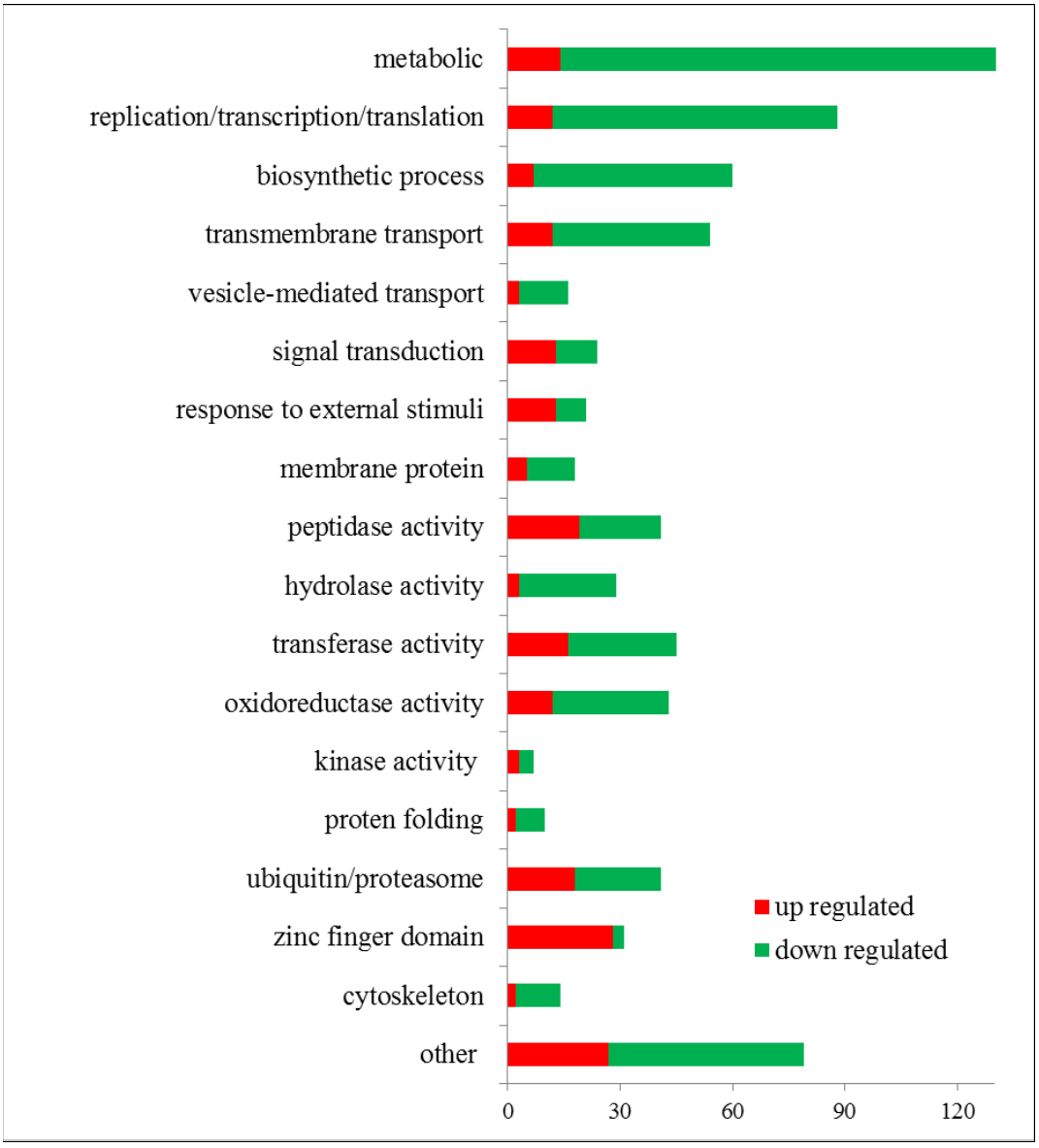


Figure 2.1: Number of up- and down- regulated transcripts in QPX cells cultivated in clam adductor muscle homogenate as compared to QPX cells cultivated in standard media; Transcripts are grouped into biological and molecular function categories based on gene ontology terminology or conserved protein domain

Table 2.1: Selected QPX transcripts up regulated in response to clam-derived factors

Sequence ID	Sequence description	Mean fold	Top BLAST result			InterProScan Results	
			Hit ACC	E-Value	Organism		
Proteolysis and hydrolysis							
qpx_c187	peptidase C1-1	2.1	XP_002507788	1.6E-75	*Micromonas sp.*	IPR000668	IPR013128
qpx_c2129	peptidase C1-2	2.5	CAX70999	2.5E-28	*Schistosoma japonicum*	IPR000668	IPR013128
qpx_c5487	peptidase C1-3	1.5	CBJ28832	1.93E-29	*Ectocarpus siliculosus*	IPR000169	IPR000668
qpx_c10523	peptidase C1-4	-1.8	EFN68284	2.32E-22	*Camponotus floridanus*	IPR000668	IPR013128
qpx_c3110	peptidase C1-5	-1.9	EAY93080	3.82E-14	*Oryza sativa*	IPR000169	IPR000668
qpx_c8710	peptidase C1-6	-1.6	XP_417483	5.22E-11	*Gallus gallus*	IPR013128	
qpx_c8663	OUT-like peptidase	-2.2	EEE25301	8.37E-09	*Toxoplasma gondii*		
qpx_c765	peptidase S8-1	-3.4	ABI79453	6.30E-30	*quahog parasite QPX*	IPR000209	IPR015500
qpx_c1822	peptidase S8-2	-2.1	ABI79453	1.93E-29	*quahog parasite QPX*	IPR000209	IPR001969
qpx_c6757	peptidase S10-1	-1.7	XP_002600751	4.93E-18	*Branchiostoma floridae*	IPR001563	
qpx_c2285	peptidase S10-2	-1.8	XP_002876119	4.99E-14	*Arabidopsis lyrata*	IPR001563	
qpx_c8079	peptidase S10-3	-4.3	XP_001747631	1.32E-06	*Monosiga brevicollis*	IPR001563	
qpx_c2535	peptidase S28	-2.4	XP_001748602	2.06E-41	*Monosiga brevicollis*	IPR008758	
qpx_c757	tripeptidyl peptidase 1	-1.8	EFA82993	1.19E-06	*Polysphondylium pallidum*	IPR000209	
qpx_c1075	tripeptidyl peptidase 2	-1.7	XP_001751904	5.89E-04	*Physcomitrella patens*		
qpx_c2515	acid phosphatase	1.7	XP_003387473	9.7E-59	*Amphimedon queenslandica*	IPR004843	
qpx_c8121	dual specif. phosphatase	1.5	ETV94297	3.0E-20	*Aphanomyces invadans*	IPR000340	IPR000387
qpx_c1349	phytanic acid oxidase	2.0	XP_004451758	1.4E-18	*Dasypus novemcinctus*	IPR008775	
qpx_lrc1228	phospholipase b-like	-1.6	XP_001745450	2.89E-54	*Monosiga brevicollis MX1*	IPR007000	
qpx_c13569	phospholipase d	-1.6	XP_002530642	2.10E-12	*Ricinus communis*	IPR001736	IPR015679
qpx_c12855	glycosyl hydrolase	1.6	ETW08658	5.0E-19	*Aphanomyces invadans*	IPR001360	IPR013781
qpx_c12309	lipase 1	2.4	EGB11221	4.9E-12	*Aureococcus anophagefferens*		
qpx_c4287	lipase 2	1.6	XP_004363103	1.3E-10	*Dictyostelium fasciculatum*		
qpx_c4756	lipase 3	2.9	XP_004992112	1.5E-22	*Salpingoeca rosetta*		
Stress-response and detoxification							
qpx_lrc6095	heat shock protein 20	2.7	YP_001997882	6.8E-15	*Chlorobaculum parvum*	IPR002068	IPR008978
qpx_c844	heat shock protein 40	2.4	GAA34756	2.8E-20	*Clonorchis sinensis*	IPR001623	IPR003095
qpx_c313	heat shock protein 70	2.0	AAU94662	9.2E-13	*Nuclearia simplex*	IPR013126	
qpx_c14576	heat shock protein 90	1.9	AAX10943	7.3E-08	*Mallomonas rasilis*		

Table 2.1 continued: Selected QPX transcripts up regulated in response to clam-derived factors

Sequence ID	Sequence description	Mean fold	Top BLAST result			InterProScan Results	
			Hit ACC	E-Value	Organism		
qpx_c1420	Zinc finger domain protein 1	2.3	XP_001765979	8.94E-04	*Physcomitrella patens*	IPR001841	IPR013083
qpx_c2126	Zinc finger domain protein 2	2.0	CBN78322	9.73E-18	*Ectocarpus siliculosus*	IPR001841	IPR013083
qpx_c4080	Zinc finger domain protein 3	2.3	XP_642978	2.38E-07	*Dictyostelium discoideum*	IPR001841	IPR006058
qpx_lrc17430	Zinc finger domain protein 4	2.1	XP_002908993	9.73E-26	*Phytophthora infestans*	IPR000058	IPR002653
qpx_c8807	cytochrome b5 1	8.9	XP_456307	3.1E-05	*Kluyveromyces lactis*		
qpx_c9641	cytochrome b5 2	1.7	XP_748717	5.6E-14	*Aspergillus fumigatus*		
qpx_c7499	cytochrome b5 3	2.3	EKV06300	5.2E-14	*Penicillium digitatum*		
qpx_c9000	cytochrome b5 4	1.8	XP_002113872	1.9E-29	*Trichoplax adhaerens*	IPR001199	PTHR21281
qpx_c2790	cytochrome P450-1	2.2	XP_006341238	1.9E-31	*Solanum tuberosum*		
qpx_c2669	cytochrome P450-2	2.2	AGN52759	9.0E-13	*Laodelphax striatella*	IPR001128	IPR002403
qpx_c1667	glutathione peroxidase	2.6	XP_007160053	3.4E-50	*Phaseolus vulgaris*		
qpx_c7351	glutathione s-transferase 1	2.4	XP_002900509	7.7E-06	*Phytophthora infestans*	IPR004046	IPR010987
qpx_c4046	glutathione s-transferase 2	2.1	XP_005848634	7.9E-17	*Chlorella variabilis*	IPR001129	IPR023352
qpx_c956	iron-sulfur cluster protein	2.3	CCA27067	2.0E-43	*Albugo laibachii*	IPR000361	IPR016092
Recognition and transport							
qpx_c1776	ABC transporter 1	5.4	XP_005094340	3.7E-15	*Aplysia californica*	IPR013525	PTHR19241
qpx_c9322	ABC transporter 2	1.5	XP_007342100	1.6E-37	*Auricularia delicata*	IPR003439	IPR017871
qpx_c1244	MSF-transporter	1.9	XP_006676685	2.1E-06	*Batrachochytrium dendrobatidis*	IPR016196	
qpx_c857	rab-type small g protein 1	1.8	XP_002177632	4.4E-47	*Phaeodactylum tricornutum*	IPR001806	IPR013753
qpx_c107	rab-type small g protein 2	1.6	CBK20771	4.4E-24	*Blastocystis hominis*	IPR001806	IPR013753
qpx_c169	rab-type small g protein 3	1.9	XP_001878287	2.7E-54	*Laccaria bicolor*	IPR001806	IPR013753
qpx_c3365	receptor tyrosine phosphatase	3.1	EQC38629	3.1E-19	*Saprolegnia diclina*		
qpx_c898	roh-type small g protein	1.5	XP_006628040	1.7E-08	*Lepisosteus oculatus*	IPR000198	IPR008936

***In vivo* virulence**

The QPX infection experiment was conducted to evaluate changes in QPX virulence due to its exposure to biomolecules present in clam adductor muscle homogenate (CAMH). Two different QPX isolates were used in the experiment: one continually sub-cultured for approximately 5 years (NY-03) and the other freshly isolated (NY-12). In addition, NY-12 isolate was prepared in media supplemented with clam pallial mucus. No clam mortalities were recorded during the 3 months of QPX challenge experiment for any of the treatments. However, the presence of QPX cells within clam tissue was estimated using quantitative PCR (Liu et al 2009) for 12 randomly selected clams from each treatment showing the presence of QPX in 75-92% of clams injected with the 2003 isolate and in 92-100% of clams injected with the 2012 isolate (Figure 2.2). The mean QPX infection intensity in clams injected with the NY-03 isolate was 800 cells×g^{-1} (MEM FBS) and 200 cells×g^{-1} (CAMH) (**Figure 2.2**). For clams injected with NY-12 isolate the mean QPX intensity was 6700 cells×g^{-1} for MEM FBS treatment, 9300 cells×g^{-1} for CAMH treatment and 8400 cells × g^{-1} for pallial mucus treatment (**Figure 2.2**). Ten clams with the highest QPX intensities revealed by qPCR (from the 2012 isolate) were checked for the presence of QPX lesions using histopathology techniques (Dahl & Allam 2007). No signs of disease (characteristic tissue lesions) were detected on the ten examined histological slides of clam tissue (data not shown).

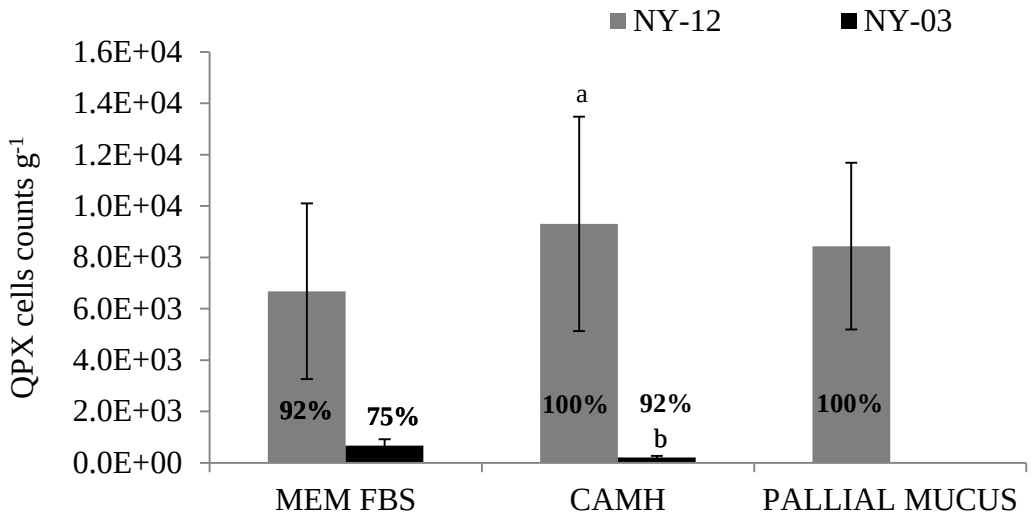


Figure 2.2: QPX presence in clam tissue (QPX cells g^{-1}, mean ± std. error, n=12) detected by quantitative PCR three months following injection of QPX cells from three different cultures: minimal essential medium supplemented with fetal bovine serum (MEM FBS) clam adductor muscle homogenate (CAMH) and clam pallial

mucus . Present numbers represent QPX prevalence of clams in each treatment. Letter labels represent significant difference (t test, $p<0.05$)

Discussion

The goal of the present study was to evaluate the effect of culture conditions on QPX virulence. One experiment was designed to identify transcriptional changes in QPX cells in response to contact with molecules present in the clam adductor muscle homogenate. The second was to test QPX virulence *in vivo* using a laboratory challenge experiment (Dahl et al 2008). Results from the oligoarray experiment suggest that QPX cells cultured in clam adductor muscle suffered from cellular stress in comparison to QPX cells cultured in standard media conditions. Previously reported virulence-related genes (Rubin et al 2014) were down regulated for QPX cultivated in clam tissue suggesting that this media does not enhance QPX virulence. In addition, the experimental challenge experiment showed no difference in virulence for QPX cells derived from the two culture media.

QPX response to the clam tissue homogenate was associated with stress and cell detoxification indicated by the up-regulation of heat shock proteins, ubiquitin-proteasome complex-related proteins, cytochrome P450, cytochrome b5, and A20/An1-type zinc finger proteins. Heat shock proteins are best known as molecular chaperones protecting other proteins from denaturing and misfolding during temperature shifts (Gophna and Ron 2003). They are also induced by many other environmental conditions such as pollution, starvation and interaction with eukaryotic hosts (Gophna and Ron 2003). For infectious organisms, heat shock proteins help the pathogens to cope with the stressful environment inside the host (Neckers and Tatu 2008). The up-regulation of QPX heat shock proteins while it was cultivated in CAMH suggests the hsps fulfill a similar role in QPX interaction with the clam host. In addition, 18 transcripts coding for proteins associated with the ubiquitin-proteasome proteolytic pathway were also up-regulated in QPX cells grown in CAMH. The ubiquitin-proteasome pathway eliminates intracellular proteins which are no longer needed and ensures adequate protein turnover during many biological processes such as cell cycle and differentiation, and it also plays a major role in the cellular response to stress (Glickman & Ciechanover 2002). A total of 28 transcripts coding for zinc finger proteins were also up-regulated in QPX cells from CAMH. These included 19 different transcripts (97% nucleotide similarity) belonging to the A20/An1 type zinc fingers which have only been well characterized in animals in which they regulate immune responses;

however, in plants these proteins are involved in stress response by unknown mechanisms of action (Vij &Tyagi 2008). Their cellular function in protists is unknown but from the current study it appears that they may function in a similar way in labyrinthulomycetes.

Cytochrome P450 proteins are detoxifying enzymes found in all organisms (Hrycay & Bandiera 2012, Jung 2011). Two cytochrome P450 were up regulated in QPX cells grown in CAMH> In addition, four different cytochrome b5 proteins, which are part of the CYB5/CYP450 oxidation/reduction detoxification system, were also up-regulated in QPX cells derived from clam adductor muscle. The CYB5/CYP450 system metabolizes both endogenous and exogenous compounds but their most important function is oxidation of xenobiotic substances and protection of cells, tissues and organs against their harmful effects (Cresnar & Petric 2011, Lewis & Wiseman 2005, Seliskar & Rozman 2007, Uno et al 2012). Further, in pathogenic organisms, the CYTB5 AND CYTP450 have been considered to be virulence factors because they provide mechanisms to metabolize antimicrobial compounds produced by animal and plant hosts (Coleman et al 2011, Kim et al 2007, Lin et al 2011, Zhang et al 2012). Two QPX transcripts coding for different cytochrome P450 proteins were found to be up-regulated when QPX cells were cultivated in clam adductor muscle homogenate, suggesting that they may be involved in protection of QPX cells against antimicrobial compounds produced by the clam immune system.

Change in the expression of two different ABC-type transmembrane transporters in QPX cells exposed to clam-derived factors was also found. The ABC (after ATP binding cassette) transmembrane transporters are involved in the export and import of substances from cells. In pathogenic organisms they remove noxious or toxic substances (Dassa & Bouige 2001). The expression pattern of the ABC-transporter suggests that they might perform a similar function for QPX cells which are in contact with clam tissue. Clam plasma and different clam tissue types have an inhibitory effect on QPX growth, suggesting clam factor toxicity to QPX cells (Perrigault et al 2009a, Anderson et al 2003). The current study shows that QPX possesses many molecules which might enable the parasite to defend itself from potentially toxic compounds produced by the clam host.

Other transcripts for which expression was up-regulated when QPX cells were grown in clam tissue homogenate suggest QPX acclimation to the food source. Three different lipases were also up-regulated in QPX cells cultivated in CAMH which suggests QPX acclimation to the

lipid content in the media. Similarly, the changes in the expression of QPX peptidases might be related to the adjustment of proteolytic enzymes to the available protein source in the media. Another example is a phytanic acid (a branched fatty acid) hydrolase which was also up-regulated in response to clam tissue homogenate. Tissues from some species of mollusks have been found to contain small amounts of this essential fatty acid (Ackman et al 1971), so it might be a component of clam tissue, thus explaining this transcriptional change in QPX cells.

The QPX challenge experiments failed to induce overt disease, as the usual signs of active infection (mortality and/or histopathological lesions) were not detected after 3 months of incubation under temperature conditions considered to favor disease development in Florida clams (Perrigault et al 2011). The qPCR technique allowed for the detection of QPX DNA in clam tissue. However, it cannot be determined if this reflects an active infection or just the presence of parasite cells in the clam tissue because no lesions or viable QPX cells were detected during histopathology examinations. On the other hand, histological observations only cover a small fraction of the clam tissues, so there is a possibility that QPX cells were missed in the ten examined clams. Alternatively, it has been previously suggested that laboratory conditions can promote clam healing and that disease development is primarily controlled by host resistance (Dahl & Allam 2007). In fact, results in the current study show that cold water temperature was not sufficient to induce overt disease in the current experimental settings, suggesting that there might be other environmental factors influencing either the host or the pathogen or both Finally, the exposure of QPX cells to clam-factors did not appear to increase its virulence, suggesting that QPX is a weak pathogen capable of causing mortalities only in clams with high levels of immunosuppression.

It is also possible that QPX disease development is mostly controlled by the genetic resistance of clams, and that the clam strain obtained for this experiment exhibited high resistance to QPX. Even though the clams were obtained from a commercial source from Florida previously shown to be suitable for QPX disease (Dahl et al 2008), it is possible that the strain has been out-crossed with northern stock. It is also possible that any possible increase in QPX virulence might be triggered by its contact with clam-derived factors other than those tested in this study. Since QPX infections are mainly found in clam connective tissue (Ragone Calvo et al 1998, Smolowitz et al 1998), it is possible that clam factors such as collagen or fibronectin trigger production of QPX virulence factors. Perrigault et al (2009) showed that extracts from

different clam tissue types have different effects on the *in vitro* growth of QPX, with the gill and mantle tissue homogenates significantly inhibiting, and foot and adductor muscle enhancing QPX growth. The effect of the clam tissue types on QPX virulence characteristics could be different than the effect on its growth.

Higher numbers of QPX cells were detected in tissue of clams which were injected with the more recent QPX isolate. This is consistent with other studies which show that freshly isolated pathogens are more likely to cause disease than isolates kept in culture for longer time (Ford et al 2002a). However, the results of this study are inconclusive because no signs of overt disease were detected by histopathology. Further investigations which would include more QPX-susceptible clam broodstocks (e.g. South Carolina, Kraeuter et al 2011), or more stressful environmental conditions and freshly isolated QPX cells could be applied to improve this experiment.

Chapter 3: Differential gene expression in five isolates of quahog parasite unknown (QPX)

Manuscript in preparation for Journal of Eukaryotic Microbiology

Abstract

Transcriptomic profiles of five geographically-distinct QPX isolates were generated using custom 15K 60-mer oligonucleotide arrays. A total of 1,263 transcripts were differentially expressed (DE) among the five QPX isolates. The DE genes detected between different QPX isolates were used to investigate the molecular basis for previously demonstrated variations in growth and virulence among QPX isolates. The hierarchical clustering of gene expression profiles showed that the QPX isolates from Raritan Bay (RB, NY) and from Provincetown Harbor (MA) were more similar to each other and diverged from QPX isolates from Peconic Bay (PB, NY) and Old Plantation Creek (VA) which had more similar gene expression profiles. The most prominent difference was based on 78 transcripts coding for heat shock proteins DE between the five QPX isolates. The study generated contrasting transcriptomic profiles for QPX isolated from northern (MA) and deeper (RB, NY) locations as compared to southern (VA) and shallower (PB, NY) areas, suggesting the adaptation of the parasite to local environmental, in particular temperature, conditions.

Introduction

QPX disease affects hard clam populations in different geographic locations in the United States and in Canada. One of the first published reports of the QPX organism infecting clams comes from 1989 from shellfish hatcheries on Prince Edward Island in Canada (Whyte et al 1994). In the summer of 1995, QPX disease outbreaks were reported to occur on the coast of Massachusetts near Provincetown (Smolowitz et al 1998). Between 1996 and 1997, QPX was reported in clams from the coast of Virginia, and in the summer of 2002 severe mortality events were documented from Raritan Bay, New York (Dove et al 2004). Since the clam mortalities occurred in distant geographic locations, at first it was unclear if they were caused by multiple and similar organisms. It was noted that QPX cells on histology slides of clam tissues from Massachusetts and Virginia differ in morphological characteristics, including retention of Groccott's methanamine silver and Alcian blue stains by the extracellular material around the QPX cell wall (Ragone Calvo et al 1998). In the search for strain differences, Stokes et al (2002)

compared partial DNA sequences of the small ribosomal subunit (SSU) of QPX cells preserved in the histology samples from the above studies and concluded that all QPX isolates (Canada, MA, VA and NJ) are most likely the same species but the similarity is based on a small fragment (about 411 bp) of the SSU gene, thus not resolving possible strain differences. Clonal QPX cultures derived from NY and MA exhibit differences in growth rate in response to different temperature with the NY isolates having optimal growth at 23 °C and MA isolate at 20 °C (Perrigault et al 2010). Clam challenge experiments also showed that three different QPX isolates (two from NY and one from MA) caused different clam mortality rates thus suggesting variability in their virulence characteristics (Dahl et al 2008). Despite these phenotypic differences, no molecular markers exist to distinguish between these possible QPX strains.

Transcription profiling using cDNA microarrays has been applied to identify the molecular basis of phenotypic differences between isolates of the same species for many protozoan pathogens (Baptista et al 2006, Bozdech et al 2008, Davis et al 2007, Llinas et al 2006). These studies show that gene expression patterns differ between highly related isolates and strains and reveal genes responsible for their different phenotypes. The most differentially expressed transcripts between isolates are usually investigated for their genetic polymorphisms within the gene coding sequence (e.g. Baptista et al 2006), or within the 5'-UTR or 3'-UTR which determine the binding of transcription factors and gene expression regulation (e.g. Rodrigues et al 2010). For example, transcription profiles of six *Trypanosoma cruzi* isolates revealed 30-fold transcription regulation of the gene coding for NADH dehydrogenase subunit 7 (Baptista et al 2006) which then revealed a large deletion within the coding sequence in some of the isolates.

The aim of this study was to identify possible transcriptome-wide gene expression differences among QPX isolates, and to identify genes with the highest expression differences among QPX isolates. Specific focus was given to exploring the potential molecular basis for virulence differences among QPX isolates and for possible genetic adaptations to local environmental conditions.

Material and Methods

QPX cultures

Five QPX isolates were used in this study: three from New York, one from Massachusetts, and one from Virginia. The NY QPX included two isolates from Raritan Bay previously shown to display different virulence characteristics (NY0313808BC7 and NY0314220AC6, here designated NY1 and NY2 respectively Dahl et al. 2008) as well as an isolate cultured from a diseased clam collected from the Peconic Bay (NY070348D, or NY3). All isolates were grown in minimal essential medium supplemented with fetal bovine serum (MEM-FBS, Kleinschuster et al 1998) media in triplicate cultures. Each culture flask (25-ml culture flasks, Falcon) contained 5 ml sterile media inoculated with 100 µl of an exponentially growing 4-days-old QPX culture (2×10^5 cells ml^{-1}) cultivated at 23 °C. Cultures were incubated at 23 °C for six days to reach the exponential growth phase (6×10^6 cells ml^{-1}). At the end of the incubation period, the cultures were diluted with equal volumes of filtered artificial seawater and passed several times through a syringe to facilitate liquefaction of QPX mucus secretion. The mixtures were then transferred into 15-ml conical tubes and centrifuged at 1000 g for 20 minutes at 4°C. The supernatant was discarded and cell pellets were collected and kept on ice for immediate RNA extraction. Trizol reagent (Molecular Research Center, Inc.) and the manufacturer's protocol were used to isolate RNA from all samples and RNA quality and quantity were estimated spectrophotometrically by Nandorop and using an Agilent 2100 Bioanalyzer.

