Flinders
UNIVERSITY
ADELAIDE • AUSTRALIA

Evolution, gene expression and enzymatic production of Tyrian purple: A molecular study of the Australian muricid *Dicathais orbita* (Neogastropoda: Muricidae)

Patrick Laffy

Faculty of Science and Engineering

School of Biological Sciences

Flinders University

November 2011

Declaration

I certify that this thesis does not incorporate, without acknowledgment, any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge it does not contain any material previously published or written by another person except where due reference is made in the text

…….. ……..

Patrick William Laffy Date

Acknowledgements

Where to begin? First and foremost I would like to thank my supervisors Associate Professor Cathy Abbott and Dr Kirsten Benkendorff for their guidance and support throughout this project and the last 5 years of work. Your feedback, musing of ideas and scientific knowledge has made me into a much better writer and scientist than I was 5 years ago. Not only did you conceptualize this project before I started my candidature, but you provided the financial funding to make it all possible.

I also need to thank several people who were lent their scientific knowledge and support to me. To Dr Melissa Pitman, Dr Melanie Sulda, Dr Ana Glavinic, Dr Chantel Westley and Associate Professor Michael Schwarz, I have appreciated your feedback, advice and your tireless support to the work that I have produced. I wouldn't have been able to complete this thesis without you. Special thanks to Melissa, Melanie, and Ana for the friendship you offered at the same time. And another thank you to Melissa (my, she is popular!) and Dr Simon Schmidt, who over the last year or so have turned Thursday night dinner into much needed impromptu lab meetings. We have a very supportive group of students (and now post-docs) in the School of Biological Sciences at Flinders, and I have learnt a lot from all of you, in particular, Dr Chevaun Smith, Dr Melissa Gregory, Lexi Young, Dr Simon Williams, Emma Decourcey–Ireland, Nick Warnock, Pradeep Sornaraj, Bianca Kyriacou, Dr Peter Bain, Drew Sutton and Crystal Sweetman. I am glad to call you my friends and you have all made coming to work everyday a wonderful thing, and I am really going to miss you guys when I'm gone (fingers crossed it will be soon). Special thanks to Lexi for all her help editing and formatting my thesis.

To two great labs at Flinders who have welcomed me in, despite the fact that my project didn't exactly conform to the typical lab experiments. To A/Prof Cathy Abbott, Dr Tong Chen, Dr Roger Yazbek, Dr Melissa Pitman, Dr Melanie Sulda and Dono Indarto, thank you for exposing me to the world of dipeptidyl peptidases and lending me your molecular knowledge. And to all the members of the Mollusc lab; Dr Kirsten Benkendorff, Dr Chantel Westley, Dr Cassandra McIver, Dr Ana Glavinic, Dr Skye Woodcock, David Rudd, Warwick Noble, Tom Stewart, Casey Campleman and Ryan Baring, your marine biology know-how (and essential skills in maintaining marine aquaria) have contributed greatly to my project. Many thanks also go to Dr Chevaun Smith, Associate Professor Kathleen Soole and Professor David Day for keeping me gainfully employed over the last year, I really appreciate it.

Thanks also to my friends Andrew Cook, Ana dela Cruz, Johan Velleman, Natasha Pietsch, Andrew Harding, Jamie Bowles and Jodi Sargent. You have no idea how much I have appreciated having a group of understanding people around me who accept me for me and can tolerate endless scientific ramblings without loosing your cool, or falling asleep with boredom. And to my gorgeous godchild Isabella, who can lighten up even the darkest mood with a giggle or a smile.

To my sisters Bridget and Siobhan Laffy, who have put up with me over the last few years. I have been moody, messy and generally disagreeable and I appreciate your support, tolerance and love.

Finally, thank you to my parents, Pat and Catherine Laffy. You have shaped me into the person I am today. In particular to Dad, who has supported me financially over the course of the last 5 years (and, well the rest of my life) and who has never questioned my need to pursue a career in research. It hasn't been easy, but I hope you are proud of the achievements I have made. I promise to get a real job as soon as I can.

Abstract

Tyrian purple is the traditional source of purple pigmentation used in the textile industry since ancient times, sourced from the Muricidae family of neogastropod molluscs. Brominated indole derivatives of tryptophan, the precursors to Tyrian purple are potent anticancer and antibacterial compounds which may have potential for pharmaceutical development. In addition to their production within the hypobranchial gland, some members of the Muricidae invest Tyrian purple precursors within their egg capsules. The first aim of this thesis was to investigate the evolution of Tyrian purple precursor investment within the egg masses of muricid molluscs using a molecular phylogenetic analysis of 18s and 28s ribosomal RNA sequences. The second aim of this thesis was to investigate the gene expression of the hypobranchial gland of the Australian muricid *Dicathais orbita*, in an effort to uncover the enzymes involved in the production of Tyrian purple.

The investigation into the evolution of Tyrian purple precursor investment within the egg capsules of muricid molluscs identified that the capacity for adults to invest these compounds in their egg capsules is a trait that was not ancestral and has arisen at least twice in the evolution of the Muricidae. Molecular analysis confirmed the monophyly of the Rapaninae and Ocenebrinae muricid subfamilies members and supports Tan's 2003 classification of a new muricid subfamily, the Haustrinae. These findings also support the use of *D. orbita* as a representative of the Rapaninae in which to study Tyrian purple synthesis and investment.

Suppressive subtractive hybridization (SSH) was used to identify genes that were up-regulated or uniquely expressed in the hypobranchial gland of *D. orbita*. A total of 438 sequences were identified to be differentially expressed in the hypobranchial gland, including an arylsulfatase gene. Arylsulfatase activity is known to be involved in the formation of Tyrian purple from precursors in muricid molluscs. The full length arylsulfatase sequence was amplified and recombinantly expressed in a mammalian expression system. No active enzyme was produced from these experiments suggesting an incompatibility between molluscan arylsulfatase and mammalian expression systems.

Initial manual sequence analysis indicated that over 65% of sequences expressed in the hypobranchial gland showed no homology to known database sequences. The subset of genes

that did show sequence matches to genes in the database showed homology to a wide variety of taxa, including chordate, molluscan and ciliate sequences. Our investigation into the gene expression of the hypobranchial gland of *D. orbita* enabled the functional assignment of 110 sequences using BLAST2GO automated sequence annotation. The hypobranchial gland plays a key role in muricid biology as a site of chemical interaction and biosynthesis. Manual sequence annotation also identified a number of sequences within our cDNA library that would only be functional if translated using an alternate codon translation system used by ciliate protozoans. Histological analysis of the hypobranchial gland identified intracellular ciliate protozoans present within the gland. Ciliate abundance varied in accordance to the reproductive condition of the host snail and 57 ciliate protein coding genes were identified within our cDNA library. Analysis of ribosomal RNA sequences from our expression library confirmed the presence of ciliate protozoans within the hypobranchial gland of *D. orbita* belonging the ciliate class Phyllopharyngea and possibly from another unidentified ciliate class. A novel use of SSH is proposed for the investigation of symbiont gene expression in other biological systems.

In summary, this thesis uses molecular techniques to explore the synthesis and evolution of Tyrian purple and hypobranchial gland gene expression in the muricid mollusc *D. orbita*. This thesis is the first study to investigate the evolution of Tyrian purple precursor investment within the egg capsules of muricid molluscs and has revealed that this is a derived trait that has arisen at least twice since the muricids diverged from other Muricoidean species. In addition, this is the first study to investigate gene expression within the hypobranchial gland of any mollusc. This study also identified one of the gene sequences involved in the enzymatic production of these bioactive compounds. Further investigations are required in order to produce active recombinant molluscan arylsulfatase enzymes. Additional investigations are also required in order to identify the other enzymes involved in the production of Tyrian purple precursors, which would then facilitate the *in vitro* synthesis of these compounds in the future. If sustainable synthesis of these compounds is established, these bioactive compounds may find use in pharmaceutical or nutraceutical treatments.

Preface

Parts of the work presented in this thesis have been published, or are currently in preparation for submission for publication. For consistency and ease of reading, all manuscripts are presented in the required formatting for the Journal of Biological Chemistry. As all references are listed in a single reference list at the end of the thesis to eliminate repetition, Harvard referencing was used in the text.

Published papers

Chapter 3. Laffy, PW, Benkendorff, K & Abbott, CA "Trends in molluscan gene sequence similarity: An observation from genes expressed within the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae)" *Nautilus*, 123(3), 154-158.

Papers in preparation

Chapter 2. Laffy, PW, Schwarz, MP, Abbott CA & Benkendorff, K "The evolution of bioactive Tyrian purple precursors in the egg capsules of the Muricidae (Mollusca: Gastropoda)".Molluscan Research, under preparation.

Chapter 4 Laffy, PW, Benkendorff, K & Abbott, CA "Annotation and characterization of a partial transcriptome of the hypobranchial gland of *Dicathais orbita*" *Marine Biotechnology,* under preparation.

Chapter 5 Laffy, PW, Westley C, Abbott, CA & Benkendorff "Novel application of suppressive subtractive hybridization for the identification of symbionts: Discovery of ciliate protozoa in the hypobranchial gland of *Dicathais orbita* (Neogastropoda, Mollusca)" Molecular Ecology, under preparation.

Chapter 6 Laffy, PW, Chen, T, Benkendorff, K & Abbott, CA "Characterisation and expression of recombinant arylsulfatase from the marine snail *Dicathais orbita*" Comparative Biochemistry and Physiology Part B Biochemistry and Molecular Biology, under preparation.

The data published in abstract form and contributed to this thesis:

Laffy, PW, Benkendorff, K & Abbott CA (2006) "Gene expression within the hypobranchial gland of *Dicathais orbita*" Oral presentation at Molluscs 2006 meeting, University of Wollongong, Australia 6-8th December 2006.

Laffy, PW, Benkendorff, K & Abbott CA (2007) "Using molecular approaches to investigate the function of the hypobranchial gland of the marine snail, *Dicathais orbita*" Oral presentation at World Congress of Malacology, Antwerp, Belgium 15-20th July 2007

Laffy, PW, Benkendorff, K & Abbott CA (2008) "Transcriptomic analysis of the functions of the hypobranchial gland of the marine snail *Dicathais orbita*, using BLAST2GO" Poster presentation at ASMR Annual Meeting, Adelaide, Australia 4th June 2008

Laffy, PW, Benkendorff, K & Abbott CA (2008) "Annotating a partial transcriptome of the marine snail *Dicathais orbita's* hypobranchial gland using suppressive subtractive hybridization and BLAST2GO" Poster presentation at the 19th international conference on Genome Informatics, Gold Coast, Australia, 1st-3rd December 2008

Laffy, PW, Westley, C, Benkendorff, K & Abbott CA (2009) "Parallel genome identification: A novel use of Suppressive Subtractive Hybridization in systems biology" Poster presentation at the Gordon Research Conference on Evolutionary and Ecological Functional Genomics In Tilton, New Hampshire, USA, 12-17 July 2009

Abbreviations

[N-morpholino] ethanesulfonic acid	MES
Adenosine triphosphate	ATP
Amino acid	aa
Australian Genome Research Facility	AGRF
Basepair	bp
Basic local alignment search tool	BLAST
Bayesian inference	BI
Celcius	C°
Cell membrane	Cm
Complimentary DNA	cDNA
Cytochrome oxidase subunit I	COI
Cytoplasm	Cy
Daltons	Da
Deoxynucleotide triphospates	dNTPs
Deoxyribose nucleic acid	DNA
Diethyl pyrocarbonate	DEPC
Dipeptidyl peptidase	DP
Dipeptidyl peptidase IV	DPIV
Ethylenediaminetetraacetic acid	EDTA
Example	e.g.
Expressed sequence tags	ESTs
Figure	Fig
Foetal calf serum	FCS
Formylglycine Generating Enzyme	FGE
Gas chromatography-mass spectrometry	GC/MS
Gene ontology	GO
Guanosine triphosphate	GTP
H-Ala-Pro-p-nitroanilide	H-Ala-Pro-pNA
Heat shock protein	HSP
Hour	hr
Hypobranchial gland	HBG

Internal transcribed spacer region 2	ITS2
Inter-Services Intelligence	ISI
KiloDaltons	kDa
Kyoto Encyclopedia of Genes and Genomes	KEGG
Mantle cavity	Mc
Mass spectrometry	MS
Maximum parsimony	MP
Micrograms	µg
Microlitres	µl
Micrometre	µm
Micromolar	µM
Milligram	mg
millilitre	ml
milliMolar	mM
Minutes	min
Modified Harris haematoxylin and Eosin Y	H&E
Molar	M
Nanogram	ng
Nanometres	nm
National Centre for Biotechnology Information	NCBI
Nucleus	Nu
Open reading frame	ORF
Phospho-buffered saline	PBS
Phospho-buffered saline buffered Tris	PBST
p-nitrocatechol sulfate	pNCS
Polyacrylamide gel electrophoresis	PAGE
Polymerase chain reaction	PCR
Polyvinylidene Fluoride	PVDF
Posterior probability	PP
Rectal gland	Rg
Rapid amplification of cDNA ends	RACE
Reverse jump hyperprior	rjhp
Revolutions per minute	RPM

Ribose nucleic acid	RNA
Ribosomal RNA	rRNA
Room temperature	RT
Seconds	sec
Secretion	Sc
Sodium dodecyl sulfate	SDS
South Australian Partnership for Advanced Computing	SAPAC
Standard error	S.E.
Stress-inducible protein	STI-1
Suppressive subtractive hybridization	SSH
Tetramethylethylenediamine	TEMED
units	U
Volts	V

Table of Contents

Table of Figures

Data Tables

Chapter 1. General introduction

1.1 The biomedical potential of marine natural products

The development of pharmaceutical compounds is one of the most pressing and expensive industries in current biotechnology. A survey on the development of new drugs estimated a final cost of up to US $802 million for each drug developed, taking into account the final clinical success rate of 21.5% (DiMasi *et al.* 2003). Given the astronomical cost of developing such compounds, it is logical to adopt evidence-supported methods when identifying bioactive compounds for use in disease treatment. Products identified in nature have been used for thousands of years for the treatment of a variety of ailments. Over 60% of approved pharmaceutical agents are either natural products, are derived from natural products or utilize natural products as lead compounds in their production and synthesis (Demain 2009). While history supports the use of terrestrial natural products in the development of pharmaceuticals, the marine environment is an almost untapped resource with amazing potential for future drug development. This review highlights the potential of marine natural products in the treatment and management of cancer and microbial infection and highlights the potential muricid molluscs have in the development of new pharmaceutical and nutraceutical treatments.

1.1.2 Marine-derived cancer treatments

Over 100,000 new cases of cancer were reported in Australia in the year 2005, and the incidence of cancer diagnosis was predicted to increase by 3,000 people every year in this country until the year 2010 (AIHW (Australian institute of Health and Welfare) & AACR (Australasian Association of Cancer Registries) 2008). In addition, 39,000 people died of cancer in Australia in 2005, with an additional 800 dying each year until 2010 (AIHW (Australian institute of Health and Welfare) & AACR (Australasian Association of Cancer Registries) 2008). The increases in cancer diagnoses, as well as the limitations of current treatment options, have prompted the discovery of new treatment alternatives in the clinical realm. Traditional chemotherapeutic agents act on rapidly dividing cells but are unable to differentiate between healthy and cancerous cells (Ma & Wang 2009). A more recent focus of anticancer research entails the identification of chemotherapeutic agents that specifically target cancer cells (Sawyers 2004). One of the most effective sources of target-specific anticancer compounds is the natural environment. Natural products are typically more effective in their anticancer activity while being less toxic than traditional chemotherapeutic compounds or chemically synthesized compounds (Ma & Wang 2009). While traditional naturally derived pharmaceutical compounds are almost always of terrestrial origin, a burgeoning bioprospecting

field focusing on marine organisms as a source of novel and bioactive compounds has developed (Haefner 2003).

The sea squirt-derived Trabectedin (also known as yolendis and ET-743) has been developed and marketed as an anticancer compound used to target soft tissue sarcoma (Carter & Keam 2007). The macrocyclic lactone compound Bryostatin 1 from the bryozoan *Bugula nerita* has protein kinase C inhibitory qualities and has shown potential for use in dendritic cell-based anticancer treatments (Do *et al.* 2004). The sea sponge derived calyculins, originally identified in *Discodermia calyx,* may show potential in cancer treatments due to their protein phosphatase inhibition activity (Fagerholm *et al.* 2010). Heteronemin, a sesterterpene isolated from the sponge *Hyriotis* sp. has apoptosis inducing ability against myeloid leukemia cell lines (Schumacher *et al.* 2010). Kahalahide F is a depsipeptide derived from the mollusc *Elysia rufescens,* which is currently in phase II clinical trials for the treatment of prostate cancer (Faircloth & Cuevas 2006). Dolastatin 10, a peptide from the sea hare *Dolabella auricularia,* binds to tubulin causing cell cycle arrest and has been shown to inhibit cancer growth in leukemic cells (Pettit *et al.* 1987). While dolastatin 10 proved too problematic in phase I clinical trials (Pitot *et al.* 1999), its chemical analogue TZT-1027 has shown promise in phase I solid tumor clinical trials (Schoffski *et al.* 2004). Another depsipeptide, Aplidine from the sea squirt *Aplidum albicans*, inhibits vascular endothelial growth factor, which causes cell cycle arrest and has shown potential for the treatment of breast cancer and leukemia (Erba *et al.* 2002, Cuadrado *et al.* 2003). Brominated indoles from the of the marine mollusc *Dicathais orbita* have exhibited cytotoxic activity against cancer cell lines (Vine *et al.* 2007, Benkendorff *et al.* 2011), with preclinical studies for colon cancer showing the extracts induce apoptosis *in vivo* (Westley *et al.* 2010c). The variety of compounds derived from marine organisms that have potential as cancer treatments supports the trend into surveying marine organisms for more pharmaceutically active compounds. The surveying of marine organisms for antimicrobial activity has also identified a variety of different compounds which have the potential in the treatment of microbial infections.

1.1.3 Marine-derived microbial infection treatments.

The increase in prevalence of multidrug resistant bacteria is becoming one of the most prominent medical issues in hospitals today. Nosocomial infections, also known as hospital-acquired infections, infect approximately two million people every year in the United States of America and are responsible for approximately 100,000 deaths annually (Balaban & Dell'Acqua

2005).There has been a marked decrease in antibiotic drug discovery in the pharmaceutical industry since the late 1980's (Demain 2009), caused by the low cost of recovery from their development together with a declining rate of discovery of novel chemical structures displaying antibacterial activity. This has reduced treatment options for infections and is partially responsible for the prevalence of drug resistant bacteria in our medical institutions. Hospitals around the world are faced with the financial burden of prolonged stays and additional treatment costs due to the abundance of drug resistant nosocomial infections (Croft *et al.* 2007). In addition to the increase in hospital based infections, the prevalence of community derived drug resistant bacteria place more and more of the community at risk. Without the development of new treatment methods for drug resistant bacteria, we are likely to see an increase in nosocomial infections and may start to witness the rise of multidrug resistant bacteria, resulting in increased mortality rates from infections and an inflated financial burden on healthcare systems worldwide.

Bacteria are the traditional source of the majority of antibiotics that are used in medicine (Demain 2009). This is due to the frequency of bacterial interactions and environmental competition. The marine environment is one of the most bacteria rich environments on the planet, with some studies estimating the number of bacteria per milliliter of sea water as high as 10^6-10^9 (Haug *et al.* 2004). The world's oceans span up to 70% of the planet's surface and contain an amazing diversity of organisms (Faulker 2000). Many invertebrate species in the marine environment, particularly those living in or on the benthic substrata, such as members of the Porifera, Mollusca, Cnidaria and Echinodermata, have developed potent antibacterial secondary metabolites in order to defend themselves from microorganism attack (Shaw *et al.* 1974). While traditional drug development studies rely on terrestrial species for the identification of new antibacterial compounds, the marine environment is emerging as an ideal place to identify new and novel antibacterial compounds for clinical use.

Gorgonian corals have been shown to exhibit antimicrobial activity against the marine pathogen *Aspergillus sydowii* and the human pathogen *Aspergillus flavus* (Kim *et al.* 2000). Marine steroids and sesquiterpenes isolated from these gorgonian corals were shown to also display potent antimicrobial against several human bacterial pathogens (Roussis *et al.* 2001). Other marine invertebrates including the sea urchin *Strongylocentrotus driebachiensis*, the sea star *Asterias rubens* and the sea cucumber *Cucumaria frondosa* all exhibit antibacterial activity against several Gram positive bacteria via an unidentified mechanism of action (Haug *et al.*

2002b). The polyhydroxylated fucophlorethol isolated from the marine alga *Fucus vesiculosus* exhibits antibacterial activity against Gram positive and Gram negative human pathogenic bacteria (Sandsdalen *et al.* 2003). Secondary metabolites with antibacterial activity have been identified in several marine sponges including *Didiscus oxeata* and *Hippospongia communis. Didiscus oxeata* produces at least two secondary metabolites, (+)-curcuphenol and (+)-curcudiol, which show antifungal activity against the human pathogenic fungus *Trichiphythton mentagrophytes*. *Hippospongia communis* produces (-)-untenospongin B, a C21 bisfuranoterpene with antimicrobial activity against human pathogenic fungi, bacteria and yeast (Rifai *et al.* 2004). Marine decapods have also shown antibacterial (Haug *et al.* 2002a, Haug *et al.* 2002b).

Marine molluscs have also shown potential as a source of antimicrobial secondary metabolites. Antimicrobial peptides and other unidentified compounds within the horse mussel *Modiolus modiolus* inhibit the growth of bacteria including the human pathogen *Staphylococcus aureus* (Haug *et al.* 2004). Antibacterial and antiviral activity has been detected in chemical extracts from the common Cockle *Cerastoderma edule*, the common whelk *Buccinum undatum*, the Japanese carpet shell *Ruditapes philippinarum* and the European flat oyster *Ostrea edulis* (Defer *et al.* 2009). Tribromoimidazole was isolated from the egg masses of the muricid molluscs *Trunculariopsis trunculus*, *Ceratosoma erinaceum* and *Trophon geversianus* and was shown to exhibit antibacterial activity (Benkendorff *et al.* 2004a). Antimicrobial peptides have also been identified in several bivalve species (Mitta *et al.* 2000, Cellura *et al.* 2007, Zhao *et al.* 2007, Li *et al.* 2009). Brominated indoles from the egg capsules of the Muricidae also exhibit antibacterial activity (Benkendorff *et al.* 2000, Benkendorff *et al.* 2001a). A survey of antimicrobial activity from marine organisms in California in 1974 suggested that 9 out of 14 different phyla tested displayed measurable activity (Shaw *et al.* 1974), and a survey into the antimicrobial activity of molluscan egg masses identified that 18 out of 23 species tested displayed antimicrobial activity (Benkendorff *et al.* 2001b). The high incidence of bioactivity detected from marine organisms and the diversity of life in the marine environments highlights the potential of marine bioprospecting for the identification of new lead compounds which may find use in fighting infections and treating cancer in the future. Marine molluscs are one such group of organisms that show great potential for the development of new anticancer and antibacterial treatments.

1.1.4 The therapeutic potential of marine molluscs

The Mollusca is the second largest animal phylum, consisting of approximately 7% of living animals. The marine environment is home to as many as 200,000 different species of molluscs. Members of the Mollusca are essentially soft bodied invertebrates that can be found in almost every environment on the planet, with members occupying nearly every possible trophic niche (Benkendorff 2010). A summary of the current phylogeny for the Mollusca has recently been published (Fig 1.1) utilizing phylogenomic techniques and has for the first time produced a well supported topology for the Mollusca (Kocot *et al.* 2011). In addition to being one of the most speciose animal phyla in the world's oceans, marine molluscs have developed chemical strategies to effectively fight infections and bacterial attacks (Benkendorff *et al.* 2001b, Benkendorff 2010); in an environment that has as many as 10^6 microbial cells per ml of water (Whitman *et al.* 1998). The abundance and diversity of marine molluscs make them ideal candidates for marine bioprospecting investigations.

Previous investigations into the chemical ecology and secondary metabolites produced by marine molluscs have seen a major focus on the soft bodied opisthobranch gastropods (Benkendorff 2010). Most members of this family display either a reduced shell or no shell at all, implying that, as they display no physical defences against predators, they would need to develop other means of anti-predatory defence including chemical defences (Cobb & Willan 2006). The opisthobranchs belong to the molluscan class Gastropoda and subclass Orthogastropoda, which makes up to 90% of all marine molluscs (Ponder & Lindberg 2008). Gastropod molluscs can be found in terrestrial, freshwater, marine benthic, pelagic and infaunal habitats and include herbivores, scavengers and predators (Benkendorff 2010). Yet despite this diversity and abundance, it has been reported that less than one percent of gastropods have been investigated for their chemical diversity (Avila 2006, Benkendorff 2010). Over 91% of all chemical studies on orthogastropods have been performed on members of the Heterobranchia (including opistobranchs and the air-breathing pulmonates), despite the abundance of other species within the subclass. Of the chemical studies that have been performed on orthogastropods, it has been shown that members of the Caenogastropoda show the highest number of chemical compounds, but are typically overlooked for chemical investigations due to the presence of external shells in these species (Benkendorff 2010). Benkendorff (2010) suggested future chemical studies in shelled molluscs is warranted, based on previous studies,

the species diversity that is available and the historical use of shelled molluscs in natural therapies.

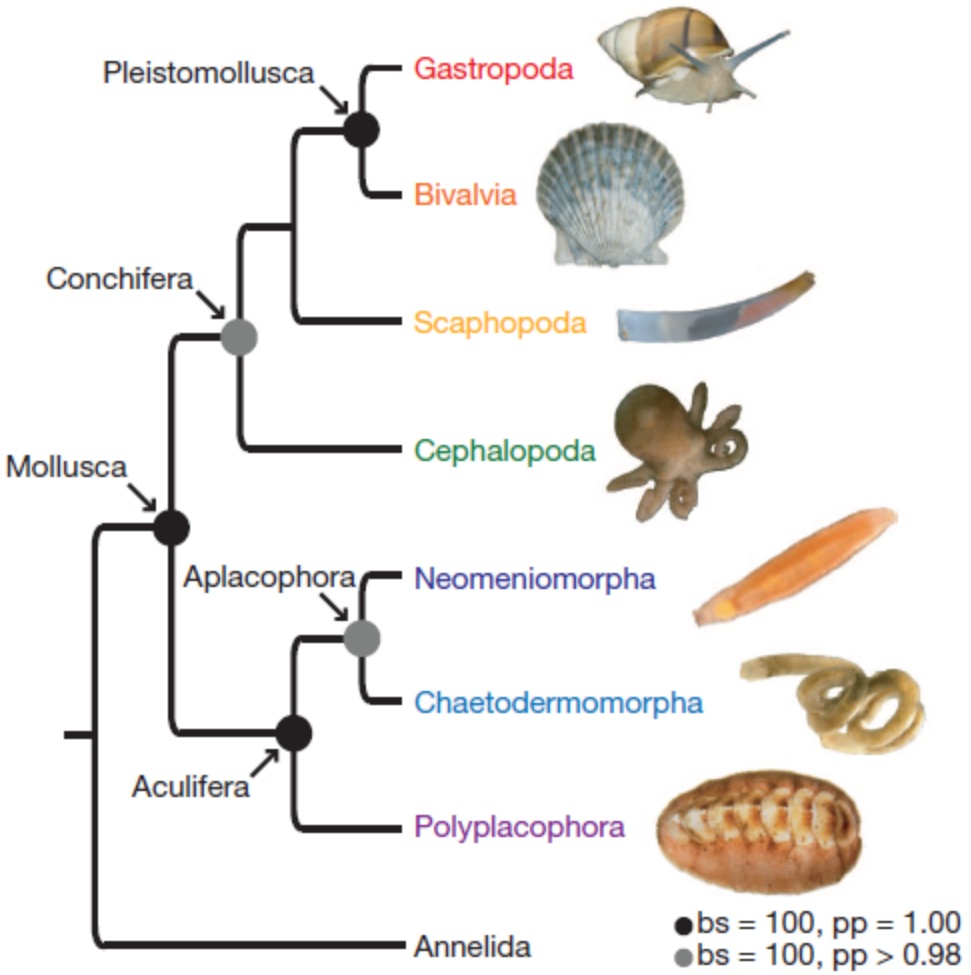


Figure 1.1 Molluscan phylogeny presented in Kocot *et al* 2011. This phylogenetic analysis was performed by compiling transcriptome data to review the evolution of the Mollusca and has been constructed from Bayesian and maximum parsimony analysis of a combined protein sequence dataset containing information from 308 protein sequences (Kocot *et al.* 2011). Reprinted by permission from Macmillan Publishers Ltd: Nature 477(7365): 452-U101.Kocot, K.M., J.T. Cannon, *et al.* (2011). "Phylogenomics reveals deep molluscan relationships."

It seems logical that members of the Caenogastropoda that are historically used in natural medicines would be an ideal source of chemical compounds, which may also find use in the pharmaceutical industry. The Murex homeopathic remedy, derived from extracts from the Muricidae, has been used since the 1800s to treat a number of ailments, including uterine and breast cancer (Boericke 1999). Brominated indoles isolated from the hypobranchial glands and egg masses of the muricid mollusc *Dicathais orbita*, have been shown to display antibacterial activity against human and marine pathogens (Benkendorff *et al.* 2000) and have exhibited

potent anticancer activity against human cancer cell lines (Vine *et al.* 2007, Benkendorff *et al.* 2011) and in a rodent model of colorectal cancer (Westley *et al.* 2010c). This historical use of murex extracts to treat illness, as well as the supportive evidence of bioactive compounds being present within the hypobranchial glands of these molluscs, prompts a deeper investigation into the medical potential of the Muricidae.