Oligoarray design and data analysis

The oligoarray platform and protocols used in this study were described earlier (Rubin et al., 2014; see Chapter 1 and Chapter 2). Briefly, previously generated transcriptomic data were used for the production of a 8×15 k 60-mer oligonucleotide array using the Agilent eArray application (https://earray.chem.agilent.com/earray/). These included 6,781 curated annotated sequences and 8,297 non-annotated sequences (Rubin et al., 2014). Probes (1 probe/sequence) were synthesized in situ along with positive and negative controls using 8x15K-feature Agilent format slides. Labeled (Cy3 or Cy5) complementary RNA (cRNA) was synthesized from 150 ng of RNA purified from cultures using the Two-Color Microarray-Based Gene Expression Analysis Protocol (Quick Amp Labeling) following manufacturer's protocol. Labeled cRNA was purified using Illustra CyScribe GFX Purification Kit (GE Healthcare) and used for array hybridization (300 ng of each Cy3- and Cy5-labeled cRNA). Array hybridization and scanning

were conducted as previously described (Rubin et al., 2014; Chapter 1). LIMMA package in R software was used to normalize the intensity data and to remove within-array (method: global lowess) and between-array (method: quantile) non-biological variation (Smyth & Speed 2003). After normalization, the intensities in separate color channels were exported into an excel spreadsheet for further data quality control and trimming. The spot intensities which were less than two-fold of the background intensities were eliminated from further analysis. The relative mRNA expression levels for each transcript were calculated as the ratio of the fluorescent intensity in individual samples and the mean fluorescent intensity of that transcript in all strains (n=3 per strain except for MA n =2)). Hierarchical clustering of samples and genes (Pearson correlation) and the determination of statistically significant differentially expressed genes (Anova, p-value < 0.01) were completed in the TM4-suite using MeV program (Saeed et al 2006, Saeed et al 2003). The final criteria for differential gene expression were significance by Anova analysis and a one and one-half fold increase (up-regulation) or one and one-half fold decrease (down-regulation) from the mean (Rubin et al 2014).

Results

A total of 1263 transcripts were found to be differentially expressed (DE) among five QPX isolates (ANOVA, $p<0.01$ in conjunction with 1.5-fold change (Rubin et al 2014), **Figure 3.1**). Of these, 547 sequences gave a positive blastx hit in the NCBI non redundant (nr) data base. However, only 477 transcripts could be annotated with a putative function based on either gene ontology (GO) terminologies or conserved protein domains (InterPro). The 477 transcripts were grouped into 17 different categories representing their biological or molecular function (**Figure 3.2**). This classification resulted in 78 DE transcripts representing response to temperature, 70 DE transcripts categorized into metabolic processes, 60 DE transcripts related to transmembrane transport, and 50 DE belonging to cellular biosynthetic processes (**Figure 3.2**).

Multiple sequence alignments of the 78 transcripts coding for heat shock proteins revealed that 55 sequences represented multiple isoforms of one heat shock protein (93% pair-wise identity) with a predicted molecular weight of 15 kDa. Another 12 transcripts encode for a different small heat shock protein of around 18 kDa, and the remaining 11 sequences encode two 16 kDa, one 40 kDa, three 70 kDa, one 90 kDa and one 96 kDa heat shock proteins. . In addition, the differentially expressed transcripts were sorted by the highest-fold change, which revealed

that the expression of 131 transcripts changed by five or more fold. Nineteen of these sequences had predicted functions and are presented in **Table 3.1.** These included: the two heat shock proteins (15 and 18 kDa), two serine peptidases (one from the S8 and one from the S10 family), a non-ribosomal peptide synthase, a catalase, a g-protein-coupled receptor and three transmembrane transporters.

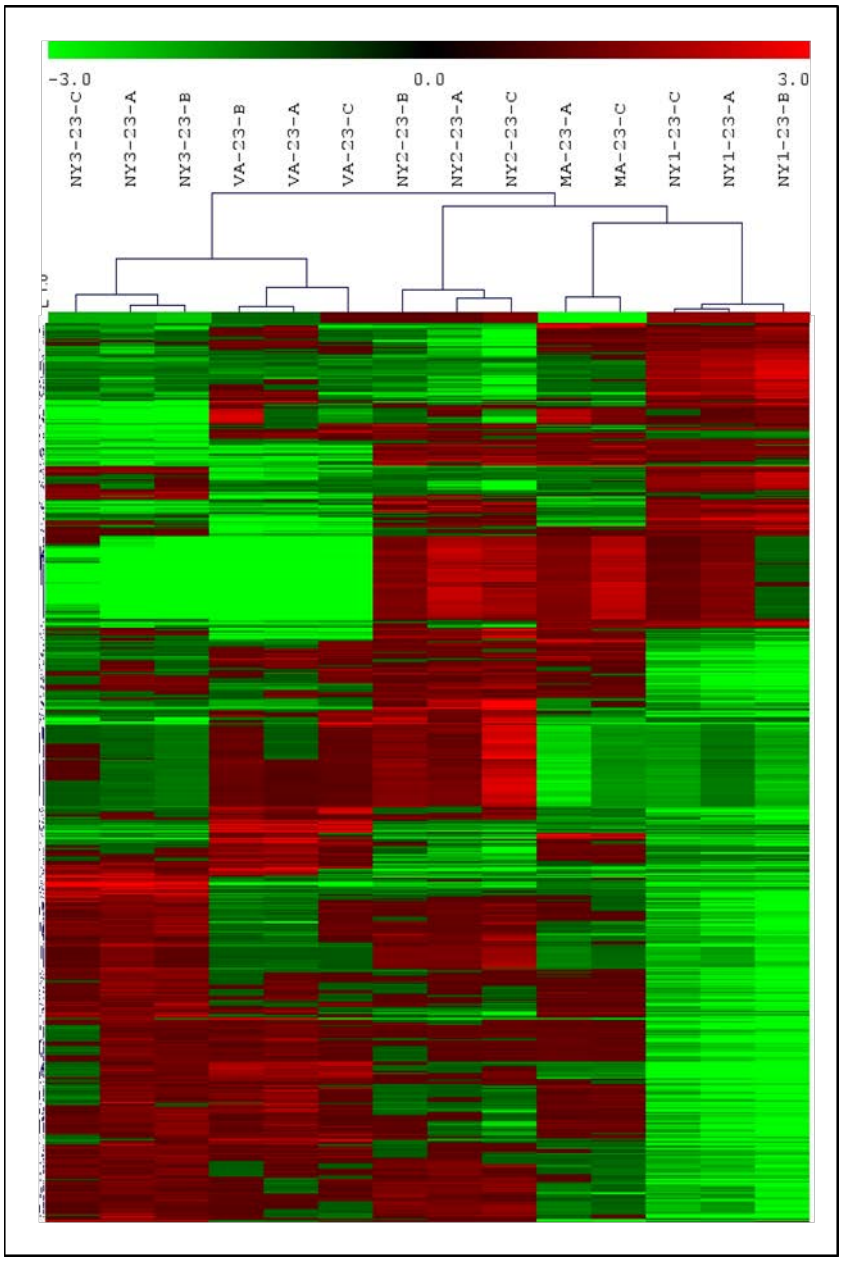


Figure 3.1:Hierarchical clustering of 1,263 QPX transcripts (Pearson's correlation centered) identified to be differentially expressed among five QPX isolates (NY1 and NY2 – from Raritan Bay, NY; NY3 – from Peconic Bay, NY; MA – from Provincetown Harbor, MA; VA – from Old Plantation Creek, VA)

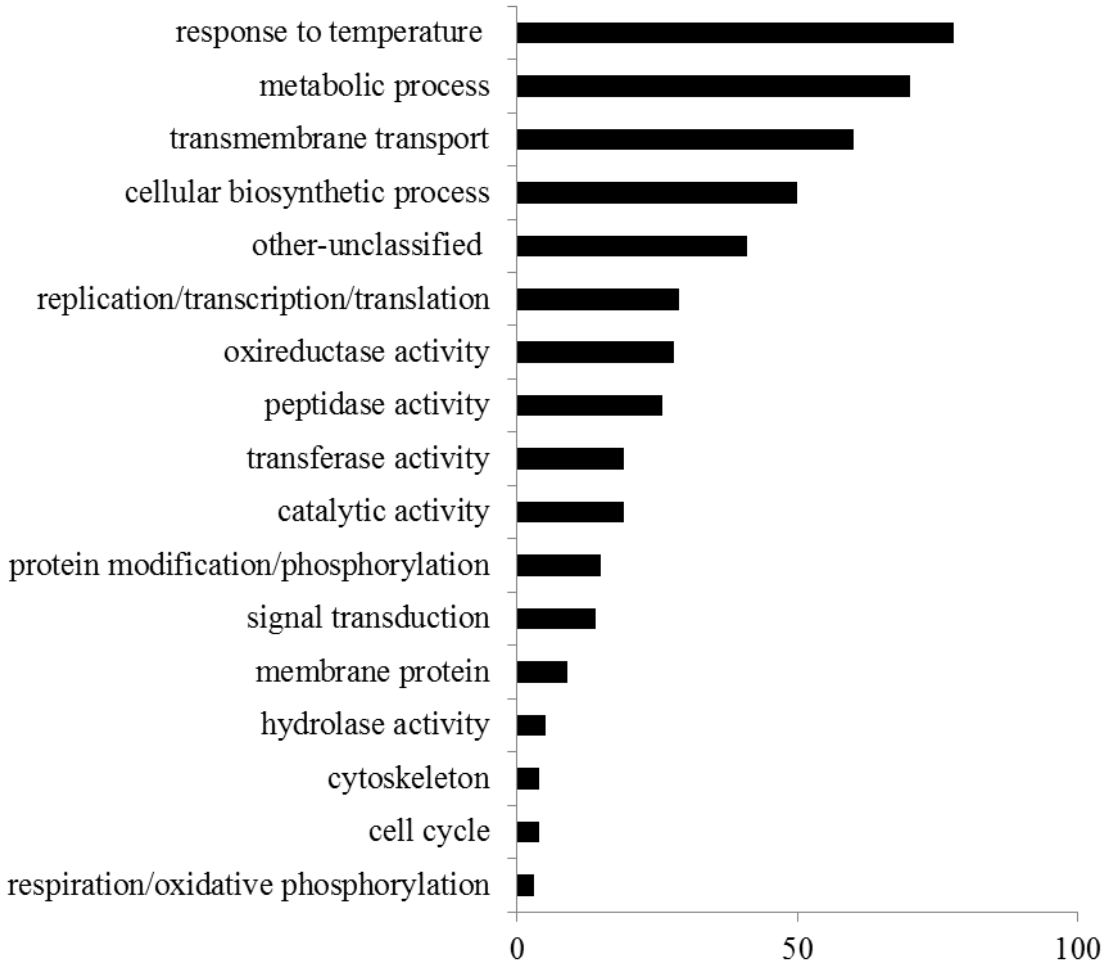


Figure 3.2: Number of differentially expressed transcripts among five QPX isolates, grouped into biological and molecular function categories based on gene ontology terminology or conserved protein domain

Table 3.1: Annotated transcripts differentially expressed by more than five-fold difference among five QPX isolates (ANOVA, p<0.01)

Seq ID	Sequence desc.	Top blast results			Mean fold change				
		Hit ACC	E-Value	Organism	NY-1	NY-2	NY-3	VA	MA
qpx_c11918	ammonium transporter	XP_003062934	2.4E-21	*Micromonas pusilla*	0.3	1.9	0.4	-5.4	-2.1
qpx_c1462	ABC transporter	EJK46845	2.1E-87	*Thalassiosira oceanica*	-6.6	1.9	-0.5	0.4	0.0
qpx_c3745	ABC transporter	XP_001023596	3.9E-13	*Tetrahymena thermophila*	1.3	1.1	-1.2	-5.1	1.7
qpx_c1535	calcium transporting ATPase	XP_002184696	0.0E+00	*Phaeodactylum tricornutum*	-2.0	2.0	0.5	-6.9	0.2
qpx_c1451	non-ribosomal peptide synthetase	WP_004272016	4.8E-07	*Azospirillum amazonense*	-2.9	1.2	2.3	-1.3	-21.6
qpx_c1660	Zinc finger, CCHC-type	XP_003103852	1.9E-05	*Caenorhabditis remanei*	1.8	1.4	-6.8	-0.3	-2.2
qpx_c2153	Von Willebrand factor domain	ETV79340	1.8E-79	*Aphanomyces astaci*	-5.2	-0.5	1.6	-0.1	1.1
qpx_c2181	catalase	AEX91749	6.4E-34	*Acanthamoeba castellanii*	-5.4	1.2	-2.4	1.3	2.3
qpx_c3119	bacitracin synthetase	AHG22620	7.7E-13	*Serratia fonticola RB-25*	-17.6	-6.1	-0.6	0.8	-5.3
qpx_c3768	lipase	EFA75426	1.5E-09	*Polysphondylium pallidum*	-0.6	0.9	1.8	-7.5	-4.7
qpx_c4899	long chain fatty acid synthetase	YP_006821641	2.9E-23	*Alcanivorax dieselolei*	2.5	1.4	-1.8	-9.0	-9.4
qpx_c8108	vacuolar protein sorting-associated protein	EXX57160	2.0E-10	*Rhizophagus irregularis*	0.2	2.2	-7.9	-2.3	1.4
qpx_lrc660	Zinc finger, ZZ-type	XP_007299136	1.8E-13	*Stereum hirsutum*	-8.2	1.2	0.4	1.9	-1.5
qpx_c14573	tubulin alpha chain	XP_006433708	3.1E-88	*Citrus clementina*	-5.9	-0.6	1.3	1.3	1.5
qpx_c1944	histone acetyltransferase	XP_001417549	9.4E-26	*Ostreococcus lucimarinus*	-6.2	0.5	-1.2	1.7	1.2
qpx_lrc5272	heat shock protein 15 kDa	BAK07374	7.0E-21	*Hordeum vulgare*	0.6	1.8	-6.6	-11.9	2.0
qpx_c56	heat shock protein 18 kDa	WP_003690914	4.0E-20	*Lactobacillus mali*	0.5	-0.7	-5.4	-0.3	1.8
qpx_c2560	carboxypeptidase y	XP_002550523	7.4E-10	*Candida tropicalis*	-1.8	0.4	-5.5	2.4	-1.4
qpx_c2644	subtilisin kexin type peptidase	ABI79453	6.0E-56	*Quahog parasite X*	-4.8	-7.4	-7.3	1.2	3.4

Discussion

Oligoarray methodology was applied to generate transcription profiles of five QPX isolates originating from locations along the geographic range of distribution of the QPX disease in the United States. The study identified 1263 transcripts to be differentially expressed (DE) among the five isolates, with the highest number of DE transcripts belonging to heat shock response and metabolic processes. These changes reflect adjustments of QPX metabolic rate in response to the temperature of the experimental cultures (Rubin et al 2014). For example, three low molecular weight heat shock proteins (15, 16 and 18 kDa) have been previously shown to be up-regulated in the New York isolate NY1 grown at the warm temperature of 27 ℃ (Rubin et al 2014). On the other hand, the high molecular weight heat shock proteins (70, 90 and 100 kDa) were shown to be up-regulated in QPX cells grown at cold (13 ℃) temperatures (Rubin et al 2014). In the current study, all experimental cultures were incubated at 23 ℃, which is confirmed to be the optimal temperature for NY-1 isolate, but thermally stressful for MA isolate as highlighted by the up-regulation of two small heat shock proteins (15 and 18 kDa). These results are in agreement with the observations of Perrigault et al (2010) who demonstrated maximal growth for the Raritan Bay isolates of QPX (NY-QPX-1 and NY-QPX-2) at 23 ℃ while the MA isolate (MA-QPX) displayed a 20% reduction in growth when the temperature increased from 20 to 23 ℃. In addition, the VA isolate of the most southern origin, showed strong down-regulation (12 fold decrease) of the 15 kDa heat shock protein, suggesting divergence of that gene based on latitudinal distribution of QPX. Interestingly, NY--3 isolate from Peconic Bay exhibited down-regulation of all three small heat shock proteins by 3, 5 and 7-fold for the 16, 18 and 15 kDa heat shock proteins, suggesting its adaptation to warm temperatures of that shallow water embayment. Differential expression of heat shock proteins has been documented for many aquatic organisms living in shallow, stagnant waters (Feder & Hofmann 1999). In addition, expression of hsp varies along the temperature gradient related to geographic distribution across degrees of latitude (Feder & Hofmann 1999). In fact, hsp-related sequences have been used to study differences among strains within the same species. For example, the nucleotide sequence polymorphisms of the hsp16 gene allow for assignment of *Trypanosoma cruzi* strains into two main genetic groups

(Perez-Morales, 2009). In addition, genetic polymorphism within the 5' or 3' untranslated regions (UTRs) of hsp70 is responsible for the regulation of gene expression and the difference among strains of *T. cruzi* (Rodrigues et al 2010). The strong variation in the expression of the 15 kDa heat shock protein (qpx_lrc_5272) reported here across different QPX isolates suggests that the coding region or the genomic regions involved in the regulation of this transcript could represent a genetic marker for the study of intraspecific QPX variation.

The present study also suggests that QPX isolates from different locations exhibit adaptation to the salinity of local waters. The VA-QPX isolate used in the current study originated from a clam mortality outbreak which took place in 2009 at an aquaculture farm located at Old Plantation Creek (OPC), a small embayment connected to Chesapeake Bay, VA. The water salinity of OPC varies between 15-25 ppt which makes it an unusual location for QPX disease because previous QPX disease outbreaks in VA were only reported for high salinity (30-35 ppt locations of the Atlantic coast (Ragone Calvo et al 1998). In addition, NY1, NY2 and MA isolates where shown to have optimal growth salinity around 34 ppt (Perrigault et al 2010, Ragone Calvo et al 1998). The cultures of all five isolates in this study were prepared in minimal essential medium supplemented with fetal bovine serum and adjusted to 34 ppt with salts (Kleinschuster et al 1998). Interestingly, only VA QPX cells were found to have higher expression of an aquaporin (qpx_c9002) and lower expression of two different sodium/hydrogen ion antiporters (qpx_c242, qpx_458, Interpro protein domain id: IPR01846). Aquaporins are transmembrane proteins which export and import water out of and into the cell and participate in osmoregulation of microbial, plant and animal cells (Echevarria & Ilundain 1998, Fu & Lu 2007). The sodium/hydrogen ion antiporters belong to the major facilitator superfamily (MFS) and they exchange sodium and hydrogen ions across the cell membrane, allowing them to move along the concentration gradient (Pao et al 1998). The down-regulation of the sodium transporter in VA-QPX in response to the high sodium concentration outside the cell suggests a phenotypic plasticity and acclimation ability of that isolate to the high salinity of the culture media.

A main objective of the present study was to gain insight into the molecular basis of virulence variation among three QPX isolates previously shown to cause different mortality rates in clams during experimental challenge (Dahl et al 2008). The DE transcripts were screened for

sequences coding for proteins commonly associated with virulence in other pathogenic organisms. In the study by Dahl et al (2008), the MA isolate caused highest percent mortality in clams from VA, NY2 isolate caused the highest percent mortality in MA clams, the NY1 isolate caused the highest percent mortality to NY clams, and finally both NY1 and NY2 caused the highest percent mortality in clams from Florida. Overall the study concluded that both the genetic susceptibility of clams and the virulence of QPX isolates determine the disease outcome (Dahl et al 2008). In the current study, all five QPX isolates were found to have higher expression of some molecules, previously classified as putative virulence factors, than other isolates (Rubin et al 2014). NY1 QPX cells exhibited higher the expression of three different peptidases (qpx_c1221, qpx_c2535, qpx_c8079) as compared to mean expression of all isolates. Among these three peptidases, only one belongs to a family of peptidases associated with virulence: the papain-like peptidase family C1; while the other two peptidases of the carboxypeptidase family, S28 and oligopeptidase family, S10 are less commonly involved in infection processes. On the other hand, NY2 QPX exhibited higher expression of two virulence-related peptidases: a S8 or subtilisin-like peptidase (qpx_765) and a C1 or papain-like peptidase (qpx_c60). Interestingly, contrasting expression of papain-like cysteine peptidases has been documented for two different strains of *Entamoeba histolytica* exhibiting different virulence phenotypes (Davis et al 2007). The actualfuntion of the specific QPX peptidase coded by qpx_c60 transcript has to be further investigated to confirm its role in QPX infectivity.

Further, the MA QPX cells were found to have more than a three-fold higher expressionof another subtilisin-like peptidase (qpx_c2644). QPX possesses at least seven different sequences encoding subtilisin-like peptidases with amino acid similarities ranging from 36 to 79% (this thesis, chapter 5). A recent sequencing project of four transcriptomes from four different isolates (NY1, NY3, MA-QPX and VA-QPX) shows that the two NY isolates lack the sequence for the one subtilisin-like peptidase: qpx_c2644 (Allam, unpublished). This suggests that QPX isolates may vary in the number of different genes coding for substilisin-like peptidases. It is important to note that QPX disease is endemic in MA and has been causing regular severe mortality to clams since the 1990s (Allam, personal communication). In contrast, QPX epizootics in NY are rarer and the organism is often detected at background levels. Therefore, it can be speculated

that the MA QPX isolate is more virulent and the subtilisin peptidase gene (qpx_c2644) could be responsible for that difference. However, the final QPX disease presentation and the clam mortality are not exclusively dependent on QPX virulence but also on genetic susceptibility of clam broodstock (Dahl et al 2008, this thesis Chapter 2). In another study, when three different clam aquaculture strains were kept in Barnstable Harbor, MA and in Dry Bay, NJ, all strains suffered from highest mortalities and the highest QPX prevalence at the MA location (Kraeuter et al 2011). Of course, environmental conditions, specifically any stress factors, have to be considered as well. QPX is an opportunistic pathogen which appears to cause disease when the clams are stressed by environmental factors. Thus the QPX disease outcome is the results of a combination of stressors, clam susceptibility, and QPX virulence, and all of them have to be accounted for in analysis of clam mortality.

Subtilisin-like peptidases are common virulence factors of many pathogenic organisms including bacteria, fungi, and protozoa (Bonifait et al 2010, McKerrow et al 2006, Monod et al 2002). However, further studies aimed at understanding the function and role of substilisin peptidases in QPX virulence and its infection mechanism in clams are required to determine which increase in expression contributes most to the variation in virulence among QPX isolates. The expression of these transcripts, however, was also found to be influenced by temperature changes (Rubin et al 2014), and since the different isolates have different optimal temperatures, the expression of these peptidases might also be a result of adjustment of QPX metabolic rate in response to temperature.

Another noticeable difference between QPX isolates was the expression of several enzymes related to carbohydrate metabolism, some of which are likely associated with extracellular polysaccharide production. For example, dTDP-4-dehydrorhamnose reductase (qpx_c4448, IPR005913) is an enzyme which catalyzes the production of L-rhamnose. The carbohydrate composition of QPX extracellular muco-polysaccharide secretion is unknown and rhamnose is not a major component of extracellular polysaccharides of two other thraustochytrids (Jain et al 2005). However, this does not exclude the possibility of its presence in QPX mucus and variability of that sugar in the composition of mucus from different isolates. Similarly, the expression of UDP-glucose/GDP-mannose dehydrogenase (qpx_c1232, IPR017476), a QPX

enzyme most likely involved in polysaccharide biosynthesis (Rubin et al 2014), was also differentially expressed between QPX isolates. Differences in the thickness and stain affinity of the mucus layer between MA-QPX and VA-QPX cells examined on histological section of clam tissue has been previously noted (Ragone Calvo et al 1998). In addition, the differences in mucus viscosity among QPX isolates are very noticeable during routine preparation of QPX subcultures (Rubin, personal observation). The carbohydrate and protein composition of mucus secretion of the studied QPX isolates needs to be completed to test the above hypotheses.

In conclusion, this study shows that the gene expression profiles of QPX isolates originating from distant geographic locations reflect their physiological differences and their acclimation to local environmental conditions, specifically temperature and salinity ranges. In addition, the expression of some genes might reflect other genetic variation among the studied isolates. However, the actual nucleotide polymorphism of these genes among QPX isolates would have to be determined to demonstrate their suitability for QPX genotyping or population level studies. This study also identifies potential virulence-related genes which are differentially expressed among different QPX isolates. These molecules are excellent subjects for further studies on molecular bases of virulence variation among QPX strains.

Chapter 4: Characterization of QPX secretome

Manuscript submitted to International Journal for Parasitology

Abstract

Secreted and cell surface associated molecules play a major role in disease development processes and host-pathogen interactions, and usually determine virulence of invading organisms. In this study, we investigated proteins secreted by Quahog Parasite Unknown (QPX), a thraustochytrid protist that infects the hard clam, *Mercenaria mercenaria*. *In silico* analysis of QPX transcripts predicted over 1200 proteins to possess an amino-terminal signal peptide which directs proteins into the classical eukaryotic secretory pathway. Proteomic analysis using LC-MS technology identified 56 proteins present in the extracellular products (ECP) of QPX cells grown *in vitro*, including six mucin-like molecules, four glycosyl hydrolases and eight peptidases. Transcription levels of 19 QPX extracellular proteins were investigated in clam tissue lesions *(in vivo)* using qPCR. The overexpression of six of these extracellular molecules in clam tissues as compared to *in vitro* cultures suggests that they are involved in the infection process, virulence and/or interaction with the clam host.

Introduction

Quahog parasite unknown (QPX) is a thraustochytrid protist infecting the hard clam *Mercenaria mercenaria* along the northwestern coasts of the Atlantic Ocean. Histopathological examination of lesions within diseased clam tissue typically shows QPX cells surrounded by a thick layer of muco-polysaccharide secretions as it can be revealed by positive Alcian blue staining (Ragone Calvo et al 1998, Smolowitz et al 1998c). In addition, QPX cells cultivated *in vitro* produce a very dense mucus secretion containing the parasite's extracellular products (QPX ECP) (Perrigault et al 2009a). It has been suggested that the mucus layer surrounding QPX may function as a virulence factor by facilitating the parasite's infection of clam tissue and providing protection against the clam immune response by limiting clam hemocyte phagocytosis and encapsulation (Smolowitz et al 1998b). *In vitro* experiments on QPX showed that the mucoid layer surrounding the parasite provides protection against the antimicrobial activity of clam

plasma (Anderson et al 2003b). A preliminary study also showed that QPX ECP contain unidentified peptidases capable of breaking down proteins, and possibly used by the parasite to destroy clam connective tissue (Anderson et al 2006). In addition, exposure to QPX ECP causes death of clam hemocytes (Perrigault et al 2009a), suggesting that QPX ECP may also contain cytotoxic molecules. The specific biochemical composition of QPX ECP and the role of its specific components in QPX virulence remain unknown and require further investigation.

It is well documented that pathogenic organisms secrete molecules which enable them to invade their hosts, establish infection sites and cause damage to the host tissue (Kale & Tyler 2011, Kamoun 2006). These extracellular and surface molecules include hydrolyzing enzymes (peptidases, glycosidases, lipases), receptors, adhesins and other recognition and attachment proteins. *In silico* predictions of pathogenic secretomes are often used to evaluate the pathogenic capabilities of microorganisms (e.g. Raffaele et al 2010, Denoeud et al 2011, and Jiang et al 2013). Proteomics approaches such as mass spectrometry are also being successfully used to investigate extracellular proteins of pathogenic protozoans (e.g. *Trypanosoma brucei* (Geiger et al 2010) or *Leishmania donovani* (Silverman et al 2008)). In this study, we use similar approaches – bioinformatics prediction and mass spectrometry – to identify proteins secreted by QPX and to provide new insight into the molecular basis of QPX pathogenicity. We also investigated transcription levels of QPX putative virulence factors within the clam infection sites (aka clam nodules) to reveal their possible importance during the infection process.

Materials and Methods

In silico prediction of QPX secretome

The QPX transcriptome and peptide sequence libraries were generated as a part of the Marine Microbial Eukaryotic Transcriptome Project under the sample ID numbers: MMESTP0098 and MMETSP0100 (http://marinemicroeukaryotes.org/). The two transcriptome libraries were generated for QPX cells cultivated in 1) minimal essential medium (MEM) supplemented with 10% (v/v) fetal bovine serum (MMESTP0098) and 2) MEM supplemented with 10% FBS and 2% clam adductor muscle homogenates of 7 mg/ml protein concentration (MMETSP0100). For sample MMESTP0100, 12,579 transcriptomic sequences were used to find

the longest open reading frames (ORF) using the online Virtual Ribosome Tool Version 1.1 (Wernersson 2006). The ORFs were scanned for N-terminal signal peptide sequences using Signal P version 4.1 server: http://www.cbs.dtu.dk/services/SignalP/ (Bendtsen et al 2004b, Petersen et al 2011), and then the ORFs with a signal peptide sequence were searched for cell localization using the Protein Subcellular Localization Prediction tool WoLF PSORT, available from http://wolfpsort.org/ (Horton et al 2007). The sequences predicted to be secreted outside the cell were annotated using blastp search, gene ontology and conserved protein domains tools within the Blast2Go server tool (Gotz et al 2008). For sample MMETSP0098, peptide sequences provided by the MMETSP were used to find peptides possessing an N-terminal signal peptide sequence using the SignalP tool and to predict extracellular localization using the WoLF PSORT tool. Local blastp between the two datasets was performed using the Geneious bioinformatics tool (Drummond et al 2011) to find the final number of unique proteins secreted by QPX.

Mass spectrometry of QPX ECP

QPX cells isolated from a diseased clam collected from Raritan Bay, NY in 2003 (isolate NY0313808B, Qian et al 2007) were grown in 100 ml MEM supplemented with 0.3% of yeastolate (BD 255772). In addition, the media contained salts, HEPES and antibiotic solution according to previously described methods (Kleinschuster et al 1998). The culture was incubated at room temperature in a 400-ml BD Falcon tissue culture flask on an orbital shaker at 80 rpm for six days. At the time of collection the QPX cell concentration had reached 8×10^7 cells ml^{-1}. The culture was centrifuged at 2500 g and 4 °C for 70 min and cell-free supernatant or QPX extracellular product (QPX ECP) was collected. Because the QPX ECP includes a dense muco-polysaccharide secretion, a few steps were applied to facilitate mucus solubility. First, EDTA (10 mM final concentration) was added to the supernatant and incubated for two hours at room temperature to dissolve the EDTA-soluble fraction (Smith et al 2009). Second, the QPX ECP was further liquefied by passing it several times through a 10-ml syringe without a needle. Next, the suspected large (>100 kDa) polysaccharide fraction was separated from a smaller protein fraction by ultrafiltration through a 100 kDa column (Amicon, UFC910008). A five-ml subsample of the collected filtrate was further concentrated by a second ultrafiltration step

through 3 kDa columns (AmiconUFC500324) and a final lyophilization step. The final protein sample, a 50x concentrate of the original QPX ECP (by volume), was visualized by sodium dodecyl sulphate 12% polyacrylamide gel electrophoresis (SDS PAGE) stained with silver nitrate (**Figure 4.1**). A second gel stained with Coomassie blue G-250 (Bio Rad) was used for mass spectrometric identification of the protein content. The sample was submitted to the Stony Brook Proteomics Center for liquid chromatography/mass spectrometry analysis.

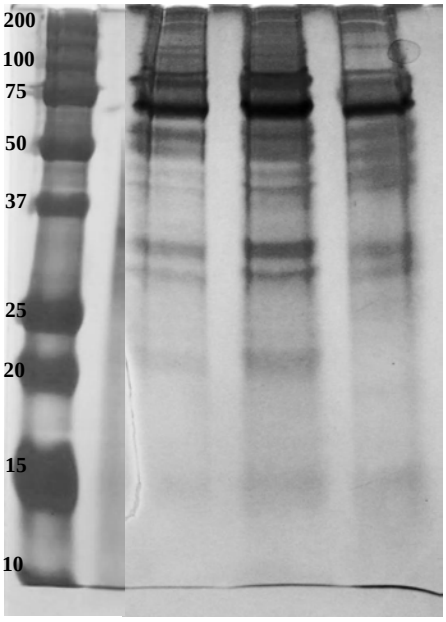


Figure 4.1: SDS PAGE silver staining of extracellular products (n=3) of quahog parasite unknown grown in minimal essential medium supplemented with 0.3% (w/v) yeastolate. One sample was analyzed by liquid chromatography / mass spectrometry.

In vivo vs. *in vitro* QPX transcript levels

Several of the identified extracellular proteins were investigated for their relative mRNA levels inside QPX-induced lesions within clam tissue. For this purpose, diseased clams were obtained from an aquaculture supplier located in Wellfleet (MA) and were collected during a mortality event in spring 2012. Five infection nodules were excised from the mantle edge of diseased clams (**Figure 4.2**) and kept frozen at -80°C until RNA extraction. To test the QPX primer specificity, one sample piece of mantle from a healthy clam was treated as a negative

control for QPX gene amplification. For the *in vitro* comparison, seven different QPX isolates (four NY and three MA isolates) were cultivated in clam adductor muscle homogenates adjusted to a total protein concentration of 2 mg/ml (Perrigault et al 2009a). The QPX strain selection included 2 NY isolates and 1 MA isolate which have been subcultured since 2003, and 2 NY and 2 MA isolates obtained from diseased clams collected from the field in 2012.

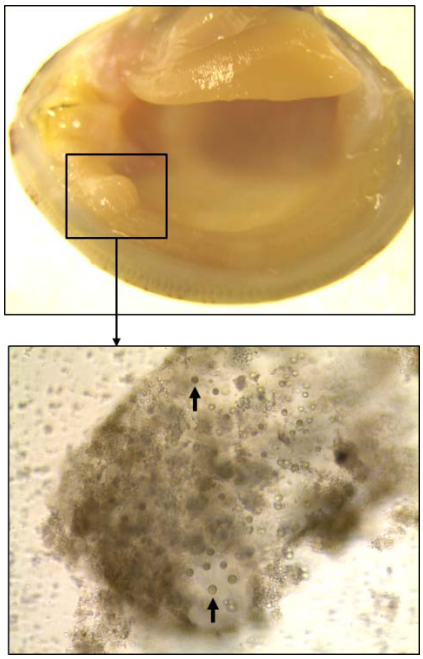


Figure 4.2: A QPX infection nodule at the edge of the mantle of *Mercenaria mercenaria;* arrows point to QPX cells viewed microscopically in a fresh biopsy preparation

Infected clam tissues (nodules) containing QPX cells and QPX cell pellets from *in vitro* cultures were separately homogenized in Trizol (Invitrogen) reagents using a mechanical homogenizer. RNA extractions were performed following the manufacturer's protocol (Invitrogen). Total RNA was precipitated using 7 M LiCl and kept at -20°C overnight. The RNA pellets were washed using 70% molecular grade RNase-free ethanol. cDNA was synthesized using Moloney-murine leukemia virus reverse transcriptase and its buffer (Promega) in reactions containing 0.5 µg of oligodT primers, 2 µg of total RNA, and 2mM final dNTPs concentration. Primers (all with the annealing temperature of 60°C) for the selected QPX transcripts were designed using the PrimerQuest design tool available from the Integrated DNA Technology

website. All possible primer pairs were mapped onto the corresponding full cDNA sequences within the Geneious software and then local blastn searches were performed against a clam cDNA library (Allam unpublished) to find possible clam sequences matching the QPX transcripts and to avoid selecting primers which would have more than 50% nucleotide sequence identity to any clam sequence. In addition, to ensure QPX primer specificity, the selected primers were tested using the PrimerBLAST NCBI tool against the non-redundant nucleotide database and were confirmed not to amplify any sequence other than the QPX transcripts of interest. Only primers which showed amplification of QPX genes and no amplification of uninfected clam cDNA were chosen, and the final list of selected genes and nucleotide sequences of their corresponding primers is presented **in Table 4.1.** Quantitative PCR reactions were carried out in 96-well plates using 10 µl volume reactions containing 1x Brilliant II Syber Green master mix (Agilent Technologies), 0.2 mM concentration of each primer, and 15 ng of cDNA. The PCR reactions were performed using the Mastercycler ep realplex PCR machine (Eppendorf) using the following temperature cycling program: an initial denaturation step at 95 °C for 10 min, 40 cycles of denaturation at 95 °C for 30 s and annealing plus extension at 60 °C for 1 min, and a 20 min melting curve cycle. The mRNA levels of the selected transcripts were normalized to the glyceraldehyde 3-phosphate dehydrogenase (GADPH) gene, and relative expression levels were calculated using the delta delta Ct method (Livak & Schmittgen 2001).

Table 4.1: Nucleotide sequences of primers used for qPCR estimate of transcript levels in QPX cells found *in vivo* relative to QPX cells grown *in vitro*

Gene name	Gene ID	Transcript ID		Primer Sequences
Glyceraldehyde 3-phosphate dehydrogenase	GADPH	MMETSP0100_6578	F:	GGAGATAACGGTGTGGAATAC
			R:	TTGACTCCCATAACGAACATAG
Mucin-related 1	Muc-1	MMETSP0100_3380	F:	CTTGCGACTGTGTGTAGTAG
			R:	CACATGCACTGGTGAGAATA
Mucin-related 2	Muc-2	MMETSP0098_3037	F:	GGAGATGCTGGTGCTAAAC
			R:	GTGGGTTGTAGTGGGATTTC
Mucin-related 3	Muc-3	MMETSP0100_3085	F:	AACTACACTTGCACCGACTAC
			R:	TGGAGAATGGGTTGGAGAATG
Mucin-related 4	Muc-4	MMETSP0098_18146	F:	GGTGCTACCAACTGGATTAGAG
			R:	GGAAGTCGTAAGGGTACTTTGG
Integrin	Integrin	MMETSP0098_115	F:	CAGTCGTGCAAATGGTAGGT
			R:	GTCAGCAGGGTTCTCAATAGTC
Lectin	Lectin	MMESTP0100_3133	F:	TTGAAGCAGGGCATGTTCACAAGG
			R:	ACCTTCGTGGAAACGGTAATGGGA
Heat shock protein 70	Hsp70	MMETSP0100_1067	F:	AACGGTAAGGAGCCAAGTAAG
			R:	GGGCAACATCGAGAAGAAGA
Hemolysin	Hemolysin	MMETSP0098_300	F:	CCACGCTCAACTCAACTATTA
			R:	ATTTCGTCATAGCCGTTCTC
Elicitin-like 1	El-like 1	MMETSP0098_3214	F:	ATGCTCCAACACCCAATC
			R:	CAACACTCGGGTTTCTAGTT
Elicitin like 2	El-like 2	MMETSP0098_17402	F:	CTTCACCAACAGGGTCTGTATC
			R:	CACAGTCGTAGCCCAGAAAG
Endotoxin	Endotoxin	MMETSP0098_1515	F:	ACTCATCATCGGACCTATCT
			R:	GACCCGTACCTCACATTAAC
Necrosis inducing protein	Nec Ind P	MMETSP0100_1986	F:	TAACAGTGGTGGGAGAGTAG
			R:	GCCGCCATGTTGGTAATA
Subtilase 1	S8-1	MMETSP0100_1574	F:	AAGAGCGGTTGGGAATATGGGAGT
			R:	AACAACAAGCATGCCCTCTTCTGC
Subtilase 2	S8-2	MMETSP0100_4754	F:	TATGGCTACTCCATTTGTCGCTGG
			R:	ACGAGCAAATTGGGAGATTCGTGC
Subtilase 3	S8-3	MMETSP0100_1744	F:	TTGATGCAAGGCGTTTCCGATGTC
			R:	GCGTCGATCGAAATGGCAAGTTGT
Serine carboxy-peptidase	S10	MMETSP0100_3910	F:	AAACAGCTTGCCGGGTTCATTGAG
			R:	AACTACAGCCTGGACCACCATTCA
Cysteine peptidase 1	C1-1	MMETSP0100_2357	F:	AAGGAGCAGGAAGCACTCGAGAAA
			R:	GCCAACTAAAGTTCGCAGGCAAGT
Cysteine peptidase 2	C1-2	MMETSP0098_1002	F:	AGCACAAGCCATGTAATCTCGGTG
			R:	TACCTTGGCAAAGCCGTTGTTTCC
Metallopeptidase	M35	MMETSP0100_1424	F:	ATCCACCAACAAGTACGCCAGAGA
			R:	GGTTGTTCAATGCAAGGGAGCGAA

Results

Predicted secretome of QPX

After scanning QPX peptide sequences using the SignalP bioinformatics tool, a total of 806 and 641 proteins with an N-terminal secretory signal peptide were identified in MMETSP0098 and MMETSP0100 libraries, respectively. Among these, 192 sequences were found to have 95% or higher amino acid identity between the two transcriptome libraries, representing the same proteins or highly similar isoforms. When the 806 signal peptide containing proteins from sample MMETSP0098 were tblasted against the entire MMETSP0100 transcriptome database (and vice versa) all sequences were found in both samples. Thus differences between the two samples reflect different numbers of transcript contigs with complete 5' UTR rather than different genes being expressed during growth in the two culture media. For samples MMETSP0098 and MMETSP0100, 522 and 400 proteins, respectively, were also predicted to be extracellular by WoLF PSORT. The blastp homology searches against the NCBI database returned blastp hit matches for 347 sequences (54%) for MMESTP0100 and 520 sequences (64%) for sample MMETSP0098. A summary of predicted extracellular proteins of QPX is presented in **Table 4.2**. The two groups represented by the highest number of proteins in the predicted QPX secretome were hydrolases and peptidases (**Table 4.2**). These two groups included enzymes commonly associated with pathogenicity such as lipases, phospholipases, hemolysins, subtilases, cathepsins, and disintegrins (**Table 4.2**). QPX was also predicted to secrete antioxidants such as thioredoxins, ascorbate (plant-like) peroxidase and glutathione peroxidase, and one superoxide dismutase. Other QPX proteins which contain an N-terminal signal peptide included heat shock proteins, lectins, cell membrane-bound receptors, glycotransferases, cyclophilins, fibronectin- and integrin-related proteins. In addition, the predicted secretome included five proteins possibly related to QPX virulence: elicitin-like proteins, endotoxins and a necrosis-inducing protein (**Table 4.2).**

Table 4.2: Number of proteins predicted to be secreted by QPX grouped by categories of common virulence factors

Group Name	SignalP v. 4.0	SignalP and WolfPSORT II	CONSERVED PROTEIN DOMAINS (InterPro database)		
Hydrolases					
Glycosyl hydrolases	34	18	IPR001547	IPR001139	
			IPR001382	IPR013781	
Lipases	14	13	IPR002921	IPR008139	
Phosphatases	14	12	IPR018946	IPR017849	IPR004843
Phospholipases	5	5	IPR017946	IPR007000	IPR002641
Hemolysins	5	4	IPR027018		
Chitinases	3	3	IPR011583		
Pectinase	1	0	IPR012334		
Peptidases					
serine	27	22	IPR000209	IPR001563	IPR001314
cysteine	23	22	IPR000668	IPR001096	IPR011697
metallo	15	11	IPR000834	IPR001506	IPR024079
aspartic	8	5	IPR001461	IPR021109	
theorine	2	2	IPR000426	IPR001353	
Antioxidants					
Thioredoxins	9	7	IPR005746		
Ascorbate peroxidases	2	2	IPR000823	IPR012336	
Thioredoxin reductase	1	1	IPR013027		
Glutathione peroxidases	1	1	IPR000889		
Superoxide dismutase	1	1	IPR001424		
Receptors					
GPCR receptors	11	0	IPR002455	IPR001828	IPR000337
Mannose-6-phosphate receptor	3	2	IPR009011		
Transmembrane receptor	2	0	IPR009637		
Rhodopsin-like GPCR	2	0	IPR019336		
MD-2 lipid-recognition domain	2	2	IPR003172		
Heat shock proteins					
heat shock proteins 40	7	4	IPR001623		
heat shock proteins 70	3	1	IPR013126		
heat shock protiens 20	1	1	IPR002068		
Cyclophilins	8	6	IPR002130	IPR001179	
Glycotransferases	7	3	IPR021067	IPR007657	
Fibronectin-related	6	2	IPR003961		
Integrin-related	5	5	IPR013519	IPR002035	IPR013517
Lectins	5	3	IPR005052	IPR001304	IPR013320
Elicitin like proteins	2	1			
Endotoxin	2	1	IPR005639		
Necrosis inducing protein	1	1	IPR008701		

Extracellular proteins identified by liquid chromatography / mass spectrometry

Fifty-six proteins present in QPX extracellular products (QPX ECP) were identified by mass spectrometry ($p<0.01$; minimum of 2 peptides matching predicted sequence, **Table 4.3**). Based on *in silico* predictors, 31 of the identified proteins possess the signal peptide for secretion via the classical vesicle-mediated pathway, and 14 proteins are possibly secreted via the non-classical pathway (**Table 4.3**). The remaining 11 proteins represent either intracellular contamination, open reading frames with missing N-terminal sequence, or proteins for which the currently available bioinformatics tools cannot predict their secretion pathway. Twelve of the extracellular proteins are unique to QPX as there are no similar sequences found in the NCBI database. Several proteins known to be involved in host-pathogen interactions were identified in QPX ECP using LC/MS analysis. These include six mucin-like or glycosylated proteins, eight proteolytic enzymes, five glycosyl hydrolases, an integrin-related protein, a lectin and a hsp70.

Table 4.3: Proteins identified in QPX extracellular secretions using liquid chromatography coupled to mass spectrometry

Protein name	Mass spectrometry results			Length	Predicted MW (kDa)[a]	SignalP D score[b]	SecretomeP 2.0
	peptides	spectra	coverage				
MUCUS COMPONENT							
mucin-like 1	13	94	25.4%			0.407	0.609
mucin-like 2	22	62	30.0%			0.400	0.151
mucin-like 3	7	67	17.7%	277	27.87	0.554	
mucin-like 4	16	74	14.3%	760	80.83	0.691	
mucin-like 5	9	79	12.2%	784	80.82	0.439	0.585
mucin-like 6	9	24	26.8%	331	34.70	0.542	
extracellular matrix protein	10	17	20.1%	726	78.22	0.288	0.523
ADHESION AND RECOGNITION							
integrin-related	20	96	51.8%	440	47.25	0.819	
lectin	3	3	11.5%	434	49.40	0.558	
coagulation factor	2	3	3.4%	591	64.99	0.683	
ANTIOXIDANT ACTIVITY							
glutathione reductase	4	6	9.9%	482	51.53	0.171	0.685
oxireductase	15	36	35.5%	537	58.35	0.679	
peroxidase	2	3	9.8%	348	38.18	0.151	0.851
PEPTIDASE ACTIVITY							
subtilase 1	18	140	48.3%	400	41.46	0.537	
subtilase 2	5	9	17.5%	407	43.73	0.777	
subtilase 3	3	3	17.2%	402	42.38	0.722	
serine carboxypeptidase 1	15	36	36.1%	457	50.78	0.702	
papain like peptidase 1	10	15	11.2%	546	60.82	0.688	
papain peptidase 2	4	8	12.6%	556	60.95	0.655	
metallopeptidase M35	3	6	4.9%	508	55.99	0.800	
metallopeptidase M12B	6	14	10.5%	902	99.58	0.567	
HYDROLASE ACTIVITY							
glycosidase beta-amylase	19	53	41.4%	464	52.61	0.612	
beta-glucosidase/glucanes	15	81	56.6%	318	34.80	0.481	
glycosyl hydrolase 1	14	43	46.8%	449	50.26	0.682	
glycosyl hydrolase 2	2	7	4.0%	447	49.25	0.703	
acetylesterase	8	12	21.8%	499	55.86	0.458	
aminohydrolase	2	3	6.3%				
adenosylhomocysteinase	4	7	10.2%	481	52.86	0.098	0.483
chitolectin/chitinase	2	7	1.7%	1259	139.22		

Table 4.3 continued. Proteins identified in QPX extracellular secretions using liquid chromatography coupled to mass spectrometry

Protein name	Mass spectrometry results			Length	Predicted MW (kDa)[a]	SignalP D score[b]	SecretomeP NN score[c]
	peptides	spectra	coverage				
OTHER							
aminotransferase 1	9	18	31.6%	395	42.67	0.106	0.514
aminotransferase 2	2	2	2.2%	420	46.82	0.180	0.628
aminotransferase 3	11	33	34.9%	397	43.17	0.223	0.550
dihydrodipicolinate synthase	10	14	50.2%	305	32.22	0.235	0.585
heat shock protein 70	3	5	6.8%	664	72.61	0.659	
triosephosphate isomerase 1	13	39	57.2%	258	27.74	0.105	0.546
triosephosphate isomerase 2	6	7	17.3%	256	27.69	0.156	0.549
2-hydroxyacid dehydrogenase	15	28	49.7%	426	46.96	0.179	0.724
inorganic pyrophosphatase	3	5	13.6%	274	31.23	0.135	0.621
chitin binding domain	15	48	28.9%	398	43.24	0.324	0.231
WD repeats (G-protein)	8	16	33.1%	315	35.16	0.110	0.433
dihydrolipoamide dehydrogenase	20	44	58.2%	502	53.58	0.135	0.442
ribose-5-phosphate isomerase	4	5	22.9%	319	34.13	0.316	0.321
phosphoenolpyruvate carboxykinase	4	6	10.3%	575	63.53	0.140	0.415
transaldolase	4	4	19.8%	323	36.02	0.104	0.286
unknown 1	12	40	47.7%			0.603	
unknown 2	5	27	22.8%	347	36.87	0.483	
unknown 3	7	13	26.3%	413	46.08	0.766	
unknown 4	4	5	7.9%	850	94.79	0.824	
unknown 5	2	3	10.2%	381	40.40	0.606	
unknown 6	11	20	69.2%	246	27.74	0.690	
unknown 7	5	13	9.1%	616	69.11	0.566	
unknown 8	15	58	52.0%	352	38.75	0.478	
unknown 9	13	45	14.0%	1178	130.95	0.530	
unknown 10	3	6	6.7%	476	53.12	0.709	
unknown 11	3	4	1.6%				
unknown 12	4	24	6.4%	422	46.17	0.390	

[a]- molecular weight predictions using Geneious v. 5, missing values for incomplete coding sequences
[b]- SignalP v. 4.1; classically secreted proteins, with D score above 0.45; missing values for sequences without 5' end
[c]- SecretomeP v. 2.0 - non-classically secreted proteins should obtain an NN-score exceeding the normal threshold

of 0.5, but not at the same time be predicted to contain a signal peptide; missing values for sequences with D score above 0.45

Six extracellular proteins identified in the QPX ECP had amino acid motifs known to be sites for the attachment of carbohydrate chains. They all contain repeating sequence motifs rich in proline, threonine, serine and alanine residues which comprise between 36 and 58% of the total amino acids in these proteins (data not shown). The blastp results show similarity of the proline threonine (PT) rich fragments to many different molecules in a variety of organisms including: a LPXTG-motif cell wall anchor domain protein of *Lactobacillus reuteri* (WP_003664141), an elicitin-like protein of the oomycete *Phytophthora medicaginis* (ABH11745), a proline-threonine-rich repeat protein of the fungus *Trichophyton rubrum* (XP_003237106), and a mucin of *Dictyostelium fasciculatum* (XP_004360105) (**Figure 4.3**).

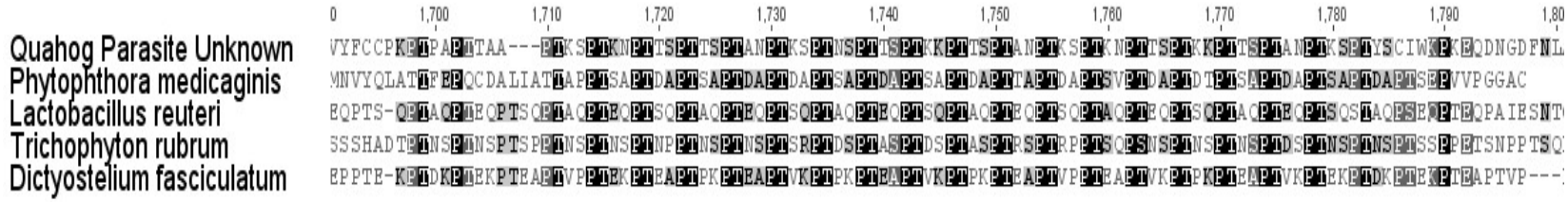


Figure 4.3: Conserved PT - proline, threonine residues in a partial amino acid sequence alignment of mucin-related proteins from four different organisms QPX mucin, *Phytophthora medicaginis* (ABH11745), *Lactobacillus reuteri* (WP_003664141), *Trichophyton rubrum* (XP_003237106) and *Dictyostelium fasciculatum* (XP_004360105)

The mass spectrometry data of QPX ECP also revealed that QPX secretes a legume type lectin, an integrin-related protein, and a heat shock protein of about 70 kDa molecular weight (**Table 4.3**). All three proteins were predicted to be secreted via the classical ER/Golgi vesicle mediated secretion pathway (**Table 4.2**). QPX extracellular lectin is similar to the leguminous plant lectins named L-type lectins (InterPro protein domain number: IPR005052). The integrin related protein contains the conserved protein repeat (IPR013519, IPR013517 protein domains) found in alpha integrins. QPX ECP also was shown to contain three proteins having oxido-reductase activities which might participate in antioxidant processes: flavin dependent oxireductase, peroxidase and glutathione reductase (**Table 4.3**).

QPX was also found to secrete eight different peptidases including four serine type peptidases, two cysteine peptidases and two metallopeptidases (**Table 4.3**). Three of the serine peptidases belong to the subtilase family (S8 family, IPR015500 and IPR000209), one serine peptidase is a carboxypeptidase (S10 family, IPR001563), and the two cysteine peptidases belong to the papain family (C1 family, IPR000668). The two metallopeptidases (IPR024079) could not be assigned into specific families. Eight different types of hydrolases were also identified in the QPX secretome within which the three most abundant (based on number of spectra) were QPX extracellular beta-amylase (or glycosyl hyrolase family 14), beta-1,3-glucanase (glycosyl hydrolase family 17) and unassigned family glycosyl hydrolase (**Table 4.3**).