1.2 The family Muricidae

The Muricidae are a family of shelled caenogastropods, also known as murex snails. They are found all over the world and display unique shells, which are often prized by shell collectors. Muricid molluscs are predatory species that prey on other gastropods, bivalves, barnacles and ascidians (Taylor *et al.* 1980). A recent phylogenetic analysis has been performed on the Muricidae, investigating nine of the ten muricid subfamilies currently considered (Fig 1.2). The monopholy of six of these subfamilies (Ergataxinae, Rapaninae, Coralliophilinae, Haustrinae, Ocenebrinae, and Typhinae) was confirmed, as was the Monophyly of the family as a group in the Neogastropoda (Barco *et al.* 2010). Members of the Muricidae are the traditional source of the ancient pigment compound Tyrian purple.

1.2.1 Tyrian purple production

Tyrian purple, also known as royal purple, shellfish purple or purple of the ancients describes a pigmented compound that has been used since ancient times to dye garments and fabrics purple (Baker 1974). This compound was produced from the extracts and secretions of muricid molluscs and evidence of this textile industry dates back as far as the 13th century B.C in the Mediterranean (Naegel & Alvarez 2005). The biosynthesis of Tyrian purple precursor compounds occurs in the hypobranchial gland of muricid molluscs, although extracts from the gland only develop the intense purple colouring under sunlight-stimulated oxidation (Cooksey & Sinclair 2005). The chemical structure of Tyrian purple was first identified as 6,6'-dibromoindigo in 1909 by Friedlander from *Murex brandaris* (Friedlander 1909). Four Tyrian purple precursors or prochromagens were identified in *M. trunculus* and a fifth was also identified in *D. orbita* (Baker & Sutherland 1968, Baker 1974). The combination of precursors produced by each snail varies between species (Baker 1974), however all are derived from the tryptophan derivative indoxyl sulfate (Cooksey 2001).

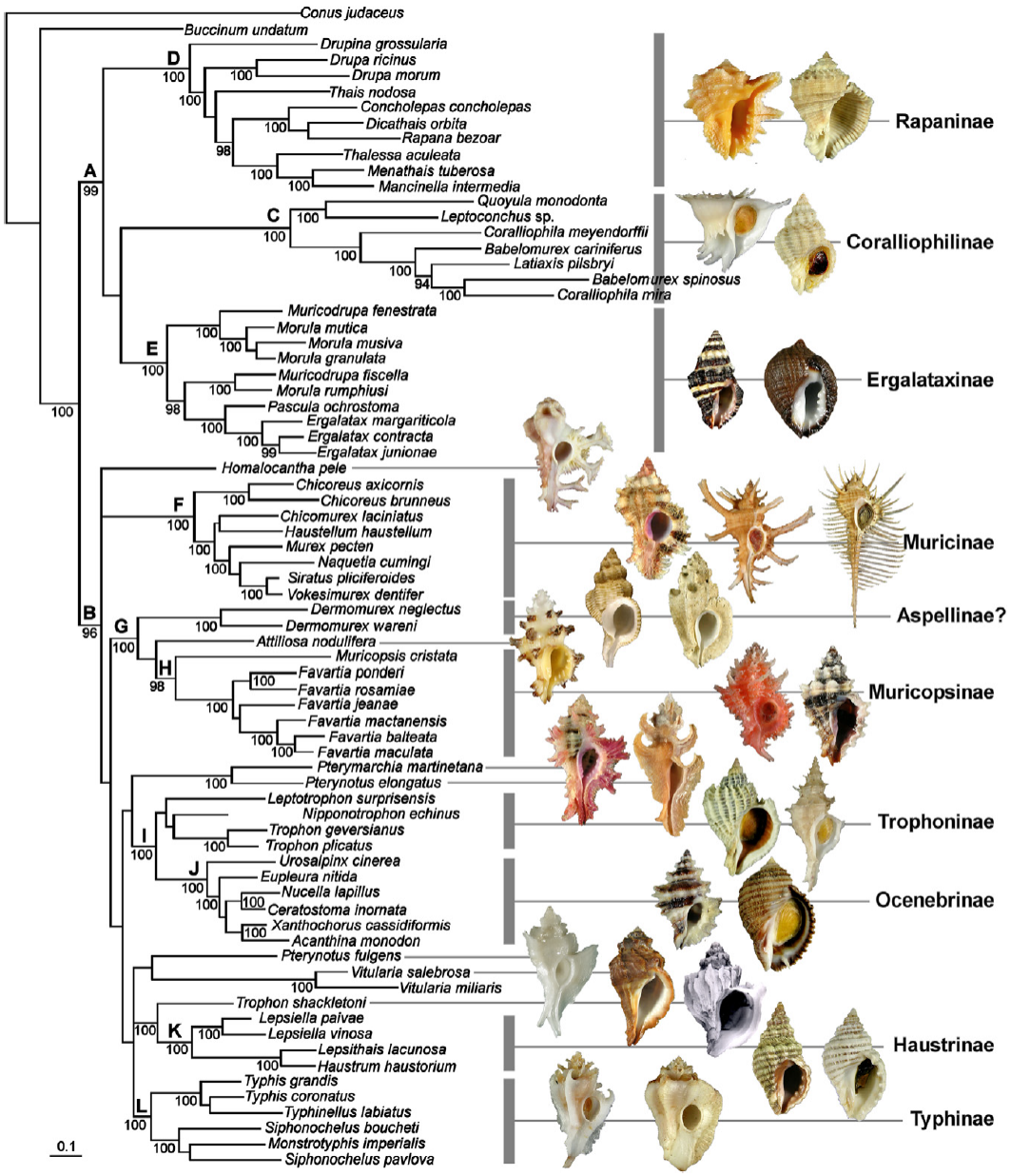


Figure 1.2 Bayesian inference of the Muricidae as reported by Barco *et al.* 2010. Phylogenetic analysis was constructed from a combined dataset of 12s, 16s and 28s ribosomal RNA sequences and Cytochrome oxidase subunit I (Barco *et al.* 2010). Bayesian analysis confirms the monophyly of subfamilies Coralliophilinae (C), Rapaninae (D), Ergalataxinae (E), Muricinae (F), Muricopsinae (H) and Haustrinae (K). Reprinted from Molecular Phylogenetics and Evolution, Volume 56(3), Barco, A., M. Claremont, *et al.* "A molecular phylogenetic framework for the Muricidae, a diverse family of carnivorous gastropods." 1025-1039 (2010).

The chemical formation of Tyrian purple in *D. orbita* is summarized in Figure 1.3. In order to produce the initial precursor tyrindoxyl sulfate, it has been suggested that different enzymes are required to convert the amino acid tryptophan into a brominated indole. Westley *et al.* (2006) suggested the most likely method of indole production is the *de novo* conversion of

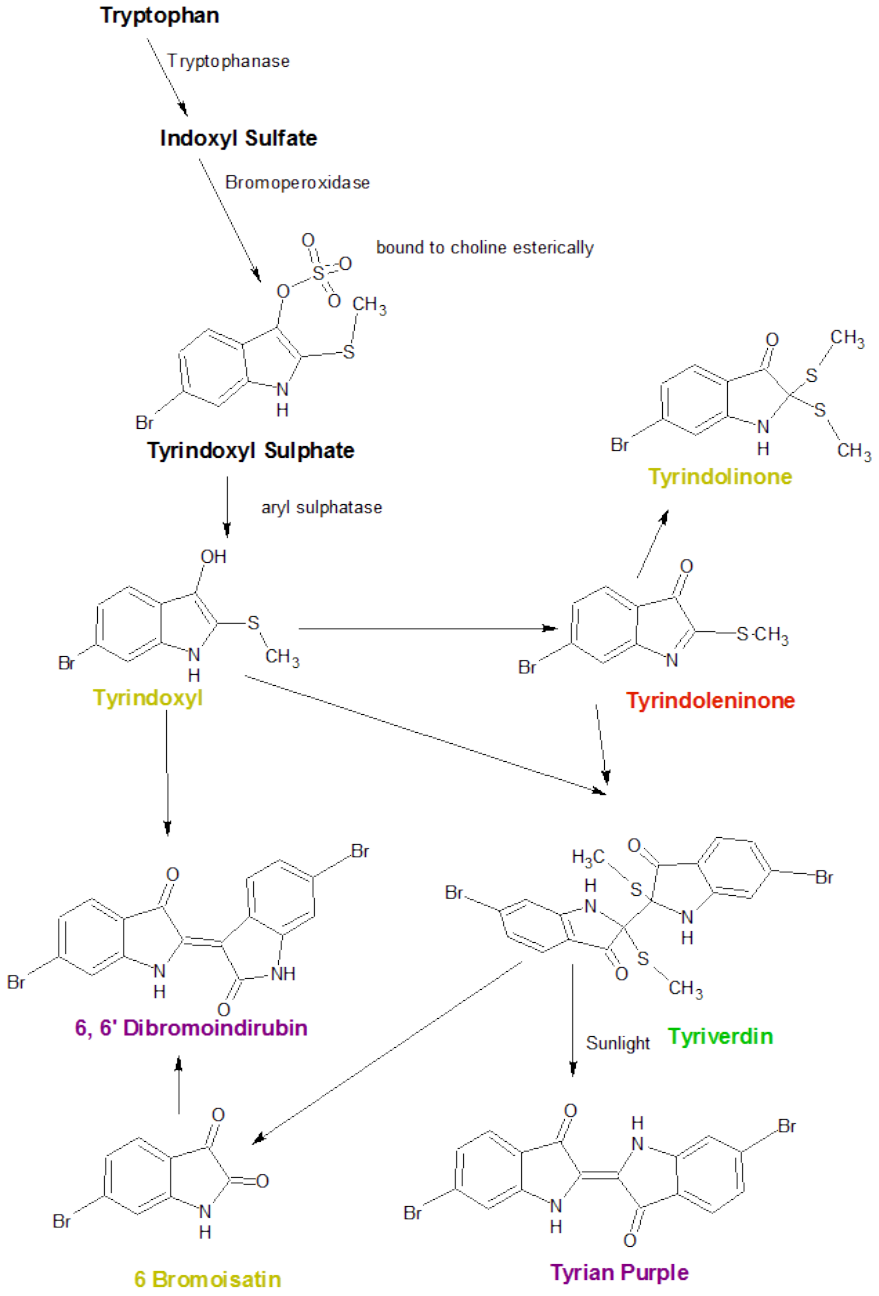


Figure 1.3 Chemical pathway involved in the formation of Tyrian purple.
This figure shows the various colour changes seen in the formation of Tyrian purple. The amino acid tryptophan undergoes degradation from a (hypothetical) tryptophanase enzyme forming indoxyl sulfate, and bromoperoxidase then reacts with the indole ring, forming tyrindoxyl sulfate (1), with the addition of methane thiol from unknown enzymes. The white compound tyrindoxyl sulfate (1), found in the hypobranchial gland is cleaved by an arylsulfatase enzyme, forming the yellow compound tyrindoxyl (2). Tyrindoxyl (2) forms the orange tyrindoleninone (3) which either reacts with a methane thiol group to form the green/yellow tyrindolinone (5) or reacts with tyrindoxyl (2) to form the ultimate precursor of Tyrian purple (6), tyriverdin (4). The green compound tyriverdin (4) is cleaved or degraded under sunlight to form Tyrian purple (6). Tyriverdin (4), tyrindoleninone and tyrindoxyl can also oxidize to form the compound 6-bromoisatin (7), a compound yellow in colour. The reaction of 6-bromoisatin with tyrindoleninone generates 6.6' dibromoindirubin (8), a structural isomer of Tyrian purple, 6,6, dibromoindigo. (Adapted from Cooksey, 2001 & Westley *et al.* , 2006)

dietary derived tryptophan to indole or indoxyl sulfate using a tryptophanase enzyme. Additional investigations identified significant concentrations of tryptophan in the hypobranchial gland region based on histochemical examination (Westley & Benkendorff 2008, Westley & Benkendorff 2009). The incorporation of the bromine in the 6-position of the indole ring of the prochromagens (1-4) implies that a bromoperoxidase or other haloperoxidase is also involved in the processing of tryptophan or indole. Bromoperoxidase activity has been identified in hypobranchial gland homogenates in *M. trunculus* (Jannun & Coe 1987) and histological sections of the hypobranchial gland of *D.orbita* (Westley & Benkendorff 2009), but the protein responsible for this activity has yet to be isolated. The ultimate precursor to Tyrian purple (6), tyrindoxyl sulfate (1), an indoxyl sulfate which is esterically bound to choline esters in the hypobranchial gland (Baker & Duke 1976), is oxidized by an arylsulfatase enzyme to form tyrindoxyl (2) (Baker & Sutherland 1968). The resulting prochromagen, which is in turn oxidized to form tyrindoleninone (3) was isolated from diethyl ether extracts of the gland identified, along with small amounts of the methanethiol adduct tyrindolinone (5) (Baker & Duke 1973b, Baker 1974). The prochromagens tyrindoleninone (3) and tyrindoxyl (2) combine to form the intermediate prochromagen tyriverdin (4) (Christopherson *et al.* 1978), which is externally oxidized to form the commercially important Tyrian purple (6). An oxidative artifact to Tyrian purple, 6-bromoisatin (7), has also been identified as a degradation product of tyriverdin (4) (Baker 1974), tyrindoleninone or tyindoxyl (2) (Cooksey, 2001).

1.2.2 Muricid bioactive compounds,

As discussed previously, several of the precursors of Tyrian purple have been found to be biologically active, exhibiting anticancer and antibacterial activity. The precursors tyrindoleninone, tyriverdin and the artefact 6-bromoisatin, isolated from the muricid *D. orbita*, have all been shown to exhibit varying levels of antibacterial activity (Benkendorff *et al.* 2000). These compounds are produced in the hypobranchial gland (Baker & Sutherland 1968, Baker & Duke 1976), but have also been reported from the egg capsules and reproductive organs (Benkendorff *et al.* 2000, Westley & Benkendorff 2008). It has been suggested that *D. orbita* invests tyrindoxyl sulfate inside egg capsules where the bioactive compounds are generated, protecting the developing embryos from bacterial infections (Benkendorff *et al.* 2000, Westley & Benkendorff 2008). Furthermore, the artefact 6-bromoisatin and the precursor tyrindoleninone have been shown to be cytotoxic to colon and breast carcinoma and lymphoma cell lines (Vine *et al.* 2007, Benkendorff *et al.* 2011), as well as inducing apoptosis in response to a carcinogen

in an *in vivo* rodent model for colorectal cancer (Westley *et al.* 2010c). Additional brominated indoles form as minor pigments of Tyrian purple, 6- dibromoindirubin and 6,6'-dibromoindirubin. These were isolated from the hypobranchial glands of *D. orbita* (Baker & Duke 1973a, Baker 1974) and in *Hexaplex trunculus,* where its specific glycogen-synthase-kinase-3 (GSK3) inhibitory activity has been observed (Meijer *et al.* 2003). The inhibition of GSK3 in humans reduces the rate of apoptotic cell death and increases neuronal survival by facilitating the regulation of Tau and β-Catenin molecules, implicated in neurodegeneration (Culbert *et al.* 2001). A traditional Chinese remedy, Ganggui Luhui Wan, that is used to treat chronic myelogenous leukemia in China, contains plant derived indirubin as an active ingredient, responsible for the treatments' antiproliferative and apoptosis inducing mechanism (Xiao *et al.* 2002). In addition to the bioactivity that several Tyrian purple precursors and pigments display, the ultimate precursor to Tyrian purple is a salt of a choline ester, that displays potent neuromuscular blocking action and muscle relaxing properties (Baker & Duke 1976, Roseghini *et al.* 1995). Tyrian purple precursor synthesis varies between muricid species (Baker 1974), and a better understanding of muricid evolution is required in order to understand the mechanisms involved in the synthesis of these bioactive compounds.

1.2.3 Muricid taxonomy and phylogeny

Australia has close to 150 different species of muricids and approximately 1600 species exist internationally (Tan 2003, Barco *et al.* 2010). However, the phylogenetic classification of species within this family is a complex field, with limited studies dedicated to resolving muricid classifications at the subfamily level. Molluscan taxonomic classifications have historically been presented based on shell morphology and anatomical characters (Kocot *et al.* 2011), however disparity between morphological characters and of family members has made Muricidae taxonomy difficult to resolve. While the application of molecular phylogenetics in the field of systematic classification was initially met with significant trepidation (Lipscomb *et al.* 2003, Seberg *et al.* 2003, Will & Rubinoff 2004), molecular taxonomy is an informative and useful tool in systematic classifications, particularly in areas where traditional morphological data is inconclusive (Barco *et al.* 2010, Kocot *et al.* 2011).

The taxonomic analysis of the Rapaninae subfamily by Kool (1993a), based on gross anatomical, radular, opercular and protoconch morphology combined with shell ultrastructure, reported that an updated systematic classification was required (Kool 1993a). Species that were previously classified as belonging to the subfamily Thaididae/nae were reclassified due to

the morphologically derived paraphyletic grouping of the clade, and it was proposed that the species be reclassified to belong to either the newly declared Rapaninae subfamily or the Ocenebrinae subfamily (Kool 1993a). An additional study by Kool (1993b) was performed to investigate the classification of the muricid genera *Nucella*, *Trophon* and *Ocenebra* based on anatomy, radular, protoconch, operculum and shell ultrastructure. Kool (1993b) concluded that these three genera appeared more closely related to each other than to representatives of the Rapaninae subfamily, and suggested that members of the *Nucella* genus be classified in the subfamily Ocenebrinae and that members of the *Trophon* genus be classified in the subfamily Trophoninae (Kool 1993b). Vermeij and Carlson (2000) reviewed the subfamily Rapaninae classification performing a phylogenetic analysis on the subfamily, looking at shell characteristics incorporating fossil evidence as well as gross morphological characters (Vermeij & Carlson 2000). Their results conclude that shell characteristics alone are not effective in resolving phylogenetic relationships, but they do have their place in the phylogenetic analysis, particularly in supporting lower-level relationships (Vermeij & Carlson 2000). The revision of taxonomy of Australian and New Zealand muricid species by Tan (2003) supported the classification of subfamilies Ocenebrinae and Rapaninae, and proposed a new subfamily Haustrinae (Tan 2003).

In addition to resolving the taxonomic classification of species, a robust phylogeny can be used to investigate the evolution of specific character traits within this family. A study into the development of labral spines in Ocenebrinae muricids utilized cytochrome oxidase and 12s ribosomal RNA sequences to construct the molecular phylogeny and found that the development of labral spines from marine gastropods is a trait that has developed more than once in evolutionary history (Marko & Vermeij 1999). In 2001, Oliverio and Mariottini investigated the molecular framework and phylogeny of the Corallilophila subfamily of muricids, which confirmed the monophyletic relationship observed in this subfamily (Oliverio & Mariottini 2001). Oliverio *et al.* (2002) investigated the molecular phylogeny of Muricidae family members utilizing ITS2 (internal transcribed spacer region 2) and identified the Rapaninae as a closely related sister clade to the Corallilophilae subfamily, but the bootstrap values from maximum parsimony analysis showed weak support for several clade formations (Oliverio *et al.* 2002). The largest molecular investigation into muricid phylogeny was published by Barco *et al* (2010) and resolved the monophyly of six of the ten currently classified muricid subfamilies, as well as confirming the monophyly of the Muricidae family of Neogastropods (Barco *et al.* 2010). This phylogenetic analysis of the Muricidae was only possible by performing a multigenetic study

using three mitochondrial sequences (12s, 28s ribosomal RNA and cytochrome oxidase subunit I), as well as one nuclear sequence (16s ribosomal RNA). While the relationship between all muricid subfamilies was not resolved using this dataset, Barco *et al.* (2010)'s study is, to date, the most thorough investigation into muricid taxonomy that has been presented and will greatly influence future classification of this cosmopolitan family of Neogastropods.

1.2.4 *Dicathais orbita* as a model for Tyrian purple biosynthesis

Dicathais orbita (Fig 1.4) is a marine predator found on shallow subtidal and intertidal rocky reefs across the southern coasts of Australia and New Zealand (Woodcock & Benkendorff 2008). *D. orbita* is an ideal candidate for gene expression studies due to the extensive work that has been performed on Tyrian purple production of this species. The Tyrian purple indole precursors were first identified in *D. orbita* (Baker & Duke 1976). The bioactivity of these indole compounds was also first identified from *D. orbita* extracts (Benkendorff *et al.* 2000, Vine *et al.* 2007, Benkendorff *et al.* 2011). Preliminary investigations into the aquaculture of the snail (e.g. Woodcock and Benkendorff, 2008; Noble *et al.* , 2009), as well as a detailed anatomical and histochemical investigation into the enzymatic synthesis of bioactive Tyrian purple precursors has been performed (Westley 2008, Westley & Benkendorff 2008, Westley & Benkendorff 2009, Westley *et al.* 2010a, Westley *et al.* 2010b). The strong supportive information regarding the activity of enzymes involved in the formation of these bioactive compounds makes *D. orbita* the ideal model organism in which to investigate Tyrian purple biogenesis from a gene expression standpoint.

1.3 Sustainable supply of marine bioactives

While marine compounds may hold potential for the development of new pharmaceuticals, nutraceuticals and medical treatments, without a sustainable supply development will cease.

When trying to obtain commercial quantities of marine bioactive compounds, several strategies can be employed, including aquaculture, fisheries development and synthetic production (Benkendorff 2009). In order to harvest commercial quantities of bioactive compounds directly from wild harvested marine organisms, you must ensure that fishing of the target species is a sustainable industry that can withstand the commercial demand. Alternatively, the development or implementation of aquaculture industry for the target organism could be employed; ensuring adequate commercial quantities are produced. Finally, it may also be possible to produce bioactive compounds using a synthetic route, using current advances and knowledge of chemistry and biochemistry to synthesize compounds of interest.

Figure 1.4 *Dicathais orbita* laying egg capsules in the marine aquaria.
Two female *D. orbita* laying egg capsules on rock substrate housed within the marine aquaria at Flinders University, South Australia. Egg capsules are visible on the substrate in varying stages of development. The purple pigmentation of older capsules is due to the presence of Tyrian purple.

1.3.1 Chemical synthesis

Chemical synthesis is always the preferred option for large scale supply for the pharmaceutical industry (Benkendorff 2009). The chemical structure of several marine natural compounds allows chemical synthesis in order to obtain commercial quantities. Ara-A and Ara-C are antiviral and antileukemic compounds that were the first marine-derived pharmaceuticals on the drug market and are synthesized via microbial fermentation of a chemical analogue and additional chemical synthesis (Sipkema *et al.* 2005). A synthetically produced analogue of the cone shell toxin ω-conotoxin MVIIA, also known as Ziconotide, is marketed as a potent painkiller (Prialt), due to its potent neuromuscular blocking action (Garber 2005). Ecteinascidin 743, a drug with anticancer activity derived from the Caribbean tunicate *Ecteinascidia turbinata* was only further developed as a viable pharmaceutical once a large scale semi-synthetic production was developed. Eribulin mesylate, a synthetic isomer of the marine sponge anti-tumor agent Halichondrin B is currently involved in phase III clinical trials in the treatment of breast cancer in Europe and USA (Molinski *et al.* 2009). If chemical synthesis is not possible,

other techniques, such as harvesting from the wild, may be implemented in order to supply the pharmaceutical industry with the required bioactive compounds.

1.3.2 Wild harvest

The development of a sustainable fisheries industry may provide marine bioactives for the pharmaceutical or nutraceutical industry, depending on the target species in question. Sufficient quantities of the marine sponge derived compounds manoalide, bryostatin-1 and avarol were isolated from sponges obtained from their natural environments for preclinical trials (Schaufelberger *et al.* 1991, Sipkema *et al.* 2005). However, it is unlikely that any of these compounds could be sustainably harvested from the wild should the pharmaceutical industry require commercial quantities. It has been reported that for preclinical and clinical trials on natural products derived from marine organisms, up to 1000kg of the source organism is required and thousands of metric tonnes would be required per annum to maximise investment returns from from a patented product, to cover licensing costs (Benkendorff 2009). With such large quantities required for a sustainable marketable agent, it is clear that wild harvesting could have long-term impacts on the sustainability of target populations and therefore the biodiversity of the oceans. One notable exception may be the northern pacific sea star, *Asterias amurensis* which exhibits anti-inflammatory and anti-cancer activity (Fernandez *et al.* 2005). It has previously been shown that 6.5kg of the starfish *A. amurensis* are required to produce anywhere from five to 12mg of bioactive asterosaponins from chemical extracts (Hwang *et al.* 2011). This marine star is considered a pest in the waters off the coast of Melbourne in Australia, and due to its prolific nature and abundance in coastal Australian waters it may actually sustain viable commercial harvesting of bioactives if a market arises in the future. Glucosamine, used in the treatment of joint inflammation and rheumatoid arthritis (Towheed *et al.* 2005), is almost entirely harvested from crustaceans and shark cartilage, taking advantage of the chitin found in the exoskeletal and cartilage waste products involved in the fishing industry (Maria *et al.* 2008). However, some shark fisheries are now considered overfished and the blackmarket practice of removing fins from live sharks is ethically questionable. Aquaculture may be an alternative that will facilitate the adequate production of marine bioactive compounds if wild harvesting and chemical synthesis fails to provide adequate supplies for the pharmaceutical industry.

1.3.3 Aquaculture

As the global population gets larger and larger, human dependency on the fisheries industry to provide adequate nutrition is resulting in the worldwide collapse of fish populations, due to overharvesting and ineffective management (Zeller & Pauly 2005). The aquaculture industry has been established in order to sustain not only the growing seafood demand, but in order to supply other marine based products that are required in the healthcare industry. The nutraceutical Lyprinol®, produced from the farmed green-lip mussel, is a stable lipid extract that is used to minimize arthritis symptoms and severity (Brien *et al.* 2008) and has the potential to reduce ameliorating symptoms of inflammatory bowel disease (Tenikoff *et al.* 2005). Microalgal culture is also being employed to produce omega-3 fatty acids due to the stability, quality and low cost of production in comparison to marine animal lipids (Lebeau & Robert 2003). Sea-based or land-based aquaculture may also be used to obtain sponge-based bioactives in the future, with several sponge species being successfully grown in aquaculture systems, although not all sponge species can be commercially produced in aquaculture systems (de Caralt *et al.* 2003, van Treeck *et al.* 2003).

1.3.4 Microbial biosynthesis

It is interesting to note that several bioactive compounds from marine sponges are produced by bacterial and fungal symbionts living within the sponge (Thomas *et al.* 2010). The aquaculture of host species is then used as a culture media, supporting the proliferation of symbionts and increasing the synthesis of target compounds (Hentschel *et al.* 2006). Although there are problems with current methods for culturing marine sponges for the purposes of harvesting bioactive secondary metabolites (Proksch *et al.* 2003), our current capacity to culture these bioactive producing microbes *ex situ* has been estimated to have a success rate of <1% (Fortman & Sherman 2005). If we are unable to obtain these secondary metabolites by direct microbial culture methods, the culturing of these secondary metabolites within the host sponge is the only alternative method of ensuring adequate sustainable supply (Thomas *et al.* 2010).

Standard microbial techniques are successful at culturing marine bacterium *Micromonospora sp*. from Indonesian marine sponges leading to the large-scale vat fermentation of manzamine-producing cultures (Taylor *et al.* 2007). Manzamine compounds from *Micromonospora* sp. exhibit antimalarial activity (Ang *et al.* 2000). Actinomycete symbionts of the marine sponge *Craniella australiensis*, are responsible for the production of broad spectrum antimicrobial

agents within their sponge hosts, and have been successfully isolated and cultured in the laboratory in order to maintain sustainable supplies (Li & Liu 2006).

Recent studies have focused on metagenomic experiments, profiling the microbial communities present within marine sponges, identifying the bacterial enzymes responsible for the production of bioactive secondary metabolites, then using recombinant expression systems to facilitate synthesis using freely available source metabolites (Taylor *et al.* 2007). The antibiotic Cyanosafracin B, is produced via a large scale fermentation of bacteria *Pseudomonas fluorescens* and is utilized as a starting material to produce commercial quantities of the marine-derived antitumor agent Ecteinascidin ET-743 (Cuevas *et al.* 2000). The partial or complete synthesis of biologically active marine compounds is crucial in the further development of these compounds as pharmaceutical agents, highlighting how important sustainable supply is in the successful application of marine natural products for medical applications. While it is not currently known whether Tyrian purple precursor synthesis is influenced by microbial symbiosis, further studies are required to fully understand the synthesis of these compounds in the Muricidae if a sustainable supply of these compounds is to be achieved.

1.3.5 The development of muricid bioactives as pharmaceutical or nutraceutical agents.

When attempting to develop indole compounds from muricid molluscs into pharmaceutical or nutraceutical products it is imperative that commercial quantities of the compounds are produced. It is unlikely that fishing of target species would provide enough bioactive compounds for the health care industry without decimating wild populations. Populations of the Southern American Mucicidae *Concholepas concholepas* have been decimated in Chile resulting in fisheries closure (Disalvo 1994). Abalone farms and oyster leases in South Australia provide adequate numbers of the muricid *D. orbita* in order to perform preclinical trials investigating the anti-cancer effects of muricid bioactive indole extracts (Benkendorff 2009). However, the pharmaceutical or nutraceutical market would require substantially more source materials than current suppliers could provide. In order to farm muricid species for use in medical treatments, the breeding cycle of muricids needs to be replicated in an aquaculture system. Pilot hatching and rearing techniques for culturing the South American *Concholepas concholepas* (Manriquez *et al.* 2008) and *Hexaplex trunculus* (Vasconcelos *et al.* 2004) have been performed, but an economically viable muricid aquaculture industry has yet to be

developed. Some of the bioactive precursors of Tyrian purple are commercially available (e.g. indole and 6' bromoisatin via TCI chemicals and API chemicals) whereas others cannot be produced synthetically in commercial quantities (i.e. tyrindoleninone (Vine *et al.* 2007)). One possible solution for the commercial production may be a semi-synthetic process, where mollusc-derived enzymes are recombinantly produced and used to facilitate synthetic chemical production. By identifying the individual genes responsible for the production of muricid bioactive indoles we may be able to facilitate large-scale production of these compounds for a commercial market.