In vivo vs. *in vitro* QPX transcript levels

To reveal the possible involvement of selected extracellular QPX molecules during interactions with its clam host, relative abundances of their mRNA were compared between parasite cells in clam tissue lesions (*in vivo*) and *in vitro* cultures. Transcript levels of 19 QPX extracellular proteins which showed similarity to virulence factors of other pathogenic microorganisms or were hypothesized to be involved in host-pathogen interactions were selected and estimated using quantitative real time PCR (**Table 4.1**). The expression of six extracellular proteins was shown to be significantly higher for QPX cells inside the infection nodules than for QPX cells cultivated *in vitro* (t-test, $p < 0.05$, **Figure 4.4**). Higher expression was recorded for two mucin-like proteins, two peptidases, the hemolysin E and the elicitin-like protein (**Figure 4.4**). The expression of one additional peptidase (S10) was also higher for the *in vivo*

QPX,however the difference was not statistically significant due to high variability between the nodule samples (**Figure 4.4**).

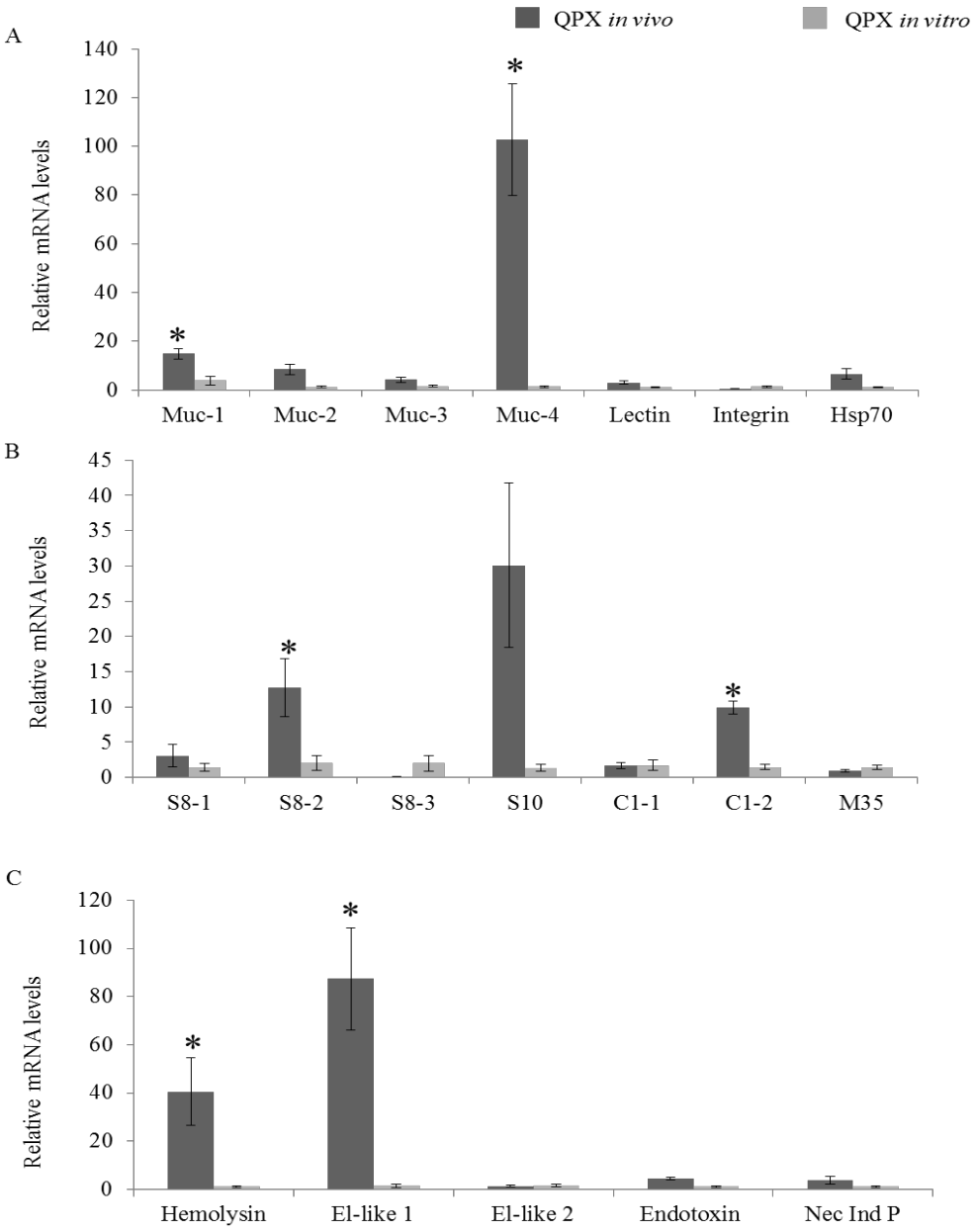


Figure 4.4: Relative mRNA levels of selected protein of QPX cells *in vivo* vs. QPX cells *in vitro* (A, B - selected proteins found to be extracellular by mass spectrometry; C – *in silico* predicted to be extracellular). Plotted are means ± standard error (n=5 for *in vivo* samples, n=7 for *in vitro* samples); statistically significant difference ($p < 0.05$; t-test) labeled by asterisks.

Discussion

Quahog parasite unknown (QPX) was predicted to secrete 1255 proteins outside its cell based on the presence of a eukaryotic signal peptide at the N-terminal of proteins. The number of potentially secreted proteins by QPX is within the same range as predictions for other stramenopilan pathogens such as *Phytophtophtora infestans*, *Saprolegnia parasitica* and *Pythium ultimum* which respectively have 1415, 970 and 747 proteins predicted to be secreted based on the presence of eukaryotic signal peptide (Jiang et al 2013, Levesque et al 2010, Raffaele et al 2010). To predict the secretome of QPX, we used SignalP v. 4, which in addition to the previous version's cleavage site score, C, and the signal peptide score, S, contains an additional discriminant score, D (Petersen et al 2011). The D score has been created to discriminate between secreted and non-secreted sequences (D score $\geq$0.450) and also exclude sequences which are predicted to contain N-terminal transmembrane helices (Petersen et al 2011). Despite these improvements, the predicted secretome of QPX still contains many cell membrane-bound proteins including G-protein coupled receptors and transmembrane transporters. In the present study we also used an additional tool, WoLF PSORT to predict localization of QPX protein sequences possessing a signal peptide. Based on this tool, only 548 proteins were predicted to be extracellular and include many hydrolytic enzymes which comprised about 21% of QPX predicted secretome. Using LC/MS analysis only 31 out of the 1255 predicted proteins containing signal peptides were found in culture supernatant. This most likely reflects QPX's response to the particular media conditions at which it was cultivated for the analysis. Media composed of only hydrolyzed yeast proteins was added to the QPX culture to avoid foreign protein contamination during sample analysis. It is likely that the number and type of proteins actually secreted by QPX cells changes in response to different environmental or growth conditions occurring outside QPX cells. These 31 extracellular proteins represented over half of the proteins found by LC/MS analysis, suggesting that a large fraction of QPX extracellular proteins is secreted via the classical eukaryotic mechanism. Some extracellular proteins are known to be transported via a non-conventional secretion pathway which does not require the presence of a signal peptide (Nickel & Rabouille 2009). For some kinetoplastid pathogens, the majority of extracellular proteins does not possess a signal peptide and are released by a micro-vesicle based secretion system (Cuervo et al 2009, Geiger et al 2010, Silverman et al 2008). SecretomeP software (Bendtsen et al 2004a) was used to determine if some of the QPX proteins

which were found to be extracellular by LC/MS analysis but did not contain an N-terminal signal peptide were secreted via a non-classical pathway (Geiger et al 2010). The analysis revealed that 14 QPX extracellular proteins might be transported via non-classical mechanisms. However, SecretomeP has been created based on conserved amino acid motifs found in mammalian proteins (Bendtsen et al 2004a) and might not reflect the secretion mechanisms in protists.

QPX has a very similar composition of its predicted secretome to other pathogenic stramenopiles which have been described to secrete elicitin-like proteins, necrosis-inducing proteins, phospholipases, glycosyl hydrolases, peptidases, acid phosphatases, peptidyl-prolyl-cis-trans isomerases, lectins and glycotransferases (Denoeud et al 2011, Jiang et al 2013, Levesque et al 2010, Raffaele et al 2010). Interestingly, *S. parasitica* is the only oomycete predicted to secrete hemolysins (Jiang et al 2013), which were also predicted to be secreted by QPX and both of these stramenopiles are infective to animal hosts. QPX's hemolysins show 18.8% amino acid sequence similarity to hemolysin E described from bacterial species (Wallace et al 2000). A number of different types of unrelated hemolysins are known from bacteria and protozoan pathogens and they exhibit cytolytic and/or cytotoxic activity against a wide range of host cells (e.g. erythrocytes, granulocytes, monocytes and endothelial cells). The expression of one QPX hemolysin was up-regulated in QPX cells derived from infection sites inside clam tissue showing that this enzyme plays a role during the infection process. However, its molecular function and possible cytolytic activity against clam hemocytes requires additional, targeted studies.

The elicitins and the necrosis-inducing proteins of pathogenic oomycetes are known to cause necrotic cell death or evoke an immune response when they are administered to their plant hosts. Necrosis-inducing protein sequences are found in bacteria, fungi and oomycetes, but the elicitins and elicitin-like protein are apparently exclusive to the oomycetes (Jiang et al 2006). In this study, we have identified two QPX sequences sharing between 33% -36% amino acid identities with the elicitin RAM6 from *Phytophthora ramorum* (ABB55989). Some of the elicitin-like proteins have C-terminal extensions rich in threonine, serine and proline residues as an indication of extensive glycosylation (Jiang et al 2006). The higher expression of one QPX elicitin-like protein inside the infection nodules indicates its possible involvement in QPX pathogenicity.

QPX was predicted to secrete a protein containing a conserved domain coding for delta endotoxin N (IPR005639). The only endotoxin N which has been characterized in the literature

is derived from *Bacillus thuringiensis;* the toxin causes lysis of the epithelial cells of the gut of infected insects (Boncheva et al 2006, Ito et al 2006, Srinivasan 2008). The presence of delta endotoxin in the QPX transcriptome might be a consequence of horizontal gene transfer (HGT) from a prokaryotic donor. There are examples of HGT facilitating evolution of parasitism in oomycetes (Richards et al 2011). QPX also possesses a predicted extracellular protein which contains a conserved protein domain named necrosis-inducing protein (IPR008701, NPP1). The NPP1 purified from a few *Phytophthora* pathogens has been shown to cause necrotic cell death on tested plant host species (Fellbrich et al 2002, Qutob et al 2002). The transcripts of QPX's endotoxin and the necrosis-inducing protein were not up-regulated inside the infection nodules. These results might reflect either a lack of importance of these factors in QPX virulence, their need at different time of the infection, or a lack of transcriptional regulation of these molecules. Post-translational protein modification or other means of protein activation might, for instance, be more important than the transcriptional regulation of these factors.

The present study used mass spectrometry to identify 56 different proteins in QPX mucus secretion. Important components of QPX secretion are mucin-like glycoproteins which can be identified by the presence of proline and threonine-rich motifs. The PT or SPT rich protein motifs are sites of extensive glycosylation, or in other words attachment of glycans or sugars (Lang et al 2007). Mucins is a general term referring to a type of glycoprotein of high molecular weight and heavy glycosylation which shows similarity to mammalian glycoproteins secreted by epithelium cells and forming a mucous protective layer (Jain et al 2001). Three of QPX's extracellular glycoproteins have 50% of the amino acid content composed of proline, threonine, serine and alanine which is a characteristic of mammalian mucins (Jain et al 2001b). Mucin-like molecules found in protozoan cells are either secreted into extracellular space or attached to the cell membrane by glycosylphosphatidylinositol (GPI) anchors (Guha-Niyogi et al 2001, Hicks et al 2000, Jain et al 2001). The main role of mucins and mucin-like molecules in parasitic protozoa such as the kinetoplastid (*Trypanosoma*, *Leishmania*), apicomplexan (*Cryptosporidium*) and amoebic (*Entamoeba*) parasites, is cell surface protection facilitating establishment of infection sites (Jain et al 2001). In the present study, the expression of two (out of four tested) QPX mucin-like molecules was increased under *in vivo* conditions in comparison to *in vitro* cultures, suggesting that QPX mucins may also play a role during the infection process and are likely involved in the protection of QPX cells against clam hemocyte phagocytosis or

encapsulation. These conclusions are in agreement with previous findings reporting that the mucus layer surrounding QPX cells provides protection to the parasite against clam defense factors (Anderson et al 2003).

Mucin-like molecules are also involved in other host-parasite interactions such as attachment to host cells and tissues but the means of carbohydrate chain interactions (Hicks et al 2000). Some parasites which are known to secrete mucin also secrete mucin degrading enzymes which might be used by the parasite to degrade the protective lining of epithelial tissue (Hicks et al 2000). Extracellular glycosyl hydrolases and extracellular glycosyltransferases have been hypothesized to participate in cross-linking between glycoproteins and polysaccharides of the pathogen and the host (Denoeud et al 2011). In this study QPX has been shown (LC/MS) to secrete five glycosyl hydrolases and has been predicted (*in silico*) to secrete over 30 glycosyl hydrolases and 7 glycotransferases, which should be subjected to further investigation for their potential roles in QPX infection of clam tissue.

The transcript level of the extracellular QPX integrin-related protein was investigated using qPCR (**Figure 4.4A**). Integrin-mediated adhesion is the main adhesion mechanism of metazoan cells (Hynes 2002). However, integrin-like molecules have also been shown to mediate adhesion in the protozoan *Dictyostelium discoideum*, which secretes a protein containing integrin and a von Willebrand factor type A domain (Cornillon et al 2006). Investigation of the *Entamoeba histolytica* beta-integrin receptor showed that it functions in adhesion (function similar to the integrins of metazoans) to the human fibronectin of the extracellular matrix (Talamas-Rohana et al 1998). Similarly, a number of apicomplexan pathogens possess adhesive thrombospondins, surface molecules containing integrin-related protein domains (Harper et al 2004, Pereira et al 2011, Yuda et al 1999). QPX is predicted to secrete five different proteins which contain conserved amino acid motifs commonly found in integrins. Additional studies are required to test the role of QPX integrins and their possible roles in QPX adhesion or recognition of clam cells and extracellular matrix.

Similarly to the integrin-related protein, the expression of one QPX extracellular lectin was not regulated in QPX cells from tissue lesions in comparison to QPX cells from cultures Lectin is a very general term describing carbohydrate-binding proteins. They bind either mono- or oligosaccharides reversibly and with high specificity and each lectin molecule contains one or more carbohydrate-binding sites, i.e., they are mono-, di- or polyvalent (Lis & Sharon 1998).

They are structurally diverse proteins whose main function is binding to carbohydrates, fulfilling the role of recognition (as receptors) or of attachment (as adhesins). Lectins of pathogenic organisms are often virulence factors for their role in the initiation of infections. In pathogenic bacteria, some fimbriae (or pili) which are important for the attachment and initiation of infection sites are partially composed of lectins (Sharon 2006). In addition to the one lectin molecule investigated in this study by qPCR, QPX has four more lectin-like molecules predicted to be secreted. Pathogenic protozoans secrete lectin-like molecules that interact with host cells and host carbohydrates (Hicks et al 2000). The predicted extracellular lectins are excellent targets for future investigations of factors involved in the adhesion to, and interactions with, host tissues.

The LC/MS analysis showed that eight different peptidases are secreted by QPX when it is cultivated in one type of media. The mRNA levels of two of these peptidases (one subtilisin and one papain-like peptidase) were higher for QPX cells present in infection nodules suggesting possible involvement of these peptidases in QPX interactions with the host. In fact, subtilisins are universal virulence factors secreted by bacterial, fungal and protozoan pathogens (Kennan et al 2010, Monod et al 2002, Muszewska et al 2011, Withers-Martinez et al 2004). The papain-like cysteine peptidases are the most studied peptidases of parasitic protozoa and the family contains some of the best described virulence factors such as falcipain of *Plasmodium falciparum* and cruzain of *Trypanosoma cruzi* (Atkinson et al 2009, Lecaille et al 2002). The specific mechanisms of the extracellular subtilisin and papain-like peptidases of QPX infection inside the clam tissue would have to be investigated but the data from this study suggests that they are important in disease development. Extracellular peptidases are major constituents of QPX predicted secretome (74 different peptidases have been predicted to possess a signal peptide) so it is very likely that additional peptidases also play a role in QPX virulence, warranting further studies.

Chapter 5: Identification and characterization of extracellular peptidases secreted by quahog parasite unknown (QPX), the protistan parasite of hard clams

Manuscript in preparation for Molecular and Biochemical Parasitology

Abstract

Quahog parasite unknown (QPX) is a protistan parasite causing deadly infections in the hard clam, *Mercenaria mercenaria*, one of the most valuable shellfish species in the USA. It is an extracellular parasite found mostly in the connective tissue of clam mantle and, in severe cases of infection, other clam organs. Histopathologic examinations report that QPX cells within clam tissues are typically surrounded by hollow areas which have been hypothesized to be a result of extracellular digestion of clam proteins by the parasite. The present investigation documents peptidase activity in QPX extracellular secretions revealed using sodium dodecyl sulfate-polyacrylamide gels containing gelatin as a co-polymerized substrate. Multiple peptidase activity bands of molecular weights ranging from 20 to 100 kDa can be detected in QPX secretions derived from a variety of culture media. One major band of 30 kDa is composed of subtilisin-like peptidases which are released by QPX cells in all studied media, showing that these are the most common peptidases used by QPX for nutrient acquisition. PCR quantification of mRNA encoding QPX subtilisins showed that their expression is regulated at the transcriptional level. A FPLC-purified fraction containing a subtilisin type serine peptidase was able to digest clam plasma proteins suggesting that this peptidase could serve as a virulence factor of the parasite.

Introduction

Peptidases are enzymes that hydrolyze proteins and represent important virulence factors of pathogenic and parasitic organisms (Klemba & Goldberg 2002, McKerrow et al 2006, McKerrow et al 1993). Both extracellular and lysosomal peptidases have been shown to be important factors for pathogenicity of bacterial (Miyoshi & Shinoda 2000), fungal (Monod et al 2002) and protozoan invaders (Alvarez et al 2012, Carruthers 2006). Their major function is to digest food proteins for nutrients, but as the pathogenic organisms usually live inside host tissue or cells, these peptidases damage host-derived proteins, thus facilitating infection and spread. Pathogens can secrete peptidases and hydrolyze host proteins extracellularly or take up host-derived proteins via endocytosis and use lysosomal peptidases for digestion. In pathogenic

protozoa, peptidases break down the components of the host extracellular matrix, such as collagen or fibrin, and components of blood and the immune system, such as hemoglobin, proteins of the complement system and immunoglobulins (Klemba & Goldberg 2002, Lecaille et al 2002, McKerrow et al 1993).

QPX (quahog parasite unknown) belongs to the thraustochytrids, which are mostly osmo-heterotrophic and saprophytic organisms (Raghukumar 2002). It has been speculated that QPX mainly uses extracellular digestion to acquire nutrients, and possibly uses extracellular peptidases to digest clam tissues during infection (Anderson et al 2006). Our previous investigations showed that the QPX transcriptome contains over 200 sequences encoding for a variety of peptidases of all major peptidase types: serine (SPs), cysteine (CPs), aspartic (APs), threonine (TPs) and metallopeptidases (MPs) (Rubin et al 2014). Over 70 of these peptidases have been shown to possess an N-terminal signal peptide thus are directed into the secretory pathway (the present thesis, Chapter 4). In addition, eight of these peptidases were detected in QPX extracellular products (ECP) using liquid chromatography coupled with mass spectrometry (LC/MS) (the present thesis, Chapter 4).

The present study was designed to characterize the extracellular peptidases of QPX. The investigations included the identification of extracellular peptidases using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS PAGE) in the presence of gelatin substrate co-polymerized in the gels. Peptidase activities in QPX secretions derived from a variety of QPX cultures were compared to determine if peptidase secretion varies in response to external protein sources. In addition, extracellular peptidase activity was compared among three different QPX isolates. Ultrafiltration and ion exchange chromatography were used partially purify QPX secretion and the protein fractions were examined using clam plasma substrate polyacrylamide gels.

Materials and Methods

In silico identification of QPX extracellular peptidases

The identification of peptidases predicted to be secreted by QPX cells was accomplished as described previously (this thesis, Chapter 4). Briefly, translated open reading frames from two QPX transcriptome libraries (11,005 and 12,579 ORFs) were screened for the presence of an N-terminal signal peptide which is known to direct the peptide into the secretory pathway (Petersen

et al 2011). A total of 1255 unique proteins were predicted to constitute QPX secretome and these included 74 putative extracellular peptidases (this thesis, Chapter 4). In the current study, the translated amino acid sequences of these 74 peptidases were further grouped into peptidase families based on classification by the peptidase database, MEROPS (Rawlings et al 2012). Amino acid sequence alignments of peptidases belonging to the same family were generated using ClustalW algorithm plug-in within the Geneious v. 7.1.4 software (Drummond et al 2011). Percent amino acid sequence identities between peptidases belonging to the same families, as well as their expected molecular weights, were calculated using the same software.

QPX cultures and media

To compare peptidase transcription expression of QPX cells in response to different protein supplements, the following four media were prepared: minimal essential medium supplemented with fetal bovine serum (MEM FBS), MEM supplemented with yeastolate (MEM YSTO), MEM supplemented with gelatin (MEM GEL), and MEM supplemented with clam adductor muscle homogenate (MEM CAMH). The protein concentrations in all protein supplements were estimated using Pierce BCA Protein Assay Kit (Thermo Scientific) and adjusted to the concentration of 2 mg/ml with sterilized artificial seawater (Instant Ocean).For the experimental cultures, one volume of MEM was mixed with one volume of one of the four protein supplements resulting in a final protein concentration of 1 mg/ml. The cultures were prepared in 12-well plates filled with 4 ml media in triplicate for each treatment. The QPX cells inoculum (NY1, isolate NY0313808BC7, Qian et al 2007) was prepared from an exponentially growing culture in MEM and FBS. The cells were collected by centrifugation (3000 g for 15 min at RT), washed twice with sterile artificial seawater and re-suspended in un-supplemented MEM. Cell concentration of QPX inoculla was determined with a hemocytometer and the starting concentration of QPX was 8×10^3 cells ml^{-1} in all cultures. QPX growth was monitored by cell counts using a hemocytometerevery other day, and QPX cells were collected when they reached similar cell concentrations: MEM FBS-2.3 $\times10^6$ cells ml^{-1}, MEM GEL-2.1×10^6 cells ml^{-1}, MEM YST-2.0 $\times10^6$ cells ml^{-1}, MEM CAMH -1.7×10^6 cells ml^{-1}. To collect the cells and QPX secretion (media supernatant), one volume of QPX culture was diluted with an equal volume of sterilized artificial seawater and passed several times through a syringe to facilitate liquefaction of QPX extracellular mucus. The mixtures were transferred into 15-ml conical tubes and

centrifuged at 3000 g for 20 minutes at 20 °C. The supernatant was collected into 1.5-ml tubes as 40-µl aliquots. To each aliquot 10 µl of 5x Laemmli loading dye, containing SDS but no beta-mercapto-ethanol, was added and the aliquots were stored in a -80 °C freezer until zymogram analyses. The cells were used for RNA extractions (see below) which were conducted on the day of collection. As the zymogram analysis of QPX secretions from the described above cultures revealed low or no peptidase activities (see results below) additional cultures of the same QPX isolate were prepared but with a higher concentration of protein:: MEM supplemented with 10% (v/v) FBS (standard media) and in clam adductor muscle homogenate (CAMH), adjusted to 3 mg/ml of protein. QPX cultures were incubated for 7 days at 20 °C before the supernatant was collected and preserved as described above.

For the comparison of peptidase activities among different QPX isolates, three NY isolates were cultivated in MEM supplemented with 10% (v/v) gelatin hydrolysate, yeastolate, and peptone solution prepared in seawater each at 10 mg/ml (w/v) and sterilized by autoclaving (MEM GYP). This media formulation was used because it was shown to be optimal for extracellular peptidase production revealed on zymograms in preliminary experiments. Two different QPX isolates originating from Raritan Bay, NY (NY0313808BC7 and NY0314220AC6, here designated NY-1 and NY-2 respectively) and one QPX isolate (NY070348D, or NY-3) originating from a different NY embayment (Peconic Bay) were selected for the peptidase expression comparison. QPX cultures (in triplicate) were established in 25-ml canted neck culture flasks (Falcon) with 7 ml of media and 0.3 ml of parasite cells suspended in sterile artificial seawater at 8.6×10^5 cells*ml^{-1}. All cultures were incubated at 20 °C for 11 days when the cell concentration reached 3.2×10^6 for NY-1, 3.5×10^6 for NY-2 and 1.4×10^6 for NY-3 isolate (mean, n=3). The cells and supernatants were collected as described above.

Zymography

SDS PAGE was carried out in the presence of gelatin as a substrate copolymerized in the gels (zymography, Lantz and Ciborowski, 1994) to detect proteolytic activity in QPX secretions. Gelatin type A from porcine skin (G9136, Sigma) was incorporated (0.1 %) in either 10 or 12% polyacrylamide gels. Supernatant samples containing the loading dye (see previous section) were thawed and not boiled to retain non-denatured structure of the peptidases. 20µl of each sample

was loaded into each lane. The gel electrophoresis was run at a constant 125 V in a standard Leammli Tris-glycine SDS buffer. To remove the SDS, the gels were washed twice in 2.5% Triton-X100 in Tris-HCl buffer (pH = 8.0) for 30 minutes on a rotating platform (80 rpm) at room temperature. To allow for the enzyme digestion of gelatin, the gels were incubated in 0.1 M Tris-HCl, 2 mM Mg_2Cl, 2 mM $CaCl_2$ buffer (pH = 8.0) at 37 °C overnight. For inhibition assays, the gels were incubated overnight in the same 0.1 M Tris-HCl buffer (pH = 8.0) containing one of the following eight inhibitors: ethylenediaminetetraacetic acid (EDTA, 10618973, Fisher Scientific), trans-epoxy succinyl amido (4-guanidino) butane (E-64, E3132, Sigma), chymostatin (C7268, Sigma), phenylmethylsulfonyl fluoride (PMSF, P7626, Sigma), aprotinin (A1153, Sigma), tosylphenylalanylchloromethane (TPCK, T4376, Sigma), N- A-tosyl-l-lysine Chloromethyl Ketone Hydrochloride (TLCK, T7254, Sigma), or antipain (291907, Santa Cruz Biotechnology). All inhibitors were used at a final concentration equal to the maximal effective concentration suggested by the manufacturers: 10 mM PMSF, 1 mM EDTA, 10 μM E-64, 100 μM chymostatin, 0.8 μM aprotinin, 100 μM TPCK, 100 μM TLCK, 100 μM antipain as. After incubation, the gels were stained for 10-15 minutes with a solution containing 0.23 % (w/v) coomassie blue, 5.8% (v/v) glacial acetic acid and 30% methanol (v/v) in water and destained for 1 hour in 10% acetic acid and 10% methanol..

Differential transcript expression of QPX peptidases

A total of 11 sequences were selected to investigate their expression patterns in response to the four media and among the three QPX isolates. These included five different sequences coding for peptidases belonging to the subtilisin-like family (labeled S8-1 to S8-5) and six different sequences coding for peptidases from the papain-like family (labeled C1-1 to C1-6). These two families were selected because their members are known virulence factors in other pathogenic organisms. In addition, previous data shows that some of the S8 and C1 peptidases are secreted by QPX cells (this thesis, Chapter 4). The housekeeping gene beta actin was used for mRNA normalization because its expression was previously shown not to change with different growth conditions (Rubin et al 2014). All primers were designed using the Primer 3 plug-in in the Geneious software to amplify a PCR product between 100 and 200 bp in length (**Table 5.1**). For genes having similar nucleotide sequences (max identity 78% between S8-3 and S8-5), the

final primer pairs were selected to be located in the least conserved regions of the nucleotide alignments.

Table 5.1 Nucleotide sequences of PCR primers used to quantify mRNA levels of QPX peptidases

Gene ID	qPCR primer sequences			
S8-1	F:	TCGTGCTGGACACATAGTTGTCGT	R:	TATCGGTGGCTCCAACGCTTATCA
S8-2	F:	TATGGCTACTCCATTTGTCGCTGG	R:	ACGAGCAAATTGGGAGATTCGTGC
S8-3	F:	AAGAGCGGTTGGGAATATGGGAGT	R:	AACAACAAGCATGCCCTCTTCTGC
S8-4	F:	TTGCCGGTGTATTGGCTACGCTTT	R:	CTCGAGGTTTGCACCAACCAGTTT
S8-5	F:	TTGATGCAAGGCGTTTCCGATGTC	R:	GCGTCGATCGAAATGGCAAGTTGT
C1-1	F:	ACGGCAATGTTACCGAAGAGGCTA	R:	TAGTTATCCAATGGGCCCGCGTTA
C1-2	F:	ACTGGAGCAAGAAGGGAGCAGTAA	R:	AAGACCACCTGTGGTGGAGAAACT
C1-3	F:	ACACAAGTCCGAAACATGCTCTGC	R:	TACCGTGAGCGGTCCGTATTTGAT
C1-4	F:	TGCAGGTCGTCGTTGCTTTAGTCT	R:	TAGCCAACGATTGAAACTGCGTGG
C1-5	F:	AAGGAGCAGGAAGCACTCGAGAAA	R:	GCCAACTAAAGTTCGCAGGCAAGT
C1-6	F:	AGCACAAGCCATGTAATCTCGGTG	R:	TACCTTGGCAAAGCCGTTGTTTCC
ACTIN	F:	TGAAGATCTTGACCGAGCGTGGTT	R:	AGCGGTCTTCATCTCCTGGTCAAA

Trizol reagent (Molecular Research Center, Inc.) and the manufacturer protocol were used to isolate RNA from all samples, and RNA quality and quantity were estimated spectophotometrically by Nandrop. Total RNA (2.0 μg) from each sample was used to synthesize cDNA using the Moloney Murine Leukemia Virus Reverse Transcription kit (MMLV-RT, Promega) and 0.5 μg of oligo dT18 primers following the manufacturer protocol. Relative quantification was carried out in 10 μl reactions with Brilliant II SYBR green qPCR master mix (Agilent), 100 nM final primer concentration, and 5 μg of RNA-equivalent cDNA. The PCR reactions were performed using a Mastercycler ep realplex PCR machine (Eppendorf). The peptidase expression levels were normalized to the beta actin gene and relative transcript levels were calculated using the delta delta Ct method (Livak & Schmittgen 2001).

Purification of QPX subtilisin

The QPX isolate NY070348D (NY3) was cultivated in minimal essential medium (MEM) supplemented with 10% (v/v) gelatin hydrolysate, yeasolate, and peptone solution (each at 10 mg/ml (w/v)) in sterilized artificial seawater. Six 70-ml flasks were filled with 15 ml of medium and inoculated with 1 ml of starting culture (8-days-old QPX culture in the same medium). The cultures were incubated at 20 °C for 10 days. For mucus collection, 10 ml of

sterilized artificial seawater was added to each 16 ml culture and mixed with a syringe to facilitate liquefaction of muco-polysaccharides. The total of 156 ml seawater and QPX culture mixture was distributed into 15-ml conical Falcon tubes and centrifuged for 15 min at 3000 g at room temperature. The supernatant was collected into new 15 ml tubes and centrifuged for another 15 min to remove any remaining cells. Cell-free supernatant was passed through 15-ml 100 kDa ultrafiltration devices (Amicon) to remove QPX muco-polysaccharides, and the flow-through was concentrated using 3 kDa ultrafiltration devices (Amicon). The concentrated sample of 2 ml volume was dialyzed three times against 20 mM Tris HCl buffer at pH = 8.0. The sample was then fractionated by anion exchange fast-flow chromatography using an FPLC (AKTA system). The sample was loaded onto a Q-Sepharose fast-flow column (Q-FF 1 ml, Amersham Biosciences), previously equilibrated with 20 mM Tris-HCl, pH 8.0 (buffer A). After washing the column with buffer A, the elution was carried out with an increasing gradient of NaCl mixed with Tris HCl buffer (buffer B). The gradient conditions were as follows: 20% buffer B for 15 min, 45% buffer B for 40 min and 100% buffer B for 10 min. FPCL fractions identified by protein spectrophotometric signals were separated using (SDS-PAGE) and peptidase activity of each fraction was revealed on gelatin and clam plasma substrate zymograms. The clam plasma substrate gels were prepared by substituting water in the gel mix with clam plasma pool which was withdrawn with a needle and a syringe from adductor muscles of three different clams collected from Raritan Bay, NY.

Results

In silico identification of QPX extracellular peptidases

Over 200 transcripts coding for a variety of peptidases were identified in two different QPX transcriptome libraries (Keeling et al 2014, Rubin et al 2014). Based on the classification of the MEROPS comprehensive peptidase database (Rawlings et al 2012), QPX possesses intra- and extracellular peptidases belonging to38 different peptidase families and all major catalytic types including metallo-, serine, cysteine-, aspartic- and threonine peptidases. Further amino acid sequence analyses and alignments of all peptidase transcripts revealed that QPX is predicted to possess 74 peptidases with an N-terminal signal peptide possibly secreted outside of its cell

(**Table 5.2**). The predicted extracellular peptidases of QPX belong to 16 different peptidase families (**Table 5.2**) with three main families represented by the highest number of transcripts: C1, S8 and S10 (**Table 5.2**). Eight of the 74 predicted peptidases have been experimentally confirmed to be secreted by QPX cells cultivated in minimal essential medium supplemented with 0.3% (w/v) yeastolate (this thesis, Chapter 4). These eight extracellular peptidases include: three subtilisin-like peptidases (MEROPS family: S8), two papain-like peptidases (C1), one carboxypeptidase (S10), and two metallopeptidases (M35 and M12B) (this thesis, Chapter 4, marked in grey -**Table 5.2**).

Table 5.2: Seventy four QPX peptidases containing an N-terminal secretion peptide sequences, grouped into families based on classification by the MEROPS peptidase database; members of families shaded in grey were previously detected by LC/MS

	MEROPS FAMILY	FAMILY NAMES	NO.	Calculated molecular weight (kDa)[a]	INHIBITORS
SERINE	S1	Trypsins/chymotrypsins	6	29, 30, 36, 41, 50, 74	Aprotinin, TLCK, PMSF
	S8	Subtilases/subtilisins	7	39-43 (6), 77	DFP, PMSF
	S10	Carboxypeptidases	11	41, 49-52 (7), 55, 62, 67	Chymostatin, antipain, PMSF, DFP
	S28	Pro-Xaa carboxypeptidase	3	52, 54, 75	Lys thiazolidide
	S54	Rhamboid peptidases	1	31	TPCK and 3,4-DCI
CYSTEINE	C1	Papain-like peptidases	18	N/A, 22, 35, 36, 39, 45, 47, 48, 49, 52, 53(2), 56, 60, 61(2), 70, 75	E-64, leupeptin
	C13	Legumain-like peptidases	2	N/A, 49	iodoacetamide, N-ethylmaleimide
	C26	Gamma-glutamyl hydrolase	3	34, 37, 38	azaserine, 6-diazo-5-oxo-L-norleucine
METALLO	M6	Immune inhibitor A peptidase	1	83	1,10-phenanthroline, EDTA
	M10	Metzincins	1	N/A	1,10-phenanthroline, EDTA
	M12	Astacins/ADAMS/adamalysins	6	41, 47, 53, 57, 86, 100	1,10-phenanthroline, EDTA
	M14	Carboxypeptidases A	3	44, 52, 83	1,10-phenanthroline, EDTA
	M20	Aminohydrolases	1	55	1,10-phenanthroline, EDTA
	M35	Deuterolysins	2	49, 56	1,10-phenanthroline, EDTA
ASPARTIC	A1	Pepsins	6	46, 50, 53, 66, 97, 173	Pepstatin
	A22B	Presenilin	1	42	Pepstatin
THREONINE	T1	Proteasome peptidases	2	26(2)	No inhibitors

[a] – molecular weight (kDa) calculated using Geneious v. 7.1.4

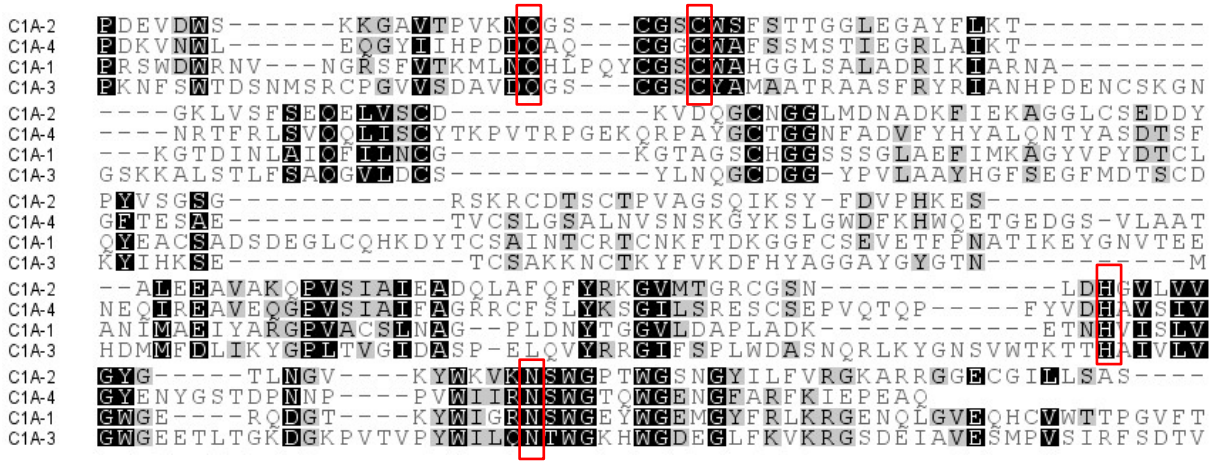
() – values between parentheses represent the number of peptidases with the preceding molecular weight

N/A – complete open reading frame not available

Based on currently available QPX transcriptome libraries, the highest number of QPX putative extracellular peptidases (18 different transcripts, **Table 5.2**) belongs to the C1 family, also known as the papain-like peptidase family. The assignment of peptidase family is based on sequence similarity matches, however not all QPX peptidases have the expected active site residues. Peptidases in the C1 family contain three conserved amino acid residues known as the catalytic triad: cysteine, histidine and asparagine (C-H-N) (Atkinson et al 2009, Lecaille et al 2002, Sajid & McKerrow 2002). The sulfhydryl group (-SH) of the cysteine residue is involved in the hydrolysis (Sajid & McKerrow 2002). The three active residues were only detected in 14 out of the 18 C1 peptidases, suggesting that the remaining four might be non-peptidase homologs. Ten of the C1 peptidases of QPX were classified into the C1A subfamily (also known as cathepsins), which is characterized by four active residues with an additional glutamine preceding the active triad (Q-C-H-N, **Figure 5.1A**) (Sajid & McKerrow 2002).

The second highest number of predicted extracellular peptidases belongs to the serine carboxypeptidase family (S10 family). QPX possesses eleven different S10 peptidases (**Table 5.2**) of predicted molecular weight in a range from 41 to 67 kDa (**Table 5.2**).The third group of extracellular peptidases predicted from QPX transcriptome databases is the S8 serine peptidase family. Peptidases of the S8 family have a catalytic triad in the order aspartic acid (D), histidine (H) and serine (S) (**Figure 5.1B**), which is a different order to that of other serine peptidases such as the S10 family which has the same catalytic triad in the order S, D and H (alignment not shown). QPX possesses seven subtilisin-like peptidases (S8) containing a signal peptide (**Table 5.2**), and the calculated MW of six S8 peptidases was between 39 and 43 kDa. One of the S8 peptidases has a C-terminal extension rich in proline-threonine residues, suggesting it is anchored to the cell surface or is glycosylated. The amino acid sequence alignment of five QPX subtilisins revealed between 36 and 79% pairwise nucleotide identity (**Figure 5.1B**).

A)



B)

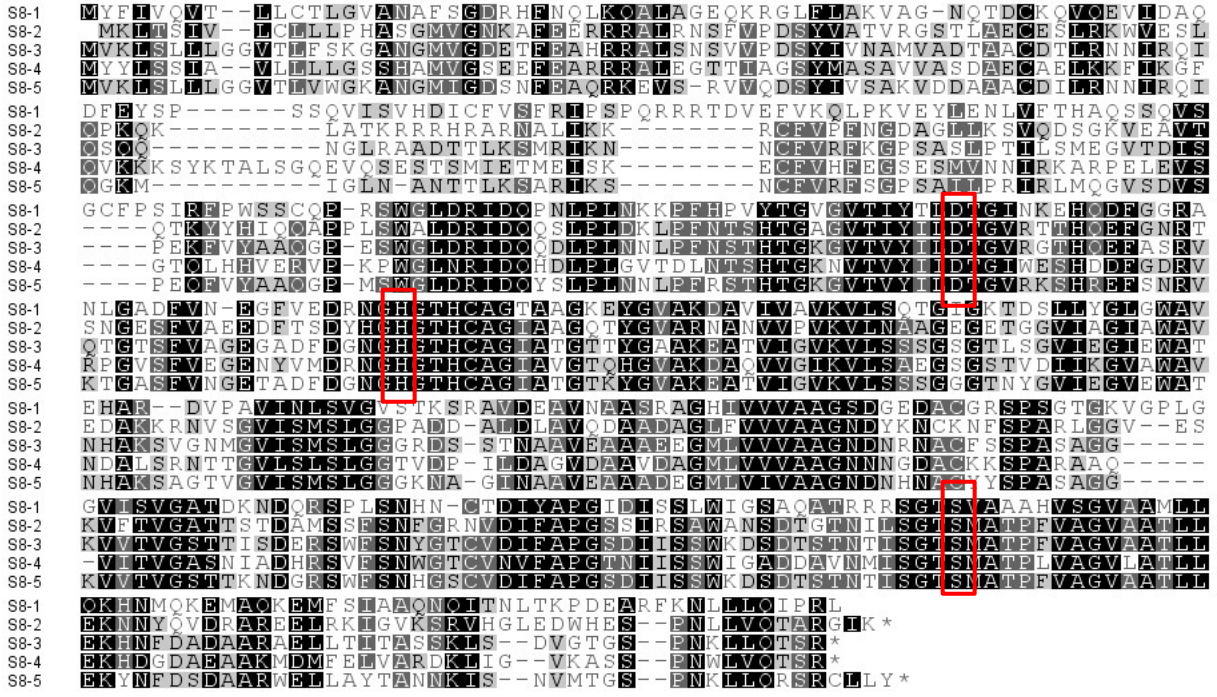


Figure 5.1: Amino acid alignments of predicted QPX extracellular peptidases A) papain-like family (C1) and B) subtilisin-like family (S8); Red boxes mark the catalytic residues in each family: Q-C-H-N for the C1 family and D-H-S for the S8 family

Peptidase activity in QPX secretions

The supernatants from the experimental cultures were analyzed for peptidase activity by gelatin substrate SDS PAGE. Different proteolytic activity bands were detected in the secretions of QPX cells grown in different media for one isolate (NY1) and among three different QPX isolates cultured in one media **(Table 5.3)**. When QPX cells were cultivated in MEM supplemented with one of four different protein sources (gelatin, yeastolate, FBS and CAMH) at the final protein concentration of 1 mg ml^{-1}, only one peptidase activity band (30 kDa) was revealed for only one media, MEM and FBS (**Figure 5.2A**). Two bands of peptidase activity (30 and 40 kDa) were detectable in QPX secretion derived from CAMH cultures (without the MEM, 3 mg ml^{-1} final concentration) (**Figure 5.2B**). The maximum number of peptidase bands (6 in total) was revealed when QPX (NY1) was grown in MEM FBS. However, different MW bands were revealed from two separate sets of cultures: subculture A - five peptidase bands from MEM FBS culture adjusted at 3 mg ml^{-1} of total protein concentration (**Figure 5.2B**) and subculture B - 3 bands from MEM FBS culture at 4 mg ml^{-1} (**Figure 5.2C**). The difference in the subcultures could have been caused by the usage of newly purchased FBS stock.

The most pronounced/thickest band in MEM FBS media was a band around 40 kDa molecular weight (MW) which was partially inhibited by two serine peptidase inhibitors phenylmethylsulfonyl fluoride (PMSF) and chymostatin (**Figure 5.2C**). This band was not inhibited by the cysteine peptidase inhibitor, E-64 (**Figure 5.2D**) or the metallo-peptidase inhibitor, EDTA (**data not shown**) suggesting that it is primarily composed of serine peptidases. The band is most likely composed of multiple peptidases because it contains at least one peptidase resistant to PMSF and chymostatin inhibition (**Figure 5.2C**). In addition, the band was not inhibited by any other tested serine peptidase inhibitors (TLCK, TPCK, aprotinin and antipain; data not shown), suggesting the band is mostly composedof subtilisin-type (S8 family) peptidases. The other two bands (20 and 30 kDa) which were present in QPX secretion from MEM FBS (at 4 mg/ml protein concentration) cultures were also inhibited by the serine peptidase inhibitors PMSF and chymostatin (**Figure 5.2C**). These two low MW peptidase bands were markedly stronger when the media was supplemented with higher concentration of FBS (3 mg ml^{-1} and 4 mg ml^{-1}, **Figure 5.2B vs. 5.2C**). The two peptidase bands of lowest MW were not revealed when frozen supernatant was used, preventing their further identification using inhibitors.

When QPX cells were cultivated in minimal essential medium supplemented with gelatin, yeastolate and peptone, each at the final concentration of 1 mg/ml or 3 mg/ml total (w/v) the extracellular peptidase production was at its highest, represented by the brightest activity bands on the zymograms (**Figure 5.3A**). Three different QPX isolates were grown in this medium and the ECP derived from those cultures differed in the peptidase activity band pattern (**Table 5.3 and Figure 5.3A**). NY2 QPX had a peptidase band running at approximately 75 kDa MW which was not present in ECP from NY1 and NY3 QPX isolates, but possibly corresponding to high molecular bands of about 90 kDa and 100 kDa were present in the secretion of NY1 and NY3 isolates, respectively (**Figure 5.3B**). All peptidases were partially degraded after freezing (**Figure 5.3A vs. 5.3B**) and the two peptidase bands of MW just below 37 kDa – present for all three isolates - were inhibited by PMSF (**Figure 5.3B vs. 5.3C**) but not by cysteine peptidase inhibitor E-64 or metallopeptidase inhibitor EDTA (**data not shown**).

Table 5.3: Summary of peptidase activity bands detected in QPX secretion on multiple zymograms; QPX secretion was derived from different media for NY1 isolate and from three different QPX isolates (NY1, NY2, NY3) grown in the same media; MEM – minimal essential medium, YST – yeastolate, FBS – fetal bovine serum, GEL – gelatin, CAMH – clam adductor muscle homogenate, GYP- mix of gelatin hydrolysate, yeastolate, and peptone; protein concentration in the media given in mg/ml

MW (kDa)	MEM with				CAMH	MEM FBS		MEM GYP		
	YST	FBS	GEL	CAMH				NY1	NY2	NY3
	1mg/ml	1mg/ml	1mg/ml	1mg/ml	3 mg/ml	3 mg/ml	4 mg/ml	3 mg/ml	3 mg/ml	3 mg/ml
20							∨	∨	∨	∨
30		∨			∨	∨	∨	∨	∨	∨
40					∨	∨	∨	∨	∨	
75						∨			∨	
90						∨		∨		
100						∨				∨

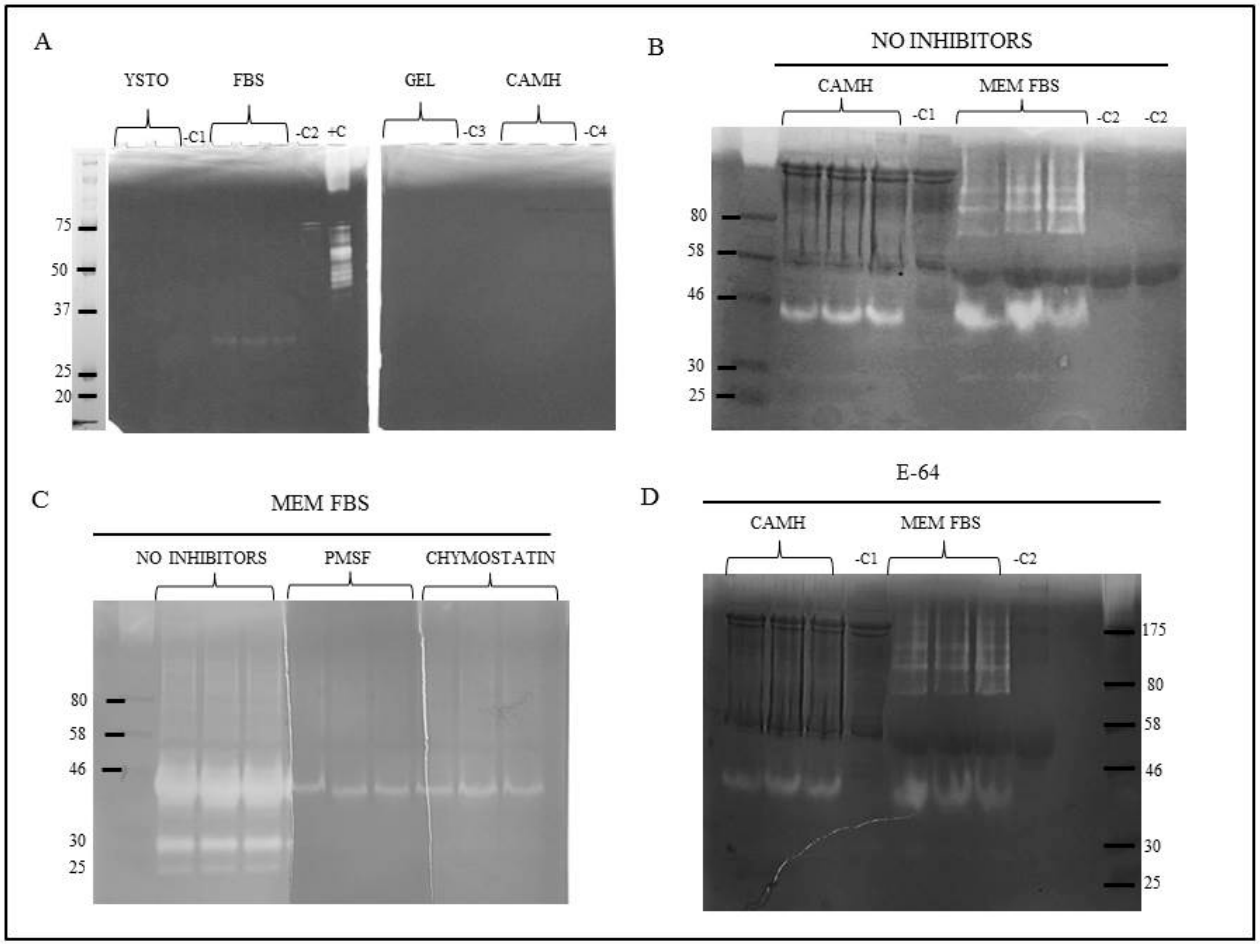


Figure 5.2: Zymograms showing extracellular peptidases produced by QPX (NY1-isolate NY20038BC7) grown in: A) minimal essential medium (MEM) supplemented with either yeastolate (YSTO), fetal bovine serum (FBS), gelatin (GEL) or clam adductor muscle homogenate (CAMH) each to the final protein concentration of 1 mg ml^{-1}; B) in MEM FBS and CAMH with protein concentration adjusted to 3 mg ml^{-1}; C) Effect of peptidase inhibitors (10 mM PMSF and 100 μM chymostatin on extracellular peptidases produced by QPX cells grown in MEM FBS at total protein concentration of about 4 mg ml^{-1} (in triplicates); D) Effect of E-64 (cysteine peptidase inhibitor) at 10 uM final concentration on peptidases in QPX ECP samples from part B; -C1, -C2, -C3 and –C4 are media incubated in parallel with the cultures but without QPX cells

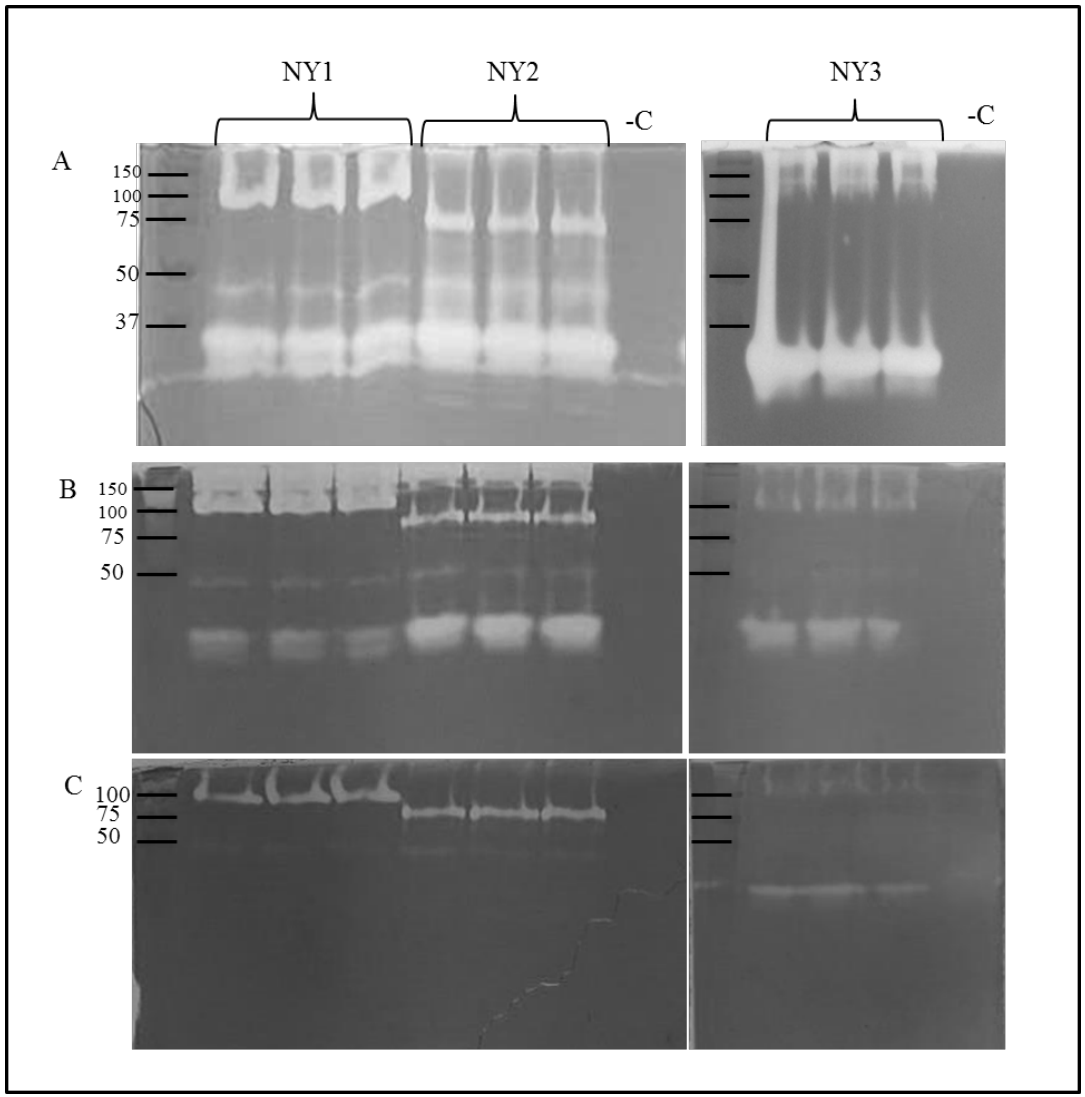


Figure 5.3: Extracellular peptidase activity on 0.1% gelatin zymograms of three QPX isolates (NY1 – NY20038BC7, NY2 – NY200320AC6, NY3-NY20073408D) cultivated in minimal essential medium supplemented with gelatin hydrolysate, peptone, and yeastolate (each at 1 mg/ml final concentration); A: fresh ECP without inhibitors; B: ECP frozen for 7 days at -80ºC without inhibitors; C: with 10 mM serine peptidase inhibitor PMSF

Differential transcript levels of QPX peptidases

The relative transcript levels of 11 peptidases were examined among three QPX isolates including two from Raritan Bay, NY (NY1, NY2) and one from Peconic Bay, NY (NY3) and for one isolate (NY1) cultivated in four different media (**Figure 5.4 A-D**). The examined peptidases included six papain-like peptidases (C1-1 to C1-6) and five subtilisin-like peptidases (S8-1 to S8-5). Five peptidases belonging to these two families (S8-2, S8-3, S8-5, C1-5 and C1-6) were shown to be secreted into the culture media by QPX cells by LC/MS methodology (this thesis, Chapter 4), and the other genes were selected for comparison. The expression of four subtilisins varied between media with generally a few-fold higher expression in QPX cells grown in minimal essential medium supplemented with fetal bovine serum (**Figure 5.4B**). However, the data was statistically significant for only one subtilisin (S8-1, **Figure 5.4B**) because of the high variability between the replicates. In addition, the expression of QPX extracellular papain-like peptidase (C1-5) was higher in QPX cells grown in the media supplemented with clam adductor muscle homogenate (**Figure 5.4A**, ANOVA, $p<0.05$ Tukey's pairwise comparisons). The abundance of mRNA encoding wo papain-like (C1-3 and C1-4) and three subtilisin-like peptidases (S8-2, S8-3 and S8-4) differed among three QPX isolates (ANOVA, $p<0.05$, Tukey's pairwise comparisons; **Figure 5.4C-D**).