1.4 Molluscan genomics and bioinformatics

While traditional biological experiments such as histology, biochemical analysis, enzyme kinetics and field studies allow us to gain an insight into how biological systems function, there are few studies that can produce the volume of information that is generated from large scale gene sequence analysis. Since the completion of the human genome project in 2001 (Venter *et al.* 2001), the field of genomics has grown exponentially and there are currently (as of October 2011) 1786 prokaryotic and 698 eukaryotic genome sequence projects in progress, under annotation or completely sequenced in the National Centre of Biotechnology Information (NCBI) public database. Furthermore, the field of transcriptomics is a burgeoning area of research, allowing researchers to not only identify species-specific global gene expression, but to compare global expression levels within different tissues or under different environmental conditions or disease states. Molluscan sequencing projects have been fairly under-represented in the current genomics climate, with only one genome in draft assembly stage (Moroz *et al.* 2004), and three genome projects in progress (www.newswise.com 2005, Raghavan & Knight 2006). Nevertheless, there have been several studies investigating the transcriptomes of specific molluscs, largely focussing on three areas of research: Biomineralization (Jackson *et al.* 2006, Joubert *et al.* 2010), investigations into molluscan interactions with the Schistosome parasite (Oliveira *et al.* 2004, Lockyer *et al.* 2007) and the use of molluscs as model organisms in neurological research (Moroz *et al.* 2006, Feng *et al.* 2009). There have also been a handful of transcriptome studies investigating environmental and immune responses in molluscs, as well as developmental processes and specific cellular and physiological mechanisms. A summary of molluscan transcriptome information has been listed in Table 1.1.

Table 1.1 Summary and location of Molluscan transcriptome data. Total number of EST sequences are listed, or where next generation sequencing was used, the total number of assembled contigs is listed. Locations are defined as follows; NCBI EST pertains to the EST databse in Genbank (http://www.ncbi.nlm.nih.gov/nucest), NCBI SRI pertains to the short read archive database (http://www.ncbi.nlm.nih.gov/sra). DDBJ pertains to the DNA database of Japan short read archive (http://www.ddbj.nig.ac.jp/). MG RAST pertains to the metagenomics RAST server (http://metagenomics.anl.gov/).

Species	# of ESTs/ Contigs	Location	Reference
Aplysia californica	267411	NCBI EST	(Moroz *et al.* 2006, Walters & Moroz 2009, York *et al.* 2010)
Lottia gigantea	252093	NCBI EST	http://genome.jgi.doe.gov/Lotgi1/Lotgi1.home.html
Biomphalaria glabrata	86936	NCBI EST	(Guillou *et al.* 2007, Lockyer *et al.* 2007)
Plakobranchus cellatus	77648	NCBI SRA	(Waegele *et al.* 2011)
Mytilus galloprivincialis	67942	NCBI EST	(Pantzartzi *et al.* 2010)
Radix balthica	54450	published as additional material in journal	(Feldmeyer *et al.* 2011)
Crepidula fornicata	39897	NCBI SRA	(Henry *et al.* 2010)
	29682	DDBJ SRA	(Kinoshita *et al.* 2011)
Elysia timida	24200	NCBI SRA	(Waegele *et al.* 2011)
Concholepas concholepas	19219	MG RAST	(Cardenas *et al.* 2011)
Laternula elliptica	18290	NCBI SRA	(Clark *et al.* 2010)
Pinctada margaritera	15606	NCBI EST	(Berland *et al.* 2011)
Lymnaea stagnalis	12287	NCBI EST	(Feng *et al.* 2009)
Hyriopsis cumingii	10156	NCBI EST	(Bai *et al.* 2010)
Mizuhopecten yessoensis	9100	NCBI EST	(Meng *et al.* 2010)
Haliotis asinina	8341	NCBI EST	(Jackson *et al.* 2005, Jackson & Degnan 2006, Jackson *et al.* 2006, York *et al.* 2010)
Pinctada maxima	7099	NCBI EST	(Jackson *et al.* 2010)
Ruditapes decussatus	4646	NCBI EST	(Gestal *et al.* 2007)
Chlyamys farreri	3537	NCBI EST	(Wang *et al.* 2009)
Pinctada fucata	1374	NCBI EST	(Fang *et al.* 2011)

The development of nacre and shell formation in molluscs holds particular importance due to the commercial pearl industry, and the potential of molluscan biomineralization in dental and bone regeneration studies (Atlan *et al.* 1997, Westbroek & Marin 1998), as well as its application in materials sciences (Lin & Meyers 2005). Transcriptomics was first used in 2006 by Jackson *et al.* , in order to investigate the shell secretome of the tropical abalone *Haliotis asinina* (Jackson *et al.* 2006). Since this pioneering study, there have been further investigations into pearl biomineralization in several bivalve species, including the black-lipped

pearl oyster *Pinctada margaritifera* (Joubert *et al.* 2010, Berland *et al.* 2011) the freshwater pearl mussel *Hyriopsis cumingii* (Bai *et al.* 2010), the Antarctic bivalve *Laternula elliptica* (Clark *et al.* 2010) and the pearl oyster *Pinctada fucata* (Fang *et al.* 2011). Interestingly, it has been shown that transcript information varies greatly between gastropods and bivalves. The analysis of the transcriptomes from *H. asinina* and *Pinctada maxima* showed that nacre biomineralization is a process that has arisen via convergent evolution in these two molluscan classes (Jackson *et al.* 2010). It has even been shown that transcripts involved in pearl formation and mother-of-pearl formation in the same animal are vastly different, when significant differences in gene expression in nacre-secreting mantle tissue and the pearl sac of *P. fucata* was reported (Kinoshita *et al.* 2011). It is likely that the application of transcriptomics will be further utilised in other mollusc species in the future, and this may facilitate a much deeper understanding of how nacre and shell structures are synthesised and maintained.

Schistosomiasis, a disease caused by trematode parasites, is prevalent in over 76 countries, affecting over 200,000 individuals and causing an average of 20,000 deaths each year (Oliveira *et al.* 2004). Trematode parasites responsible for this disease, *Schistosoma mansoni* and *Schistosoma japonicum*, spend a proportion of their lifecycle in their secondary host, freshwater snails belonging to the *Biomphalaria* genus, which plays a significant role in the spread of the disease (Reeves *et al.* 2008). A genome sequencing project is currently under way for the host snail *Biomphalaria glabrata*, with the aim of gaining a better understanding of the genetic processes involved in parasite lifecycles within this snail (Raghavan & Knight 2006). In order to further our understanding of snail-parasite interactions, several transcriptome projects have been performed on *B. glabrata*, investigating cell-signalling and transcriptional control (Lockyer *et al.* 2007) as well as anti-parasitic response mechanisms (Guillou *et al.* 2007). It is hoped that these studies will lead to new control mechanisms being developed that will reduce the spread of this disease, and future transcriptome investigations into this and other molluscan hosts will be vital in eradicating this problem for future generations.

The large size of some molluscan neurons, combined with the relative simplicity of mollusc nervous systems has made it an ideal model system for investigating neuronal interactions and neurodegenerative disease conditions (Sattelle & Buckingham 2006). It is primarily because of these implications in neuroscience that the *Aplysia* genome project was initiated in 2004 (Moroz *et al.* 2004). While still in the assembly stage, considerable sequence data is available from *Aplysia californica*, with 5165 nucleotide and 267,411 EST entries submitted to NCBI (as

of October 2011). The neuronal transcriptome of *A. californica* was first published in 2006 and provided functional data about genes expression in the central nervous system and in individual neurons, as well as the gene expression involved in sensory response mechanisms in this mollusc (Moroz *et al.* 2006). A recent investigation into genes expressed in the nervous system of *A. californica* identified that the molluscan nervous system shares the greatest sequence similarity with human neuronal sequences compared to other invertebrate model systems, and as such, makes this model organism the ideal species in which to investigate gene signalling in neurological disorders (Walters & Moroz 2009). The freshwater snail *Lymnaea stagnalis* is another model species used to investigate neuron signalling, and has undergone transcriptome sequencing, which identified a surprising amount of gene expression variation to *A. californica* (Feng *et al.* 2009). The study of molluscan neuronal systems is not limited to *A. californica* and *L. stagnalis*, as the transcriptome of the neuronal ganglia of *H. asinina* has been investigated in order to better understand growth regulation mechanisms (York *et al.* 2010). Because molluscs are a good model for neurological studies, it is likely that we will see a greater focus on large scale molluscan transcriptome and genome analysis in the field of neuroscience in the future.

While biomineralization, Schistisome parasite-interactions and neurological modelling have been the dominant areas of molluscan transcriptome research, there have been a variety of other studies which have used these techniques in order to better understand molluscan physiology, processes and populations. The immune system of molluscs has been investigated under parasite-infection conditions in the Shistosome intermediate-host *Biomphalaria glabrata* (Guillou *et al.* 2007), the Zhikong scallop *Chlamys farreri* (Wang *et al.* 2009) and in the hemocytes of the carpet-shell clam *Ruditapes decussatus* (Gestal *et al.* 2007). The transcriptomic effect of environmental stressors in *Mytilus galloprivincialis* has been investigated, where genes that are overexpressed with heat shock proteins were identified (Pantzartzi *et al.* 2010). In an effort to understand what genetic factors effect flesh quality in the Japanese scallop *Mizuhopecten yessoensis*, the transcriptome of the adductor muscle was sequenced and annotated (Meng *et al.* 2010). A considerable study investigating the transcriptome of early development in the common slipper shell *Crepidula fornicata*, was performed in order to better understand asymmetrical cell division (Henry *et al.* 2010). Transcriptomic developmental studies have also been performed on *H. asinina* both at the fundamental level (Jackson & Degnan 2006), as well as under starvation conditions (York *et al.* 2010). The opportunistic sequestering of photosynthetic plastids from algae in sacoglossan sea

slugs has been investigated using transcriptomics techniques and has identified that both *Elysia timida* and *Plakobranchus cellatus* do not import nuclear information from algal sources to maintain plastid activity, but partial digestion of plastids, as well as signal peptide interactions are likely responsible for the longevity of these photosynthetically active plastids in slug physiology (Waegele *et al.* 2011). A lot of these recent sequencing projects have been possible due to the use of next generation sequencing techniques, and to that end, the transcriptome of a non-model snail species (*Radix balthica*) was sequenced using Illumina technology, in order to compare sequence assembly software, and describe that a combined analysis using several software tools to create a meta-assembly was the best method to use when assembling sequence information from non-model organisms (Feldmeyer *et al.* 2011). Transcriptomics sequencing has also been a useful tool in determining molecular markers, for use in population genetics and genotyping studies, including in the Chilean muricid species *Concholepas concholepas* (Cardenas *et al.* 2011) and several bivalve species (Tanguy *et al.* 2008). To date the second largest molluscan transcriptome that is available is from the gastropod *Lottia gigantia*, which was produced in an effort to obtain a full length sequence, however, there has been no publication regarding the transcriptom of *L. gigantea* however EST sequences are available in the NCBI EST database and its genome is available through the JGI portal (http://genome.jgi.doe.gov/Lotgi1/Lotgi1.home.html).

Current molluscan sequencing projects typically focus on the commercially important species (Hedgecock *et al.* 2005, Jackson *et al.* 2005, Jackson & Degnan 2006, Saavedra & Bachere 2006) or those with an application or relevance to human health (Raghavan & Knight 2006, Gestal *et al.* 2007, Zhao *et al.* 2007). The identification of functionally expressed genes is not only beneficial to increase the general pool of knowledge on molluscs, but in the case of *D. orbita*, may allow us to identify key enzymes involved in the formation of bioactive Tyrian purple precursors. The potential of *D. orbita* as both a new aquaculture species (Woodcock & Benkendorff 2008) and in the development of bioactive compounds (Westley *et al.* 2010c, Benkendorff *et al.* 2011) makes it an ideal candidate for genomics research.

1. 5 Thesis aims, significance, structure and objectives

1.5.1 Thesis aims and significance

Through the use of molecular biology techniques and bioinformatics applications, this thesis aims to investigate and identify enzymes involved in the production of bioactive Tyrian purple precursors in the marine snail *D. orbita*. An investigation into Muricidae phylogenetics was undertaken in order to determine whether the presence of Tyrian purple precursors within the egg capsules of muricid molluscs is an ancestral or derived trait. In order to identify the enzymes involved in Tyrian purple biosynthesis, a transcritome approach was used to identify genes that are upregulated or uniquely expressed in the hypobranchial gland of *D. orbita* using suppressive subtractive hybridization. In order to best interpret our findings, a survey into the molluscan sequence homology trends was investigated, before thorough analysis and classification of genes was performed. The arylsulfatase enzyme that was identified from our transcriptome analysis was further studied, to obtain full length expressed sequences and studies were performed to investigate their potential recombinant expression. In terms of molluscan biology, it is hoped that the information gained from this study will further explain the role of the hypobranchial gland in molluscan biology, identify key enzymes involved in Tyrian purple production and determine whether indole transfer to egg capsules has occurred within a single lineage of the Muricidae, when mapped to a phylogenetic classification of Muricidae subfamilies. Ultimately it is hoped that the results of this thesis will assist in the development of Tyrian purple precursors as pharmaceutical agents, gaining both a better understanding of how these compounds are produced and facilitating the future production of bioactive compounds on a commercial scale.

1.5.2 Thesis structure

This thesis is presented in manuscript format. Although each chapter is intended for independent publication, or in the case of chapter three has already been published in a peer reviewed journal, the underlying concepts comprise a progressive body of research. To maintain continuity in presentation, all chapters with the exception of chapter three have been formatted in a consistent manner and have adopted the referencing format outlined in the Journal of Biological Chemistry. Chapter three is presented as published in Volume 123, issue 3 of *The Nautilus*. Whilst I am the first author on all chapters and personally responsible for the experimental design, conducting the research and preparing the manuscripts, the contribution

of additional authors on published manuscripts or those currently under review are outlined in the acknowledgements of each chapter. The following outlines the objective of each chapter and indicates the publication status.

1.5.3 Chapter objectives

Chapter two. The evolution of bioactive Tyrian purple precursors in the egg capsules of the Muricidae (Mollusca: Gastropoda).

Objective: Chapter two aims to identify whether the presence of brominated indoles in the egg capsules of muricid molluscs is a trait that is specific to individual subfamilies or is an ancestral trait that has been subsequently gained and/or lost in some lineages.

Chapter three. Trends in molluscan gene sequence similarity: An observation from genes expressed within the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae). *Nautilus (2009) 123(3): 154-158.*

Objective: Chapter three investigates the trends observed from BLAST sequence homology analysis from a molluscan expressed sequence library. Patterns are reported based on e-values of the pairwise sequence matches. The taxa of the most significant blast sequence match for each sequence are investigated.

Chapter four. Annotation and characterization of a partial transcriptome of the hypobranchial gland of *Dicathais orbita*.

Objective: Chapter four uses suppressive subtractive hybridization to create a cDNA library that contains genes that are expressed within the hypobranchial gland of *D. orbita*. Genes are annotated based on sequence homology results using BLAST2GO and the identified biological processes are further discussed. A subset of sequences identified from this study display an expression profile similar to ciliate protozoans which is further investigated in Chapter five. A partial sequence for an arylsulfatase enzyme was also identified from this study which is further investigated in Chapter six.

Chapter five. Novel application of suppressive subtractive hybridization for the identification of symbionts: Discovery of ciliate protozoa in the hypobranchial gland of *Dicathais orbita* (Neogastropoda, Mollusca).

Objective: Chapter five investigates the surprise finding of ciliate protozoa present within the hypobranchial gland of *D. orbita* based on gene expression patterns, phylogenetic analysis of 18s ribosomal RNA sequences and histological analysis of the hypobranchial gland. This investigation, while not an expected result of our sequence analysis, is important in understanding the biology of the hypobranchial gland of *D. orbita* and in explaining complex symbiotic relationships that exist within the gland.

Chapter six. Characterisation and expression of recombinant arylsulfatase from the marine snail *Dicathais orbita.*

Objective: Chapter six investigates the identification of a full length arylsulfatase gene isolated from the hypobranchial gland and our attempts to express this in HEK293t mammalian tissue culture expression systems. This investigation aims to clone one of the enzymes involved in the production of bioactive Tyrian purple precursors, allowing synthesis of these bioactive compounds *in vitro*.

Chapter seven. The use of molecular techniques for the investigation of Tyrian purple precursor synthesis in the Muricidae: Outcomes and future directions.

Objective: This chapter provides a summary of the research findings of this thesis and expands on individual chapter discussions by investigating gene expression as a whole in the hypobranchial gland of *D. orbita*. This chapter also highlights future directions for the investigation of remaining Tyrian purple enzymes, the production of bioactive precursors and how our findings relate to other muricid species.

Three Appendices are included in this thesis; Appendix I is supplementary to Chapter two and lists all additional phylogenetic analyses performed; Appendix II is supplementary to Chapter six and details recombinant arylsulfatase purification and peptide sequencing; Appendix III details additional investigations performed in order to amplify additional arylsulfatase, bromoperoxidase and tryptophanase gene sequences from *D. orbita*.

Chapter 2. The evolution of bioactive Tyrian purple precursors in the egg capsules of the Muricidae (Mollusca: Gastropoda).

2.1 Abstract

The evolution of egg capsule chemical investment within the Muricidae family of neogastropods was investigated in this study. Tyrian purple and the related brominated indole precursors are produced within the hypobranchial glands of muricid molluscs. These Tyrian purple precursors have been proposed to be involved in the chemical defences of the developing mollusc embryos when transferred to the egg capsules, although it is not fully understood why only some members of the Muricidae invest these compounds into their egg capsules. Molecular phylogenetic analysis on 18s and 28s ribosomal RNA gene sequences were analysed from a selection of muricid molluscs whose egg chemistry has been investigated. Bayesian and maximum parsimony analyses were used to construct phylogenetic relationships and BayesTraits analysis was used to infer the trait evolution of Tyrian purple bioactive precursor investment into the egg capsules of muricid molluscs. The findings of this study indicate that this trait has evolved twice in divergent muricid subfamilies and suggests that the development of this trait did not occur prior to the divergence of the Muricidae family. This analysis supports the recent classification of a new muricid subfamily, the Haustrinae, as a monophyletic subfamily separate to the Rapaninae and Ocenebrinae, and supports the use of *D. orbita* as a suitable representative of the subfamily Rapaninae in which to study Tyrian purple production and investment.

2.2 Introduction

Muricid molluscs have a history of commercial use in the production of the ancient pigment Tyrian purple, once the only known source of purple pigment in the textile and garment industry (Cooksey 2001). Tyrian purple comes from a class of compounds called indoles, which are biosynthesized by a wide range of organisms (Prota 1980, Christopherson 1983). The main pigment in Tyrian purple is 6,6'dibromoindigo (Friedlander 1909), but a range of colours can be produced from the Muricidae, including blue indigo and red indirubins (Cooksey 2001, Cooksey & Sinclair 2005). The blue pigment, indigo, is much more commonly derived from plants, including woad (*Isatis tinctoria*), *Indigofera* spp., *Polygonum tincorium* and others (Balfours-Paul 1998). The biosynthesis of indigos by plants and the Muricidae family of neogastropod molluscs thus presents an interesting case of convergent evolution (Max 1989). The Muricidae are unique in that they are the only organisms known to biosynthesize the brominated indigotins of Tyrian purple.

The role of Tyrian purple in the hypobranchial glands of the Muricidae has remained a mystery (Prota 1980, Max 1989, Naegel & Cooksey 2002, Westley *et al.* 2006). It has been suggested that the pigments merely represent excretory products from the breakdown of tryptophan (Fox 1974, Ziderman 1990). However, the metabolic expense invested into the biosynthesis of these secondary metabolites in the Muricidae strongly suggests a function shaped by natural selection (Prota 1980, Westley *et al.* 2006). In particular, the addition of a bromine to the indole ring would not be a product of degradation and requires energy and specific biosynthetic enzymes. Bromoperoxidase activity, required for synthesizing 6' brominated indoles has been demonstrated in the hypobranchial glands of two muricid species (*Trunculariopsis (Murex) trunculus*, (Jannun & Coe 1987); and *Dicathais orbita*, (Westley & Benkendorff 2009)). Bromoperoxidase enzymes have also been reported from a number of bacteria and plants (Pelletier *et al.* 1994, Weyand *et al.* 1999, Isupov *et al.* 2000), but have not been found in any other molluscs, thus suggesting independent evolution within this family.

It has been suggested that muricid molluscs produce bioactive Tyrian purple precursors in order to defend against bacterial pathogens as a part of the innate immune defences of the snail (Westley *et al.* 2006). The chemical precursors of Tyrian purple display antibacterial activity against several marine and human bacterial pathogens (Benkendorff *et al.* 2000, Benkendorff *et al.* 2001b). The localization of the bioactive precursors within the egg capsules

of some muricid species may therefore be a form of maternal investment for the purpose of protecting developing embryos against bacterial attack. While all muricids studied to date are known to synthesize Tyrian purple precursors within their hypobranchial glands (Cooksey 2001, Benkendorff *et al.* 2004b), their capacity to produce these compounds within their egg capsules is a trait that has primarily been identified in members of the subfamilies Rapaninae and Muricinae (Benkendorff *et al.* 2001a). In the same study, no members of the subfamily Ocenebrinae were found to contain brominated Tyrian purple precursors within their egg capsules, suggesting that the capacity to invest these compounds in egg capsules is not a trait common to all members of the Muricidae (Benkendorff *et al.* 2001a, Benkendorff *et al.* 2004b).

The capacity for these molluscs to transfer bioactive compounds into their egg capsules is likely dependent on the location of biosynthetic cells throughout the gonoduct and the hypobranchial gland. Nine different cells types have been identified within the hypobranchial gland of *D. orbita* (Westley *et al.* 2010b), and some of the cells involved in Tyrian purple production are also present in the capsule gland of this species (Westley & Benkendorff 2008, Westley & Benkendorff 2009). The differentiation of these biosynthetic cells in the different organs during embryonic development may be facilitated by their anatomical proximity. In *Dicathais orbita* the hypobranchial gland lies adjacent to the gonoduct (Westley & Benkendorff 2008, Westley & Benkendorff 2009, Westley *et al.* 2010b). However the anatomy and physiology of the hypobranchial gland and the gonoduct varies between different muricids (Westley *et al.* 2010a, Westley *et al.* 2010b). Thus the development of biosynthetic capacity of the gonoduct may not be anatomically possible in all members of the Muricidae. It remains to be seen whether the ability to chemically protect their egg capsules with brominated indoles is a derived trait specific to some muricid subfamilies, or whether it has arisen multiple times in muricid evolution. The resolution of this question requires a good understanding of the phylogeny and evolution of the Muricidae.

As the development of different methods in the area of taxonomy, the muricids have been the focus of several investigations which have influenced the taxonomic classification of this family. Gross anatomical, radular, operculum and protoconch morphology, have been used in conjunction with shell ultrastructure and egg capsule structural morphology, as well as biogeographical data, in an effort to revise the taxonomic classification of the members of the Rapaninae subfamily (Kool 1993a). In Kool (1993), the subfamily Thaididae/nae was reclassified into either the Rapaninae or the Ocenebrinae. Species belonging to the genus

Nucella were reclassified to the subfamily Ocenebrinae instead of the Rapaninae to which it was previously assigned. Further revision of the Rapaninae subfamily was performed in 2000, when shell characters and morphological characters were used for phylogenetic analyses of members of the Rapaninae and Ocenebrinae (Vermeij & Carlson 2000). It was found that while members of the Rapaninae formed a monophyletic clade, the phylogeny of the Ocenebrinae was poorly resolved (Vermeij & Carlson 2000). Vermeij and Carlson (2000) concluded that shell morphology traits alone, while useful in differentiating between genera, were not useful in the resolution of subfamilial classifications (Vermeij & Carlson 2000). A thorough review of the systematics of the Muricidae in Australia and New Zealand undertaken by Tan (2003) supported the classification of Ocenebrinae and Rapaninae as distinct subfamilies within the Muricidae. In addition, Tan (2003) proposed the establishment of a new muricid subfamily, the Haustrinae, endemic to Australia and New Zealand.

There have been only limited phylogenetic analyses that use molecular sequence data. Oliverio *et al.* (2002) showed that the Coralliophilinae, another subfamily of muricids, formed a monophyletic group, and was the sister clade to the Rapaninae. This was supported by a molecular phylogenetic analysis using sequences from the second internal transcribed spacer region (ITS2) of mitochondrial DNA (Oliverio *et al.* 2002). A detailed molecular analysis on 29 species of the Muricidae was undertaken in 2008 which identified the subfamilies Rapaninae and Ergalataxinae as monophyletic subfamilies, using sequence data from cytochrome oxidase (COI) and 28s ribosomal RNA gene sequences (Claremont *et al.* 2008). By far the most comprehensive phylogenetic study of the Muricidae was performed in 2010, where representatives from nine of the ten currently recognized subfamilies were investigated in a multigene study of 74 species using maximum likelihood and Baysian analysis (Barco *et al.* 2010). This study confirmed the monophyly of the family Muricidae and also confirmed the monophyly of several subfamilies, including Ergalataxinae, Rapaninae, Coralliophilinae, Ocenebrinae, Typhinae and Haustrinae, although there is little information regarding the egg chemistry of species included in this study (Barco *et al.* 2010). This improved knowledge of the Muricidae phylogeny based on molecular markers should facilitate the investigation of character trait evolution within the family. A previous study investigated the evolution of labral spines in members of the Ocenebrinae using molecular phylogenetics, where COI and 12s ribosomal RNA gene sequences were used to infer relationships within the Ocenebrinae (Marko & Vermeij 1999). A similar molecular approach should therefore facilitate investigation

of the evolution of investing bioactive brominated indoles within the egg capsules in the Muricidae.

This study investigates the evolutionary origin of capacity to invest brominated indole precursors within the egg capsules of muricid molluscs. We used 18s and 28s ribosomal RNA gene sequences to perform Bayesian and maximum parsimony analysis in order to infer the phylogeny of a subset of species in the Rapaninae, Ocenebrinae, Haustrinae and Muricinae. This phylogeny was then used to explore the evolution of chemical defences in the egg capsules of muricid molluscs, using a Bayesian technique implemented in BayesTraits (Pagel *et al.* 2004).

2.3 Methods

2.3.1 Specimen collection and egg chemistry data

Taxa and sampling localities, along with NCBI accession numbers are listed in Table 2.1. Literature and unpublished chemical analyses were summarized in order to obtain information regarding the presence of Tyrian purple within their capsules (Table 2.1). Our ingroup comprised of ten species from four subfamilies of muricid molluscs: We use the subfamily name Rapaninae (*sensu* Kool 1993a) to include the following species: *Dicathais (Thais) orbita* (Gmelin, 1791), *Agnewia tritoniformis* (Blainville, 1832) and *Concholepas concholepas* (Bruguière, 1789). Three species have been included in the subfamily Ocenebrinae: *Chorus giganteus* (Lesson, 1830), *Acanthina monodon* (Pallas, 1744) and *Xanthochorus cassidiformis* (Blainville, 1832). Two species were included from the subfamily Muricinae: *Pterynotus (Pterochelus) triformis* (Reeve, 1845) and *Bolinus brandaris* (Linnaeus, 1758). We used two species from the Haustrinae (*sensu* Tan 2003), *Lepsiella vinosa* (Lamarck, 1822) and *Lepsiella flindersi* (Adams and Angus, 1863). The Ranellidae species *Cabestana spengleri* (Perry, 1811) was used as an outgroup for phylogenetic analysis, as this species is known to definitely not contain Tyrian purple precursors in its hypobranchial gland or egg masses (Benkendorff *et al.* 2001a). Voucher specimens are housed in the collections of K. Benkendorff at Flinders University, Adelaide, South Australia.

2.3.2 DNA extraction, amplification and sequencing methods

Tissue samples of approximately 25 mg were taken from the foot muscle of each specimen. DNA extractions followed DNeasy blood and tissue kit (Qiagen, Hilden, Germany) standard protocols. One to ten nanograms of DNA template was used in each polymerase chain reaction (PCR) reaction consisting of 2.5 mM dNTPs, 1.5 mM MgCl2, 200 nM each primer, 0.1 units Platinum Taq polymerase (Invitrogen, Carlsbad, CA, USA) in 10× PCR buffer. The 18s ribosomal RNA sequences used in this study were produced using three primer sets, whose products were assembled together after sequencing. All primers used in this study are listed in Table 2.2. Primers designed in this study were based on nucleotide sequence data for the 18s rRNA sequence from *Rapana venosa* (accession number X98826) and the 28s rRNA sequence from *D. orbita* (accession number AY296898). Forward primer TimA and reverse primer 1100R, which were previously designed and used to amplify 18s ribosomal RNA sequences from several flatworms species (Noren & Jondelius 1999) were used to produce a 1100 bp 18s ribosomal RNA fragment; forward primer 18s_2_For was used with reverse primer 18s_2_Rev to produce a 650 bp fragment: Forward primer 18s_3_For was used in conjunction with reverse primer 18s2R to produce a 700 bp fragment. When all three 18s fragments were assembled together, they produced a region of the 18s ribosomal RNA gene approximately 1800 bp long. The 28s ribosomal RNA sequences used in this study were produced using the forward primer 28s_For and the reverse primer 28s_Rev and resulted in a sequence 278 bp long. Primers Cycle conditions for both genes were as follows; An initial denaturation was performed for 2 min at 94°C; 35 cycles were then performed with denaturation for 45 sec at 94°C; primer annealing for 45 seconds at 48-50°C; primer extension for 2 min at 72°C. PCR products were cleaned up using Promega Wizard SV gel and PCR clean-up system (Promega, Madison, WI, USA) following manufacturers instructions before products and were cloned using the pGEM®t-easy vector cloning system (Promega) following manufacturers instructions. M13 forward and reverse universal primers were used to sequence inserts bidirectionally at the Australian Genome Research Facility (AGRF, Brisbane, Australia). Vector sequence was edited, clipped and the remaining sequences were assembled together using Sequencher 4.1.4 (Genecodes, Ann Arbor MI, USA) with a minimum three times coverage.