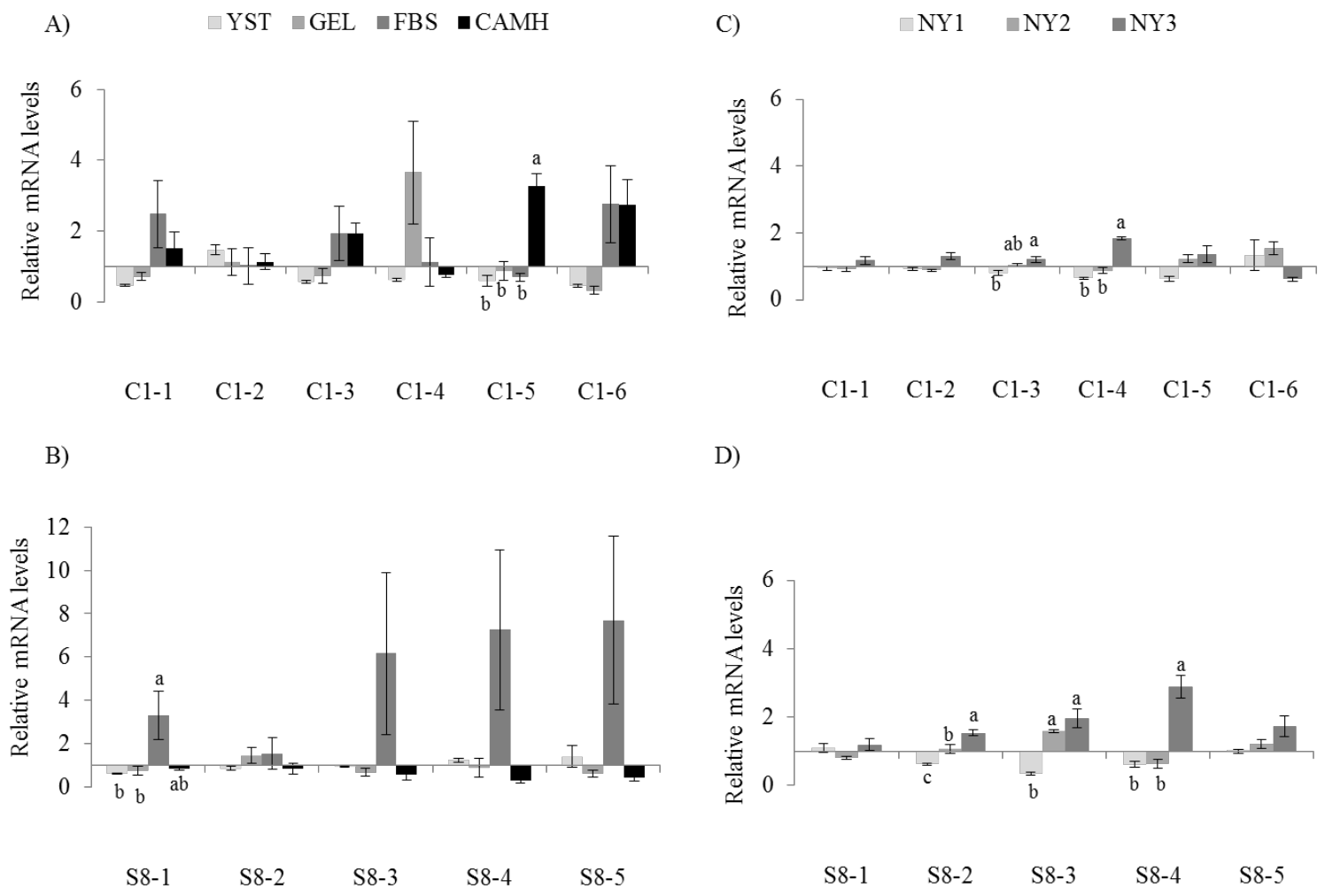


Figure 5.4: Differential mRNA levels of papain-like: C1-1 to C1-6 and subtilisin-like: S8-1-S8-5 peptidases in QPX cells grown in minimal essential medium (MEM) with addition of one of four different protein supplements (A-B: YST – yeastolate, GEL-gelatin, FBS-fetal bovine serum, CAMH-clam adductor muscle homogenate; at a final concentration of 1 mg/ml) and in QPX cells of three different isolates (C-D: NY1, NY2, and NY3) cultivated in MEM supplemented with gelatin hydrolysate, yeastolate and peptone (1 mg ml^{-1} each resulting in 3mg ml^{-1} combined protein concentration). Different letters (a, b, c) designate statistically significant differential gene expression (ANOVA, $p<0.05$, Tukey post hoc analysis).

Purification of QPX extracellular peptidase and LC/MS analysis

The peptidase activity band (35 kDa MW) which was detectable in ECP from NY3 QPX isolate was found to be resistant to PMSF inhibition and to freezing temperatures (**Figure 5.3C**) so it was chosen to be purified using ion exchange chromatography column and FPLC. Proteins in FPLC fractions 12 to19 were revealed on SDS PAGE (**Figure 5.5A**) and peptidase activities in these fractions were tested on gelatin- and clam plasma-incorporated polyacrylamide gel (**Figure 5.5B and 5.5C**). The thickest band (35 kDa) in fraction 19 was submitted for LC/MS analysis (Stony Brook Proteomics Center) which showed that the band contains a subtilisin type peptidase (**S8-3**, **Figure 5.6**)

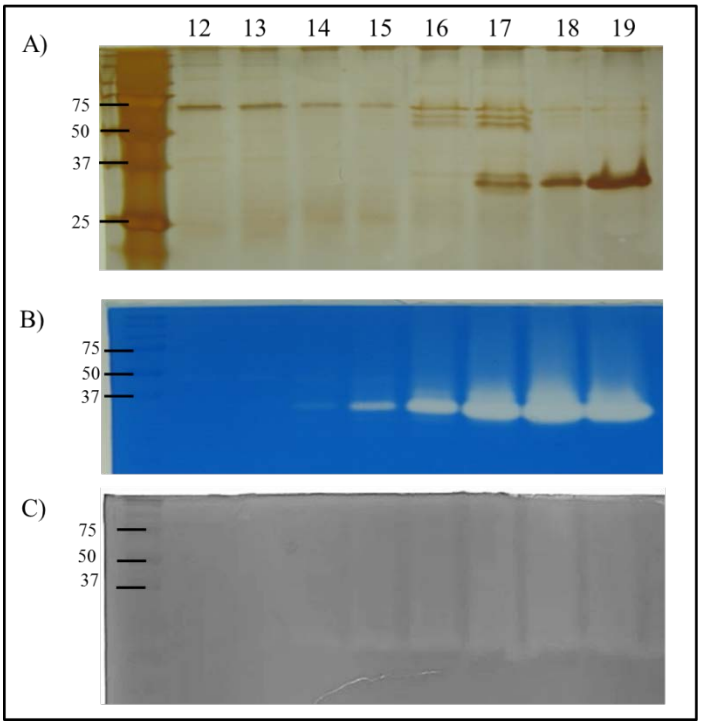


Figure 5.5 Fractions of QPX ECP purified by ion exchange chromatography (12-19) with protein content revealed on SDS-PAGE gel with silver nitrate stain (A) and peptidase activity revealed on gelatin- and clam plasma- (pool from 3 clams) incorporated gels (B and C, respectively)

QPX peptidase S8-3, length=400 aa
9 peptides, 14 spectra, 34.3% coverage

```
  1 MVKLSLLLGG VTLFSKGANG MVGDETFEAH RRALSNSVVP DSYIVNAMVA
 51 DTAACDTLRN NIRQIQSQQN GLRAADTTLK SMRIKNNCFV RFKGPSASLP
101 TILSMEGVTD ISPEKFVYAA QGPESWGLDR IDQQDLPLNN LPFNSTHTGK
151 GVTVYILDTG VRGTHQEFAS RVQTGTSFVA GEGADFDGNG HGTHCAGIAT
201 GTTYGAAKEA TVIGVKVLSS SGSGTLSGVI EGIEWATNHA KSVGNMGVIS
251 MSLGGGRDSS TNAAVEAAAE EGMLVVVAAG NDNRNACFSS PASAGGKVVT
301 VGSTTISDER SWFSNYGTCV DIFAPGSDII SSWKDSDTST NTISGTSMAT
351 PFVAGVAATL LEKHNFDADA ARAELLTITA SSKLSDVGTG SPNKLLQTSR
```

Figure 5.6: Number of peptides identified by LC/MS for the QPX serine peptidase, subtilisin (S8-3)

Discussion

In this study, bioinformatics tools and zymography were used to identify and describe peptidases secreted by quahog parasite unknown (QPX), the thraustochytrid parasite infecting the hard clam, *Mercenaria mercenaria*. Bioinformatics tools predict that QPX is able to secrete 74 different peptidases, all of which possess an N-terminal signal peptide sequence for the classical eukaryotic secretory pathway, classified into 16 families: (Table 5.2, MEROPS database; Rawlings et al (2012)). Members of at least six peptidase families, including A1, S1, S8, S10, C1,and M35 are known virulence factors of pathogenic organisms (Klemba & Goldberg 2002, McKerrow et al 2006, McKerrow et al 1993, Monod et al 2002). Three of these families are represented by the highest number of transcripts: C1, the papain-like peptidases family; S8, the subtilisin-like peptidase family; and S10, also known as the serine carboxypeptidase family.

The cysteine peptidase family C1 (papain-like peptidases) is the most studied family in the context of virulence factors. Most known parasite-derived peptidase sequences belong to the C1 peptidase family (Atkinson et al 2009). These include hemoglobinolytic falcipains from *Plasmodium* spp. and cruzain from *Trypanosoma cruzi* (Atkinson et al 2009). Secreted C1 peptidases also determine pathogenesis of *Enteamoeba histolytica*, which uses these enzymes to digest the extracellular matrix of human tissue, including collagen, elastin, fibrinogen, and laminin as well as immunoglobulin A (Que & Reed 2000). The S8 family, or subtilisins, are universal virulence factors of many pathogenic bacteria (Bonifait et al 2011, Brown et al 2000, Kennan et al 2010), protozoa (Hruzik et al 2011, Jean et al 2003, Miller et al 2001, Swenerton et al 2010, Wanyiri et al 2009) and fungi (Huang et al 2004, Moser et al 1994, Withers-Martinez et

al 2004). The S10 carboxypeptidases are secreted serine peptidases of some pathogenic fungi, including *Aspergillus niger* and *Aspergillus oryzae* (Monod et al 2002), but are only lysosomal in pathogenic kinetoplastids such as *T. cruzi* (Alvarez et al 2012, Parussini 2003).

In this study, QPX wasshown to secrete multiple extracellular peptidases which were revealed by gelatin digestion incorporated in SDS polyacrylamide gels. A peptidase band of approximately 35 kDa MW was produced by all three QPX isolates and in all types of media, suggesting that it is composed of the primary peptidases used by QPX for nutrient acquisition. Based on the effect of peptidase inhibitors, the band is composed mostly of subtilisin-type peptidases, as it was inhibited by phenylmethylsulfonyl fluoride (PMSF) and chymostatin, but not inhibited by other types of serine peptidase inhibitors. The inhibition assay was in agreement with the analysis by LC/ MS of that band from NY3 isolate which showed the presence of one subtilisin. Three different subtilisins were also detected in NY1 isolate in previous mass spectrometry analysis of ECP (this thesis, Chapter 4).

Based on QPX transcriptome sequences, QPX is predicted to also secrete aspartic, cysteine and metallo-peptidases which were not revealed on the zymogram used in this study. Only one type of substrate was used (gelatin) and the peptidase activity was developed in pH=8 buffer. Preliminary tests using different pH of the developing buffer showed no peptidase activity except for the slightly alkaline buffer. Gelatin is a commonly used substrate to detect multiple peptidases because it is composed of partially digested collagen so it contains small peptides with different amino acid sequences (Lantz & Ciborowski 1994). However, additional peptidases might be revealed on gels containing other protein substrates such as casein, fibrinogen, and elastin (Bertolini & Rohovec 1992, La Peyre et al 1995). It is also possible that the other peptidases are secreted in low quantities, not detectable by this technique. Previous LC/MS analysis of QPX secretion revealed an additional serine carboxypeptidase (S10), two papain-like peptidases (C1) and two metallopeptidases (M12B, M35) secreted by QPX cells cultivated in MEM supplemented only with yeastolate (this thesis, Chapter 4).

This study also shows that QPX cells produce different types and amounts of peptidases in response to different protein content in media. In general, QPX cells cultivated in minimal essential medium supplemented with a mix of gelatin hydrolysate, yeastolate and peptone had overall higher extracellular peptidase production than QPX cells cultivated in MEM supplemented with fetal bovine serum or clam adductor muscle homogenate. In addition, QPX

cells cultivated in MEM supplemented with 3 or 4 mg ml^{-1} of FBS showed higher production of extracellular peptidases than cultures with 1 mg ml^{-1} of FBS for which only one peptidase band was detected.

This study also shows that the production of subtilisin-type peptidases is regulated on the transcriptional level. The transcript levels of subtilisins were higher in QPX cells cultivated in the media supplemented with FBS in comparison to other media. However, one of the extracellular papain like peptidases (C1-5) had increased transcript levels in QPX cells cultivated in media containing clam adductor muscle homogenate. The expression of the same peptidase was previously shown not to be higher in QPX cells inside clam tissue (this thesis, Chapter 4, papain-like peptidase C1-1), suggesting that it is not involved in QPX virulence. One can argue that this enzyme might be important at a different stage of the infection process, e.g. during the late phase of infection when QPX takes over the clam body and produces the peptidase to digest proteins present in clam connective tissue. This would assume that the QPX growth in clam adductor muscle homogenate would simulate that late stage of the infection process.

In general, these results suggest that QPX is a well-adapted saprophytic organism capable of secreting extracellular digestive enzymes to use protein-rich organic matter of animal origin (collagen, peptone, bovine serum). It is thought that QPX can survive in the environment outside its host because it has been detected in some environmental samples including seawater, marine sediment, algae and organic matter scraped from shell surfaces of marine invertebrates (Gast et al 2008, Liu et al 2009). No studies have been conducted to investigate how long QPX can survive without a sufficient protein source. Based on the data of peptidase activity from this study, QPX cells should be able to thrive on decomposing animal matter including dead or dying clam tissue.

Subtilisins are secreted when the parasite is grown in clam adductor muscle homogenate (Figure 5.2B), which indicates their ability to digest major proteins in muscular tissue, for example myosin. They are also secreted when the media contains gelatin (partially digested collagen), yeastolate and peptone Extracellular subtilisins of other pathogenic organisms show the ability to digest gelatin (Bonifait et al 2011) or have strong activity towards the extracellular matrix of animal connective tissue (Hong et al 2000). The ion exchange purified fraction containing one subtilisin peptidase was shown to digest some unidentified proteins in freshly collected clam plasma suggesting that it might play an important role in QPX infection *in vivo*.

In addition, subtilisins possess cytotoxic activities towards host epithelial cells (Bonifait et al 2011, Hong et al 2000). Preliminary assays of the FPLC fraction containing QPX subtilisin exposed to clam hemocytes revealed moderate cytotoxic effect with 53% dead hemocytes in the treatment versus 38% dead hemocytes in the control (40% increase, n=1). Additional studies with a larger number of biological replicates are required to confirm these results and to conduct other tests to evaluate QPX subtilisin cytotoxicity.

Chapter 6: Summary and significance of major findings

Occurrence of infectious diseases is driven by multiple factors associated with the host, the pathogen and the habitat in which they both live. Environmental factors impact host abundance, physiological condition, and susceptibility to the disease. Similarly, the environment influences pathogen availability, fitness and infectivity towards the host. There is also an underlying genetic variability among individuals of the host and pathogen populations. To understand the dynamics of any disease, all of the above parameters and their interactions have to be investigated. Combined knowledge from multiple disciplines is usually required to understand disease occurrence and its outcomes. The understanding of QPX infections in hard clams lags because of the limited information on the parasite's biology, ecology and virulence.

Population ecology studies have contributed to our understanding of hard clam abundance, population density and other parameters which determine the size of clam population, such as availability of adequate food sources, reproductive success and recruitment of young individuals to the population (Kraeuter et al 2009, MacKenzie Jr et al 2006). Further, population genetic studies allow for the understanding of genetic predisposition to cause (for the pathogen) and to be affected by (for the host) a disease. A significant number of studies have been examining clam broodstock (aka clam strains) and clam geographic origin in relation to possible genetic resistance to QPX infections (Dahl et al 2008 and 2010, Ford et al 2002c, Kraeuter et al 2011, Ragone Calvo et al 2007).

Several recent studies have been directed to understanding hard clam immunity and the influence of environmental factors on clam immune parameters during QPX infections (Dahl et al 2011, Hegaret et al 2010, Perrigault & Allam 2012, Perrigault et al 2012, Perrigault et al 2011, Perrigault et al 2009b). On the other hand, very little has been done regarding QPX genetic diversity, QPX virulence factors, and molecular mechanisms of QPX infection**.** The studies in the present dissertation were designed to help fill in the gaps in knowledge about this parasite's virulence, its molecular interactions with the host, and its molecular adaption to the environment (Figure 6.2), thus contributing to a better understanding of QPX infections and mechanisms of disease development in *Mercenaria mercenaria*.

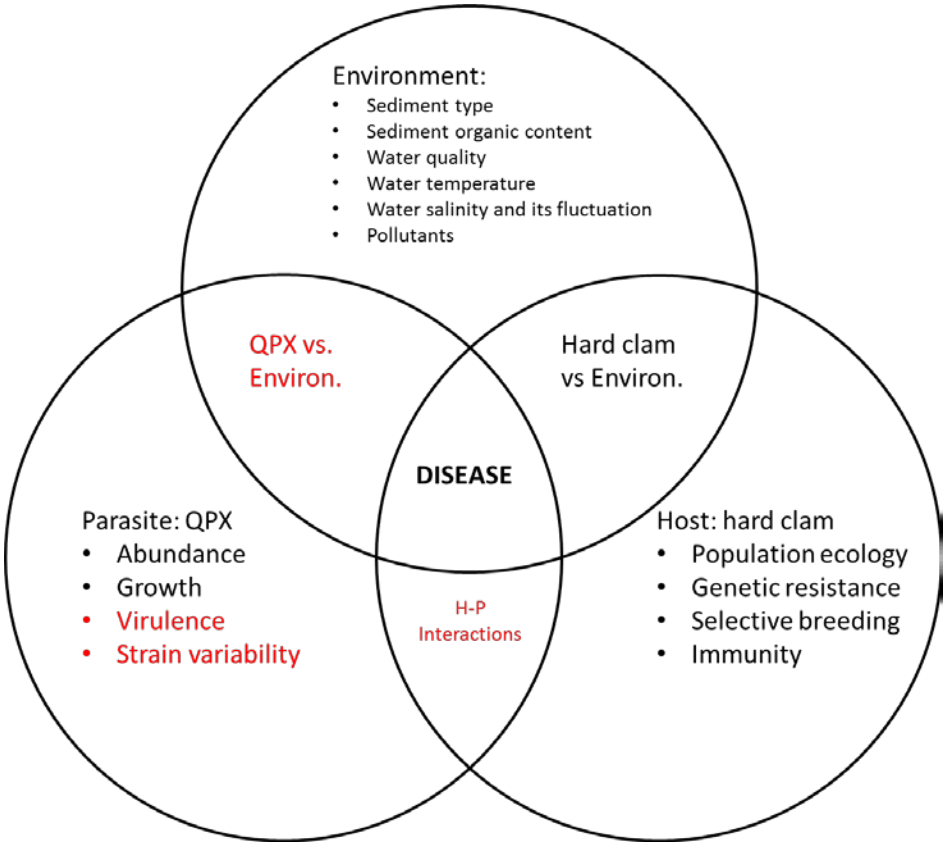


Figure 6.2: Contribution of the dissertation studies (in red) to the understanding of QPX disease development and its interactions with the clam host.

QPX putative virulence factors

The main objective of this dissertation was to examine the molecular basis of QPX virulence. The results presented in this dissertation identify and describe molecules that may act as virulence factors for QPX invasion and spread in clam tissues. These putative virulence factors likely facilitate QPX infection, allow for host tissue degradation, and manipulation of host immune response. The sequenced and annotated QPX transcriptome library generated during this research is an extensive molecular database of QPX putative virulence factors (Chapter 1). Some of the highlighted putative virulence factors are QPX surface and extracellular molecules involved in recognition and attachment, hydrolytic enzymes, antioxidants, and enzymes involved in muco-polysaccharide production. This research provides fundamental information for future studies on QPX infectiveness at the molecular level. Some of the most important virulence factors of pathogenic organisms are factors which are released outside the

cell, where they interact with host-derived molecules. Chapter 4 of this dissertation was designed to identify molecules secreted by QPX cells into the external environment. This study used a combination of complementary bioinformatics tools and LC/MS analyses to explore and identify proteins secreted by QPX to the surrounding environment. The proteomics analysis revealed 56 proteins secreted into the culture media, including several mucin-like molecules, glycosyl hydrolase, and peptidases (**Figure 6.1**). In addition, results of the gene expression experiment suggest that some of the mucin-like molecules, serine and cysteine peptidases, and hemolysins may be directly involved in the infectious process.

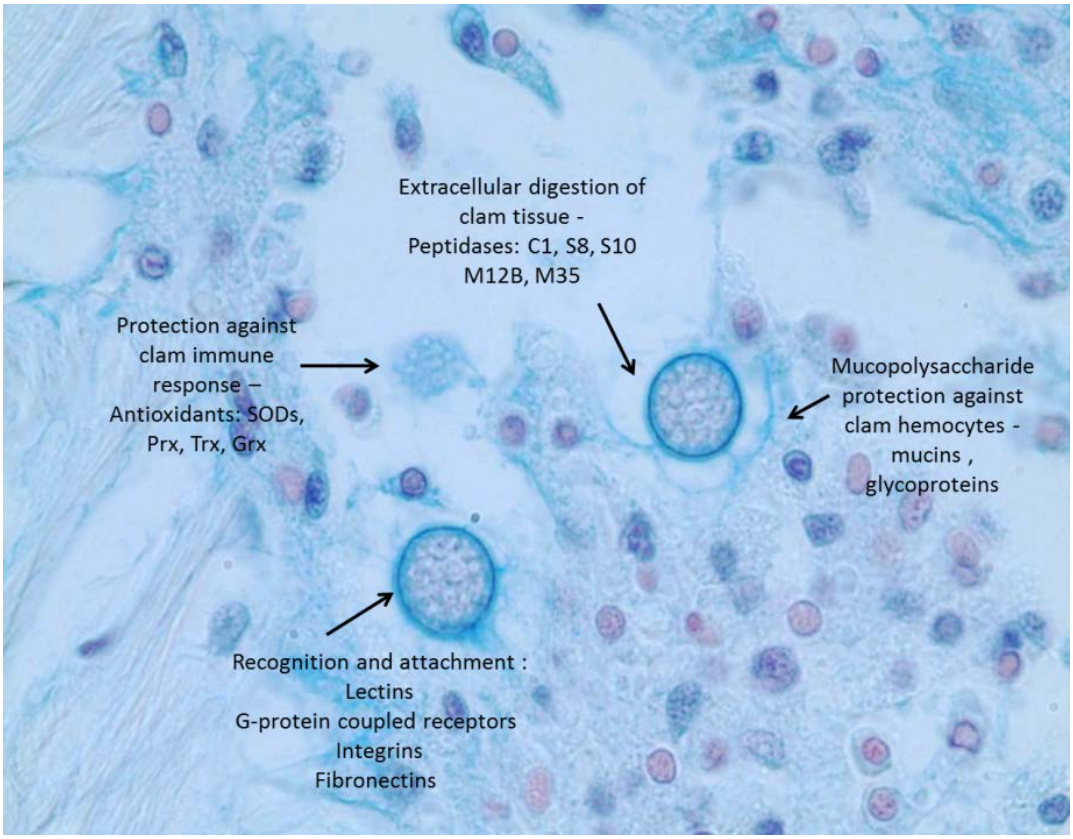


Figure 6.1: Major groups of putative QPX virulence factors reported in Chapter 4 of the dissertation: C1-papain-like peptidases, S8-subtilisins, S10-serine carboxypeptidases, M12B-adamalysins, M35 – deuterolysins, SODs - superoxide dismutases, Trx- thioredoxins, Prx -peroxidases, Grx-glutaredoxins

The experiments in this dissertation also aimed to descrive the proteomic characterization of molecules which are most likely to play a role in QPX virulence, the extracellular peptidases (Chapter 5). The functional annotation of three different QPX transcriptome libraries, one generated by 454-techonology (Chapter 1) and two others generated as part of the Marine

Eukaryotic Microbial Transcriptome Project (Chapter 4), revealed that QPX possesses 11 different nucleotide sequences coding for subtilisin-like peptidases (family S8). Seven of these sequences contained complete open reading frames, including the 5' and 3' untranslated regions (5' and 3' UTRs). These open reading frames were used in cell localization searches which revealed that all seven possess an N-terminal signal peptide which directs these peptidases into the eukaryotic secretion pathway (Chapter 4). The MS/LC analysis of QPX extracellular proteins showed that three subtilisin-like peptidases are secreted by QPX cells into the extracellular space (Chapter 4). The zymogram (gelatin substrate polyacrylamide gels) analysis confirmed that the majority of QPX extracellular proteolytic enzymes are subtilisins (Chapter 5). These subtilsin-like peptidases are released by QPX cells in all studied media, suggesting that they are the major peptidases used by QPX to obtain nutrients (Chapter 5). The expression of QPX subtilisin-like peptidases is regulated at the transcriptional level. The relative mRNA levels of six QPX subtilisins was measured using qPCR for QPX cells cultivated at four different temperatures (Chapter 1). Three QPX subtilisins were shown to be differentially expressed in response to temperature based on both the qPCR and oligoarray results. The highest expression of those three peptidases was measured at 23°C, which corresponds to QPX's optimal growth temperature. The maximal growth of QPX cells at 23°C can be explained by the efficient extracellular protein degradation and nutrient acquisition system associated with high expression of QPX major digestive enzymes. The expression of QPX subtilisin-like peptidases was also altered in QPX cells grown in media supplemented with different proteins sources (Chapter 2 and 5). The oligoarray analysis revealed lower amounts of transcripts coding for subtilsins in QPX cells cultivated in media containing clam adductor muscle homogenate in comparison to standard media containing fetal bovine serum (FBS), (Chapter 2). The results were consistent with the protein expression of subtilisins examined by zymography, which also revealed higher activity of subtilisins in QPX secretion derived from standard culture media containing FBS. In contrast, the FPLC purified fraction containing QPX subtilisin showed peptidase activity on clam plasma incorporated polyacrylamide gels, revealing that the subtilisin digests proteins present in clam plasma.

The work in this dissertation also investigated if QPX virulence and the expression of its putative virulence factors are influenced by exposure of parasite cells to clam-tissue homogenate (Chapter 2). A significant number of transcripts related to stress response and detoxification were

expressed at a higher level in QPX cells cultivated in clam adductor muscle homogenate. In contrast, the results from that study showed that certain genes commonly related to virulence, including many extracellular peptidases, were expressed at lower levels in QPX cells grown in clam adductor muscle homogenate, suggesting that the parasite cells were not more virulent than QPX cells grown in standard media. The results were consistent with the QPX challenge experiment which showed no difference in disease development in clams injected with QPX cells grown in clam adductor muscle in comparison to clams injected with QPX cells grown in standard media. Based on the results from this study, future exploration of QPX virulence should be focused on additional media factors which might enhance its infectiveness.

Differences among QPX isolates driven by its molecular adaptation to local environmental conditions

Temperature has been shown to be one of the most important environmental factors influencing QPX disease development in clams. For this reason, the QPX transcriptome library was used to design an oligoarray platform to evaluate transcriptional changes in QPX cells which were exposed to different temperatures (Chapter 1). Results from this experiment demonstrated the ability of QPX to cope with a wide range of environmental temperatures, including low temperatures which are considered to be suboptimal for clam immunity. For example, QPX cells cultivated at low temperatures: 10 and 13 °C exhibited up-regulation of high molecular weight heat shock proteins (hsp40, hsp70s, hsp90s and hsp100), showing that the parasite is well-adapted to withstand stressful conditions of cold temperature. QPX disease is considered to be a cold water disease because it occurs in coastal waters of the North Atlantic Ocean and the cold water stress chaperones show QPX molecular adaption to its geographic range. On the other hand, small molecular weight heat shock proteins (hsp15, hsp16, hsp18) are QPX molecular chaperones for heat stress conditions. These results provide a possible explanation for QPX disease distribution associated with latitude and for high disease prevalence and intensity at low temperatures.

Comparative transcriptomic analysis of QPX isolates originating from different geographic locations showed variation in gene expression which might correlate with adaptation of QPX isolates to local environmental conditions (Chapter 3). In particular, adaption to

temperatures typical in their original environments was consistent with differential expression of small cytosolic heat shock proteins at different temperatures. QPX isolates originating from the southern-most location, Old Plantation Creek in VA and from a shallow water embayment, Peconic Bay on Long Island, NY, did not exhibit heat stress at 23 °C incubation temperature, while QPX isolates from the northern-most location, Cape Cod, MA and from deep water of Raritan Bay, NY exhibited heat stress response. The combined results of gene expression profiles generated from the temperature and from the multiple isolates experiments suggest that QPX genes encoding these small heat shock proteins are promising biomarkers to be used in the study of genetic variation among QPX strains.

Besides the highly DE small heat shock proteins, a list of other highly regulated genes among different QPX isolates was generated as potential markers for population level studies (Chapter 3). These included: a catalase, a lipase, and two serine-like peptidases which could also be responsible for virulence differences among different QPX isolates. For example, one subtilisin-like peptidase had a three-fold higher expression level in the MA QPX and from five to seven-fold lower expression in the NY QPX isolates, suggesting a strong regulation of that particular peptidase among QPX isolates from different geographic locations and a possible basis for virulence differences. In addition, other transcripts coding for peptidases were also DE among QPX isolates (Chapter 3) which corresponded to the differential peptidase activity profiles revealed on the zymograms (Chapter 5). The results from the two experiments show that QPX isolates from different geographic locations differ in the peptidase expression profiles and the genes coding for peptidases could also be potential markers in population level studies. In other pathogenic protozoans, peptidases encoding genes also exhibit high intraspecific genetic variation (Cortez et al 2009, Reece 2001). The high nucleotide polymorphism of these genes is caused by relatively quick evolution due to natural selection pressure caused by a disease (Aguileta et al 2009). The identification of QPX virulence related genes presented in this study might help to select for markers to study genetic diversity among QPX populations in different geographic locations, strain genotyping, and answer questions about QPX spread or possible human introduction to different areas.

References

Ackman RG, Hooper SN, Ke PJ. 1971. The distribution of saturated and isoprenoid fatty acids in the lipids of three species of molluscs, Littorina littorea, Crassostrea virginica and Venus mercenaria. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* 39: 579-87

Acorn AR, Clark KF, Jones S, Despres BM, Munro S, et al. 2011. Analysis of expressed sequence tags (ESTs) and gene expression changes under different growth conditions for the ciliate Anophryoides haemophila, the causative agent of bumper car disease in the American lobster (Homarus americanus). *Journal of Invertebrate Pathology* 107: 146-54

Addadi L, Joester D, Nudelman F, Weiner S. 2006. Mollusk shell formation: a source of new concepts for understanding biomineralization processes. *Chemistry* 12: 980-7

Aguileta G, Refregier G, Yockteng R, Fournier E, Giraud T. 2009. Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious diseases* 9: 656-70

Alvarez VE, Niemirowicz GT, Cazzulo JJ. 2012. The peptidases of Trypanosoma cruzi: digestive enzymes, virulence factors, and mediators of autophagy and programmed cell death. *Biochimica et biophysica acta* 1824: 195-206

Anderson RS, Kraus BS, McGladdery S, Smolowitz R. 2003. QPX, a pathogen of quahogs (hard clams), employs mucoid secretions to resist host antimicrobial agents. *Journal of Shellfish Research* 22: 205-08

Anderson RS, Luskey TM, Strauss MA. *Abstracts of the 98th Annual Meeting of the National Shellfisheries Association, Monterey, California, USA., 2006.*

Atkinson HJ, Babbitt PC, Sajid M. 2009. The global cysteine peptidase landscape in parasites. *Trends in Parasitology* 25: 573-81

Avrova AO, Whisson SC, Pritchard L, Venter E, De Luca S, et al. 2007. A novel non-protein-coding infection-specific gene family is clustered throughout the genome of Phytophthora infestans. *Microbiology* 153: 747-59

Baptista CS, Vencio RZ, Abdala S, Carranza JC, Westenberger SJ, et al. 2006. Differential transcription profiles in Trypanosoma cruzi associated with clinical forms of Chagas disease: Maxicircle NADH dehydrogenase subunit 7 gene truncation in asymptomatic patient isolates. *Molecular and Biochemical Parasitology* 150: 236-48

Belaunzaran ML, Wilkowsky SE, Lammel EM, Gimenez G, Bott E, et al. 2013. Phospholipase A1: a novel virulence factor in Trypanosoma cruzi. *Molecular and Biochemical Parasitology* 187: 77-86

Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. 2004a. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Engineering, Design & Selection : PEDS* 17: 349-56

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004b. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* 340: 783-95

Bergmann N, Fricke B, Schmidt MC, Tams V, Beining K, et al. 2011. A quantitative real-time polymerase chain reaction assay for the seagrass pathogen Labyrinthula zosterae. *Molecular Ecology Resources* 11: 1076-81

Bertolini JM, Rohovec JS. 1992. Electrophoretic detection of proteases from different Flexibacter columnaris strains and assessment of their variability. Disease of Aquatic Organisms 12: 121-28

Bigelow DM, Olsen MW, Gilbertson RL. 2005. Labyrinthula terrestris sp. nov., a new pathogen of turf grass. *Mycologia* 97: 185-90

Blagojevic DP, Grubor-Lajsic GN, Spasic MB. 2011. Cold defence responses: the role of oxidative stress. *Front Biosci (Schol Ed)* 3: 416-27

Bogdan C. 2007. Oxidative burst without phagocytes: the role of respiratory proteins. *Nature Immunology* 8: 1029-31

Boncheva R, Dukiandjiev S, Minkov I, de Maagd RA, Naimov S. 2006. Activity of Bacillus thuringiensis delta-endotoxins against codling moth (Cydia pomonella L.) larvae. *Journal of Invertebrate Pathology* 92: 96-9

Bonifait L, de la Cruz Dominguez-Punaro M, Vaillancourt K, Bart C, Slater J, et al. 2010. The cell envelope subtilisin-like proteinase is a virulence determinant for Streptococcus suis. *BMC Microbiology* 10: 42

Bonifait L, Vaillancourt K, Gottschalk M, Frenette M, Grenier D. 2011. Purification and characterization of the subtilisin-like protease of Streptococcus suis that contributes to its virulence. *Veterinary Microbiology* 148: 333-40

Bower SM. 1987. The life cycle and ultrastructure of a new species of thraustochytrid (Protozoa: Labyrinthomorpha) pathogenic to small abalone. *Aquaculture* 67: 269-72

Bozdech Z, Mok S, Hu G, Imwong M, Jaidee A, et al. 2008. The transcriptome of Plasmodium vivax reveals divergence and diversity of transcriptional regulation in malaria parasites. *Proceedings of the National Academy of Sciences of the United States of America* 105: 16290-5

Breton C, Imberty A. 1999. Structure/function studies of glycosyltransferases. *Current Opinion in Structural Biology* 9: 563-71

Brown GD, Dave JA, Gey van Pittius NC, Stevens L, Ehlers MR, Beyers AD. 2000. The mycosins of Mycobacterium tuberculosis H37Rv: a family of subtilisin-like serine proteases. *Gene* 254: 147-55

Brown SP, Cornforth DM, Mideo N. 2012. Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control. *Trends in Microbiology* 20: 336-42

Bugge DM, Allam B. 2005. A fluorometric technique for the in vitro measurment of growth and viability in quahog parasite unknown (QPX). *Journal of Shellfish Research* 24: 1013-18

Bugge DM, Hegaret H, Wikfors GH, Allam B. 2007. Oxidative burst in hard clam (Mercenaria mercenaria) haemocytes. *Fish & Shellfish Immunology* 23: 188-96

Calderone RA, Fonzi WA. 2001. Virulence factors of Candida albicans. *Trends in Microbiol*ogy 9: 327-35

Carruthers VB. 2006. Proteolysis and Toxoplasma invasion. *International Journal for Parasitology* 36: 595-600

Casadevall A, Pirofski L. 1999. Host-Pathogen Interactions: Redefining the Basic Concepts of Virulence and Pathogenicity *Infection and Immunity* 67: 3703-13

Cheung F, Win J, Lang JM, Hamilton J, Vuong H, et al. 2008. Analysis of the Pythium ultimum transcriptome using Sanger and Pyrosequencing approaches. *BMC Genomics* 9: 542

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WE, et al. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research* 14: 1147-59

Chia-Ta T, Wen-Lin H, Shinn-Jang H, Li-Sun S, Shinn-Ying H. 2009. Virulent-GO: Prediction of Virulent Proteins in Bacterial Pathogens Utilizing Gene Ontology Terms. *World Academy of Science, Engineering and Technology* 53: 80-87

Coetzer TH, Goldring JP, Huson LE. 2008. Oligopeptidase B: a processing peptidase involved in pathogenesis. *Biochimie* 90: 336-44

Coleman JJ, White GJ, Rodriguez-Carres M, Vanetten HD. 2011. An ABC transporter and a cytochrome P450 of Nectria haematococca MPVI are virulence factors on pea and are the major tolerance mechanisms to the phytoalexin pisatin. *Molecular Plant-Microbe Interactions : MPMI* 24: 368-76

Cornillon S, Gebbie L, Benghezal M, Nair P, Keller S, et al. 2006. An adhesion molecule in free-living Dictyostelium amoebae with integrin beta features. *EMBO Reports* 7: 617-21

Cortez AP, Rodrigues AC, Garcia HA, Neves L, Batista JS, et al. 2009. Cathepsin L-like genes of Trypanosoma vivax from Africa and South America--characterization, relationships and diagnostic implications. *Molecular and Cellular probes* 23: 44-51

Coutte L, Antoine R, Drobecq H, Locht C, Jacob-Dubuisson F. 2001. Subtilisin-like autotransporter serves as maturation protease in a bacterial secretion pathway. *The EMBO Journal* 20: 5040-8

Cox GM, McDade HC, Chen SC, Tucker SC, Gottfredsson M, et al. 2001. Extracellular phospholipase activity is a virulence factor for Cryptococcus neoformans. *Molecular Microbiology* 39: 166-75

Cresnar B, Petric S. 2011. Cytochrome P450 enzymes in the fungal kingdom. *Biochimica et biophysica acta* 1814: 29-35

Cuervo P, De Jesus JB, Saboia-Vahia L, Mendonca-Lima L, Domont GB, Cupolillo E. 2009. Proteomic characterization of the released/secreted proteins of Leishmania (Viannia) braziliensis promastigotes. *Journal of Proteomics* 73: 79-92

Dacks JB, Doolittle WF. 2004. Molecular and phylogenetic characterization of syntaxin genes from parasitic protozoa. *Molecular and Biochemical Parasitology* 136: 123-36

Dahl SF, Allam B. 2007. Laboratory transmission studies of QPX disease in the northern quahog (=hard clam): Development of an infection procedure *Journal of Shellfish Research* 26: 383-89

Dahl SF, Perrigault M, Allam B. 2008. Laboratory transmission studies of QPX disease in the hard clam: Interactions between different host strains and pathogen isolates. *Aquaculture* 280: 64-70

Dahl SF, Perrigault M, Liu Q, Collier JL, Barnes DA, Allam B. 2011. Effects of temperature on hard clam (Mercenaria mercenaria) immunity and QPX (Quahog Parasite Unknown) disease development: I. Dynamics of QPX disease. *Journal of Invertebrate Pathology* 106: 314-21

Dahl SF, Thiel J, Allam B. 2010. Field performance and QPX disease progress in culture and wild-type strains of Mercenaria mercenaria in New York waters *Journal of Shellfish Research* 29: 83-90

Dassa E, Bouige P. 2001. The ABC of ABCs: a phylogenetic and functional classification\ of ABC systems in living organisms. *Research in Microbiology* 152: 211-29
Davis PH, Schulze J, Stanley Jr SL. 2007. Transcriptomic comparison of two Entamoeba histolytica strains with defined virulence phenotypes identifies new virulence factor candidates and key differences in the expression patterns of cysteine proteases, lectin light chains, and calmodulin. *Molecular and Biochemical Parasitology* 151: 118-28
de Miguel N, Braun N, Bepperling A, Kriehuber T, Kastenmuller A, et al. 2009. Structural and functional diversity in the family of small heat shock proteins from the parasite Toxoplasma gondii. *Biochimica et biophysica acta* 1793: 1738-48
Debnath A, Das P, Sajid M, McKerrow JH. 2004. Identification of genomic responses to collagen binding by trophozoites of Entamoeba histolytica. *The Journal of infectious diseases* 190: 448-57
Debnath A, Tashker JS, Sajid M, McKerrow JH. 2007. Transcriptional and secretory responses of Entamoeba histolytica to mucins, epithelial cells and bacteria. *International Journal of Parasitology* 37: 897-906
Denoeud F, Roussel M, Noel B, Wawrzyniak I, Da Silva C, et al. 2011. Genome sequence of the stramenopile Blastocystis, a human anaerobic parasite. *Genome biology* 12: R29
Dove ADM, Bowser PR, Cerrato RM. 2004a. Histological Analysis of an Outbreak of QPX Disease in Wild Hard Clams Mercenaria mercenaria in New York. *Journal of Aquatic Animal Health* 16: 246-50
Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, et al. 2011. Geneious v5.4.
Dubovenko AG, Dunaevsky YE, Belozersky MA, Oppert B, Lord JC, Elpidina EN. 2010. Trypsin-like proteins of the fungi as possible markers of pathogenicity. *Fungal Biology* 114: 151-59
Earnhart CG, Vogelbein MA, Brown GD, Reece KS, Kaattari SL. 2004. Supplementation of Perkinsus marinus cultures with host plasma or tissue homogenate enhances their infectivity. *Applied and environmental microbiology* 70: 421-31
Faust DM, Guillen N. 2012. Virulence and virulence factors in Entamoeba histolytica, the agent of human amoebiasis. *Microbes and Iinfection* 14: 1428-41
Echevarria M, Ilundain AA. 1998. Aquaporins. *Journal of physiology and biochemistry* 54: 107-18
Feder ME, Hofmann GE. 1999. Heat-shock proteins, molecular chaperones, and the stres response: evolutionary and ecological physiology. *Annual Review in Physiology* 61: 243-82
Fellbrich G, Romanski A, Varet A, Blume B, Brunner F, et al. 2002. NPP1, a Phytophthora-associated trigger of plant defense in parsley and Arabidopsis. *The Plant Journal for Cell and Molecular Biology* 32: 375-90
Ford SE, Chintala MM, Bushek D. 2002a. Comparison of in vitro-cultured and wild-type Perkinsus marinus. I. Pathogen virulence. *Diseased of Aquatic Organisms* 51: 187-201
Ford SE, Kraeuter JN, Barber RD, Mathis G. 2002b. Aquaculture-associated factors in QPX disease of hard clams: density and seed source. *Aquaculture* 208: 23-38
Fu D, Lu M. 2007. The structural basis of water permeation and proton exclusion in aquaporins. *Molecular Membrane Biology* 24: 366-74
Furuhashi T, Schwarzinger C, Miksik I, Smrz M, Beran A. 2009. Molluscan shell evolution with review of shell calcification hypothesis. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular Biology* 154: 351-71
Garcia-Vedrenne AE, Groner M, Page-Karjian A, Siegmund GF, Singh S, et al. 2013. Development of Genomic Resources for a thraustochytrid Pathogen and Investigation of Temperature Influences on Gene Expression. *PloS One* 8: e74196. doi:10.1371/journal.pone.0074196
Gast RJ, Moran DM, Audemard C, Lyons MM, DeFavari J, et al. 2008. Enviornmental distribution and persistence of Quahog Parasite Unknown (QPX). *Disease of Aquatic Organisms* 81: 219-29
Gaulin E, Jauneau A, Villalba F, Rickauer M, Esquerre-Tugaye MT, Bottin A. 2002. The CBEL glycoprotein of Phytophthora parasitica var-nicotianae is involved in cell wall deposition and adhesion to cellulosic substrates. *Journal of Cell Science* 115: 4565-75
Gaulin E, Madoui MA, Bottin A, Jacquet C, Mathe C, et al. 2008. Transcriptome of Aphanomyces euteiches: new oomycete putative pathogenicity factors and metabolic pathways. *PloS One* 3: e1723
Gavino PD, Fry WE. 2002. Diversity in and evidence for selection on the mitochondrial genome of Phytophthora infestans. *Mycologia* 94: 781-93
Geiger A, Hirtz C, Becue T, Bellard E, Centeno D, et al. 2010. Exocytosis and protein secretion in Trypanosoma. *BMC Microbiology* 10: 20
Ghannoum MA. 2000. Potential Role of Phospholipases in Virulence and Fungal Pathogenesis. *Clin Microbiol Rev* 13: 122-43

Giannini MS. 1974. Effects of promastigote growth phase, frequency of subculture, and host age on promastigote-initiated infections with Leishmania donovani in the golden hamster. *The Journal of protozoology* 21: 521-7
Gilchrist CA, Petri WA. 1999. Virulence factors of Entamoeba histolytica. *Current opinion in microbiology* 2: 433-7
Glickman MH, Ciechanover A. 2002. The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. *Physiological Reviews* 82: 373-428
Gobert GN, Moertel LP, McManus DP. 2005. Microarrays: new tools to unravel parasite transcriptomes. *Parasitology* 131: 439-48
Gophna U, Ron EZ. 2003. Virulence and the heat shock response. *International Journal of Medical Microbiology* : IJMM 292: 453-61
Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, et al. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 36: 3420-35
Guha-Niyogi A, Sullivan DR, Turco SJ. 2001. Glycoconjugate structures of parastic protozoa. *Glycobiology* 11: 45-59
Gupta A, Barrow CJ, Puri M. 2012. Omega-3 biotechnology: Thraustochytrids as a novel source of omega-3 oils. *Biotechnology Advances*
Gvozdeva EL, Ievleva EV, Gerasimova NG, Ozeretskovskaia OL, Valueva TA. 2004. Exoproteinases of the Oomecete Phytophthora infestans *Applied Biochemistry amd Microbiology* 40: 165-69
Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. *Nature* 461: 393-8
Haldar K, Kamoun S, Hiller NL, Bhattacharje S, van Ooij C. 2006. Common infection strategies of pathogenic eukaryotes. *Nature reviews. Microbiology* 4: 922-31
Harper JM, Hoff EF, Carruthers VB. 2004. Multimerization of the Toxoplasma gondii MIC2 integrin-like A-domain is required for binding to heparin and human cells. *Molecular and Biochemical Parasitology* 134: 201-12
He Y, Yu H, Bao Z, Zhang Q, Guo X. 2012. Mutation in promoter region of a serine protease inhibitor confers Perkinsus marinus resistance in the eastern oyster (Crassostrea virginica). *Fish & shellfish immunology* 33: 411-7
Hegaret H, Smolowitz RM, Sunila I, Shumway SE, Alix J, et al. 2010. Combined effects of a parasite, QPX, and the harmful-alga, Prorocentrum minimum on northern quahogs, Mercenaria mercenaria. *Marine environmental research* 69: 337-44
Hernandez-Romano P, Hernandez R, Arroyo R, Alderete JF, Lopez-Villasenor I. 2010. Identification and characterization of a surface-associated, subtilisin-like serine protease in Trichomonas vaginalis. *Parasitology* 137: 1621-35
Hicks SJ, Theodoropoulos G, Carrington JC, Corfield AP. 2000. The role of mucins in Host-Parasite Interactions. Part 1 - Protozoan Parasites. *Parasitology Today* 16: 476-81
Hochwimmer G, Tober R, Bibars-Reiter R, Licek E, Steinborn R. 2009. Identification of two GH18 chitinase family genes and their use as targets for detection of the crayfish-plague oomycete Aphanomyces astaci. *BMC Microbiology* 9: 184
Hong Y-C, Kong H-H, Ock M-S, Kim I-S, Chung D-I. 2000. Isolation and characterization of a cDNA encoding a subtilisin-like serine proteinase (AhSUB) from Acanthamoeba healyi. *Mol Biochem Parasitol* 111: 441-46
Horton P, Park KJ, Obayashi T, Fujita N, Harada H, et al. 2007. WoLF PSORT: protein localization predictor. *Nucleic acids research* 35: W585-7
Hruzik A, Asif AR, Gross U. 2011. Identification of Toxoplasma gondii SUB1 antigen as a marker for acute infection by use of an innovative evaluation method. *Journal of clinical microbiology* 49: 2419-25
Hrycay EG, Bandiera SM. 2012. The monooxygenase, peroxidase, and peroxygenase properties of cytochrome P450. *Archives of biochemistry and biophysics* 522: 71-89
Hu Q, Liu P, Yu Z, Zhao G, Li J, et al. 2010. Identification of a cell wall-associated subtilisin-like serine protease involved in the pathogenesis of Streptococcus suis serotype 2. *Microbial pathogenesis* 48: 103-9
Huang J, Jiang X, Zhang X, Chen W, Tian B, et al. 2008. Expressed sequence tag analysis of marine fungus Schizochytrium producing docosahexaenoic acid. *Journal of Biotechnology* 138: 9-16
Huang X, Zhao N, Zhang K. 2004. Extracellular enzymes serving as virulence factors in nematophagous fungi involved in infection of the host. *Res Microbiol* 155: 811-16
Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. 2009. InterPro: the integrative protein signature database. *Nucleic acids research* 37: D211-5
Hynes RO. 2002. Integrins: Bidirectional, Allosteric Signaling Machines. *Cell* 110: 673-87

Ito T, Bando H, Asano S. 2006. Activation process of the mosquitocidal delta-endotoxin Cry39A produced by Bacillus thuringiensis subsp. aizawai BUN1-14 and binding property to Anopheles stephensi BBMV. *Journal of invertebrate pathology* 93: 29-35
Itoh N, Xue QG, Schey KL, Li Y, Cooper RK, La Peyre JF. 2011. Characterization of the major plasma protein of the eastern oyster, Crassostrea virginica, and a proposed role in host defense. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology* 158: 9-22
Jacobs T, Erdmann H, Fleischer B. 2010. Molecular interaction of Siglecs (sialic acid-binding Ig-like lectins) with sialylated ligands on Trypanosoma cruzi. *European journal of cell biology* 89: 113-6
Jain M, Karan D, Batra SK, Varshney GC. 2001a. Mucins in protozoan parasites. *Frontiers in bioscience : a journal and virtual library* 6: D1276-83
Jain R, Raghukumar S, Tharanathan R, Bhosle NB. 2005. Extracellular polysaccharide production by thraustochytrid protists. *Mar Biotechnol (NY)* 7: 184-92
Jean L, Hackett F, Martin SR, Blackman MJ. 2003. Functional characterization of the propeptide of Plasmodium falciparum subtilisin-like protease-1. *The Journal of biological chemistry* 278: 28572-9
Jiang RH, de Bruijn I, Haas BJ, Belmonte R, Lobach L, et al. 2013. Distinctive expansion of potential virulence genes in the genome of the oomycete fish pathogen Saprolegnia parasitica. *PLoS genetics* 9: e1003272
Jiang RH, Tyler BM, Whisson SC, Hardham AR, Govers F. 2006. Ancient origin of elicitin gene clusters in Phytophthora genomes. *Molecular biology and evolution* 23: 338-51
Jones GM, O'Dor RK. 1983. Ultrastructural Observations on a Thraustochytrid Fungus Parasitic in the Gills of Squid (Illex illecebrosus LeSueur). *The Journal of Parasitology* 69: 903-11
Joseph SJ, Fernandez-Robledo JA, Gardner MJ, El-Sayed NM, Kuo CH, et al. 2010. The Alveolate Perkinsus marinus: biological insights from EST gene discovery. *BMC Genomics* 11: 228
Judelson HS, Michelmore RW. 1989. Structure and expression of a gene encoding heat-shock protein Hsp70 from the Oomycete fungus Bremia lactucae. *Gene* 79: 207-17
Jung C. 2011. The mystery of cytochrome P450 Compound I: a mini-review dedicated to Klaus Ruckpaul. *Biochimica et biophysica acta* 1814: 46-57
Kale SD, Tyler BM. 2011. Entry of oomycete and fungal effectors into plant and animal host cells. *Cellular Microbiology* 13: 1839-48
Kamoun S. 2006. A Catalogue of the Effector Secretome of Plant Pathogenic Oomycetes. *Annual Review of Phytopathology* 44: 41-60
Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biology* 12: e1001889
Kennan RM, Wong W, Dhungyel OP, Han X, Wong D, et al. 2010. The subtilisin-like protease AprV2 is required for virulence and uses a novel disulphide-tethered exosite to bind substrates. *PLoS Pathogen* 6: e1001210
Kim D, Cryle MJ, De Voss JJ, Ortiz de Montellano PR. 2007. Functional expression and characterization of cytochrome P450 52A21 from Candida albicans. *Archives of Biochemistry and Biophysics* 464: 213-20
Kleinschuster SJ, Smolowitz R, Parent J. 1998. In vitro life cycle and propagation of quahog parasite unknown. *Journal of Shellfish Research* 17: 75-78
Klemba M, Goldberg DE. 2002. Biological roles of proteases in parasitic protozoa. *Annual Review of Biochemistry* 71: 275-305
Kraeuter JN, Flimlin G, Kennish MJ, Macaluso R, Viggiano J. 2009. Sustainability of northern quahogs (= hard clams) Mercenaria mercenaria, Linnaeus in Raritan Bay, New Jersey: assessment of size specific growth and mortality. *Journal of Shellfish Research* 28: 273-87
Kraeuter JN, Ford S, Bushek D, Scarpa E, Walton W, et al. 2011. Evaluation of three northern quahog (hard clam) Mercenria mercenaria (Linnaeus) strains grown in Massachusetts and New Jersey for QPX-resistance *Journal of Shellfish Research* 30: 805-12
Kvingedal R, Owens L, Jerry DR. 2006. A new parasite that infects eggs of the mud crab, Scylla serrata, in Australia. *Journal of Invertebrate Pathology* 93: 54-59
Lagal V, Binder EM, Huynh MH, Kafsack BF, Harris PK, et al. 2010. Toxoplasma gondii protease TgSUB1 is required for cell surface processing of micronemal adhesive complexes and efficient adhesion of tachyzoites. *Cellular Microbiology* 12: 1792-808
Laity JH, Lee BM, Wrigth E. 2001. Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology* 11: 39-46
Lang T, Hansson GC, Samuelsson T. 2007. Gel-forming mucins appeared early in metazoan evolution. *Proceedings of the National Academy of Sciences of the United States of America* 104: 16209-14

Lantz MS, Ciborowski P. 1994. Zymographic Techniques for Detection and Characterization of Microbial Proteases In Methods is Enzymology Volume 235: Bacterial Pathogenesis, Part A: Identification and Regulation of Virulence Factors, ed. JN Abelson, MI Simon, VL Clark, PM Bavoil, pp. 563. San Diego, CA: Academic Press, Inc.