2.3.3 Phylogenetic analysis

We carried out Bayesian inference (BI) analysis on the combined 18s and 28s ribosomal RNA dataset with MrBayes 3.1 (Ronquist & Huelsenbeck 2003). All BI analyses were run with

default priors as follows: rate matrix: 0-100, branch lengths: 0-10, Gamma shape: 0-1. We ran two parallel BI runs, each comprising one cold and five heated chains. Markov chains were started from random trees and all chains ran simultaneously with a tree saved every 1000 generations for 10,000,000 generations. The first million trees were discarded as burn-in, after examination of log likelihood plots to see when stationary was reached using Tracer v1.4 (Rambaut & Drummond 2007). The two gene fragments (18s and 28s ribosomal RNA) were partitioned and separate unlinked GTR + G + I models were applied to each. The resulting 18,000 post-burnin trees were used to calculate the posterior probabilities (PP) for each clade. Maximum parsimony (MP) analysis was also conducted on the combined 18s and 28s ribosomal RNA gene sequence dataset to identify whether broad topological features were recovered using a very different approach to BI. One thousand random sequence stepwise additions were performed in the MP analysis, conducted using PAUP b4.10 (Swofford 1999), holding 10 trees at each step and with tree bisection and reconnection for searching tree space. Node support was estimated using 100 bootstrap pseudoreplicates. Single gene comparisons using our nuclear 28s ribosomal RNA sequences and 28s ribosomal RNA sequences from Barco *et al.* 2010's paper was attempted, but poor resolution and statistical support for these analyses meant this comparison could not yield any useful information (data not shown), without the addition of multigene data for our species. Furthermore, without additional egg chemistry data for the species investigated in Barco *et al.* 2010, it is unlikely that any additional insight into the evolution of Tyrian purple precursor investment into egg capsules would have been achieved.

2.3.4 Egg chemistry phylogenetic analysis

BayesMultiState was implemented in BayesTraits (Pagel *et al.* 2004, Pagel & Meade 2006) to infer ancestral states in reference to egg capsule chemistry. Various priors were explored, with a criterion that acceptance rates had to be bounded by 20% to 40% (Pagel & Meade 2006). We used a rate deviation prior of 25 with an exponential (0.0, 5) reverse jump hyperprior (rjhp) and this yielded the required range in acceptance rates. We used 3×10^7 iterations with a burn-in of 1×10^7, sampling every 1000^{th} generation. We assigned the egg chemistry information collated in Table 2.2 based on several literature sources, where studies were performed to determine whether Tyrian purple precursors could be detected in organic extracts of the egg capsules of the species by gas chromatography mass spectrometry (Benkendorff *et al. ,* 2000; Benkendorff *et al.* , 2001). In addition, egg chemistry data was collected from visual identification of purple pigmentation from species whose hypobranchial glands also contained Tyrian purple

precursors (Benkendorff *et al.* 2004b). In order to determine whether there was support for one state over the other for various internal nodes, we used a BayesFactor test, where both possible states were fixed for each node in turn, using the 'fossil' command in BayesTraits, and twice the difference of the average harmonic means for each set fossil state indicated the support level for the alternative ancestral states. A BayesFactor value greater than 2 supports the more likely ancestral trait with "positive" evidence; a BayesFactor greater than 5 provides "strong" evidence, and a value greater than 10 supports the ancestral trait with "very strong" evidence. Bayes Factor tests do not provide significance values that are equivalent to probability values used in frequentist statistics.

2.4 Results

2.4.1 Phylogenetic analysis

The BI consensus phylogram is shown in Fig. 2.1 with posterior probability (PP) support indicated for each node, adhering to the 50% concensus rule. All subfamilies form monophyletic clades. The monophyly of Ocenebrinae and Rapaninae clades was strongly supported (0.98 PP), whereas the monophyly of the Muricinae was not strongly supported (0.67 PP). Our analysis grouped the Ocenebrinae and Rapaninae as sister taxa, although posterior probabilities for this node was 0.69 (PP). The analysis also suggested that the Haustrinae and Muricinae are sister clades, although again, posterior probabilities were very low (0.54). Our analysis found strong support for the Haustrinae + Muricinae as sister clade to the Ocenebrinae + Rapaninae (0.94 PP). A maximum parsimony recovered the same topology as for Bayesian analysis, although low bootstrap support values were found for every node, with the exception of the Rapaninae clade, which was strongly supported with a bootstrap value of 98 (Appendix A.I).

Bayes factor tests suggested there was strong evidence that the transfer of Tyrian purple precursors to the egg capsules is the ancestral trait for the Haustrinae and Rapaninae. There was also positive evidence that this same ancestral trait was present in the most recent common ancestor of the Muricinae (Fig 2.1). BayesTraits analysis suggested the most recent common ancestor of the Muricinae + Haustrinae clade invested Tyrian purple precursors within their egg capsules, although the Bayes Factor test did not support this. There was strong evidence that the most recent common ancestor of the Ocenebrinaes did not invest Tyrian purple precursors within their egg capsules.There was also strong support that the most recent

common ancestor of the Ocenebrinae + Rapaninae clade did not invest Tyrian purple precursors within their egg capsules, and BayesTraits analysis gave positive evidence that the transfer of Tyrian purple precursors to the egg capsules was not an ancestral trait for the Muricidae family. The findings of the BayesTraits analysis therefore indicate that the investment of Tyrian purple precursors in the egg capsules of muricid species is a synapomorphic trait in the Muricidae.

2.5 Discussion

BayesTraits analysis indicated that the presence of Tyrian purple precursors in muricid egg capsules is not an ancestral trait of the muricids as a group, providing positive evidence that both the Haustrinae + Muricinae clade and the Rapaninae clades independently acquired this ability. The fact that all muricid molluscs are already producing these compounds within their hypobranchial glands may explain why two separate incidents could have arisen that enable adults to transfer these compounds into their egg capsules. Additionally, BayesTraits analysis indicates that the evolution of brominated Tyrian purple precursor deposition within the egg capsules of the Muricidae is a trait that developed after their family diverged from the superfamily Muricoidea. Oxygen reactive pigment precursors have been reported from egg capsules of members of the Mitridae, although chemical analysis has identified that these compounds are not Tyrian purple or any of the related compounds (Benkendorff 1999), Nevertheless, it is possible that the capacity to transfer other types of secondary metabolites from the hypobranchial gland to the egg capsules is a trait that has developed independently in other neogastropod species under similar selective pressures. Antimicrobial activity is widespread in the egg masses of molluscs, with reports of bioactivity from the Cephalopoda and at least three families of Gastropoda (Benkendorff *et al.* 2001b). The deposition of benthic egg masses is thought to have evolved separately in these two classes of the Mollusca (Benkendorff 1999), providing further evidence for the multiple evolution of chemical defense for egg capsules within this phylum.

This study used molecular phylogenetics to infer the ancestry of transferring bioactive Tyrian purple precursors to the egg capsules in different muricid subfamilies. The Bayesian analysis based on 28s and 18s rRNA supported monophyletic clades for the subfamilies Rapaninae, Ocenebrinae, Muricinae and Haustrinae. These findings concur with the findings of Vermeij and Carlson (2000), Claremont *et al* (2008) and Barco *et al* (2010), who all identified the Rapaninae

subfamily as a monophyletic clade and the findings of Marko and Vermeij (1999) and Barco *et al.* (2010) who identified the monophyly of the Ocenebrinae. Against this phylogeny, it is clear that the transfer of Tyrian purple precursors to the egg capsules is a derived condition, but the species that do this do not form a single clade. That this trait has evolved more than once in the Muricidae family supports the fact that it is likely to be functionally selected for. This is most likely due to the added protection provided to the developing embryos by the bioactive Tyrian purple precursors providing defence against microbial infection (Benkendorff *et al.* 2000, Benkendorff *et al.* 2001b, Westley *et al.* 2006).

BayesTraits analysis suggests the capacity that muricid molluscs have to invest Tyrian purple precursors into their egg capsules is a trait that has arisen at least twice in the evolution of muricid species (Fig 2.1). Rapaninae species appear to have developed this trait after they diverged from their most recent common ancestor with Ocenebrinae species. None of the Ocenebrinae species we investigated in our study were shown to produce Tyrian purple precursors in their egg capsules (Fig 2.1). The previous report of Tyrian purple precursors in the egg capsules of the mediteranean Ocenebrinae *Ceratostoma erinaceum* (Benkendorff *et al.* 2004b) was in error, because only nonbrominated indoles were in fact reported from the egg capsules of this species (Benkendorff *et al.* 2001a, Benkendorff, personal communication). Benkendorff *et al's (*2001) paper indicated that the Ocenebrinae species *Acanthina monodon* also contains nonbrominated indoles but not the brominated Tyrian purple precursors within chemical extracts of their egg capsules, and no purple pigmentation had been observed in these capsules (Benkendorff *et al.* 2004b). *Acanthina monodon* has been shown to have a cartilaginous protrusion separating the hypobranchial gland and the reproductive gland, whilst in the species of Rapaninae studied, the reproductive organs lie adjacent to the hypobranchial gland (Benkendorff *et al.* 2004b). In the Rapaninae *D. orbita*, Westley and Benkendorff (2008) demonstrated that the reproductive organs, including the capsule gland of females and prostate glands of males, also contain Tyrian purple precursors. For this reason, it can be speculated that the proximity of the hypobranchial gland to the reproductive organs of adult snails influences their capacity to invest these brominated indoles within their egg capsules and may explain why this trait is not observed in members of the Ocenebrinae. However, the presence of nonbrominated indoles in Ocenebrinae egg capsules suggests that these species may still have the ability to transfer secondary metabolites to their egg capsules (Benkendorff *et al.* , 2001), but that they do not have the capacity to brominate these compounds. On the other hand, histochemical investigations by Westley *et al.* (2009) demonstrated that

bromoperoxidase activity does occur in the capsule gland of *D. orbita*. Thus, the biosynthetic capacity of the female reproductive organs, rather than the hypobranchial gland, may be the primary factor influencing the presence of Tyrian purple in the egg capsules of Muricidae.

Benkendorff *et al* (2004) indicated that the Ocenebrinae were polymorphic with respect to the presence of Tyrian purple within their egg capsules. However, this was due to the misclassification of the species *Lepsiella reticulata,* which contains brominated indoles within their hypobranchial glands and in their egg capsules. However, if we can consider Tan's revision of the Ocenebrinae and his inclusion of *L. flindersi* and *L. vinosa* within the new subfamily Haustrinae (Tan 2003) then the Ocenebrinae appear to be monophlyletic in our study. Our molecular study also supports the separation of *Lepsiella* spp. into the subfamily Haustrinae, which forms a distinct clade from the Ocenebrinae. This monophyly of the Ocenebrinae and the Haustrinae subfamilies was also supported by a molecular analysis of the Muricidae (Barco *et al.* 2010). Taking this species reclassification into account, there is no evidence in either this study or Benkendorff *et al* (2004) that any members of the Ocenebrinae to date contain Tyrian purple precursors in their egg capsules. Conversely, all members which were observed in the Haustrinae + Muricinae clade have been shown to produce these brominated indoles within their egg capsules, and as the classification of members of the Haustrinae is such a recent development, it highlights how egg capsule chemistry may be helpful in any additional revisions of muricid subfamiles.

BayesTraits indicated that the most recent ancestor of the Muricinae + Haustrinae clade may have had the ability to invest brominated indoles into their egg capsules; however due to the small sample size of only two species from each of these clades, the BayesFactor test did not provide statistical support for this being an ancestral condition (Fig 2.1). The members of the Haustrinae in this study appear to have shown the least amount of nucleotide differences from the outgroup *C. spengleri*, which may suggest this subfamily is the most ancestral of the Muricidae within this study. The phylogram also indicates a large amount of nucleotide differences are observed between the Mediterranean Muricinae species *B. brandaris* and all other muricids, which were collected from the southern hemisphere (Australia and Chile). The separation of northern and southern hemisphere ocean currents (Levinton 2001) may therefore have contributed to the divergent evolution of this Muricinae species. A greater diversity of Muricinae from both bioregions would need to be sampled to confirm this pattern, and may result in greater support for the classification of Muricinae as sister taxa to Haustrinae. The

recent multigene phylogenetic study by Barco *et al* (2010) confirmed that Muricinae is not a monophyletic subfamily, and further indicated that members of the genus *Pterynotus* are also not monophyletic and their actual classification needs to be revised (Barco *et al.* 2010). Further investigation will be required to determine where *Pterynotus triformis* and other members belonging to this genus reside in the taxonomy of the Muricidae.

It has been indicated that members of the Haustrinae can be defined by their geographic spread, method of larval development, single offset adventitious layer of the operculum, inner denticles of the rachidian teeth, large aggregations of pigment grains, an unpaired accessory salivary gland, a long convoluted duct connecting the mid-oesophagus to the gland of Leiblen, motile spermatozoa in males and the bursa copulatrix locality in reference to the capsule gland in females (Tan 2003). Our findings also suggest that 18s and 28s rRNA gene sequences, and possibly the presence of brominated indoles in egg capsules are character states important to the classification of the Haustrinae. However, studies into the egg capsule chemistry of other members of the Haustrinae will be required to confirm whether this is a characteristic trait for the entire subfamily.

The phylogenetic distribution of Tyrian purple precursors in muricid egg masses does not show a simple pattern whereby species or subfaimilies that do not posess this trait are ancestral nor do species who can invest these compounds belong to a single derived clade (Fig 2.1). Members of the Ocenebrinae are the only subfamily within our study that do not produce these compounds in their egg capsules, but our analysis has indicated that this subfamily is not an ancestral subfamily of the Muricidae, a finding that is supported by Barco *et al* 2010. Consequently, there are two hypotheses that would be consistent with the distribution of the trait. The first hypothesis is that the trait has evolved multiple times in different Muricidae lineages. This is supported by Bayestraits analysis and appears to be the most likely on the basis of a naturally selected role in the defense of the egg masses. The alternative hypothesis would be that the trait evolved once at an early stage of Muricidae evolution and was subsequently lost in some species. Based on our data, this loss would have occurred only once, in the Ocenebrinae family. It is not clear why this trait would have been lost in these species, however, it could be related to a rearrangement of the anatomy that prevented biosynthetic activity in a region where these compounds could be easily transferred to the egg capsules. The cartilaginous protrusion that separates the hypobranchial gland and reproductive

glands of *A. monodon* (Benkendorff *et al.* 2004b) could be one such anatomical rearrangement that may prevent the transfer of these compounds to the egg capsules.

This study is the first to investigate the evolution of the capacity that muricid molluscs have to protect their developing embryos via the investment of bioactive brominated indoles within their egg capsules. In addition to identifying that the Tyrian purple precursors egg capsule investment is a trait has arisen twice in the Muricidae, this study supports Tan's (2003) and Barco et al (2010)'s classification of the Haustrinae as a monophyletic clade, separate from the Ocenebrinae. None of the Ocenebrinae species investigated in this study produce brominated indoles in their egg capsules, whereas the Haustrinae do. Further investigations will need to be performed to investigate whether this trend is followed in all members of the Haustrinae and Ocenebrinae. Nevertheless, future systematic studies may benefit from the use of egg capsule chemistry in the revision of muricid species taxonomic affinities, due to the differentiation of this trait between different subfamilies. The capacity to produce Tyrian purple precursors in the egg capsules may be influenced by the reproductive anatomy of the different species and the distribution of bromoperoxidase enzymes. This further highlights the potential that the well studied Australian species *D. orbita* has as a model organism for representing the Rapaninae in Tyrian purple synthesis investigations and hypobranchial gland gene expression studies.

Acknowledgements

This work was funded by an anonymous philanthropic foundation. Patrick Laffy was supported firstly by a Flinders University Faculty of Science and Engineering Research scholarship, followed by a Flinders University Postgraduate Research scholarship. This manuscript was prepared with the assistance of Associate Professor Michael Schwarz, who contributed to the Bayesian analysis and will be a co-author on this paper. We would also like to thank Dr Ana Glavinic for her assistance and support completing phylogenetic analysis. We would like to thank Dr Carlos Gallardo for collection of the Chilean specimens and Dr Rachel Przeslawski for collection of the Eastern Australian specimens.

Table 2.1 Collection of species used in phylogenetic analysis, their 18s and 28s ribosomal RNA sequences and egg chemistry data.
DNA extractions and PCR reactions were performed to obtain sequence information. The presence of Tyrian purple precursors in the egg masses is derived from literature sources and/or unpublished gas chromatography/mass spectrometry data. All sequences were produced from this study and were submitted to Genbank with the accession numbers which are listed, with the exception of sequences from B. brandaris which were already available on Genbank.

FAMILY Subfamily	Species	Location/Reference	Date	Tyrian purple in egg mass	18s ribosomal RNA sequences	28s ribosomal RNA sequences
MURICIDAE						
Rapaninae	*Concholepas concholepas*	Mehuin Bay, Chile #	Oct-03	Yes[2]	HM486913	HM486923
	Dicathais orbita	Seaford, SA, Australia	Jun-06	Yes[3]	HM486916	HM486926
	Agnewia tritoniformis	Bass Point, NSW, Australia*		Yes[2]	HM486912	HM486921
Ocenebrinae	*Acanthina monodon*	Mehuin Bay, Chile	Nov-03	No[4]	HM486911	HM486922
	Chorus giganteus	Metri Bay, Chile #	Oct-03	No[5]	HM486914	HM486924
	Xanthochorus cassidiformis	Metri Bay, Chile #	Nov-03	No[5]	HM486920	HM486930
Haustrinae	*Lepsiella vinosa*	Seaford, SA, Australia		Yes[4]	HM486918	HM486927
	Lepsiella flindersi	Pearson Is., SA, Australia		Yes[2]	HM486917	HM486928
Muricinae	*Pterynotus triformis*	Seacliff, SA, Australia	Oct-03	Yes[4,5]	HM486919	HM486929
	Bolinus brandaris	Not reported[1]	unknown	Yes[4]	DQ279944	DQ279986
RANELLIDAE						
	Cabestana spengleri	Seaford, SA, Australia	Nov-06	No[2]	HM486915	HM486930

1. Giribet *et al.* , 2006. 2. Benkendorff *et al.* , 2001; 3. Benkendorff *et al.* , 2000; 4. Benkendorff unpublished data; 5. Benkendorff *et al.* , 2004

* Specimens were collected by Dr. Rachel Prezlawski. #Chilean samples were provided by Dr Carlos S. Gallardo, Instituto de Zoologia, Universidad Austral de Chile

Table 2.2 Primers used for 18s and 28s ribosomal RNA sequence amplification. 18s ribosomal sequences were assembled together prior to phylogenetic analysis to produce a single sequence approximately 1800 bp long.

Name	Primer sequence 5'-3'	Orientation	Gene of interest	Publication	Fragment size
TimA	amctggttgatcctgccag	Forward	18s rRNA	(Noren *et al.* 1999)	1100 bp
1100R	gatcgtcttcgaacctctg	Reverse	18s rRNA	(Noren *et al.* 1999)	
18s_2_For	gccccgtaattggaatgag	Forward	18s rRNA	this study	650 bp
18s_2_Rev	cgtcaattccttaagtttcagc	Reverse	18s rRNA	this study	
18s_3_For	gcttccgggaaaccaaagttt	Forward	18s rRNA	this study	700 bp
18s2R	accttcttagctgttttacttcctc	Reverse	18s rRNA	this study	
28s_For	cggcgactcaagcgggatcagccc	Forward	28s rRNA	this study	278 bp
28s_Rev	ctcacggtacttgtccgctatcgg	Reverse	28s rRNA	this study	

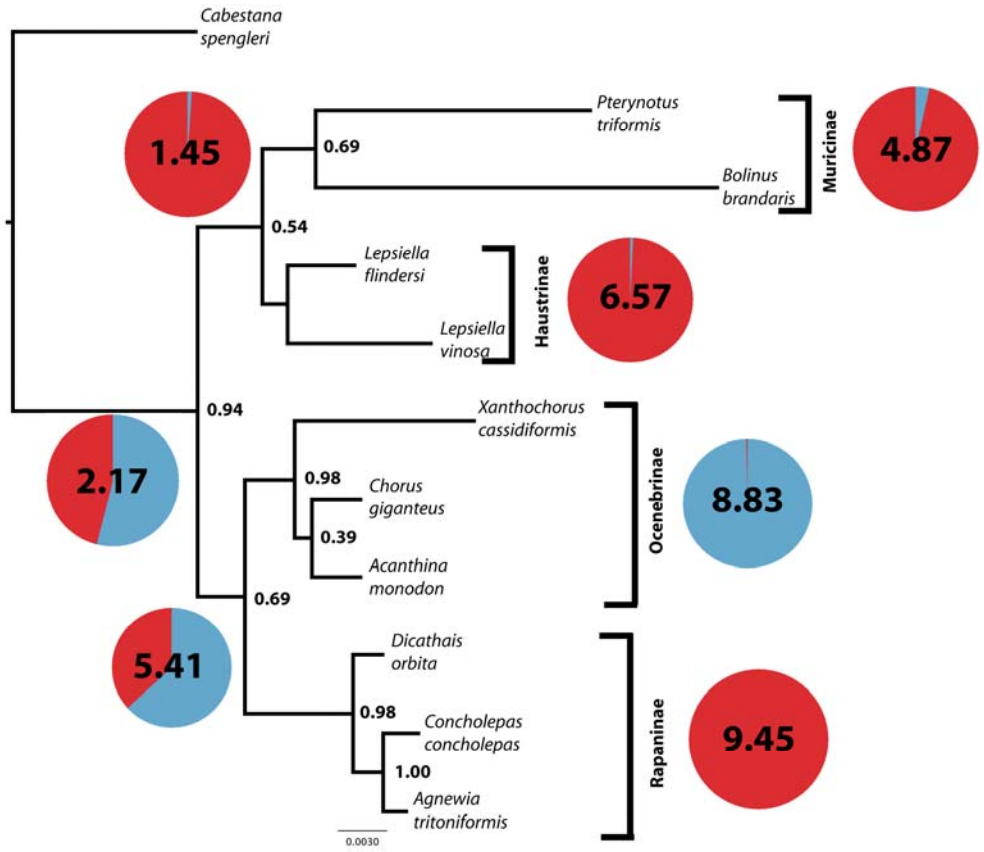


Figure 2.1 Consensus phylogram for Bayesian analysis of 18s and 28s ribosomal RNA combined sequences from muricid molluscs. Subfamily classifications of each muricid species are listed on the right of the phylogenetic tree. The trumpet shell *Cabestana spengleri* (Raneillidae) was used as an outgroup for phylogenetic analysis. Bayesian PP are presented at the site of each node formation. BayesTraits was used to investigate the presence of Tyrian purple precursors within their egg capsules. Pie charts indicate the likelihood that the most recent common ancestor of the node below displayed the trait (red represents the presence of Tyrian purple precursors within egg capsules: Blue represents a lack of Tyrian purple within their egg capsules). Bayes factor tests were performed on each investigated node and the results are displayed within the corresponding pie chart. Any bayes factor test value greater than 2 is considered positive evidence for the findings displayed, and any bayes factor greater than 5 is considered strong evidence for the findings displayed. Scale bar indicates the branch length representing a likelihood that each nucleotide has diverged from the ancestral sequence (0.0030 or 3 substitutions in every 1000 nucleotides).

Chapter 3. Trends in molluscan gene sequence similarity: An observation from genes expressed within the hypobranchial gland of *Dicathais orbita (Gmelin, 1791)*.

Laffy, PW, Benkendorff, K & Abbott CA. (2009) Trends in molluscan gene sequence similarity: An observation from genes expressed within the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791).*The Nautil*us 123 (3) 154-158.

3.1 Significance

Sequence representation within public databases varies greatly in different taxa, and the less sequence information that is available for specific taxanomic groups, the harder it is to annotate biological function to cDNA sequences produced from those taxa. In order to effectively assign function to genes expressed within the hypobranchial gland of *Dicathais orbita*, preliminary tBLASTx analysis was performed on our dataset, and the phylogenetic distribution of sequence matches was reported (performed December 2007). A quantitative survey on nucleotide sequence composition was performed on GenBank sequences and the total number of sequences available in in this database for several Phyla was reported. This study evaluates the complexities in assigning sequence annotations on sequences from an under-represented organism within public databases and describes the sequence homology patterns observed.

3.2 Results presented as a published article

Work presented in this chapter was published in Volume 123, issue 3 of the journal "The nautilus" in September 2009, under the title Trends in molluscan gene sequence similarity: An observation from genes expressed within the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791), Authored by Laffy, PW, Benkendorff, K & Abbott, CA. This publication was published as proceedings of the symposium "Neogastropod Origins, Phylogeny, Evolutionary Pathways and Mechanisms" held during the 2007 World Congress of Malacology, Antwerp, Belgium, 15-20 July 2007. This article was peer reviewed before publication

Whilst the work presented in this article represents the combined efforts of all those listed, all experiments and analysis were performed by myself. I would like to thank my supervisors Dr Kirsten Benkendorff and Dr Catherine A Abbott for their respective assistance with preparation of the manuscript and development of the project.

Trends in molluscan gene sequence similarity: An observation from genes expressed within the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae)

Patrick W. Laffy
Kirsten Benkendorff[1]
Catherine A. Abbott
School of Biological Sciences
Flinders University
GPO Box 2100 Adelaide
SA 5001 AUSTRALIA

ABSTRACT

This study investigates the phylogenetic distribution of homology to *Dicathais orbita* hypobranchial gland genes based on tBLASTx pairwise sequence alignments from the Genbank database. Suppressive subtractive hybridization was used to obtain 417 non-redundant genes that were up-regulated or uniquely expressed in the hypobranchial gland relative to mantle tissue. Of these, 133 sequences revealed matches to the database with the remaining 68% of genes appearing as apparently novel sequences. Homologous sequence matches were observed for a wide range of evolutionarily divergent taxa, encompassing animals, protozoans, plants, fungi, bacteria, and viruses. The highest frequency of homology was found towards chordate sequences, followed by the Mollusca, which highlights the current bias in availability of vertebrate versus invertebrate sequences in the database. An unexpectedly high proportion of matches were also found toward the Ciliophora, indicating a possible symbiotic relationship, as well as the Ascomycota and Streptophyta, which share the ability to biosynthesize indole derivatives with Muricidae such as *Dicathais orbita*. Overall, these results reveal the usefulness of undertaking sequence comparisons in gene expression and highlight the current paucity of knowledge of molluscan genomes.

Additional keywords: Gastropoda, DNA

INTRODUCTION

The development of genomic technologies has had a dramatic effect on all fields of biological sciences (Collins et al., 2003). Since the completion of the human genome project in 2003 (Collins et al. 2003), the number of genomes available has grown dramatically. As of November 2007, a total of 426 eukaryotic (24 complete, 164 undergoing assembly and 238 in progress) and 599 bacterial genomes were available on the Genbank database (NCBI, 2007). The increased number of genomes available enhances our understanding of the biology of the species in question and provides a basis for comparative studies in functional biology. Despite this increase in data, trends in comparative genomics favor the analysis of mammalian sequences (Barnes et al. 2004), and often the homologous identification and classification of non-vertebrate sequences is more challenging.

The Mollusca has been identified as the second most diverse and speciose phylum in the animal kingdom, with members present in marine, freshwater, and terrestrial environments (Pechenik, 2000). Despite their abundance and the economic importance of many species (Beesley et al., 1998), the genome of mollusks remains relatively uncharted. So far, the complete genome has only been sequenced for the Californian sea hare *Aplysia californica* Cooper, 1863, and this is yet to be annotated (NCBI, 2007). The bivalves *Argopecten irradians* (Lamarck, 1819), *Crassostrea virginica* (Gmelin, 1791), and *Spisula solidissima* (Dillwyn, 1817), which are all important fisheries resources, and the medically important freshwater snail *Biomphalaria glabrata*, are currently undergoing sequencing (NCBI, 2007). Nevertheless, a major hurdle in molluscan genomics lies in defining the functions of sequences identified. Sequence homology has been used heavily to assign functions in mammalian genomes. However, the lack of currently available molluscan and invertebrate sequences limits the ability to assign gene function using comparative classifications drawn from existing invertebrate genomic sequence information. Nevertheless, broader comparisons to more distantly related organisms could yield novel information about well conserved genes or genes that have independently evolved convergent functions in distinct taxa.

The hypobranchial gland of neogastropods is a uniquely molluscan organ (Beesley et al., 1998) of uncertain

origin and function (Westley et al., 2006). Within the family Muricidae, it is the well known source of the ancient dye Tyrian purple (Baker, 1974; Cooksey, 2001). Tyrian purple is generated by a series of chemical reactions from indoxyl sulphate precursors that are brominated secondary metabolites thought to be derived from the amino acid tryptophan (Westley et al., 2006). While the Muricidae are thought to be the only source of the purple brominated dye, the related blue dye indigo is produced by a number of other taxa including plants, bacteria and fungi (Epstein et al., 1969; Meijer et al., 2006; Mayser et al., 2007). This presents an interesting case of apparent convergent evolution in biosynthetic capabilities.

Basic Local Alignment Search Tool (BLAST) analysis is a key tool used to identify orthologous genes from different organisms, and its use has been instrumental in classifying countless sequences (Galagan et al., 2003; Venter et al., 2001). The taxonomic classifications of high scoring BLAST matches with unclassified sequences are useful in identifying sequences with specific or variable functions and may indicate key gaps in the current sequence data for members of specific phyla. This study results from a larger project that is currently underway to identify the genes expressed in the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791), a predatory marine gastropod belonging to the family Muricidae, order Neogastropoda. Here we report on our tBLASTx analysis, where sequences were translated into all possible protein translations and compared to all possible translations of every nucleotides sequence in Genbank, to observe trends in molluscan sequence similarity and assess the proportion of homologous genes expressed in this unique biosynthetic organ.

MATERIALS AND METHODS

A suppressive subtractive hybridization (SSH) (Diatchenko et al., 1999) cDNA library containing the upregulated and differentially expressed genes within the hypobranchial gland of *D. orbita,* when compared to mantle tissue gene expression, was created using a Clontech PCR-Select™ cDNA Subtraction Kit (Clontech, California, USA). The RNaqueous® RNA extraction kit (Ambion, Texas, USA), TRI Reagent® (Ambion) and DNaseI (Invitrogen, CA, USA) digestion were used to obtain RNA from the hypobranchial glands and mantle of two *D. orbita* specimens. The subtraction was performed utilizing pooled hypobranchial gland transcripts as the tester population and pooled mantle transcripts as the driver population. Subtracted cDNA produced from SSH were cloned into pGEM®-T Easy vector (Promega, Wisconsin, USA). Colonies with inserts were selected, plasmid DNA was purified and sequencing was performed by Southpath and Flinders Sequencing Facility (Adelaide, Australia) or Australian Genome Research Facility (AGRF sequencing, Brisbane, Australia). A total of 554 plasmids were sequenced, and vector sequence and adaptor regions were removed. Contigs were formed using Sequencher Version 4.1.4 yielding a non-redundant set of expressed sequence tags (EST's) differentially expressed in the hypobranchial gland of *D. orbita.* In total, 417 unique resulting sequences were submitted to tBLASTx analysis and the highest scoring matches for all sequences with an e value smaller than $1e^{-5}$ were collated. The phylum of the orthologous sequence was recorded, and in cases where the highest scoring tBLASTx matched a molluscan sequence, the class was determined. In cases where the matching sequence belonged to a member of the class Gastropoda, the family was also recorded. The total number of Genbank sequences for phyla with 5 or more sequence matches was recorded (Figure 1).

RESULTS

A total of 133 sequences out of 417 (31.9%) resulted in significant tBLASTx matches, with 23 different phyla represented from the best scoring blast match for each identified sequence. Seven of these phyla had matches to 5 or more hypobranchial gland sequences from *D. orbita*. The Chordata showed the highest number of matches, with 32 homologous sequences identified, closely followed by the Mollusca with 31 matches (Figure 1). Ciliophora were the third most abundant phylum with 15 matches, followed by the invertebrate phyla Arthropoda and Echinodermata, with 12 and 8 sequences identified, respectively (Figure 1). There were seven Ascomycota homologs identified in *D. orbita's* hypobranchial gland, as well as five from the Streptophyta (Figure 1).

Of the 31 molluscan sequence matches identified, 22 sequences matched gastropod sequences. Twelve of these gastropod sequence homologs belonged to other members of the Muricidae family (Figure 1). Further distribution of the sequence homology is detailed in Figure 1.

DISCUSSION

While 133 of the sequences produced had BLAST matches that indicated the function of the transcripts, the remaining 284 genes sequenced from the hypobranchial gland of *Dicathais orbita* appear to be novel, highlighting the limited information currently available on molluscan genomes. The high frequency of matches to chordate sequences is likely to be due to the large abundance of vertebrate sequences in the public database (Barnes et al., 2004) (Table 1).There are currently over 57 million gene sequences from the Chordata, compared to less than 600,000 molluscan sequences available (Table 1). There is clearly a bias towards a high proportion of tBLASTx matches returning matches to human and other chordate sequences, which have over 90 times the number of molluscan genes available for sequence alignment.

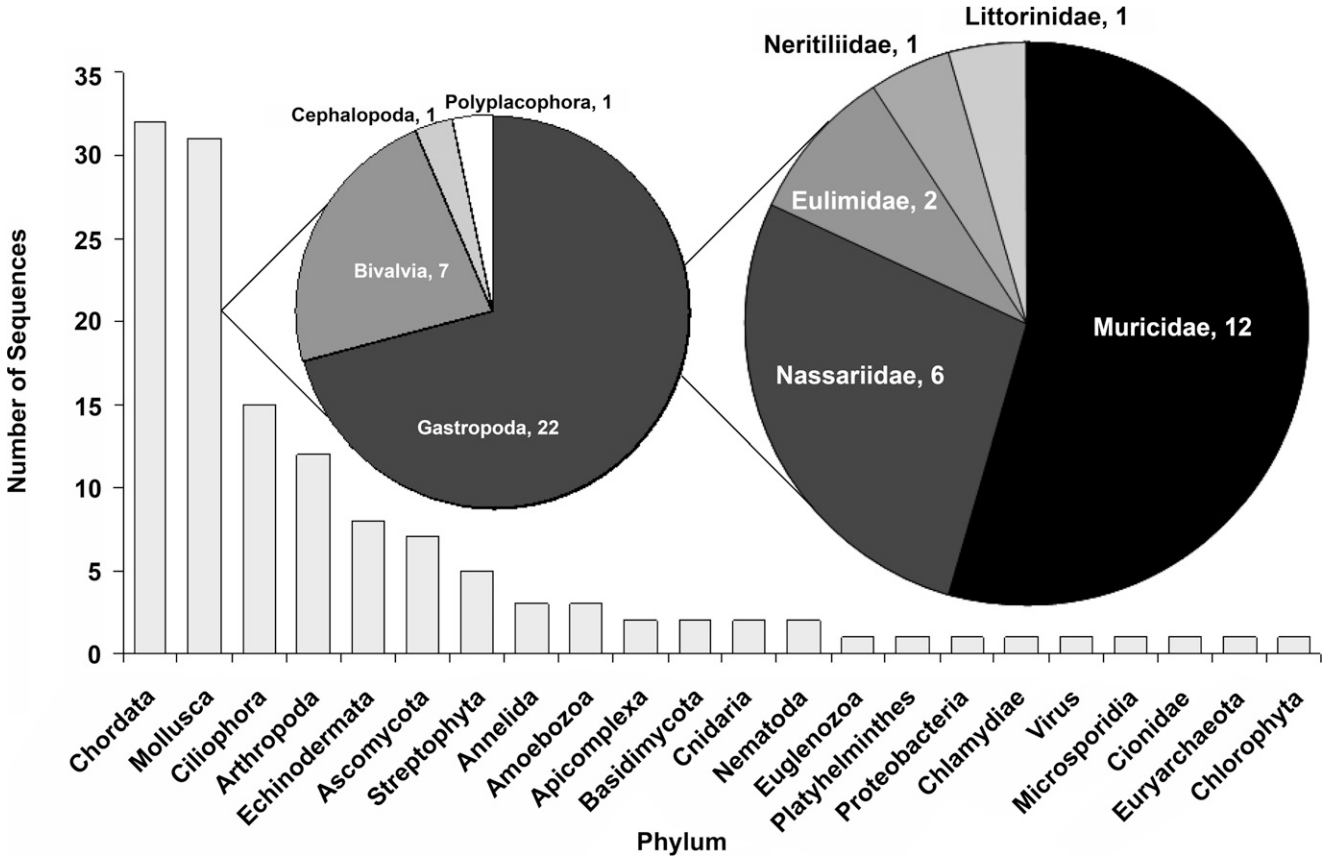


Figure 1. Phyla represented by highest scoring tBLASTx matches of genes expressed in the hypobranchial gland of *Dicathais orbita*. A total of 417 non-redundant EST sequences were analysed using tBLASTx and the resulting 133 significant matches (E value $< 10^{-5}$) were placed into 23 categories, based on the phylum grouping of the highest-scoring tBLASTx matches. Sequences grouped in the phylum Mollusca were further classified into the corresponding Class of the best tBLASTx match. Gastropod sequences were further divided according to family of the highest scoring tBLASTx matches.

The abundance of matches to sequences from the Ciliophora was unexpected, particularly since the number of ciliate sequences in public databases is just over 300,000 (Table 1). It is possible these protozoan gene matches actually result from ciliate genomes derived from endosymbionts occurring within the hypobranchial gland of *D. orbita*. Ciliates are ubiquitous protists that commonly form relationships with other species, such as the parasitic *Ichthyopthirius multifilius* (Abernathy et al., 2007) and the symbiotic *Euplotes uncinatus* (Lobban et al., 2005).

The abundance of matches to arthropod species was not unexpected due to the shared ancestral relationship between the Mollusca and Arthropoda. However, the numerous matches to Echinodermata are less expected given that this phyla occurs on the deuterostome lineage along with chordates, which diverged from the mollusks and other protostomes over 100 million years ago (Heckman et al., 2001). Notably, there were relatively few matches to the Annelida (Figure 1) despite the fact that this abundant protostome phylum occurs within the Lophotrochozoan lineage alongside the Mollusca, which form a separate clade from the Ecdyzoa, including arthropods and nematodes (Aguinaldo et al., 1997). It is likely that the small number of annelid sequences available, less than 35, 000 (Table 1), contributed to the small incidence of annelid sequence homology with our molluscan sequences. This further highlights the relatively limited genetic information that is available for so called "primitive" invertebrate phyla.

Table 1. Number of nucleotide sequences available on Genbank database for different phyla as published on the 18 December 2007. All data was compiled as published under the Taxonomy browser available on NCBI Entrez taxonomy home page http://www.ncbi.nlm.nih.gov/sites/entrez?db=Taxonomy.

Phylum	Genbank nucleotide sequences
Chordata	57,495,211
Mollusca	599,894
Ciliophora	303,304
Arthropoda	5,090,469
Echinodermata	941,561
Ascomycota	1,367,029
Streptophyta	22,413,399
Annelida	34,245

The frequency of sequence matches to the fungal Ascomycota and the plant Streptophyta was an unexpected finding. This is possibly related to the fact that members of both the Streptophyta and Ascomycota are capable of similar secondary metabolite production as is the muricid *Dicathais orbita.* Indigo is produced in *Isatis tinctoria* (phylum Streptophyta) (Epstein et al., 1967), and the production of indole compounds has been reported for *Candida glabrata* (phylum Ascomycota) (Mayser et al., 2007). These compounds are in the same chemical class of indole alkaloids as Tyrian purple, the brominated derivative of indigo secreted only from the hypobranchial gland of the Muricidae (Cooksey, 2001; Westley et al., 2006). These similarities in secondary metabolite production may influence the frequency of homology with genes expressed in the hypobranchial gland of *D. orbita.* Further analysis of the conserved genes could help reveal some key biosynthetic enzymes and/or processes. As SSH allows for amplification of only up-regulated or uniquely expressed genes in this instance, we would expect sequences involved in chemical and protein biosynthesis to be amplified. This demonstrates that it is important to consider the source of expressed genes when interpreting sequence homology.

Another key observation is the frequency and variation of molluscan gene matches observed from our tBLASTx analysis. As mentioned, a total of 31 molluscan sequence matches were identified, with 22 gastropod sequences, 12 of which belonged to the family Muricidae (Figure 1). This trend is expected as species within the same family are expected to show greater homology with our *D. orbita* sequences. The key limiting factor to the number of muricid and gastropod sequence matches is the limited amount of sequencing that has been performed on these groups, only 1994 Muricidae sequences have been published on the NCBI database as of November 2007 (NCBI 2007). The majority of sequences available for muricids are highly conserved genes involved in phylogenetic analysis such as ribosomal RNA (Colgan et al., 2007; Harasewych et al., 1997; Oliverio and Mariottini, 2001), cytochrome oxidase 1 (Colgan et al. 2007; Harasewych, et al., 1997) and histone H3 sequences (Colgan et al., 2007). The frequency of positive matches to *D. orbita* hypobranchial gland genes is likely to increase as a broader range of sequences from additional Muricidae and other gastropod species are made available on Genbank.

From tBLASTx analysis, we have identified the phylogenetic distribution of species that share homology with *Dicathais orbita* gene sequences. While less than 32% of sequences could be positively matched on the gene databases, 31 matches were found encompassing species from both invertebrates and vertebrates within the Animal Kingdom, as well as eukaryotic plants, protozoans, fungi, some prokaryotes and even viruses. Most matches pertain to chordate sequences, and this may be attributed to the abundance of these sequences within databases. Nevertheless, many of the sequences match other molluscan species and other invertebrate phyla, likely due the close evolutionary relationships leading to conserved genes. A significant proportion of sequences belong to ciliate protozoans, and it is unclear whether this is due to similarities between these protists and *D. orbita* or the addition of ciliate genes within our hypobranchial gland expressed genes. The limited number of molluscan gene matches from our dataset supports the need for a larger number of molluscan sequences to be identified and released, encompassing a broader range of functional genes. Only then will we be able to accurately view trends in gene expression within the hypobranchial gland of *D. orbita.*

ACKNOWLEDGMENTS

We appreciated funding support from an anonymous philanthropic foundation. Patrick Laffy is supported firstly by a Flinders University Faculty of Science and Engineering Research scholarship, followed by a Flinders University Postgraduate Research scholarship. We would also like to thank the South Australian Partnership for Advanced Computing, for the use of their BLAST portal, Chantel Westley for her assistance in specimen dissection and Dr. Peter Speck for his assistance in manuscript editing.

LITERATURE CITED

Abernathy, J.W., P. Xu, P. Li, D.H. Xu, H. Kucuktas, P. Klesius, C. Arias, and Z.J. Liu. 2007. Generation and analysis of expressed sequence tags from the ciliate protozoan parasite *Ichthyophthirius multifiliis*. BMC Genomics 8: 1–10.

Aguinaldo, A.M.A., J.M. Turbeville, L.S. Linford, M.C. Rivera, J.R. Garey, R.A. Raff, and J.A. Lake. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387: 489–493.

Baker, J.T. 1974. Tyrian purple: an ancient dye, a modern problem. Endeavour (Oxford). 33: 11–17.

Barnes, D.W., C.J. Mattingly, A. Parton, L.M. Dowell, C.J. Bayne, and J.N. Forrest. 2004. Marine organism cell biology and regulatory sequence discovery in comparative functional genomics 46: 123–137.

Beesley, P.L., G.J.B. Ross, and A. Wells. 1998. Mollusca: The Southern Synthesis. Fauna of Australia, CSIRO Publishing, Melbourne, Part A, xvi + 563 pp., Part B, viii + pp. 565–1234.

Colgan, D.J., W.F. Ponder, E. Beacham, and J. Macaranas. 2007. Molecular phylogenetics of Caenogastropoda (Gastropoda: Mollusca). Molecular Phylogenetics and Evolution 42: 717–737.

Collins, F.S., M. Morgan, and A. Patrinos. 2003. The Human Genome Project: lessons from large scale biology. Science 300: 286–290.

Cooksey, C.J. 2001. Tyrian purple: 6,6'-dibromoindigo and related compounds. Molecules 6: 736–769.

Diatchenko, L., S. Lukyanov, Y.F.C. Lau, and P.D. Siebert. 1999. Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes, cDNA preparation and characterization. Methods In Enzymology 303: 349–380.

Epstein, E., M.W. Nabors, and B.B. Stowe. 1967. Origin of Indigo of Woad. Nature. 216: 547–549.

Galagan, J.E., S.E. Calvo, K.A. Borkovich, E.U. Selker, N.D. Read, D. Jaffe, W. FitzHugh, L.-J. Ma, S. Smirnov, S. Purcell, B. Rehman, T. Elkins, R. Engels, S. Wang, C.B. Nielsen, J. Butler, M. Endrizzi, D. Qui, P. Ianakiev, D. Bell-Pedersen, M.A. Nelson, M. Werner-Washburne, C.P. Selitrennikoff, J.A. Kinsey, E.L. Braun, A. Zelter, U. Schulte, G.O. Kothe, G. Jedd, W. Mewes, C. Staben, E. Marcotte, D. Greenberg, A. Roy, K. Foley, J. Naylor, N. Stange-Thomann, R. Barrett, S. Gnerre, M. Kamal, M. Kamvysselis, E. Mauceli, C. Bielke, S. Rudd, D. Frishman, S. Krystofova, C. Rasmussen, R.L. Metzenberg, D.D. Perkins, S. Kroken, C. Cogoni, G. Macino, D. Catcheside, W. Li, R.J. Pratt, S. A. Osmani, C.P.C. DeSouza, L. Glass, M.J. Orbach, J.A. Berglund, R. Voelker, O. Yarden, M. Plamann, S. Seiler, J. Dunlap, A. Radford, R. Aramayo, D.O. Natvig, L.A. Alex, G. Mannhaupt, D.J. Ebbole, M. Freitag, I. Paulsen, M.S. Sachs, E.S. Lander, C. Nusbaum, and B. Birren. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422: 859–868.

Harasewych, M.G., S.L. Adamkewicz, J.A. Blake, D. Saudek, T. Spriggs, and C.J. Bult. 1997. Neogastropod phylogeny: A molecular perspective. Journal of Molluscan Studies 63: 327–351.

Heckman, D.S., D.M. Geiser, B.R. Eidell, R.L. Stauffer, N.L. Kardos, and S.B. Hedges. 2001. Molecular evidence for the early colonization of land by fungi and plants. Science 293: 1129–1133.

Lobban, C.S., L. Modeo, F. Verni, and G. Rosati. 2005. *Euplotes uncinatus* (Ciliophora, Hypotrichia), a new species with zooxanthellae. Marine Biology 147: 1055–1061.

Mayser, P., M. Wenzel, H.J. Kramer, B.L.J. Kindler, P. Spiteller, and G. Haase. 2007. Production of indole pigments by *Candida glabrata*. Medical Mycology 45: 519–524.

NCBI [National Center for Biotechnology Information]. 2007. ENTREZ Genome Project Home Website, http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj, accessed 17 December 2007.

Oliverio, M. and P. Mariottini. 2001. A molecular framework for the phylogeny of *Coralliophila* and related muricoids. Journal of Molluscan Studies 67: 215–224.

Pechenik, J.A. 2000. Biology of the Invertebrates. Fourth Edition. McGraw-Hill Higher Education, Boston, 578 pp.

Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R. J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A. J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V.D. Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.-R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z.Y. Wang, A. Wang, X. Wang, J. Wang, M.-H. Wei, R. Wides, C. Xiao, C., Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S.C. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M.L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C., Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.-H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J.F. Abril, R. Guigo, M.J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.-H. Chiang, M. Coyne, C. Dahlke, A.D. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. 2001. The Sequence of the Human Genome. Science 291(5507): 1304–1351.

Westley, C., K. Benkendorff, and K.L. Vine. 2006. A proposed functional role for indole derivatives in reproduction and the defense system of Muricidae (Neogastropoda: Mollusca), Life in Progress editions, Roscoff, France, pp. 31–44.

Chapter 4. Annotation and characterization of a partial transcriptome of the hypobranchial gland of *Dicathais orbita*.

4.1 Abstract

The hypobranchial gland present in marine and terrestrial gastropods and some bivalves, is an organ whose function is not clearly understood in molluscan biology. It is primarily responsible for mucus production and the collection of particulate material within the mantle cavity. In the Muricidae family, it is also the source of the ancient dye Tyrian purple and its bioactive precursors. The hypobranchial gland of the Australian muricid *Dicathais orbita* has been the subject of considerable histological and chemical investigations in an effort to better understand the role of this gland in muricid physiology. To gain further insights into hypobranchial gland biology, suppressive subtractive hybridisation was performed on hypobranchial gland and mantle tissue from *D. orbita* in order to create a differentially expressed cDNA library. A total of 554 clones were randomly sequenced, analysed and annotated both manually and with the BLAST2GO annotation software package. 110 sequences had their functions putatively identified. Genes involved in mucus protein synthesis, as well as Tyrian purple biosynthesis, choline ester regulation, protein and energy production, were up-regulated within our cDNA library. Quantitative realtime PCR revealed upregulation of six of these genes confirming the enrichment of differentially expressed genes within the library. To the best of our knowledge this report contains the first partial transcriptome analysis of genes transcribed by the hypobranchial gland of a mollusc. This study confirms that the hypobranchial gland is involved in the production of mucus secretion and also identifies it as a site of chemical interaction and biosynthesis. This study also identified a partial arylsulfatase sequence, one of the putative genes involved in Tyrian purple biosynthesis. This data should be useful for future studies investigating the role of specific mRNAs in hypobranchial gland functionality, and lays the foundation for a better understanding of the enzymatic production of bioactive Tyrian purple precursors within the gland.

4.2 Introduction

The function of the hypobranchial gland, present within the mantle cavity of many gastropod molluscs, has been the subject of investigation and speculation for many years. It is known that the hypobranchial gland is responsible for the amalgamation of particulate matter within the mantle cavity and it has been characterized as a predominant mucus producer (Fretter & Graham 1994). Further studies into the gland suggest varied roles for this pallial organ, including possible roles in aestivation and antiseptic production (Andrews & Little 1972), as being involved in the olfactory response during feeding (Roller *et al.* 1995), and in the production of the embedding medium into which egg capsules are deposited onto benthic substrata (Westley 2008). The hypobranchial gland has also been proposed as a chemical factory in neogastropods (Kay *et al.* 1998). Several novel secondary metabolites, including pharmacologically active choline esters (Whittaker 1960, Baker & Duke 1976, Roseghini *et al.* 1995) have been extracted from this gland. Furthermore, arylsulfatase and bromoperoxidase enzyme activity has been identified within the gland in the Muricidae (Neogastropoda(Erspamer 1946, Jannun & Coe 1987, Westley 2008) .

In the Muricidae family of neogastropods, the hypobranchial gland (Figure 4.1B) is well known as a source of the ancient dye Tyrian purple (Baker 1974, Cooksey 2001). This pigment is a brominated indole that was first identified in 1909 by Freidlander (Friedlander 1909). The dye is produced from indoxyl sulphate prochromagens, which are hydrolysed by arylsulfatase enzymes (Baker & Duke 1976, Cooksey 2001, Naegel & Cooksey 2002) to yield intermediate precursors that exhibit antibacterial and anticancer activities (Benkendorff *et al.* 2000, Benkendorff *et al.* 2001b, Westley *et al.* 2006, Vine *et al.* 2007, Benkendorff *et al.* 2011). These compounds are also found in the egg masses (Figure 1A) and their associated biological activity has led to suggestions that hypobranchial gland secretions play a role in maternal defence of the egg masses (Benkendorff *et al.* 2000, Benkendorff *et al.* 2001a), as well as an immune role within adult snails (Westley *et al.* 2006).

Bromination of secondary metabolites has also been associated with the hypobranchial gland in other families of molluscs, including at least one species in the distantly related Vetigastropoda lineage. 6-bromo-2-mercaptoptrytamine, isolated from the hypobranchial gland of the trochid mollusc *Calliostoma canaliculatum* has been characterized as a potent neurotoxin, targeting potassium channels (Kelley *et al.* 2003, Wolters *et al.* 2005). Although not

produced within the hypobranchial gland, tryptophan residues within some Conidae (Neogastropoda) venom peptides have been shown to be post-translationally modified by the addition of bromine (Jimenez *et al.* 1996, Jimenez *et al.* 1997, Jimenez *et al.* 2001, Massilia *et al.* 2001). The biosynthetic enzymes responsible for the bromination of secondary metabolites in molluscs are yet to be identified. However, bromoperoxidase activity, from an enzyme belonging to the haloperoxidase family, has been identified within the hypobranchial gland of the Muricidae *Trunculariopsis trunculus* (Jannun & Coe 1987). A recent histochemical study into the Muricidae *D. orbita* has provided further evidence for regulated bromoperoxidase and arylsulfatase activity in the hypobranchial gland (Westley 2008, Westley & Benkendorff 2009, Westley *et al.* 2010b). Activity of these biosynthetic enzymes within the hypobranchial gland suggests that one of the roles of the gland is the production of bioactive secondary metabolites, which may have key roles in the chemical defence of molluscs.

A novel approach for investigating biosynthesis and the functional role of the hypobranchial gland is to investigate gene expression. Gene expression studies in the Mollusca is a growing area of research which can help shed light on important and unique cellular interactions which occur in these organisms. The sea hare *Aplysia californica*, which has historically been studied as a model organism for neurological pathways and disorders (Moroz *et al.* 2006) has been the focus of several transcriptome studies, investigating the specific neuronal interactions, chronic pain development and neurological disorders (Moroz *et al.* 2004, Moroz *et al.* 2006, Walters & Moroz 2009). Similarly, the transcriptome of the freshwater snail *Lymnaea stagnalis* has been sequenced, in order to better understand its central nervous system (Feng *et al.* 2009). More recently, the transcriptome of *A. californica* has been further investigated, at the embryonic, larval and metamorphic developmental stages in order to better understand cellular differentiation and embryogenesis (Heyland *et al.* 2011). This investigation followed an earlier study where mRNA expression was investigated in the early development of *Crepidula fornicata* in order to better understand the asymmetrical cell division synonomous with Lophotrochazoan embryonic development (Henry *et al.* 2010). Larval development has also been investigated in abalone using EST sequence analysis (Jackson & Degnan 2006). Another area of molluscan research where transcriptomics has been applied is in the field of biomineralization and shell development and deposition. First investigated in the tropical abalone *Haliotis asinina* (Jackson *et al.* 2006), transcriptome techniques have been further applied to investigate these processes in several oyster species (Joubert *et al.* 2010, Berland *et al.* 2011, Fang *et al.* 2011), as well as the Antarctic bivalve *Laternula elliptica* (Clark *et al.* 2010)

and the freshwater mussel *Hyriopsis Cumingii* (Bai *et al.* 2010). Most recently, a transcriptomics study has been performed on the Chilean muricid mollusc *Concholepas concholepas*, describing a cursory analysis of gene expression as well as identifying potential short tandem repeats for use in environmental population studies (Cardenas *et al.* 2011). In all of these studies, transcriptomics has been used as a tool to better understand biological processes and has helped resolve how global gene expression influences physiological changes.

Research into which sequences are being expressed in hypobranchial gland tissue and the functions of these sequences could be used to infer the role this gland plays in whole organism biology. Furthermore, sequences specifically involved in the production of bioactive brominated indole compounds within the hypobranchial gland are likely to be expressed within the gland, and as such the sequencing of these transcribed genes may result in the identification of the enzymes involved in this process. Suppressive subtractive hybridization (SSH) was first developed by Diatchenko *et. al.* (Diatchenko *et al.* 1999) and has been used for a variety of different investigations. This method has been particularly useful for investigation of gene expression throughout various disease/state conditions, including bacterial attack of the carpet-shell clam *Ruditapes decussatus* (Gestal *et al.* 2007), the induction of beetle immune systems (Altincicek *et al.* 2008), and human liver cirrhosis (Shackel *et al.* 2003). It has also been a useful tool in investigating how environmental changes affect transcription, such as sublethal stressors in the copepod *Calanus finmarchicus* (Hansen *et al.* 2007), cadmium exposure in the striped seabream *Lithognathus mormyrus* (Auslander *et al.* 2008) and the difference in gene expression of the European eel *Anguilla anguilla* in migration from fresh water to salt water (Kalujnaia *et al.* 2007). SSH has also found valuable use in transcriptome studies into different developmental stages in species, including tomato fruit development (Pascual *et al.* 2007), oocyte development in the coho salmon *Oncorhynchus kisutch* (Luckenbach *et al.* 2008), the role of the cerebral ganglia in the growth, feeding behaviour and reproduction in the tropical abalone *Haliotis asinina* (York *et al.* 2010) and differences in expression patterns of the hibernating brain of the greater horseshoe bat *Rhinolophus ferrumequinum* (Yuan *et al.* 2008). The ability to compare, at the transcription level, the different genes that are expressed in a particular tissue, while at the same time selectively choosing those genes most likely responsible for phenotypical differences, makes this technique a useful transcriptomics tool, particularly when looking at gene expression in species where limited sequence data is available. So far the only known genes that have been identified from *D. orbita* are highly

conserved sequences used in phylogenetic studies (Colgan *et al.* 2000, Colgan *et al.* 2003, Colgan *et al.* 2007).

Dicathais orbita (Figure 4.1A) is a model species in which to investigate gene expression of the hypobranchial gland, due to its well documented production of bioactive compounds. We hypothesize that a number of different genes will be identified from a SSH cDNA library containing expressed sequence tags (EST) that are differentially or uniquely expressed within the hypobranchial gland, relative to the mantle tissue of *D. orbita*, corresponding to the roles this gland plays in organismal cell biology. It is predicted that sequences involved in the production of mucus and mucin-related proteins, as well as sequences involved in the production of Tyrian purple and other chemical compounds will be up-regulated and identified from our cDNA library. In order to confirm the differential expression of the sequences within our cDNA library, realtime PCR will also be used.

4.3 Methods

4.3.1 Tissue collection and RNA isolation

Pre-copulatory *D. orbita* were collected from shallow subtidal reefs along the coast of the Fleurieu and Eyre Peninsulas, South Australia, and housed within fully recirculating seawater tanks within the School of Biological Sciences at Flinders University. Dissection was performed in accordance with the methods described in Westley and Benkendorff (2008). The shell of each individual was removed by cracking in a bench top vice and removing the soft tissue by severing the columnar muscle from the shell. Soft tissue was transferred to a dissecting tray where the visceral mass was separated from the dorsal mantle by making an incision along the lateral margins of the columnar muscle. The dorsal mantle was folded back, revealing the pallial gonoduct, and was pinned in place to reveal the hypobranchial gland (Fig 4.1B). Mantle tissue was collected, and egg capsule glands were dissected away from the hypobranchial gland and the branchial and medial regions of the gland were excised (Westley & Benkendorff 2008). All tissue samples were stored in cryovials, snap frozen in liquid nitrogen and stored at -80°C until required.