La Peyre J, Schafhauser DY, Rizkalla EH, Faisal M. 1995. Production of Serine Proteases by the Oyster Pathogen Perkinus marinus (Apicomplexa) In Vitro. Journal of Eukaryotic Microbiology 42: 544-51

Latijnhouwers M, de Wit PJGM, Govers F. 2003. Oomycetes and fungi: similar weaponry to attack plants. *Trends in Microbiology* 11: 462-69

Lecaille F, Kaleta J, Bromme D. 2002. Human and Parasitic Papain-Like Cysteine Proteases: Their Role in Physiology and Pathology and Recent Developments in Inhibitor Design. *Chemical Reviews* 102: 4459-88

Lee SH, Minagawa E, Taguchi H, Matsuzawa H, Ohta T, et al. 1992. Purification and characterization of a thermostable carboxypeptidase (carboxypeptidase Taq) from Thermus aquaticus YT-1. *Bioscience, Biotechnology, and Biochemistry* 56: 1839-44

Lelivelt MJ, Kawula TH. 1995. Hsc66, an Hsp70 homolog in Escherichia coli, is induced by cold shock but not by heat shock. *Journal of Bacteriology* 177: 4900-7

Lev S, Desmarini D, Li C, Chayakulkeeree M, Traven A, et al. 2013. Phospholipase C of Cryptococcus neoformans regulates homeostasis and virulence by providing inositol trisphosphate as a substrate for Arg1 kinase. *Infection and Immunity* 81: 1245-55

Levesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, et al. 2010. Genome sequence of the necrotrophic plant pathogen Pythium ultimum reveals original pathogenicity mechanisms and effector repertoire. *Genome Biology* 11: R73

Lewis DFV, Wiseman A. 2005. A selective review of bacterial forms of cytochrome P450 enzymes. *Enzyme and Microbial Technology* 36: 377-84

Li X, Gao M, Han X, Tao S, Zheng D, et al. 2012. Disruption of the phospholipase D gene attenuates the virulence of Aspergillus fumigatus. *Infection and Immunity* 80: 429-40

Liao S, Liu Y, Jung SY, Cho PY, Zheng B, Park H. 2010. Transcriptional activity of Plasmodium subtilisin-like protease 2 (Pf-Sub2) 5'untranslated regions and its interaction with hepatocyte growth factor. *The Korean Journal of Parasitology* 48: 291-5

Lin L, Fang W, Liao X, Wang F, Wei D, St Leger RJ. 2011. The MrCYP52 cytochrome P450 monoxygenase gene of Metarhizium robertsii is important for utilizing insect epicuticular hydrocarbons. *PloS One* 6: e28984

Lis H, Sharon N. 1998. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. *Chemical Reviews* 98: 637-74

Liu Q, Collier JL, Pawagi S, Barnes D, Allam B. 2008. Quahog parasite unknown (QPX) dynamics in wild hard clams from Raritan Bay, New York. *Journal of Shellfish Research* 27: 1024-24

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402-8

Llinas M, Bozdech Z, Wong ED, Adai AT, DeRisi JL. 2006. Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. *Nucleic Acids Research* 34: 1166-73

Lushbaugh WB, Kairalla AB, Loadholt CB, Pittman FE. 1978. Effect of hamster liver passage on the virulence of axenically cultivated Entamoeba histolytica. *The American Journal of Tropical Medicine and Hygiene* 27: 248-54

Lyons MM, Smolowitz R, Dungan CF, Roberts SB. 2006. Development of a real time quantitative PCR assay for the hard clam pathogen Quahog Parasite Unknown (QPX). *Disease of Aquatic Organisms* 72: 45-52

Lyons MM, Smolowitz R, Gomez-Chiarri M, Ward JE. 2007. Epizootiology of quahog parasite unknown (QPX) diease in northern quahogs (hard clams) Mercenaria mercenaria *Journal of Shellfish Research* 26: 371-81

Lyons MM, Ward JE, Smolowitz R, Uhlinger KR, Gast RJ. 2005. Lethal marine snow: Pathogen of bivalve mollusc concealed in marine aggregates *Limnology and Oceanography* 50: 1983-88

Ma'ayeh SY, Brook-Carter PT. 2012. Representational difference analysis identifies specific genes in the interaction of Giardia duodenalis with the murine intestinal epithelial cell line, IEC-6. *International Journal for Parasitology* 42: 501-9

Maas PAY, Kleinschuster SJ, Dykstra MJ, Smolowitz R, Parent J. 1999. Molecular characterization of QPX (Quahog Parasite Unknown), a pathgogen of Mercenaria mercenaria *Journal of Shellfish Research* 18: 561-67

MacCallum GS, McGladdery SE. 2000. Quahog Parasite Unknown (QPX) in the northern quahog Mercenaria mercenaria (Linnaeus, 1758) and M. mercenaria var. notata from Atlantic Canada, survey results from three maritime provinces. *Journal of Shellfish Research* 19: 43-50

MacKenzie Jr CL, Pikanowski R, Mcmillan DG. 2006. Ampelisca amphipod tube matss may enhance abudance of northern quahogs Mercenaria mercenaria in muddy sediments *Journal of Shellfish Research* 25: 841-47
Mann BJ. 2002. Structure and function of the Entamoeba histolytica Gal/GalNAc lectin. *International Review of Cytology* 216: 59-80
Mao Y, Tyler BM. 1996. The Phytophthora sojae genome contains tandem repeat sequences which vary from strain to strain. *Fungal Genetics and Biology : FG & B* 20: 43-51
Martinez J, Perez-Serrano J, Bernadina WE, Rodriguez-Caabeiro F. 2001. Stress response to cold in Trichinella species. *Cryobiology* 43: 293-302
Masini d'Avila-Levy C, de Almeida Dias F, Nogueira de Melo AC, Martins JL, De Carvalho Santos Lopes AH, et al. 2006. Insights into the role of gp63-like proteins in lower trypanosomatids. *FEMS Microbiology Letters* 254: 149-56
McKerrow JH, Caffrey CR, Kelly BL, Loke P, Sajid M. 2006. Proteases in Parsitic Diseases. *Annual Review of Pathology: Mechanisms of Disease* 1: 497-536
McKerrow JHS, E.; Rosenthal, P.J.; Bouvier, J. 1993. The proteases and pathogenicity of parasitic protozoa. *Annual Review of Microbiology* 47: 821-53
McLean N, Porter D. 1982. The Yellow-Spot Disease of Tritonia diomedea Bergh, 1894 (Mollusca: Gastropoda: Nudibranchia): Encapsulation of the Thraustochytriaceous Parasite by Host Amoebocytes. *The Journal of Parasitology* 68: 243-52
Miller SA, Binder EM, Blackman MJ, Carruthers VB, Kim K. 2001. A conserved subtilisin-like protein TgSUB1 in microneme organelles of Toxoplasma gondii. *The Journal of Biological Chemistry* 276: 45341-8
Miyoshi S, Shinoda S. 2000. Microbial metalloproteases and pathogenesis. *Microbes and Infection* 2: 91-98
Monod M, Capoccia S, Lechenne B, Zaugg C, Holdom M, Jousson O. 2002. Secreted proteases from pathogenic fungi. *International Journal of Medical Microbiology : IJMM* 292: 405-19
Moser M, Menz G, Blaser K, Crameri R. 1994. Recombiant Expression and Antigenic Properties of a 32-Kilodalton Extracellular Alkaline Protease, Representing a Possible Virulence Factor from Aspergillus fumigatus *Infection and Immunity* 62: 936-42
Muszewska A, Taylor JW, Szczesny P, Grynberg M. 2011. Independent subtilases expansions in fungi associated with animals. *Molecular Biology and Evolution* 28: 3395-404
Neckers L, Tatu U. 2008. Molecular chaperones in pathogen virulence: emerging new targets for therapy. *Cell Host & Microbe* 4: 519-27
Nickel W, Rabouille C. 2009 Mechanisms of regulated unconventional protein secretion. *Nature Review Molecular and Cell Biology* 10: 148-55
Niemirowicz G, Parussini F, Aguero F, Cazzulo JJ. 2007. Two metallocarboxypeptidases from the protozoan Trypanosoma cruzi belong to the M32 family, found so far only in prokaryotes. *The Biochemical Journal* 401: 399-410
Olivos-Garcia A, Saavedra E, Ramos-Martinez E, Nequiz M, Perez-Tamayo R. 2009a. Molecular nature of virulence in Entamoeba histolytica. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 9: 1033-7
Osorio L, Rios I, Gutierrez B, Gonzalez J. 2012. Virulence factors of Trypanosoma cruzi: who is who? *Microbes and Infection / Institut Pasteur* 14: 1390-402
Pales Espinosa E, Corre E, Allam B. 2014. Pallial mucus of the oyster Crassostrea virginica regulates the expression of putative virulence genes of its pathogen Perkinsus marinus. *International Journal for Parasitology* 44: 305-17
Pales Espinosa E, Winnicki S, Allam B. 2013. Early host-pathogen interactions in a marine bivalve: Crassostrea virginica pallial mucus modulates Perkinsus marinus growth and virulence. *Disease Aquatic Organism* 104: 237-47
Panabieres F, Le Berre JY. 1999. A family of repeated DNA in the genome of the oomycete plant pathogen Phytophthora cryptogea. *Current Genetics* 36: 105-12
Pao SS, Paulsen I, Saier MH. 1998. Major Facilitator Superfamily. *Microbiology and Molecular Biology Reviews* 62: 1-34
Parussini F. 2003. Characterization of a lysosomal serine carboxypeptidase from Trypanosoma cruzi. *Mol Biochem Parasitol* 131: 11-23
Pereira LM, Candido-Silva JA, De Vries E, Yatsuda AP. 2011. A new thrombospondin-related anonymous protein homologue in Neospora caninum (NcMIC2-like1). *Parasitology* 138: 287-97
Perez-Morales D, Ostoa-Saloma P, Espinoza B. 2009. Trypanosoma cruzi SHSP16: Characterization of an alpha-crystallin small heat shock protein. *Experimental Parasitology* 123: 182-9

Perrigault M, Allam B. 2009a. Cytotoxicity of quahog parasite unknown (QPX) toward hard clam (Mercenaria mercenaria) haemocytes and interactions between different pathogen isolates and host strains. *Parasitology* 136: 1281-9

Perrigault M, Allam B. 2012. Differential immune response in the hard clam (mercenaria mercenaria) against bacteria and the protistan pathogen QPX (quahog parasite unknown). *Fish Shellfish Immunology* 32: 1124-34

Perrigault M, Bugge DM, Allam B. 2010. Effect of environmental factors on survival and growth of quahog parasite unknown (QPX) in vitro. *Journal of Invertebrate Pathology* 104: 83-9

Perrigault M, Bugge DM, Hao CC, Allam B. 2009b. Modulatory effects of hard clam (Mercenaria mercenaria) tissue extracts on the in vitro growth of its pathogen QPX. *Journal of Invertebrate Pathology* 100: 1-8

Perrigault M, Dahl SF, Espinosa EP, Allam B. 2012. Effects of salinity on hard clam (Mercenaria mercenaria) defense parameters and QPX disease dynamics. *Journal of Invertebrate Pathology* 110: 73-82

Perrigault M, Dahl SF, Espinosa EP, Gambino L, Allam B. 2011. Effects of temperature on hard clam (Mercenaria mercenaria) immunity and QPX (Quahog Parasite Unknown) disease development: II. Defense parameters. *Journal of Invertebrate Pathology* 106: 322-32

Perrigault M, Tanguy A, Allam B. 2009b. Identification and expression of differentially expressed genes in the hard clam, Mercenaria mercenaria, in response to quahog parasite unknown (QPX). *BMC genomics* 10: 377

Petersen TN, Brunak S, Von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8: 785-86

Porter D. 1990. 22: Phylum Labyrinthulomycota In Handbook of Protoctista. The structure, cultivation, habitats and ife histories of the eukaryotic microorganismss and therir discendants exclusive of animals, plants and fungi. A guide to the algae, ciliates, foraminifera, sporozoa, water molds, slime molds and ther other protoctists., ed. L Margulis, JO Corliss, M Melkonian, DJ Chapman. Boston, MA: Jones and Bartlett Publishers

Qian H, Liu Q, Allam B, Collier JL. 2007. Molecular genetic variation within and among isolates of QPX (Thraustochytridae), a parasite of the hard clam Mercenaria mercenaria. *Diseases of Aquatic Organisms* 77: 159-68

Qiu XB, Shao YM, Miao S, Wang L. 2006. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cellular and Molecular Life Sciences : CMLS* 63: 2560-70

Que X, Reed SL. 2000. Cysteine Proteinases and the Pathogenesis of Amebiasis. *Clinical Microbiology Reviews* 13: 196-206

Qutob D, Kamoun S, Gijzen M. 2002. Expression of a Phytophthora sojae necrosis-inducing protein occurs during transition from biotrophy to necrotrophy. *The Plant Journal for Cell and Molecular Biology* 32: 361-73

Raffaele S, Win J, Cano LM, Kamoun S. 2010. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of Phytophthora infestans. *BMC Genomics* 11: 637

Ragan MA, MacCallum GS, Murphy CA, Cannone JJ, Gutell RR, McGladdery SE. 2000a. Protistan parasite QPX of hard-shell clam Mercenaria mercenaria is a member of Labyrinthulomycota. *Disease of Aquatic Organisms* 42: 185-90

Raghukumar S. 2002. Ecology of the marine protists, the Labyrinthulomycetes (Thraustochytrids and Labyrinthulids). *European Journal of Protistology* 38: 127-45

Raghukumar S. 2008. Thraustochytrid Marine Protists: production of PUFAs and Other Emerging Technologies. *Marine Biotechnolgy (NY)* 10: 631-40

Ragone Calvo LM, Ford SE, Kraeuter JN, Leavitt DF, Smolowitz R, Burreson EM. 2007. Influence of host genetic origin and geographic location on QPX disease in northern quahogs(=hard clams), Mercenaria mercenaria. *Journal of Shellfish Research* 26: 109-19

Ragone Calvo LM, Walker JG, Burreson EM. 1998. Prevalence and distribution of QPX, Quahog Parasite Unknown, in hard clams Mercenaria mercenaria in Virginia, USA. *Disease of Aquatic Organisms* 33: 209-19

Rawlings ND, Barrett AJ, Bateman A. 2012. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* 40: D343-50

Reece KSB, D.; Hudson, K. L., Graves, J. E. . 2001. Geographic distribution of Perkinsus marinus genetic strains along the Atlantic and Gulf coasts of the USA. *Marine Biology* 139: 1047-55

Rehm BHA. 2009. Alginate Production: Precursor Biosynthesis, Polymerization and Secretion. In Alginates: Biology and Applications, Microbiology Monographs 13: 55-71, ed. Rehm BHA. Berlin. Springer-Verlag Berlin Heidelberg.

Richardsa TA, Soanesa DM, Jonesa MM, Vasievac O, Leonarda G, et al. 2011. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proceedings of National Academy of Sciences* 108: 5258-15263.
Ringqvist E, Avesson L, Soderbom F, Svard SG. 2011. Transcriptional changes in Giardia during host-parasite interactions. *International Journal of Parasitology* 41: 277-85
Robold AV, Hardham AR. 2005. During attachment Phytophthora spores secrete proteins containing thrombospondin type 1 repeats. *Current Genetics* 47: 307-15
Rodrigues DC, Silva R, Rondinelli E, Urmenyi TP. 2010. Trypanosoma cruzi: modulation of HSP70 mRNA stability by untranslated regions during heat shock. *Experimental Parasitology* 126: 245-53
Rubin E, Tanguy A, Perrigault M, Pales Espinosa E, Allam B. 2014. Characterization of the transcriptome and temperature-induced differential gene expression in QPX, the thraustochytrid parasite of hard clams. *BMC Genomics* 15: 245
S. PS, Paulsen I, Saier MH. 1998. Major Facilitator Superfamily. *Microbiology and Molecular Biology Reviews* 62: 1-34
Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, et al. 2006. TM4 microarray software suite. *Methods in Enzymology* 411: 134-93
Saeed AI, Sharov V, White J, Li J, Liang W, et al. 2003. TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* 34: 374-8
Sajid M, McKerrow JH. 2002b. Cysteine proteases of parasitic organisms. *Molecular and Biochemical Parasitology* 120: 1-21
Schärer L, Knoflach D, Vizoso DB, Rieger G, Peintner U. 2007. Thraustochytrids as novel parasitic protists of marine free-living flatworms: Thraustochytrium caudivorum sp. nov. parasitizes Macrostomum lignano. *Marine Biology* 152: 1095-104
Seals DF, Courtneidge SA. 2003. The ADAMs family of metalloproteases: multidomain proteins with multiple functions. *Genes & Development* 17: 7-30
Seliskar M, Rozman D. 2007. Mammalian cytochromes P450--importance of tissue specificity. *Biochimica et biophysica acta* 1770: 458-66
Shah FA, Allen N, Wright CJ, Butt TM. 2007. Repeated in vitro subculturing alters spore surface properties and virulence of Metarhizium anisopliae. *FEMS Microbiology Letters* 276: 60-6
Sharon N. 2006. Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochimica et biophysica acta* 1760: 527-37
Silverman JM, Chan SK, Robinson DP, Dwyer DM, Nandan D, et al. 2008. Proteomic analysis of the secretome of Leishmania donovani. *Genome Biology* 9: R35
Smith AM, Robinson TM, Salt MD, Hamilton KS, Silvia BE, Blasiak R. 2009. Robust cross-links in molluscan adhesive gels: testing for contributions from hydrophobic and electrostatic interactions. *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular biology* 152: 110-7
Smolowitz R, Leavitt D, Perkins F. 1998b. Observations of a Protistan Disease Similar to QPX in Mercenaria mercenaria (Hard Clams) from the Coast of Massachusetts. *Journal of Invertebrate Pathology* 71: 9-25
Smolowitz R, Leavitt D, Perkins F. 1998c. Observations of a Protistan Disease Similar to QPX in Mercenaria mercenaria (Hard Clams) from the Coast of Massachusetts. *Journal of Invertebrate Pathology* 71: 9-25
Smyth GK, Speed T. 2003. Normalization of cDNA microarray data. *Methods* 31: 265-73
Snook FC, Tipton PA, Beamer LJ. 2003. Crystal Structure of GDP-Mannose Dehydrogenase: A Key Enzyme of Alginate Biosynthesis in P. aeruginosa. *Biochemistry* 42: 4658-68
Srinivasan R. 2008. Susceptibility of legume pod borer (LPB), Maruca vitrata to delta-endotoxins of Bacillus thuringiensis (Bt) in Taiwan. *Journal of Invertebrate Pathology* 97: 79-81
Stanley SL, Jr., Reed SL. 2001. Microbes and microbial toxins: paradigms for microbial-mucosal interactions. VI. Entamoeba histolytica: parasite-host interactions. *American Journal of Physiology. Gastrointestinal and Liver Physiology* 280: G1049-54
Stehr F, Kretschmar M, Kröger C, Hube B, Schäfer W. 2003. Microbial lipases as virulence factors. *Journal of Molecular Catalysis B: Enzymatic* 22: 347-55
Stokes NA, Ragone Calvo LM, Reece KS, Burreson EM. 2002. Molecular diagnostics, field validation, and phylogenetic analysis of Quahog Parasite Unknown (QPX), a pathogen of the hard clam Mercenaria mercenaria. *Disease of Aquatic Organisms* 52: 232-47
Swenerton RK, Knudsen GM, Sajid M, Kelly BL, McKerrow JH. 2010. Leishmania subtilisin is a maturase for the trypanothione reductase system and contributes to disease pathology. *The Journal of Biological Chemistry* 285: 31120-9

Talamas-Rohana P, Hernandez-Ramirez VI, Perez-Garcia JN, Ventura-Juarez J. 1998. Entamoeba histolytica contains a beta 1 integrin-like molecule similar to fibronectin receptors from eukaryotic cells. *Journal of Eukaryotic Microbiology* 45: 356-60

Tenhaken R, Voglas E, Cock JM, Neu V, Huber CG. 2011. Characterization of GDP-mannose dehydrogenase from the brown alga Ectocarpus siliculosus providing the precursor for the alginate polymer. *The Journal of Biological Chemistry* 286: 16707-15

Torto-Alalibo T, Tian M, Gajendran K, Waugh ME, van West P, Kamoun S. 2005. Expressed sequence tags from the oomycete fish pathogen Saprolegnia parasitica reveal putative virulence factors. *BMC Microbiology* 5: 46

Torto TA, Li S, Styer A, Huitema E, Testa A, et al. 2003. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen Phytophthora. *Genome Research* 13: 1675-85

Tsui CK, Marshall W, Yokoyama R, Honda D, Lippmeier JC, et al. 2009. Labyrinthulomycetes phylogeny and its implications for the evolutionary loss of chloroplasts and gain of ectoplasmic gliding. *Molecular Phylogenetics and Evolution* 50: 129-40

Uno T, Ishizuka M, Itakura T. 2012. Cytochrome P450 (CYP) in fish. *Environmental Toxicology and Pharmacology* 34: 1-13

van Goor H, Melenhorst WB, Turner AJ, Holgate ST. 2009. Adamalysins in biology and disease. *The Journal of Pathology* 219: 277-86

Vij S, Tyagi AK. 2008. A20/AN1 zinc-finger domain-containing proteins in plants and animals represent common elements in stress response. *Functional & Integrative Genomics* 8: 301-7

Vines RR, Ramakrishnan G, Rogers JB, Lockhart LA, Mann BJ, Petri WA, Jr. 1998. Regulation of adherence and virulence by the Entamoeba histolytica lectin cytoplasmic domain, which contains a beta2 integrin motif. *Molecular Biology of the Cell* 9: 2069-79

Wallace AJ, Stillman TJ, Atkins A, Jamieson SJ, Artymiuk PJ. 2000. E. coli Hemolysin E (HlyE, ClyA, SheA): X-Ray Crystal Structure of the Toxin and Observation of Membrane Pores by Electron Microscopy. *Cell* 21: 265-76

Wanyiri JW, Techasintana P, O'Connor RM, Blackman MJ, Kim K, Ward HD. 2009. Role of CpSUB1, a subtilisin-like protease, in Cryptosporidium parvum infection in vitro. *Eukaryotic Cell* 8: 470-7

Wernersson R. 2006. Virtual Ribosome--a comprehensive DNA translation tool with support for integration of sequence feature annotation. *Nucleic Acids Research* 34: W385-8

Whyte SK, Cawthorn RJ, McGladdery SE. 1994. QPX (Quahog Parasite X), a pathogen of northern quahaug Mercenaria mercenaria from the Gulf of St. Lawrence, Canda. *Disease of Aquatic Organisms* 19: 129-36

Withers-Martinez C, Jean L, Blackman MJ. 2004. Subtilisin-like proteases of the malaria parasite. *Molecular Microbiology* 53: 55-63

Yeoh S, O'Donnell RA, Koussis K, Dluzewski AR, Ansell KH, et al. 2007. Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes. *Cell* 131: 1072-83

Yuda M, Sawai T, Chinzei Y. 1999. Structure and expression of an adhesive protein-like molecule of mosquito invasive-stage malarial parasite. *The Journal of experimental medicine* 189: 1947-52

Zdobnov EM, Apweiler R. 2001. InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847-8

Zhang S, Widemann E, Bernard G, Lesot A, Pinot F, et al. 2012. CYP52X1, representing new cytochrome P450 subfamily, displays fatty acid hydroxylase activity and contributes to virulence and growth on insect cuticular substrates in entomopathogenic fungus Beauveria bassiana. *The Journal of Biological Chemistry* 287: 13477-86