Total RNA was extracted by combining the HGB tissue from two individual snails, separately combining the mantle tissue from the same snails, and using the Ambion RNaqueous RNA extraction kit (Applied Biosystems, Foster City, CA). Pooled tissue was immersed in 600µl of

lysis binding solution on ice and homogenized using the Turrat T8 homogenizer (IKA-Werke, Germany) before RNA was extracted, in accordance with manufacturers' instructions with the following two modifications. Firstly, 700 µl of TRI reagent (ABgene, UK) was added to tissue in the lysis binding buffer, mixed by pipetting and then a 12,000g centrifugation for 10 min at 4°C was performed. This step improved the overall quality of the RNA as it helped to remove the mucus from the snail tissue before RNA extraction. The aqueous RNA containing fraction was then loaded on to a column as per manufacturers' instructions. The second alteration was that an on-column DNAse I digestion was performed between the second and third column wash steps, using 50 units of DNase I (Invitrogen, Carlsbad, CA), 18 µl DNase I buffer and 161 µl diethyl pyrocarbonate (DEPC) treated water. Columns were incubated at RT for 15 min before centrifugation for 30 sec at 14,000g. RNA was precipitated by treatment with 8M ammonium acetate and 100% ethanol. After overnight incubation at -20°C, 75% ethanol was added to precipitate RNA, and ethanol was removed and allowed to evaporate. RNA was resuspended in DEPC treated water, aliquoted and stored at -80°C. Total RNA quality was checked by agarose gel electrophoresis and through spectrophotometric quantification of 260/280 ratio.

4.3.2 Creation of subtracted cDNA library, plasmid isolation and sequencing

The SMART™ cDNA Library construction kit (Clontech, Mountain View, CA) and the Clontech PCR-Select™ cDNA subtraction kit (Clontech) were used to construct our cDNA library containing unique and differentially expressed genes from the hypobranchial gland of *D. orbita*, in accordance with the relevant product manuals. Mantle tissue was used as the subtractor (driver) in SSH due to the fact that there is no known association between mantle tissue and Tyrian purple biosynthesis in muricids. Histological investigations into the biosynthetic activity and the precursors involved in Tyrian purple synthesis confirmed that none of the enzymes involved were present in the mantle tissue of *D. orbita* (Westley & Benkendorff 2008, Westley *et al.* 2010b)). In addition, chemical extracts from the mantle tissue *of D. orbita* did not contain any of the precursors involved in the formation of Tyrian purple (Benkendorff *pers comm*). As these compounds or the related enzyme activities associated with their synthesis have never been identified in the mantle of *D. orbita*, it is likely that the genes responsible for these activities would be under-expressed or not expressed at all in this tissue. Two µg of total RNA was used in the SMART™ cDNA synthesis reaction, where mRNA sequences were used as templates to make cDNA. Hypobranchial gland cDNA (400 ng) was used as the tester cDNA and mantle cDNA (400 ng) was used as the reference or driver cDNA. Our subtracted cDNA population was cloned into the pGEM®T-Easy vector system (Promega, Madison, WI) using

the manufacturer's protocol. White colonies were selected and grown overnight in Luria Burtani media containing ampicillin. Colonies were stored at -80°C in 25% glycerol.

Plasmid DNA was purified at either the Australian Genome Research Facility (AGRF) (Brisbane, Australia) or using Wizard Plus SV minipreps DNA Purification system (Promega). Sequencing was performed at AGRF or Flinders and Southpath sequencing facility (Bedford Park, South Australia) using Big Dye terminator chemistry (Perkin Elmer, Waltham, MA, USA) and either M13F or M13R primer. The subsequent sequences were determined using either an ABI 3730xl 96-capillary automated DNA sequencer or an ABI 3100 sequence analyser.

4.3.3 Expressed sequence tag processing, contig assembly and preliminary analysis

Vector sequence and adaptor regions were removed from our expressed sequence tags (EST's) and clustered using Sequencher Version 4.1.4. (Genecodes, Ann Arbour, MI), resulting in the formation of a non-redundant set of EST sequences differentially expressed in the hypobranchial gland of *D. orbita.* Sequence assembly was performed to identify the amount of overlap in our sequences (Table 4.1), but all subsequent analysis was performed on unassembled EST sequences. Perl scripts were used to manipulate the formatting of sequences before BLAST analysis was performed on each EST. EST's were first manually analysed based on National Centre for Biotechnology (NCBI) nucleotide and protein database searches on the South Australian Partnership for Advanced Computing (SAPAC) BLAST portal, using the BLAST family of programs and classified in accordance with Gene Ontology defined gene functions. Manual sequence classifications were assigned where corresponding tBLASTx sequence matches displayed e values less than or equal to $1e^{-05}$. tBLASTx analysis was performed using standard nuclear translation, before *D. orbita* sequences underwent automated sequence annotation using the program BLAST2GO (Conesa *et al.* 2005). Processed, unassembled sequences were deposited at NCBI under accession numbers GD253659 to GD254033, J455778 to FJ455785, GE905829 to GE905839 and FJ476128 to FJ476170. Assembled non-redundant sequence data cannot be submitted to the NCBI EST database, therefore all subsequent sequence analysis was performed on these processed unannotated sequences.

EST sequences were also compared using BLASTn analysis to the 20 most sequenced molluscs in the EST database on NCBI as of July 2011, representing members of the taxonomic classes Gastropoda, Bivalvia and Cephalopoda. BLASTn analyses were performed

on all transcripts in addition to BLAST2GO nucleotide analysis. These EST sequences came from the following organisms, with the corresponding number of ESTs known listed in brackets: *Aplysia californica* (255605); *Lottia gigantea* (252091); *Crassostrea gigas* (206388); *Biomphalaria glabrata* (54309); *Mytilus californianus* (42354); *Euprymna scolopes* (35420); *Mytilus galloprovincialis* (19617); *Crassostrea virginica* (14560); *Lymnaea stagnalis* (11697); *Aplysia kurodai* (11445); *Hyriopsis cumingii* (10156); *Ilyanassa obsoleta* (9639); *Ideosepius paradoxus* (9079); *Haliotis asinina* (8335); *Haliotis discus* (8019) *Mizuhopecten yessoensis* (7607); *Haliotis diversicolor* (7394); *Pinctada martensi* (7130) *Tritonia diomedea* (7105) and *Pinctada maxima* (7099). Sequences from *Concholepas concholepas* produced from 454 pyrosequencing have been produced (Cardenas *et al.* 2011), but these remain largely unassembled and unannotated (average read length160-252bp) and were not used in this study.

4.3.4 Gene Ontology, KEGG (Kyoto Encyclopedia of Genes and Genomes) enzymatic pathway and InterPro annotation

D. orbita sequences were annotated in May 2011, in accordance with GO terms using the program BLAST2GO (Conesa *et al.* 2005). Using the NCBI BLAST server, BLASTx analysis was performed using the default parameters. For each sequence, the top 20 BLAST hits were analysed, and an E value threshold was set, discounting all BLASTx matches with e values greater than $1e^{-03}$. Furthermore, sequences less than 100 nucleotides were eliminated from this study in initial analysis to maximise the accuracy of annotation. The biological processes, molecular functions and cellular component were assigned to sequences with significant sequence similarity, based on sequence homology matches and previously defined terms. KEGG biochemical pathways were assigned to annotated sequences using BLAST2GO. Finally, the InterProScan suite, which is incorporated within BLAST2GO, was used to search for protein motif patterns and annotate sequences accordingly.

4.3.5 Realtime PCR gene expression levels analysis

First strand cDNA was synthesized from 1ug of total RNA from the hypobranchial gland and mantle of three female pre-copulatory individuals using Stratascript II reverse transciptase (Stratagene, La Jolla, CA) as per the manufacturers instructions. cDNA was synthesized separately on two occasions for each sample.

A quantitative real-time PCR (qRT-PCR) reaction was prepared containing 0.1 ug cDNA, 5ul 2X Absolute SYBR Green QPCR mix (Abgene, Epsom, UK), 10 ng/µl forward and 10ng/µl reverse primer, and sterile water to a final volume of 10µl and analysed in real-time on a Corbett RotorGene 3000 (Corbett Research, New South Wales, Australia). Seven sequences were selected from our EST library to be used in qRT-PCR gene expression analysis. Primers for arylsulfatase were designed from GD253910; primers for arginine kinase were designed from GD253702; primers for alphatubulin were designed from GD253991; primers for Angiotensin converting enzyme (ACE) were designed from GD253861; primers for Serine protease were designed from GD253884; Primers for Ribosomal protein P0 were designed from GD253988; Primers were designed for Cytochrome oxidase subunit I were designed from FJ476170. Preliminary analysis of primer sets identified COX1 primers were suitable for use as a housekeeper gene and were subsequently used in all gene expression analyses. Full details of primer information is listed in Table 4.1

Cycling conditions were as follows: 1 cycle of 95°C for 15 minutes for the activation of Taq polymerase followed by 40 cycles of (95 C for 30 seconds; 54 C for 30 seconds; 72 C for 30 seconds) and 1 cycle for 72 C for 1 min. Samples were run in triplicate and the run was repeated twice for each cDNA sample. All primers amplified their expected products. The external PCR standards for each gene of interest consisted of known numbers of molecules of purified PCR product. Standards were prepared and input copy number calculated according to published methods (Yin *et al.* 2001). All qRT-PCR data is represented as a ratio of the number of gene of interest copies to the number of COX1 products. Comparisons between mantle and hypobranchial gland tissue for qRT-PCR experiments were made using independent samples t-test in Prism statistical software for Graphpad (Graphpad software, La Jolla CA, USA). For all analyses, $P < 0.05$ was considered significant. All data expressed as mean ± standard error of the mean (SEM).

4.4 Results

4.4.1 EST manual analysis

A total of 554 randomly picked EST clones were sequenced from our differentially expressed hypobranchial gland cDNA library, and after performing quality control on sequence chromatographs we were left with 437 usable sequences. For the purposes of annotation of sequences and submission to Genbank's EST database, all 437 transcripts were manually

analysed in their non-assembled state, as required for database submission (Table 4.2, Table 4.3). When assemblage of sequences was performed eliminating overlap, as well as the removal of all clones with inserts less than one hundred base pairs, we were left with 311 singletons, and 37 contigs comprising, in total 348 non-redundant unique sequences (Table 4.2).

From these 437 EST sequences, 181 (41%) showed significant identity to known sequences within the NCBI database based on preliminary BLAST analysis using tBLASTx, BLASTn and BLASTp (Table 4.3). From these 181 sequences, a total of 28 ribosomal RNA sequences were found, 24 of which showed significant nucleotide identity to gastropod sequences, and four showed significant nucleotide identity with ciliate protozoa sequences, such as those belonging to *Paramecium tetraurelia* and *Cyclidium glaucoma* (Table 4.3). The remaining 153 sequences returned hits to known protein coding genes. It was observed that 57 of these sequences contained matches to known ciliate genes and would only be translated into functional proteins if the ciliate alternate codon translation system was utilized. The remaining 96 protein coding sequences, which were translated utilizing the standard codon convention, represent genes produced within the hypobranchial gland of *D. orbita.* Any sequences that were identified as ribosomal RNA or ciliate nuclear sequences from manual analysis were excluded from BLAST2GO annotation and ciliate sequences have been further investigated in chapter five. The remaining sequences, including the 96 protein coding genes together with the 256 EST sequences which did not show homology to any known nucleotide sequences were combined (352 sequences in total) for automated BLAST2GO annotation.

4.4.2 D. orbita gene BLAST2GO ontology and annotation

A total of 110 *D. orbita* unique EST's were assigned gene ontology (GO) categories from BLASTx hits in BLAST2GO, which when combined with the ribosomal RNA sequences identified, resulted in a total of 134 sequences being assigned a function from *D. orbita*. Figure 4.4 shows the distributions of GO terms from our sequences according to the GO consortium. Sequences were grouped together and functions were assigned, reporting on the biological process(es) (level 6 GO terms), molecular function(s) (level 3 GO terms) and the cellular component(s) (level 5 GO terms) that the genes are likely to be involved in, based on BLASTx sequence analysis results. Sequences of particular interest are listed in Table 4.4.

The most dominant biological process encountered from our EST library (177 biological process matches in total) was identified as translation (n=22), followed by cellular respiration (n=13), Proton transport (n=10) and ATP synthesis coupled electron transport (n=8) (Fig 4.3a). Several key genes involved in protein translation and modification were also identified; genes involved in post-translational protein modification (n=4), cellular protein complex assembly (n=18), protein folding (n=5) and phosphorylation (n=12) were all expressed in the hypobranchial gland, as well as genes involved in spindle organization and elongation (n=5 and 4 respectively), carboxylic acid, RNA, monosaccharide and hexose metabolic processes (n= 8,4,4 and 4 respectively).

When looking at the molecular function assigned to our sequences (a total of 233 assignments), genes implicated in nucleotide and nucleoside binding dominated (n=36 and 14 respectively), with a large proportion of sequences involved in protein binding also observed (n=29) as were tetrapyrole binding sequences (n=7) and sequences that are structural constituents of the ribosome (n=19). Numerous enzymatic activities were also observed, including hydrolase activity (n=33), oxidoreductase activity (n=18) and transferase activity (n=9). Sequences involved in transport were also observed, including substrate-specific transporter activity (n=9) and transmembrane transporter activity (n=10).

When analysing the cellular component assignments to sequences in our EST library, the most abundant by far, of all 284 matches produced, were those sequences within the cytoplasm or cytoplasmic part (n=58 and 48 respectively), closely followed by sequences localized in intracellular organelles (n=71). Sequences intrinsic to membranes, organelle inner membranes and plasma membranes were observed (n=15, 10 and 4 respectively). Sequences were also identified for genes that localized to the ribonucleotide protein complex (n=24) and ribosomal subunit (n=12).

To investigate the potential enzyme pathways functioning within the hypobranchial gland, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used (Kanehisa *et al.* 2004). In sequences where enzymatic pathways could be assigned, the KEGG Pathways complimented GO annotation results (Table 4.4). A total of 53 cDNA sequences were assigned to KEGG pathways, and 22 different pathways were suggested to be operating within the hypobranchial gland of *D. orbita*. Pathways of particular interest are listed in Table 4.5. In total,

only 11 different genes were identified to have KEGG enzymatic pathways, with a total of 20 different cDNA sequences involved in these classifications (Table 4.5).

InterProScan analysis within the BLAST2GO software allowed a broader range of sequences to be identified, based on protein structure and feature information. From the 352 sequences that were analysed within this study, 39 showed protein signature matches to previously defined protein domains. In addition, 122 sequences displayed conserved protein domains synonomous with signal peptides, and 61 sequences were identified to have transmembrane domains.

Interestingly one of our sequences (accession number GD253910) matched a sulfatase gene belonging to the sea urchin *Strongylocentrotus purpuratus* with an e value of $2.11e^{-39}$ according to BLASTx analysis. InterProScan analysis failed to identify any features or motif in this sequence, but KEGG analysis identified this sequence as one involved in sphingolipid metabolism. E value cutoff values were maintained at a very low threshold ($1e^{-03}$) in order to maximise the chances of picking up even distant sequence similarities, but we were unable to identify any other enzyme sequences, such as bromoperoxidases or tryptophanases, which may be involved in Tyrian purple biogenesis.

Lastly, genes involved in carbohydrate metabolism and biopolymer synthesis were identified within our cDNA library. While specific mucin genes were not identified within our cDNA library, a gene involved in the regulation of mucins (Jonckheere *et al.* 2004), Smad4, was shown to be expressed within the hypobranchial gland.

4.4.3 Molluscan EST sequence comparisons

When molluscan EST sequences were compared to our hypobranchial gland sequences we failed to identify any sequence similarity for 211 *D. orbita* sequences (Fig 4.5). When compared to our BLAST2GO annotation results, it was determined that only 23 of these sequences showed sequence similarity to known Genbank sequences, leaving 188 sequences that did not display any similarity to nucleotide or mollusc EST sequences on Genbank. 87 sequences with sequence similarity to other molluscan sequences had already had their function putatively identified from BLAST2GO analysis in the Genbank nucleotide database. An additional 54 sequences were not assigned function from BLAST2GO analysis were found to have sequence similarity with molluscan ESTs. From these results, we identified that the majority of these

sequences showed sequence similarity to known gastropod EST's, and sequences belonging to *Ilyanassa obsoleta* attribute to nearly half of these matches (Fig 4.5).

4.4.4 Gene expression comparisons

To validate that our subtractive hybridization had enriched for genes expressed in the hypobranchial gland of *D. orbita*, qRT-PCR was conducted on six genes. Arylsulfatase expression was significantly higher in hypobranchial gland tissue than the mantle (Fig 4.5A; p=0.03). Angiotensin converting enzyme (ACE) was only expressed in the HBG and not detected in the mantle (Fig 4.5B), and the trend for all other genes investigated (arginine kinase, alphatubulin, serine protease and Ribosomal protein P0) indicated that gene expression was higher in the hypobranchial gland when compared to mantle tissue (Figure 4.5). Expression levels of arginine kinase, serine protease and alphatubulin (Fig 4.5.D,E and C respectively) varied greatly between individuals.

4.5 Discussion

This study resulted in a total of 437 high quality EST sequences that are up-regulated or uniquely expressed in the hypobranchial gland of *Dicathais orbita*. When sequences underwent preliminary manual sequence analysis, the putative functions of 96 *D. orbita* protein coding sequences were identified, 28 ribosomal RNA sequences were identified and 256 novel sequences were not assigned function. A further 57 sequences were identified as having significant similarities to known ciliate protozoan genes (Chapter 5). Despite the relative small size of this EST library, this data represents the second largest collection of Muricidae sequences currently available in NCBI. Until the largely unannotated publication of the *C. concholepas* transcriptome (Cardenas *et al.* 2011), very little sequence data had been collected from the family Muricidae. Even now only 3190 annotated muricid sequences have been published on NCBI as of October 2011, and this study has increased this number by approximately 20%. Furthermore, almost every sequence currently published is used specifically in either taxonomic studies (Harasewych *et al.* 1997, Oliverio & Mariottini 2001, Oliverio *et al.* 2002, Colgan *et al.* 2007) or population genetic studies (Sanford *et al.* 2003, Martel *et al.* 2004), limiting the diversity of muricid sequences known. Thus the broad range of functional genes identified within our cDNA library greatly increases the knowledge of Gastropoda gene expression, particularly regarding gene expression within the hypobranchial gland.

53% of the *D. orbita* genes sequenced did not share similarity with any sequence in the current databases (Fig 4.4). This high level of novel and non-homologous sequences is likely to be influenced by the fact that no sequence data has ever been attained from any hypobranchial gland sample to date, and there is very little known about the gland itself. A previous SSH project investigating the gene expression in immuno-compromised clams also reported a high incidence of novel or non-homologous sequences at 54 % (Gestal *et al.* 2007). A large scale sequencing project of the eastern oyster *Crassostrea virginica* indicated that around 53% of sequences produced showed no similarity to any known sequences (Quilang *et al.* 2007). While 124,654 next generation sequences have been identified from *C. concholepas*, only 859 sequences have had functions assigned based on sequence homology analysis (0.007%), and this is likely due to the small length of each sequence (averaging 160-253 bp) (Cardenas *et al.* 2011). Several studies have identified that the success of BLAST sequence matches is dependent on query sequence length (Gotz *et al.* 2008, Jacob *et al.* 2008), and with the average length of our sequences being only 502 bases (Table 4.2), this is likely contributing to the frequency of unidentified sequences within our cDNA library. Larger e values indicate sequence matches that are less conserved, while smaller E values are assigned in sequence comparisons where larger stretches of highly conserved sequences are identified. From our analysis, the majority of the e values fall between $1e^{-60}$ and $1e^{-5}$, while the frequency of e values less than $1e^{-60}$ is low. This is symptomatic of the scarcity of molluscan sequences currently known (Laffy *et al.* 2009) (Chapter 3) and the unique functionality of sequences in our SSH cDNA library.

Realtime PCR was performed on six genes that were identified from our manual sequence analysis and was performed on both hypobranchial gland and mantle tissue in order to confirm the differential expression of these transcripts (Fig 4.5). Our results confirmed that all six genes whose expression levels were measured in the hypobranchial gland were up regulated compared to their expression in the mantle, although we were only able to measure a significant difference in expression of the arylsulfatase sequence (Fig 4.5). Variation between the gene expression of individuals investigated is likely to be responsible for the lack of statistical significance in the remaining five genes (Alphatubulin, angiotensin converting enzyme, arginine kinase, serine protease and Ribosomal protein P0). Despite this lack of statistical support for these expression level differences, the overall trend in expression levels validated our SSH results, supporting our classification of these transcripts as being differentially expressed in the hypobranchial gland.

Looking at the gene ontology results from our analysis, it is apparent that a high level of protein translation and modification is being performed within the hypobranchial gland (Table 4.4). At least 50 biological process categorizations were made for genes involved in regulation of gene expression, cellular protein complex assembly, translation or post translational protein modification (Fig 4.2A). GTPase activity and GTP binding are both intrinsic to the process of ribosomal protein translation (Voigt & Nagel 1993). The expression of genes involved in electron transport and aerobic respiration, including those with cytochrome-c oxidase, ATPase and ATP binding activity within the hypobranchial gland suggests there may be an increase in cellular activity within the gland relative to mantle tissue. Histological analysis of the hypobranchial gland of *D. orbita* has revealed seven distinct cell types, many with secretory functions (Westley *et al.* 2010b). ATP synthesis has been directly correlated with ATP demand (Hochachka 1994), and one of the most prominent consumers of ATP in cells is protein synthesis (Wieser & Krumschnabel 2001). As such, the expression of these ATP synthesis genes in the hypobranchial gland could suggest an increase in protein production.

The presence of genes involved in monosaccharide metabolic processing and hexose metabolic processing could be related to the production of mucus within the hypobranchial gland. It has been well documented that the hypobranchial gland of muricids can secrete large volumes of mucus (Naegel 2005). Several studies have investigated the chemical composition of mucous trails of gastropods (Connor 1986, Smith *et al.* 1999, Smith & Morin 2002), however few studies have investigated the mucus produced in the hypobranchial gland. Hunt and Jevon's investigations into the compositions and characteristics of the hypobranchial mucin of *Buccinum undatum* (Hunt & Jevons 1965, Hunt 1967), has given the most thorough analysis of muricid mucus currently available. It was determined that 41% of the mucus is protein, with 30% carbohydrate, 2% sugar (hexosamine), 13.2% sulphate and 0.5% calcium, with an ash composition of 9% (Hunt & Jevons 1965). An ash composition of 9% suggests a considerable proportion of hypobranchial gland mucus contains inorganic compounds. A high proportion of our identified sequences have shown ion and cofactor binding capacity (Fig3.B). The calcium binding protein calmodulin was also shown to be expressed within the hypobranchial gland of *D. orbita* (Table 4.4). While the functions of calmodulin vary greatly, it has been suggested that the presence of calmodulin within the extracellular mucus of fish is involved in the permeability of skin epithelium, influencing water and ion uptake (Flik *et al.* 1984).

It is likely that numerous proteins are involved in mucin production, both as mucus proteins, as well as in the synthesis and processing of these proteins. Recent publications investigating nacre and shell deposition in molluscs has identified a high proportion of secreted proteins are expressed in mantle tissue (Jackson *et al.* 2006, Clark *et al.* 2010). Mucus is also secreted from the pedal gland of molluscs (Wright *et al.* 1997) and these proteins might also be expressed in the mantle tissue. Thus it is possible that secreted genes from the hypobranchial gland may not be present in our cDNA library due to their expression within mantle tissue. A more thorough analysis of the hypobranchial gland transcriptome would be required to identify all sequences involved in mucus secretion in the future.

Genes associated with Tyrian purple biosynthesis were predicted to be expressed in the hypobranchial gland of *D. orbita*. Arylsulfatase has been shown to facilitate the conversion of tyrindoxyl sulphate into the biologically active precursors of Tyrian purple within the hypobranchial gland of muricid molluscs (Erspamer 1946, Cooksey 2001, Westley *et al.* 2006), including *D. orbita* (Baker & Sutherland 1968). The presence of an arylsulfatase sequence within our cDNA library (Table 4.4) is the first identification of a transcript responsible for this enzyme within the Muricidae. In addition, the identification of an arylsulfatase sequence is the first confirmation that the arylsulfatase activity previously identified in these molluscs is not obtained from dietary sources, but is in fact transcribed within the hypobranchial gland.

Some other genes involved in tryptophan metabolism and alkaloid biosynthesis were identified by BLAST2GO analysis (Table 4.5). This further supports *de novo* synthesis of the tryptophan derived indole alkaloids produced by *D. orbita* and other Muricidae. However, current analysis of our cDNA library did not identify the expression of any bromoperoxidases or any other enzyme likely to be involved in the bromination of these Tyrian purple precursors. Due to the high level of novel, unidentified sequences within our cDNA library, it is possible that we have obtained one or more of these sequences, but they are not homologous to the sequences present in the current database. It is possible that these enzymes have evolved to perform similar function without any sequence similarity to bromoperoxidases that have been previously identified from bacteria and marine algae (Pelletier *et al.* 1994, Weyand *et al.* 1999, Isupov *et al.* 2000). It is also possible that homologous bromoperoxidase genes are expressed in the hypobranchial gland, but were not sequenced in our cDNA library, although preliminary attempts to amplify bromoperoxidase sequences from hypobranchial gland DNA failed to

identify sequences (See appendix III). Further research is required to investigate any evidence for molluscan-specific bromoperoxidase enzymes.

The ultimate precursor to Tyrian purple is stored as an indoxyl sulphate choline ester salt in the hypobranchial gland of muricids (Baker & Duke 1976). Murexine and urocanylcholine are pharmacologically active choline esters present within the hypobranchial glands of muricids (Whittaker 1960). Acetylcholinesterases are vital for the control of choline esters in biological systems and play a role in cholinergic neurotransmission in molluscs and other invertebrates (Hussein *et al.* 2002). The presence of an acetylcholinesterase within the hypobranchial gland of *D. orbita* (Table 4.4), suggests a significant amount of neuronal activity is taking place within the hypobranchial gland. Recent histochemical analysis of the hypobranchial gland reported that no nerve cells were present within the hypobranchial gland of *D. orbita* (Westley *et al.* 2010b). Consequently, acetylcholinesterases present in the gland must be expressed by another cell type other than neuronal, and these proteins may be involved in the maintenance and regulation of the bioactive murexine and urocanylcholine ester compounds present within the gland.

Ciliated supportive cells have been identified as the most abundant of all cell types present within the hypobranchial gland (Westley & Benkendorff 2008, Westley *et al.* 2010b). The protein products of particular genes, namely dynein, alpha tubulin and beta tubulin, are known to be key components within cilia, and our SSH library suggests that these gene sequences are upregulated in the hypobranchial gland (Table 4.4). The ciliated supportive cells of the hypobranchial gland have been identified as the localized site of Tyrian purple precursor production, which does not occur until arylsulfatase enzymes are released from within the supportive cells (Westley 2008). The identification of these structural genes is important as it suggests that SSH can be used to differentiate between physiologically different tissue types. This can be seen in the hypobranchial gland of *D. orbita*, where highly specialized cellular functions, such as the release of Tyrian purple precursors and biosynthetic enzymes, are observed (Fig 4.3C). Hypobranchial glands from the Muricidae have been identified as containing highly specialized secretory cells (Hunt 1973, Roller *et al.* 1995, Naegel & Aguilar-Cruz 2006, Westley *et al.* 2010b). Histochemical analysis of *D. orbita* has identified seven secretory cell types present within the hypobranchial gland, most secreting intracellular granules containing mucoproteins or acidic sulphated mucopolysaccharides (Westley *et al.* 2010b). Further, the individual intracellular granules identified within the different cells

displayed varied secretions, including eoisinophilic and basophilic spherules and granules of differing dimensions. This large and varied number of secretory cells may account for the number of membrane intrinsic, cell wall and external side of the plasma membrane sequences that were identified within the gland (Figure 4.2C). This remarkable diversity in cell types and secretory phenotypes may be responsible for the large number of varied and novel sequences identified within the hypobranchial gland of *D. orbita*.

The location of the hypobranchial gland in relation to the anatomical connections to other organs may explain the expression of some other genes and gene functions that were identified in this study. The expression of sequences exhibiting peptidyl-dipeptidase A activity (Table 4.5) suggested a link between the hypobranchial gland and vasoconstriction/vasodilation (Salzet *et al.* 2001). The hypobranchial gland has been described as highly vascularised tissue (Fretter & Graham 1994), and the vascular sinus is found in close proximity to the hypobranchial gland of *D. orbita* (Westley *et al.* 2010b). Consequently these genes could be involved in regulating blood vessels and blood supply to this metabolically active biosynthetic organ.

From our SSH results, we are able to make several inferences into the biological functions and processes that occur within the hypobranchial gland of the marine snail *D. orbita*. Genes that may play a role in mucus production and secretion were expressed within the gland, as were sequences involved in protein production and modification. This supports current knowledge regarding hypobranchial gland physiology and function, including its role in the production and secretion of mucus into the mantle cavity. Clearly, high levels of protein production and processing are taking place within the gland. The role of the hypobranchial gland in Tyrian purple production is also confirmed with the identification of an arylsulfatase gene, in addition to acetylcholinesterase. Further research is required to identify the remaining enzymes involved in Tyrian purple biosynthesis. The high proportion of novel sequences present within our cDNA library suggests there is an abundance of unique *D. orbita* genes, and their expression within the hypobranchial gland may suggest an invlovement in key biochemical processes that are specific to this gland in molluscs. Further characterization of these genes, by obtaining full length sequences would be required to obtain a more thorough understanding of the transcriptome of the hypobranchial gland. This work demonstrates that the SSH approach holds great potential for the discovery and classification of novel proteins, including useful

biosynthetic enzymes, and lays a strong foundation for future studies investigating Tyrian purple biogenesis and hypobranchial gland functionality.

4.6 Acknowledgements

This work was funded by an anonymous philanthropic foundation. Patrick Laffy was supported firstly by a Flinders University Faculty of Science and Engineering Research scholarship, followed by a Flinders University Postgraduate Research scholarship. We would also like to thank the South Australian Partnership for Advanced Computing, for the use of their BLAST portal and Dr Chantel Westley for her assistance in specimen dissection.

Table 4.1 **Quantitative real-time PCR (qRT-PCR) primers.** Primers were designed from EST sequences produced from transcriptome study. In all analyses, Cytochrome Oxidase Subunit I (COXI) was used as a housekeeper gene to normalize expression levels.

Gene of interest	Primer name	Primer sequence (5'-3')
Angiotensin converting enzyme (ACE)	ACE_for	GCTGCACGCCTTCGTTCGTC
	ACE_rev	CCAGATCCTCGATGCCCTCCC
Ribosomal Protein P0	Rib_P0_for	GCCCTGCAGTCCTCGACATC
	Rib_P0_rev	GTCTCTTGAAGCCATTGACG
Serine Protease	Ser_prot_for	GCCGATGGTTTAGCAAAG
	Ser_prot_rev	GGGTGTAGGGGCCGCTG
Alpha tubulin	Alphatubulin_for	GGGCTCGAAGCAGGCGTTGG
	Alphatubulin_rev	CCACTTCCCTCTGGCCACC
Arginine Kinase	ArginineKinase_for	CCCCAACAGCCTCCTCAAAC
	ArginineKinase_rev	GCACCATTGAAGCGGGTCC
Arylsulfatase	Arylsulfatase_for	CCGTCGGGATGTGACTCCAC
	Arylsulfatase_rev	GGTCAAAGTCTTCATGGCTTGC
Cytochrome oxidase subunit I	COXI_for	GCTCCGGACATGGCCTTTC
	COXI_rev	CGCTTTCAACAGCAGCTGAGG

Table 4.2 Summary statistics of EST's generated from the hypobranchial gland of *D. orbita*

Feature	Value
Total number of clones sequenced	554
Average length of high quality EST's (bp)	502
Number of high quality sequences	437
Number of contigs	37 (containing 126 sequences)
Number of singletons	311

Table 4.3 Statistics of BLASTx searches, performing manual and automated analysis using BLAST2GO.

Feature	Manual analysis*	BLAST2GO analysis
Number of sequences analysed	437	352
Number of sequences with significant BLAST hits	181	110
Percentage of sequences with unique hits	41 %	31 %
Number of unidentified sequences	256	242
Classification of Annotated sequences from Manual sequence analysis	Number of sequences (Overall Percentage)	
D. orbita ribosomal RNA sequences	24 (13%)	
D. orbita expressed genes	96 (53%)	
Ciliate protozoan ribosomal RNA sequences	4 (2%)	
Ciliate protozoan expressed genes	57 (31%)	

*Sequences in manual analysis included sequences with significant similarity to ciliate sequences and ribosomal RNA sequences, which were not included in BLAST2GO analysis

Table 4.4 Genes expressed in the hypobranchial gland of *D. orbita* identified by BLAST2GO.
Sequences matching ribosomal RNA and ribosomal proteins were excluded from this table

Gene Function	Genbank Accession Number	Gene annotation	EC Number
Transcription	GD253679	Mad homolog 3	-
Translation	GD253962	H2A histone member V	EC:3.6.5.3
	GD253823, GD253818	Translation initiation factor subunit 1 alpha	EC:3.6.5.1-4
	FJ455780, FJ476164, FJ476158, FJ476156, FJ476138	Cytochrome B	-
Protein modification	GD253702, GD253925	Arginine kinase	EC:2.7.3.3, EC:5.3.4.1, EC:2.7.3.3., EC:2.3.2.13
	GD253956, GD253968, GD253975, GD254001, GE905837, GD253853, GD253849, GD253848, GD253837, GD253807, GD253748, GD253737	Calmodulin	-
	GD253866	dual-specific tyrosine-phosphorylation regulated kinase	EC:2.7.10-11
	GD253864, GD253881	glyceraldehyde-3-phosphate dehydrogenase	EC:1.2.1.12
	GD253884	Serine protease	-
	GD253872	heat shock protein 90	-
	GD253861, GD253664	Angiotensin converting enzyme	EC:3.4.15.1
	GD254009	sec11 protein	-
	GD254006	Hydrocephalus inducing protein	-
	GD253865	phosphatase catalytic gamma isoform	-
Electron Transport	FJ476170, FJ476167, FJ476148, FJ476141, FJ476128	Cytochrome c oxidase subunit I	EC:1.9.3.1
	GD253781	glutaryl-coenzyme a dehydrogenase	EC:1.3.99.7
	FJ455785	NADH dehydrogenase subunit I	EC:1.6.5.3
Cellular processing	GD254015, GD253983, GD253924, GD253901, GD253894, GD253867, GD253857, GD253696	Alpha-tubulin	EC:3.6.5.1-4
	GD253691, GD253871	Dynein light chain family protein	-
	GD253991, GD253936, GD253913, GD253906, GD253890, GD253885, GD253803, GD253838,	Beta Tubulin	EC:3.6.5.1-4

	GD253801, GD253693	GD253779,		
	GD253783		tbc1 domain member 13	-
	GD253713, GD253945	GD253852,	sodium potassium-dependant atpase beta-2 subunit	EC:3.6.3.9
	GD253868		pol-like protein	-
	GD253671,	GD253955	Histidine triad nucleotide binding protein 1	-
	GD253670, GD253845,	GD253772, GD253879	Actin	-
Neurotransmission	GD253726		Agrin isoform 1	-
	GD254024		Acetylcholinesterase	-
Tyrian purple Biosynthesis	GD253910		Arylsulfatase J	EC:3.1.6.1

Table 4.5 Key KEGG enzymatic pathways involved or active within the hypobranchial gland of *D. orbita* as identified through BLAST2GO analysis.
* Glutaryl-CoA dehydrogenase and alcohol dehydrogenase are involved in more than three KEGG enzymatic pathways therefore are grouped together at the bottom of this graph

KEGG enzymatic pathway	Gene(s) identified	Sequence(s) identified
Oxidative phosphorylation	ATPase	GD253852
	Cytochrome C oxidase	FJ476138, FJ476156, FJ476164, FJ476170, FJ476128
	NADH dehydrogenase	FJ455785
Vibrio cholera infection	Disulfide-isomerase	GD253702
	ATPase	GD253852
Flagellar assembly	ATPase	GD253852
Type III secretion system	ATPase	GD253852
Sphingolipid metabolism	Arylsulfatase	GD253910
Androgen and estrogen metabolism	Arylsulfatase	GD253910
Arginine and proline metabolism	Arginine kinase	FJ455783, GD253702
Glycolysis/gluconeogenesis	Glyceraldehyde-3-phosphate dehydrogenase	GD253864, GD253881
Ubiquinone biosynthesis	NADH dehydrogenase	FJ455785
Renin-angiotensin system	Peptidyl-dipeptidase A	GD254022, GD253664
Complement and coagulation cascades	Protein-glutamine gama-glutamyltransferase	GD253702
Gene identified *	KEGG enzymatic pathways	
Glutaryl-CoA dehydrogenase (GD253781)	Benzoate degradation, Lysine degradation and Tryptophan metabolism, Fatty acid metabolism	
Alcohol Dehydrogenase (GD253933)	1- and 2-methyl naphthalene degradation, drug metabolism, D-alanine metabolism, Cytochrome p450 metabolism, alanine and aspartate metabolism, 3-chloroacrylic acid degradation, Aminoacyl-tRNA biosynthesis, phenylalanine, tyrosine and tryptophan biosynthesis, aminophosphonate metabolism, Nicotinate and nicotinamide metabolism, Bile acid biosynthesis, Tyrosine metabolism and Alkaloid biosynthesis, Fatty acid metabolism	

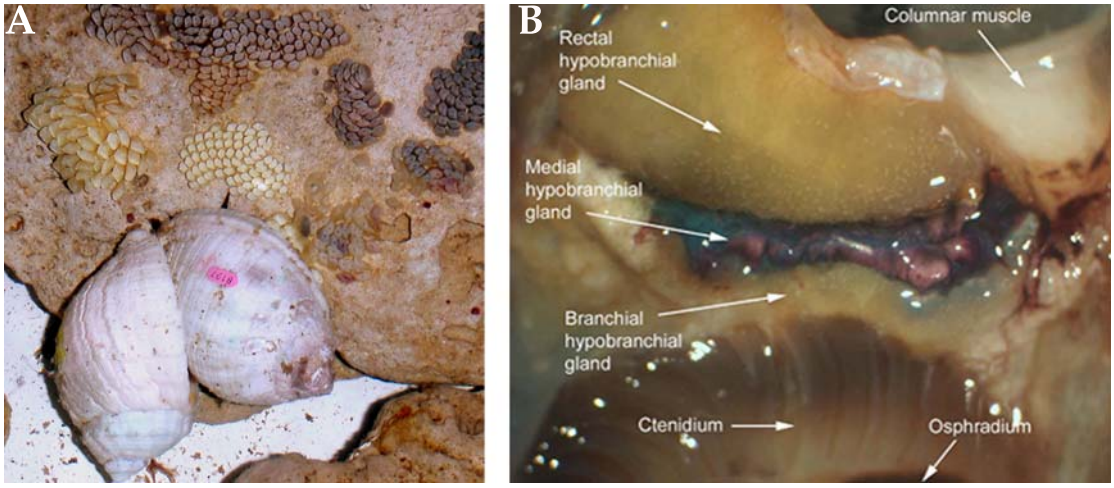


Figure 4.1 *Dicathais orbita* amongst egg capsules and the partially dissected hypobranchial gland of *D. orbita*.
A) shows two captive specimens of *D. orbita* surrounded by several egg capsule clutches. B) shows the three regions of the hypobranchial gland; branchial, medial and rectal, in reference to the ctenidium, ospharidium and columnar muscle within the mantle cavity of *D. orbita*. Photo courtesy of Chantel Westley.

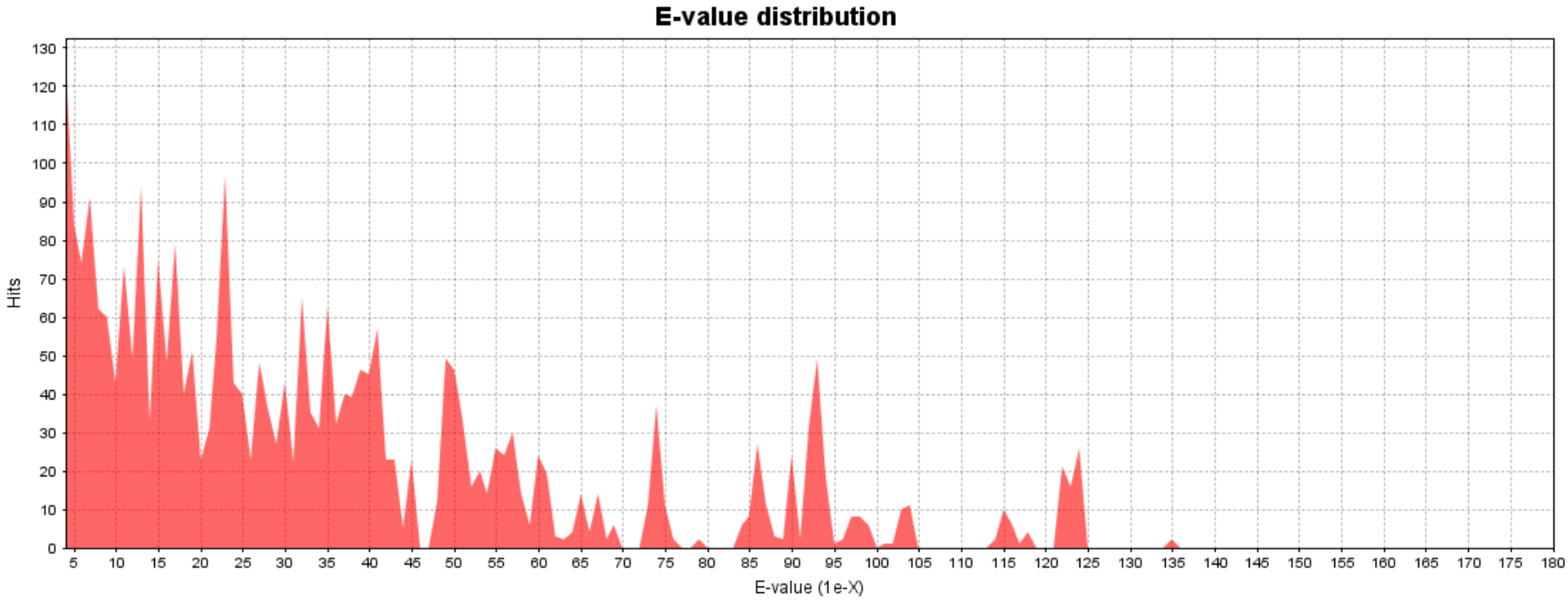


Figure 4.2 Distribution of E-values from the top hits in the NCBI protein database.
This figure was compiled in BLAST2GO, from annotation results of genes upregulated in the hypobranchial gland of *Dicathais orbita* and displays the number of sequences with corresponding significant BLASTx matches. The cutoff value for significant sequence matches was an e-value less than or equal to $1e^{-03}$

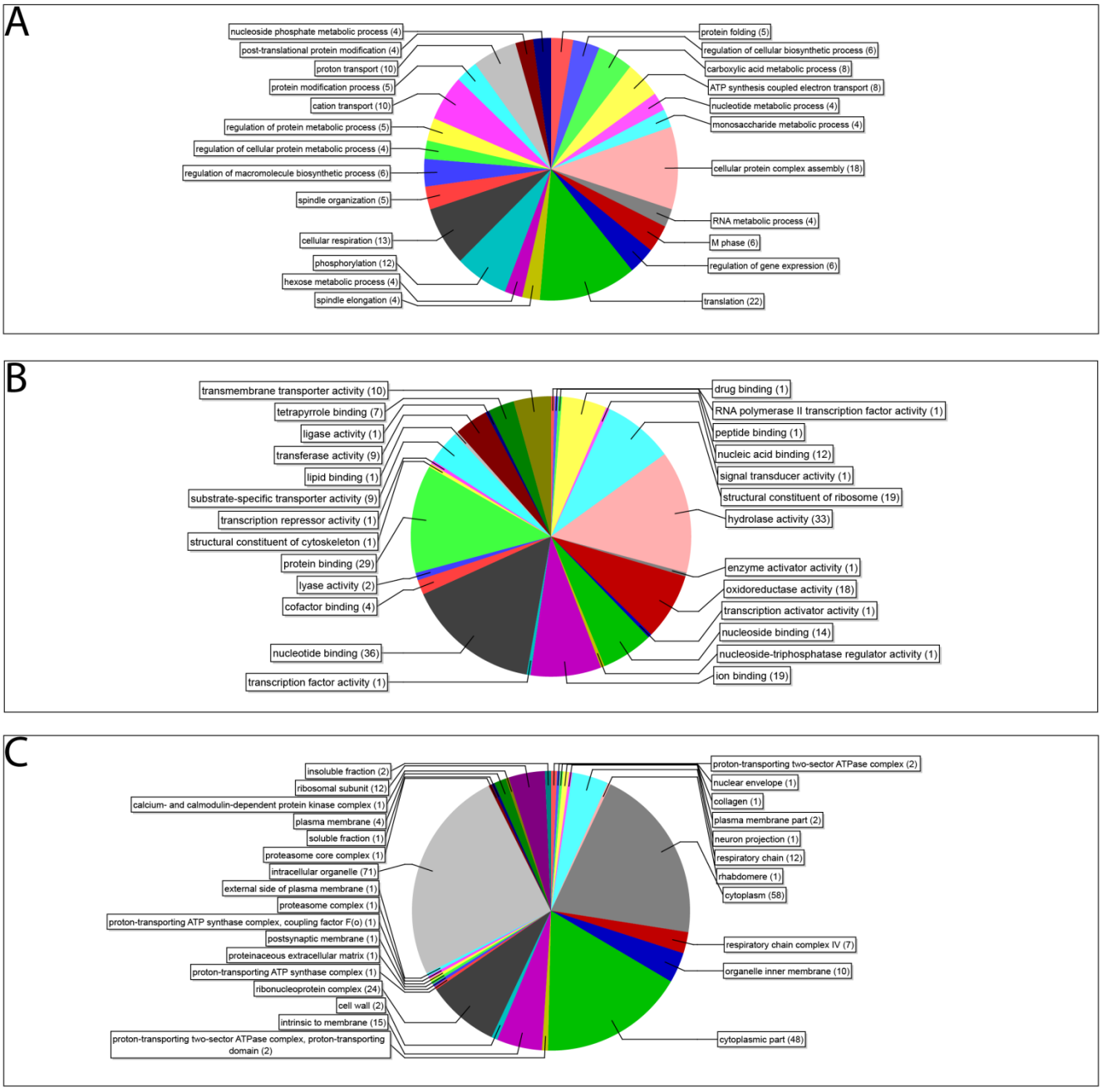


Figure 4.3 Gene ontology (GO) assignment (3rd level GO terms) of differentially expressed hypobranchial gland genes expressed in *D. orbita* using BLAST2GO automated sequence annotation.

A total of 110 sequences from our EST library, displaying BLASTx e values < $1e^{-03}$ underwent GO assignment. A) Biological process assignments, categories with more than three positive assignments are displayed. B) Molecular functions of sequences, C) Cellular compartment assignments. It is important to note that multiple assignments are possible for each individual sequence and as such sequence function assignments vary between each GO category.

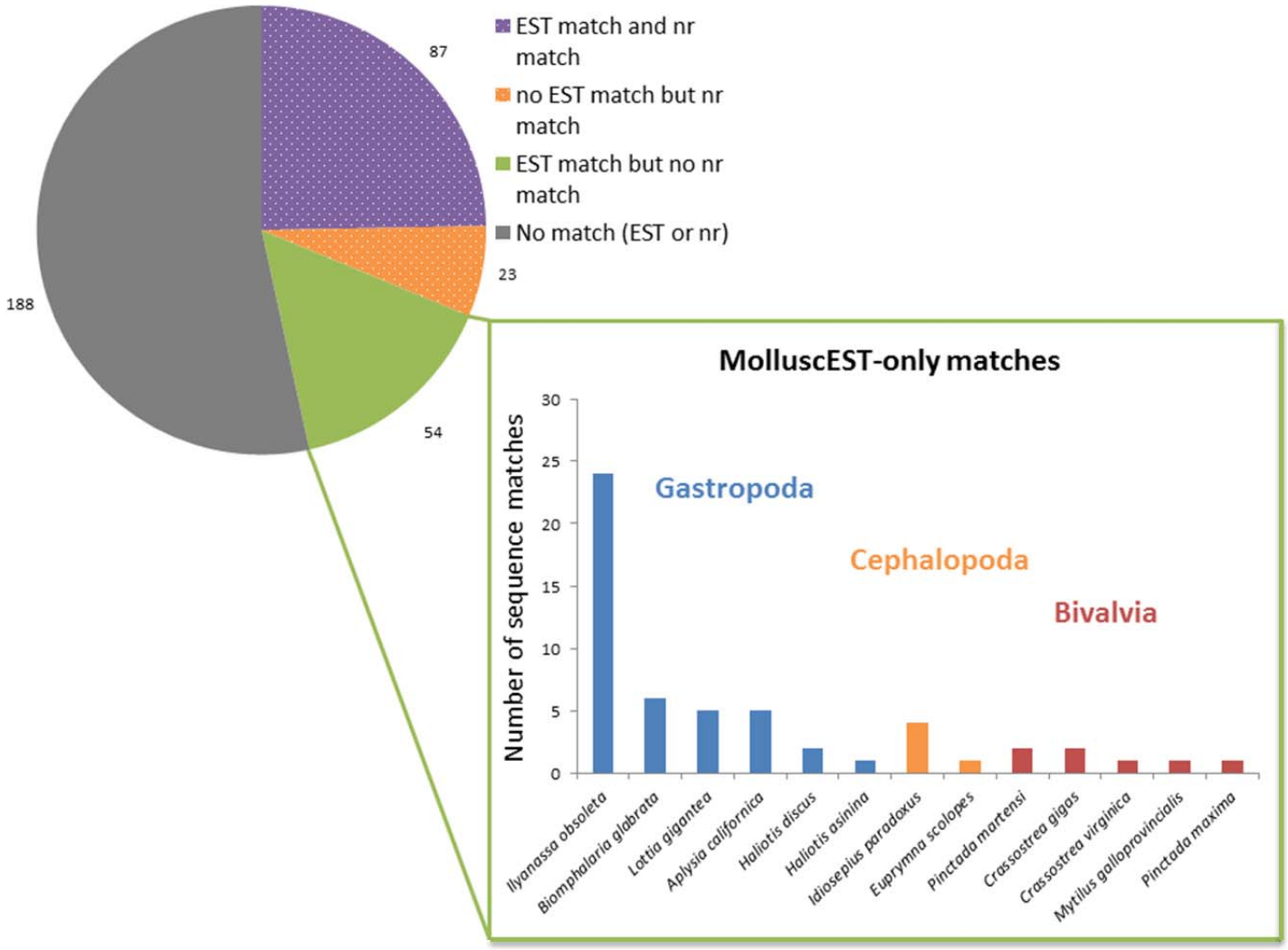


Figure 4.4 Distribution of both nucleotide and sequence matches separated by abundance and taxonomic Class.
Manual tBLASTx analysis was performed in July 2011, comparing sequences to ESTs from the 20 most sequenced molluscs in the NCBI EST database.

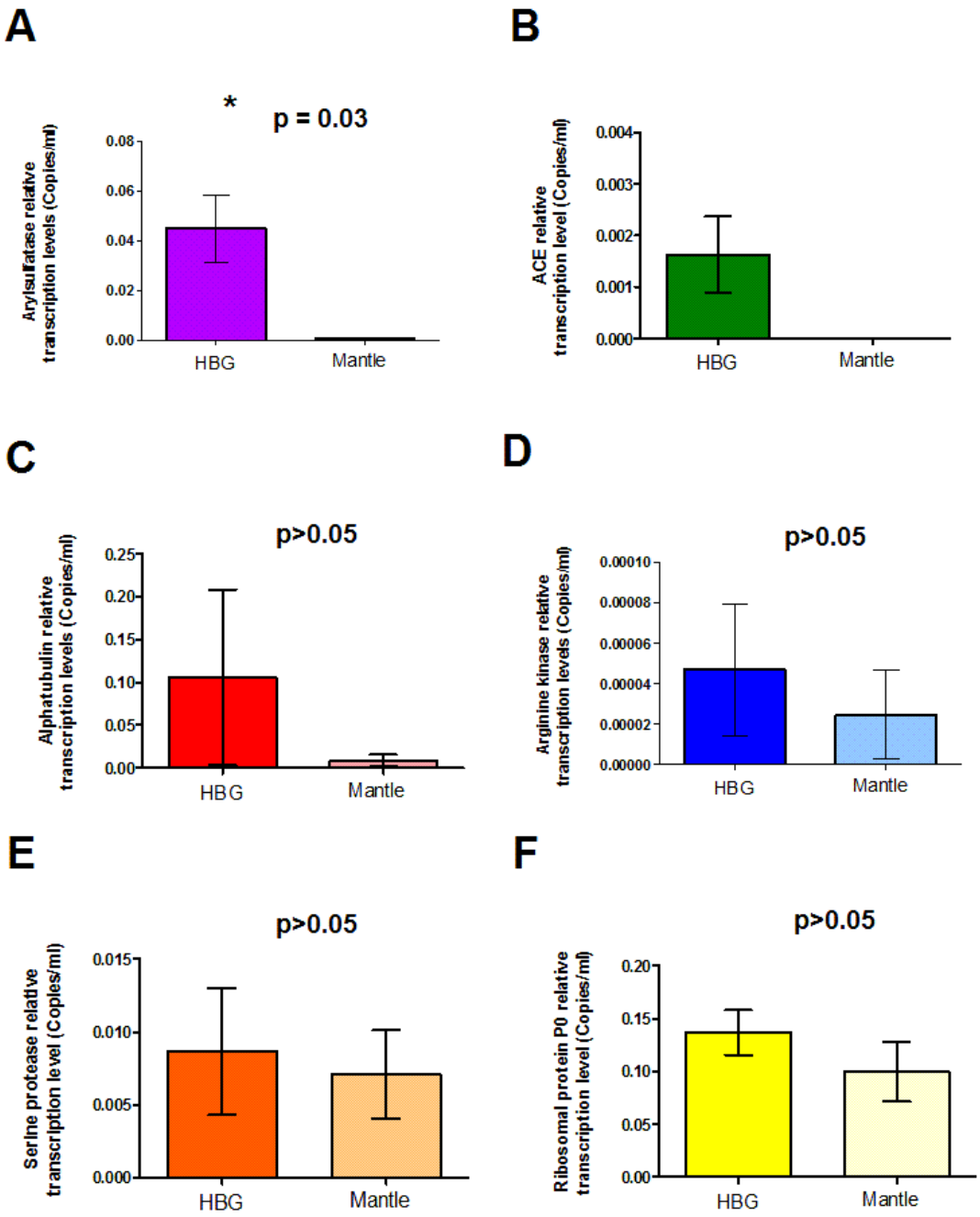


Figure 4.5 qRT-PCR results for selected genes, relative to the expression levels of housekeeper gene COXI (Cytochrome oxidase subunit I). Letters (A-F) pertain to which genes' expression level is displayed and are defined as follows; A (arylsulfatase), B (angiotensin converting enzyme (ACE)), C (alpha tubulin), D (Arginine kinase), E (Serine protease) and F (Ribosomal protein P0). All assays were performed in triplicate on RNA isolated from three pre-copulatory female *D. orbita*, and was repeated at least twice. Final values were expressed as mean ± SEM. *P < 0.05 for arylsulfatase mRNA transcripts in hypobranchial gland tissue compared to mantle tissue.

Chapter 5. Novel application of suppressive subtractive hybridization for the identification of symbionts: Discovery of ciliate protozoa in the hypobranchial gland of *Dicathais orbita* (Neogastropoda, Mollusca)

Histological analysis and associated statistical analysis presented within this manuscript were performed by Dr Chantel Westley at Flinders University, who will be credited as second author upon submission of this manuscript.

5.1 Abstract

Microbial endosymbionts are broadly distributed in the natural word, but their presence is rarely considered in annotated gene libraries. A study applying suppressive subtractive hybridization (SSH) to the identification of genes expressed in the hypobranchial gland, a biosynthetic organ of the marine snail *Dicathais orbita* and identified a symbiont. Manual analysis of 437 sequences indicated that a subset of these expressed sequence tags (ESTs) showed high sequence identity to ciliate protozoan ribosomal RNA (rRNA) sequences and some protein coding sequences used the alternate codon system typically seen in ciliate protozoans. Phylogenetic analysis of rRNA sequences from the cDNA library indicated the presence of a Suctorian ciliate belonging to the class Phyllopharangea, as well as the potential presence of a second ciliate whose taxonomic class has yet to be determined. Gene ontology revealed that the ciliates were expressing protein coding genes and were likely to be undergoing some cell replication. Histological analysis of *D. orbita* confirmed an abundance of intracellular ciliate protozoans in the hypobranchial gland, which peaked during the copulatory season. Histological analysis of the mantle tissue supported our finding that ciliate genes were differentially expressed in the hypobranchial gland. Not only is this study the first reported case of ciliate protozoans being present in the hypobranchial gland of any mollusc, but it highlights a previously unrecorded use of SSH technique, namely the identification of multiple genomes from a single biological sample.

5.2 Introduction

Suppressive subtractive hybridization (SSH) has been used since its development in 1999 (Diatchenko *et al.* 1999) to investigate and identify differing gene expression between two tissue samples. Specific expression patterns based on disease or illness (Cario-Toumaniantz *et al.* 2007, Altincicek *et al.* 2008), differences in expression patterns based on environmental variables (In *et al.* 2005, Kalujnaia *et al.* 2007) or varying developmental stages (Pascual *et al.* 2007, Luckenbach *et al.* 2008) or differences between two different tissue types (Baek *et al.* 2009, Bychenko *et al.* 2009) can all be studied using this polymerase chain reaction (PCR) based technique. Not only does this method provide vital information on which genes are up-regulated in certain biological situations, but it also ensures that the sequences which are identified from sequencing a cDNA library have more biological significance to the problem at hand, as only genes that are being actively transcribed at differing expression levels will be present in the library. By reducing and eliminating commonly expressed genes, the overall size and frequency of a cDNA library is reduced, saving time and reducing the complexity of sequence analysis.

SSH was recently applied to an investigation of the hypobranchial gland of a marine gastropod *Dicathais orbita* (Chapters 3 & 4). The hypobranchial gland is a complex biosynthetic organ (Westley *et al.* 2010b) responsible for the production of the ancient pigment compound, Tyrian purple (Baker & Sutherland 1968), in the family Muricidae (Neogastropoda; Mollusca). Studies into the bioactivity and functional role of Tyrian purple have identified that chemical precursors to the final pigmented compound have anticancer and antibacterial activity (Benkendorff *et al.* 2000, Benkendorff *et al.* 2011). These brominated indole precursors are derived from tryptophan by a number of biosynthetic enzymes, including bromoperoxidase and arylsulfatase which show activity in the hypobranchial (Baker & Sutherland 1968, Jannun & Coe 1987, Westley *et al.* 2006, Westley & Benkendorff 2009, Westley *et al.* 2010b). In the process of annotating *D. orbita* sequences in this SSH library (Chapter 4), a subset of the resulting EST sequences showed homology to ciliate protozoan genes. These included some ciliate ribosomal genes and a range of other protein coding genes.

Ciliate protozoa are protists belonging to the phylum Ciliophora in the superphylum Alveolata that are characterized by three things; the presence of cilia; nuclear dimorphism; and conjugation as a sexual process (Lynn 2008). Within this classification, ciliates can vary greatly

in size and shape. They have a global distribution, a wide variety of feeding behaviours and are often found as symbionts in a number of different organisms. Ciliates have been identified with apparently commensal relationships e.g. existing within the gills of freshwater fish (Hofer *et al.* 2005), parasitic relationships e.g. in the case of the human pathogen *Balantidium coli* (Schuster & Ramirez-Avila 2008) or in the rather complex mutualistic relationships e.g. between ruminants and the ciliates that live within their rumen, who in turn host symbiotic methanogens responsible for the breakdown of cellulose (Regensbogenova *et al.* 2004). Several mollusc species are known to engage in symbiotic relationships with ciliate protozoans, including the zebra mussel, who hosts multiple extracellular ciliates within their ctenidium (gills) and mantle cavity (Laruelle *et al.* 1999, Conn *et al.* 2008), as well as on the gills of limpet species (Van As *et al.* 1999), the shells of Ampullarid gastropods (Dias *et al.* 2008), the eyes of scallops (Harry 1977), the renal glands of cephalopods (Furuya *et al.* 2004) and both intracellularly and extracellularly in gills and digestive organs of several oyster species (Elston *et al.* 1999, Winstead *et al.* 2004, Spiers *et al.* 2008). To date, there has been no reported relationship between the hypobranchial gland of molluscs and ciliate protozoa.

This paper investigates a subset of EST sequences identified from the hypobranchial gland of the predatory whelk *Dicathais orbita* that are translated with an alternate codon translation system, and share sequence identity and similarity with known ciliate protozoan genes. A phylogenetic analysis was undertaken based on the rRNA genes showing significant matches to protozoans. Histological analysis of the hypobranchial gland of *D. orbita* was also performed to confirm that ciliate protozoans are present within this gland. This paper identifies a previously undocumented symbiotic relationship between ciliate protozoa and the hypobranchial gland of a marine gastropod, and highlights a novel use of SSH in the study of endosymbiosis.

5.3 Materials and methods

5.3.1 Creation of subtracted cDNA library, plasmid isolation and sequencing.

D. orbita specimens used in the transcriptomics analysis were collected in May 2005 prior to the reproductive season (see (Westley *et al.* 2010a). Tissue collection, RNA extraction, cDNA synthesis, SSH and cDNA library sequencing were performed as described in Chapter 4.

5.3.2 Bioinformatics and functional analysis

A non-redundant set of EST sequences was produced containing 554 genes expressed in the hypobranchial gland of *D. orbita*, as described in chapter 4. Perl scripts were used to manipulate the formatting of sequences before tBLASTx analysis was performed on each EST. Preliminary tBLASTx analysis identified 153 EST sequences showing significant amino acid identity to known protein coding sequences within the NCBI database. Several sequences had hits to known proteins, however the translated query sequences contained multiple stop codons when translated using the standard nuclear translation code. Manual sequence analysis was performed and 45 sequences were identified that contained multiple stop codons when compared to known proteins in other organisms. In addition, four ribosomal RNA sequences with conserved homology to ciliate ribosomal RNA genes were identified and included in this sequence cohort. Codon translation system six (the Ciliate, Dasycladacean and Hexamita nuclear code) was used to observe whether the 45 selected EST's formed full length translatable sequences In addition the 256 EST's which showed no similarity to NCBI sequences were retranslated using the codon translation system six, and an open reading frame was identified in 12 of these sequences that was not present previously. These 12 sequences were combined with the 45 previously identified sequences (57 in total) and resubmitted to BLAST2GO annotation, this time using codon translation system six.

BLAST2GO assigned Gene Ontology defined classifications (Table 5.1) using the National Centre for Biotechnology Information (NCBI) BLAST server, tBLASTx analysis was performed using the default parameters, except for an e value cut-off of $1e^{-03}$ and the conversion of the annotation to GOSlim view. For each sequence, the top 20 BLAST hits were analysed, and an e value threshold was set, discounting all BLASTx matches with e values greater than $1e^{-03}$. Furthermore, sequences less than 100 nucleotides were eliminated from this study in initial

analysis to maximise the accuracy of annotation. The biological processes, molecular functions and cellular component were assigned to sequences with significant homology, based on sequence homology matches and previously defined terms. KEGG biochemical pathways were assigned to annotated sequences using BLAST2GO.

5.3.3 Phylogenetic analysis of ciliate ribosomal RNA sequences

Two ribosomal RNA sequences showed significant BLASTn sequence matches to 18s ribosomal RNA sequences from ciliate protozoa, although these sequences originated from two separate regions of the 18s ribosomal RNA gene (Fig 5.1). These have been submitted to NCBI as GU199595 and GU199594. Ciliate 18s sequences that showed the closest matches to each sequence were collected from NCBI and aligned together using multiple sequence alignment program Clustalx using a Gap open penalty of 10 and a gap extension penalty of 0.2. For phylogenetic analysis of GU199594 the sequences listed in Table 5.2 were collected and aligned for analysis. A sequence from yeast species *Yarrowia lipolytica* AF156969.1 was used as an outgroup for maximum parsimony and Bayesian analysis. For the phylogenetic analysis of GU199595, ciliate sequences listed in Table 5.3 were collected and aligned. A sequence from the dinoflagellate *Prorocentrum lima* EF025381.1 was used as an outgroup in the multiple sequence alignment. The resulting multiple sequence alignments had flanking regions of sequences, not containing matches to GU199594 or GU199595, so these were removed from the alignment files using Maclade (Sinauer Associates, Sunderland MA, USA). Maximum parsimony analysis was conducted on both datasets in order to identify the phylogenetic relationships of GU199594 and GU199595 and identify what class of ciliate each sequence was derived from. A total of 100 random stepwise additions were performed in the MP analysis, conducted using PAUP b4.10 (Swofford 1999), holding 10 trees at each step and with tree bisection and reconnection for searching tree space. Node support was estimated using 100 bootstrap pseduoreplicates. We carried out Bayesian inference (BI) analysis on the both ciliate 18s ribosomal RNA datasets with MrBayes 3.1 (Ronquist & Huelsenbeck 2003). All BI analyses were run with default priors as follows: rate matrix: 0-100, branch lengths: 0-10, Gamma shape: 0-1. We ran two parallel BI runs, each comprising one cold and five heated chains. Markov chains were started from random trees and all chains ran simultaneously with a tree saved every 1000 generations for 1,500,000,000 generations. The first 150,000 trees were discarded as burn-in, after examination of log likelihood plots to see when stationary was reached using Tracer v1.4 (Rambaut & Drummond 2007). The resulting 1,350,000 post-burnin trees were used to calculate the posterior probabilities (PP) for each clade.

5.3.4 Histology of the hypobranchial gland to confirm symbionts

Triplicate females representing each of four reproductive phases (pre-copulatory, copulating, egg laying, post-reproductive (as described in Westley *et al* 2010a)) were sampled from the rocky intertidal reefs along on the Fleurieu peninsula of South Australia, Australia during 2005-2006. The shell of each *D. orbita* specimen was removed by cracking with a vice at the junction of the primary body whorl and spire, and the soft body was removed by severing the columnar muscle. The soft body was then transferred to a dissecting tray and submersed in filtered (0.22μm) seawater to reduce osmotic stress. The mantle, complete with the hypobranchial gland and rectal gland were separated from the visceral mass by incisions along the lateral margins of the columnar muscle, the gill and digestive gland. Tissue was fixed in 10% neutral buffered formalin for 6 hrs, dehydrated through an ethanol series, cleared in chloroform and embedded in paraffin. Nine transverse sections (5μm) of the mantle and hypobranchial gland from all 12 specimens were examined for the presence of extracellular and intracellular ciliates. Total intracellular ciliate abundance was quantified within each section after staining with Modified Harris haematoxylin and Eosin Y with Phloxine B (Thompson 1966). Sections were also stained with Periodic Acid Schiff (McManus 1946) and Toluidine Blue (Kramer & Windrum 1954) to allow description of ciliate biochemistry. Sections were examined at 1000x magnification under a compound light microscope (Zeiss, Axio Imager, A1) and images acquired with an AxioCam 1Cc digital camera (Zeiss) and AxioVision Release 4.8 software (Zeiss, 2006-2009).

Ciliates were first identified by Dr Judith Handingler (Fish Health Unit, Tasmanian Dept. Primary industries) and confirmed by Veterinary Pathologist Dr Celia Hooper (Gribbles Pathology, Victoria). Ciliates were identified based on their uniform ovoid shape, with a length ranging between 15-24 μm, a width of between 9-15 μm and a large stained nucleus. Total ciliate abundance was determined and statistical analyses were conducted in PASW Statistics 18 (SPSS) for Microsoft Windows. Significant differences in ciliate abundance as a function of reproductive state were determined using a non-parametric Kruskal-Wallis and Mann-Whitney pair-wise analyses as transformation failed to meet the assumptions of the Levene Statistic for homogeneous variances. Significance was determined at $\alpha = 0.008$ after Bonferroni adjustment.

5.4 Results

5.4.1 Sequence analysis

When manual tBLASTx analysis was performed on this sequence cohort, many sequences when translated, while sharing amino acid sequence identity with a protein in NCBI, contained multiple stop codons interspersed at several locations in the alignment. The presence of a glutamine residue was found in the homologous protein sequence at 37% of sites where a stop codon was found in our query sequence (e.g. Figure 5.1). Individual analysis of the codon triplets for these stop codons identified that the triplet codon was either a TAA or TAG, and never TGA. Within ciliate codon translation, codon triplets TAA and TAG code for a glutamine residue and not a stop codon (Hoffman *et al.* 1995). When translated using the Ciliate, Dasycladacean and Hexamita nuclear code, these sequences now actually coded for complete peptide sequences, suggesting that these gene sequences were being expressed by a ciliate using alternate codon translation. Ciliate, Dasycladacean and Hexamita nuclear code (Codon translation Table 6) was therefore used in the analysis of all 45 ciliate protein coding genes in order to observe the full length translations from our cDNA library.

A total of 57 ciliate sequences were identified from our cDNA library which when translated showed amino acid identity to known ciliate sequences and/or used the ciliate alternate codon translation system, and these sequences were lodged to GenBank as ciliate sequences (Table 5.1). In addition, four ciliate rRNA sequences were identified (Table 5.1) and used in a phylogenetic analysis. Upon assembly and analysis of our 57 nuclear protein coding sequences, a total of 45 individual ciliate protozoan nuclear protein coding sequences could be assigned function, while the remaining 12 sequences showed no homology to any other sequence in GenBank and therefore could not be annotated. Of the 45 remaining protein coding ciliate sequences, a total of 41 sequences were assigned Gene Ontology based functions (Table 5.1). In total 37 biological process classifications were made to our 45 identified ciliate sequences. The most common functions observed were genes involved in protein folding, microtubule based movement and protein polymerization, which were each assigned to four sequences (Fig 5.2). Genes involved in protein modification processes and embryonic development were also found, with three genes classified to be belong to each of these biological processes (Fig 5.2). Two genes were identified to be involved in nucleosome assembly, two more genes in ribosome biogenesis and assembly, and two more genes were

classified to be involved in nematode larval development. 11 more biological process classifications were made on genes which were involved in responses to light heat intensity, DNA repair, response to hydrogen peroxidase and positive regulation of growth rates (Fig 5.2). None of the ciliate genes were identified as enzymes that could contribute to biosynthesis of Tyrian purple brominated indole precursors (e.g. bromoperoxidase or arylsulfatase).

5.4.2 Phylogenetic analysis of ciliate 18s ribosomal RNA sequences

Initial BLASTn analysis of both GU199594 and GU199595 identified that these sequences showed the greatest amount of similarity with ciliate 18s ribosomal RNA sequences, confirming that these transcripts originated from ciliate protozoans. BLASTn matches for GU199594 returned e values between zero and 2e-80 and an average maximum identity of 78%, while BLASTn matches for GU199595 returned e values between $5e^{-61}$ to $4e^{-44}$ and an avege maximum identity of 79% (data not shown).

Phylogenetic analysis of the 18s ribosomal RNA sequence GU199594 indicated that this 18s ribosomal RNA sequence falls within the clade of ciliates of the Class Phyllopharyngea with a Bayesian posterior probability (PP) of 1.00 and a maximum parsimony (MP) bootstrap support value of 85 (Fig 5.3A). Ciliates belonging to the class Oligohymenophera form a monophyletic clade with a MP bootstrap support value of 80 and a PP of 0.73, a sister clade to the Phyllopharangea (PP=1.0, MP bootstrap support value=50). The 13 ciliates belonging to the class Litostomatea form a monophyletic clade supported with a MP bootstrap value of 100 and Bayesian PP of 1.00. Ciliate species belonging to the class Heterotrichia formed a monophyletic clade with a MP bootstrap value of 100 and a Bayesian PP of 1.00. The Phyllopharyngea were sister taxa to members of the Oligohymenophera, with MP bootstrap support of 50 and Bayesian PP value of 1.00. Protostomate ciliates were polyphyletic, with Protocruzia sp. (AF194409) appearing to be sister taxa to a larger clade containing the remaining Protostoma, Phyllopharyngea, Oligohymenophora, Litostomatea and solitary member of the Spirotrichea (MP bootstrap support values for this clade was 50 with a Bayesian PP fbootstrap value of 20).

In contrast to the phylogenetic analysis of GU199594 described above, the phylogenetic analysis of GU199595 indicated our ciliate sequence did not group with any of the included ciliate classes (Fig 5.3B). Our study only contained single representatives from the Litostomatea and Protostomatea and these two individuals formed a monophyletic clade

supported by a Bayesian PP value of 0.97. This was sister clade to a monophyletic grouping of all representative members of the Spirotrichea supported by a PP of 0.97 and MP bootstrap value of 98. All included Oligohymenophorea species formed a monophyletic clade with a PP of 0.56 and bootstrap support value of 77, with the exception of the Oligohymenophoran *A. marinum*, which was the most basal of all ciliate species observed. It was with this large Oligohymenophora clade and the Spirotrichea/Litostomatea/Protostomatea clade that our unidentified ciliate sequence formed a three point polytomy, supported by a PP of 0.82. All members of the Nassophora formed a monophyletic clade (MP bootstrap value=100, PP=1.00) with both representative members of the Phyllopharangea (MP boostrap value=100, PP=1.00) with a PP of 1.00 and a MP bootstrap support value of 90. Bayesian inference could not differentiate between this larger Nassophora/Phyllopharangea clade, the solitary representative of the Protostomatea and with the large polytomy described above.

5.4.3 Histology of the hypobranchial gland

Intracellular ciliates were observed within the basal cytoplasm of hypobranchial gland epithelial cells (Fig 5.4) in all specimens examined (N = 12). Extracellular ciliates were observed on the epithelial surface of the hypobranchial gland in 75% of individuals and within the sub epithelial vascular sinus adjacent to the rectal gland (Fig 5.4B) in 33.3% of females. Ciliates were not observed in association with the mantle epithelium. Many intracellular ciliates were observed to be undergoing mitosis (Fig 5.4B). The presence of ciliates within the intracellular space of hypobranchial gland epithelial cells did not appear to induce pathological effects, as the secretory function of inhabited cells was maintained (Fig. 5.4A; refer to Westley *et al.* , 2010b). However, cell membrane integrity was partially compromised when ciliate abundance in adjacent cells was comparatively high (data not shown).

Mean total intracellular abundance was greatest in copulating individuals (86.89 ± 15.27), followed by post-reproductive (23.52 ± 5.61), egg-laying (13.52 ±4.57) and pre-copulatory (1.85 ± 0.78) females (Fig. 5.5). Mean (± S.E) intracellular ciliate abundance varied significantly between all reproductive phases ($P<0.008$), except egg laying and post-reproductive individuals ($P=0.309$).

5.5 Discussion

This study has uncovered an undocumented use for SSH, the potential identification of endosymbiosis or multiple genomes within tissue samples. By utilizing ribosomal RNA sequences, typically discarded from cDNA libraries before annotation and before sequences are submitted to GenBank, we were able to identify the presence of at least one ciliate protozoan within the hypobranchial gland of *D. orbita*. These findings were supported by the presence of multiple nuclear protein coding genes that were expressed in our cDNA library that showed amino acid sequence identity and similarity to proteins from other ciliates and that used the Ciliate, Dasycladacean and Hexamita alternate codon translation system in order to produce full length proteins (Fig 5.1). Histological analysis of the hypobranchial gland of the marine snail *D. orbita* confirmed the presence of intracellular and extracellular ciliate protozoa, supporting our phylogenetic analysis and sequence annotation. This new use for SSH may find broader application in the growing field of symbiosis and has implications for the analysis of future and past studies from multicellular organisms that may have overlooked the possibility of multiple genomes in their gene libraries.

In addition to being the first recorded identification of protists living in association with the hypobranchial gland of gastropod molluscs, this study highlights a potential new application of SSH. Typical use of SSH involves comparing gene expression of tissue samples under different disease or environmental conditions (Gestal *et al.* 2007, Hansen *et al.* 2007, Kalujnaia *et al.* 2007, Altincicek *et al.* 2008), or comparisons in gene expressions between two different tissue types or individuals (Pascual *et al.* 2007, Luckenbach *et al.* 2008, Yuan *et al.* 2008). This is the first reported instance where SSH technology has been used to identify the presence of multiple transcriptomes in a tissue sample. Our histological analysis agrees with the SSH finding confirming that no ciliates were present in the mantle tissue of *D. orbita*, while they were abundant in the hypobranchial gland of the snail and thus amplified using the SSH approach. Our mantle subtracted hypobranchial gland cDNA library amplified RNA transcripts from both the snail host and the ciliates present in the tissue, with approx 30% representing ciliate protozoans (Chapter 4). The key sequence characteristic that allowed us to identify ciliate gene expression was their use of an alternate codon translation system (Fig 5.1). Differentiation of at least two transcriptomes from our cDNA library was possible through bioinformatic analysis, where the taxonomy of sequence homology matches, as well as the presence of alternate codon usage, enabled the separation of *D. orbita* and ciliate protozoan genes.

It may prove more challenging to differentiate between transcriptomes in tissues containing more than one genome source if both/all organisms present use a single translation system, but this doesn't mean that differentiation between the genomes would not be possible. Provided there is adequate genome information from a closely related species in a cDNA library made from tissue containing two or more organisms, it should be possible to identify and separate each transcriptome using this PCR based technique. A recent study, which sequenced the paired symbiont-host transcriptome of the anemone *Aiptasia pallida* and its dinoflagellate endosymbiont reported that as little as 1.6% of all ESTs in this combined library could be attributed to the endosymbiont (Sunagawa *et al.* 2009). By comparison, 57 of the 352 sequences from the hypobranchial gland SSH library have been shown to be expressed by symbiotic ciliate protozoans in *D. orbita* which is 16% of the total number of sequences produced in this study. If the complete transcriptome of the hypobranchial gland of *D. orbita* was sequenced, it is likely that symbiont expression levels would be less abundant in the transcriptome compared to the SSH library. There have recently been advances in transcriptomics studies, thanks to the application of next generation sequencing techniques, where whole transcriptome comparisons have been performed on different species or tissues (Kinoshita *et al.* 2011, Waegele *et al.* 2011) or transcriptomes from the same species under different environmental or physiological conditions (Pantzartzi *et al.* 2010, Heyland *et al.* 2011) have been compared using computational analysis. While it is possible that high throughput sequencing and computational analysis might also be used to identify symbiont gene sequences in tissue samples, as was attempted in a study of algal plastid acquisition in sea slugs (Waegele *et al.* 2011), the use of SSH to enrich for differentially expressed genes may reduce the complexity of the final dataset, making analysis of data more straightforward.

One other feature of our cDNA library that strongly supports our argument on the presence of multiple transcriptomes, was the presence of ribosomal RNA sequences in our resulting sequence data. Despite the fact that theoretically all ribosomal RNA should have been excluded from the subtractive hybridization process as they are not translated genes and do not contain a polyA tail in their nucleotide sequence (Gonzalez & Sylvester 1997), they were still present in our subtracted cDNA library. When ribosomal RNA sequences are amplified in a cDNA library, they are typically removed from the dataset as it is believed that they don't represent the expression of genes in a sample (Lockyer *et al.* 2007). In fact, it is a requirement of submission to GenBank's EST database that mitochondrial genes and rRNA are excluded

from sequence submissions (NCBI 2009). However, the removal of these sequences can result in the loss of vital information regarding the origin of transcripts. At least 60% of all RNA can be attributed to ribosomal RNA molecules (Ide *et al.* 2010) and as little as 0.5-1.5 % of the total RNA in cells pertains to polyA containing expressed genes that result in a gene product being produced (Murillo *et al.* 1995). In some cases it is understandable that this abundance of ribosomal RNA can detract from studying the expressed genes in a sample. However, the current study highlights the importance of analysing ribosomal RNA sequences present in cDNA libraries in order to identify the presence of multiple transcriptomes, particularly if symbiosis is a possible factor in the system under investigation. While we were able to identify the presence of ciliate genes based on alternate codon translation of our EST sequences, we would not have discovered the presence of two different ciliates in the hypobranchial gland that are responsible for the expression of these genes, without sequence homology matches and phylogenetic analysis of the ribosomal RNA sequences. A literature search using ISI Web of Science on the original SSH method paper (Diatchenko *et al.* 1999) identified 1992 articles that have cited this paper (search performed 23st October 2011). Because of the popularity of this transcriptomic technique and its application in the scientific community, it is imperative that every possible variable in the resulting datasets is accounted for. By including an analysis of ribosomal RNA sequences produced when this method is used, we can be sure to account for mixed cDNA populations. With the decreased price in sequencing costs, this method may become a valid technique to identify specific parasites or symbionts in a given tissue.

Phylogenetic analysis of the ciliate ribosomal RNA sequences suggests the ciliate sequences identified belonged to a member of the Class Phyllopharangea. When looking at sequence homology and taxonomic analysis of the 18s rRNA sequence GU199594 (Fig 5.3A), our *D. orbita* ciliate 18s rRNA sequence groups exclusively with members of the Phyllopharyngean subclass Suctoria. Suctorian ciliates are epibionts that are found on a variety of metazoans and prey specifically on other ciliates (Lynn 2008). Members of the Suctoria can be found on numerous host organisms, from crabs (Fernandez-Leborans *et al.* 2002) to horses (Sundermann & Paulin 1981) and can be both free-living or symbiotic. Suctorian ciliates have been shown to colonize the external bodies of crustaceans (Fernandez-Leborans *et al.* 2002, Fernandez-Leborans *et al.* 2003) and the gills of the Arctic char *Salvelinus alpinus* (Hofer *et al.* 2005). Suctorian ciliates have been identified as epibionts on the shells of molluscs, feeding on other ciliate symbionts (Dias *et al.* 2006, Dias *et al.* 2008). As our samples were taken from the hypobranchial gland tissue, these ciliates must be acting as endosymbionts in *D. orbita* rather

than typical epibionts. Both intra- and extra-cellular ciliates were identified in the hypobranchial gland by histology. It is therefore possible that these Suctoria opportunistically enter the mantle cavity of the mollusc, where they interact with the surface of the hypobranchial gland and encounter prey in the form of other ciliates.

It is unclear from our analysis of the second 18s ribosomal RNA sequence which class of ciliate protozoan this sequence belongs to (Fig 5.3B). Bayesian analysis failed to differentiate between our unidentified sequence and members of the classes Litostomatea, Armophorea, Oligohymenophorea and Spirotrichea. The monophyly of several ciliate classes has been previously confirmed by 18s and 28s ribosomal RNA, including the Phyllopharyngea, Oligohymenophoroa, Litostomatea, Prostomatea, Spirotrichea and Heterotrichea (Barointourancheau *et al.* 1992, Lynn 2003, Snoeyenbos-West *et al.* 2004, Lynn 2008, Gong *et al.* 2009, Zhan *et al.* 2009). Armophoreans are free swimming ciliates that often live in the benthic or planktonic habitats, but can also be endosymbionts commonly found in the digestive systems of a wide variety of invertebrates (Hackstein & Stumm 1994, Lynn 2008). These ciliates typically feed on bacteria and many are known to harbour methanogens (methane producing bacteria) (van Hoek *et al.* 2000). This is of particular interest given that the hypobranchial gland of female *D. orbita* provides a reducing environment with mercaptans, including dimethyldisulphide involved in the formation of Tyrian purple precursors (Westley & Benkendorff 2008). Litostomateans are predatory ciliates that that feed on flagellates and other ciliates and are endosymbionts that typically live in vertebrate metazoans (Lynn 2008). Members of the Oligyhymenophorea are a diverse group of symbiotic and parasitic protozoans that have been identified in the gills of several bivalves (Rayyan *et al.* 2006), as well as the renal gland of commercial land snails (Segade *et al.* 2009). Several different types of spirotrich ciliates have been found living in marine (Rayyan *et al.* 2006) and freshwater environments (Snoeyenbos-West *et al.* 2002), although there have been no recorded instances of their presences within marine molluscs. It is impossible to predict from this study what specific class our 18s sequence originates from, as it does not provide enough variation in order to differentiate between ciliate taxa. It is possible that it belongs to a member of Litostomatea, Armophorea, Oligohymenophorea or Spirotrichea, or it may even belong to another ciliate class. When BLASTn analysis was performed, the nucleotide sequence similarity between this sequence and the resulting matches to other ciliate 18s sequences is strong evidence that GU199595 originated from a ciliate protozoan. As this sequence doesn't overlap with GU199594, it is even possible that both 18s sequences presented in this phylogenetic analysis

belong to the same individual, although if this was the case, we would expect GU199595 to group with other Phyllopharyngea and not remain unresolved (see Fig 5.3 B). The presence of two different ciliate 18s ribosomal RNA genes within our cDNA library as well as the likelihood that the GU199594 sequence belongs to a ciliate that preys exclusively on other protists suggests that the hypobranchial gland of *D. orbita* supports an interesting microcommunity of unicellular life.

Sequence analysis of the ciliate genes provides insight into the possible biological processes being performed by the ciliates in the hypobranchial gland of *D. orbita*. Alpha and beta tubulin genes were identified within our ciliate EST dataset, and these genes are involved primarily in the formation of cilia and flagella in cells (Gull 2001). The presence of cilia is an important diagnostic and functional characteristic of all members of the ciliate protozoans, and their uses are numerous and varied (Lynn 2008). Our study suggests that the hypobranchial gland is conducive to ciliate replication as histological analysis identified many ciliates undergoing cell division within the hypobranchial gland epithelium (Fig 5.4 A). The high level of conservation between our polyubiquitin gene sequence (GD253775) and the known ciliate protozoan polyubiquitun further supports our finding that ciliate protozoan genes are being expressed in our *D. orbita* EST library. The expression of both the polyubiquitin gene (GD253775) and the ubiquitin conjugating enzyme (GD253720) indicates that ubiquitinization may be occuring in the ciliates (Sorokin *et al.* 2009). Ubiquitinization is involved in many cellular processes, including protein degradation (Hershko & Ciechanover 1998), apoptosis (Broemer & Meier 2009), DNA repair (Pickart 2002) and cellular signalling (Kirkin & Dikic 2007). The stressed induced protein STI1, which was identified from our ciliate sequences (GD253678), are known to form complexes with heatshock proteins in *Sacchryomyces cerevisiae* and effect the ligand binding specificity of interacting proteins when target proteins are folding (Chang *et al.* 1997). Furthermore, the HSP90 gene (GD253911) is involved in a variety of cellular processes including the folding and unfolding of proteins, assembly of multi-subunit complexes, buffering the expression of mutations and are invlolved in thermotolerance (Srivastava 2002). The expression of genes involved in essential cellular processes, combined with histological evidence of cellular replication, suggests that ciliates are functioning normally in a symbiotic relationship within the hypobranchial gland of *D. orbita*.

The location of the ciliates on the epithelial surface of the mantle cavity and the accumulation of ciliates within the hypobranchial gland in histological analysis suggests that the ciliates are

introduced to the mantle cavity through the inhalant stream and are immobilised in the hypobranchial gland. While there is little evidence to suggest that the ciliate symbionts identified from phylogenetic analysis pose a threat to the well being of the host, it was noted that the integrity of lateral cell membranes was partially compromised when neighbouring cells with comparatively high ciliate abundances were observed. However, the secretory function of these biosynthetic hypobranchial gland cells appears to be maintained (Westley *et al.* 2010b). It is therefore more likely that the cell membrane integrity was compromised during sectioning due to the changes in tissue density as a result of the high ciliate abundance. Overall, there is no evidence to suggest that the ciliates are endoparasitic within the hypobranchial glands of *D. orbita*. It seems more likely that these ciliates form a commensal symbiotic relationship. Of all of the sequences identified from our ciliates, only Dynein I (GD253871) is suggested to be involved in pathogenesis in the ciliate parasite *Ichthyophthirius multifiliis* (Cassidy-Hanley *et al.* 2011). However, dynein has several roles in cells, including flaggelar and ciliary beating as well as intracellular transport of organelles (Holzbaur & Vallee 1994), so it is not possible to determine whether the expression of Dynein I in hypobranchial gland ciliates is contributing to host pathogenesis. Further investigations into the role of this ciliate within the hypobranchial gland of *D. orbita* will be required in order to determine the nature of their symbiosis.

A histochemical study into the seasonal variability of biosynthetic activity within the hypobranchial gland of *D. orbita* identified an increase in the production of Tyrian purple precursor tyrindoxyl sulfate in post-reproductive and pre-copulatory females (Westley 2008). This correlates with an increase in arylsulfatase activity, an enzyme involved in modifying tyrindoxyl sulfate into the bioactive precursors of Tyrian purple, in post-reproductive and pre-copulatory females (Westley 2008). As the highest number of ciliates were found during copulation, rather than post or pre-copulation, it seems unlikely that ciliate protozoans are contributing directly to the biosynthesis of Tyrian purple precursors. However, it remains possible that their reduced abundance within post and pre-reproductive females is a consequence of increased levels of precursors within the gland at these stages of the reproductive cycle. Tyrian purple precursors display cytotoxic and antibacterial activity that has been reported to be involved in the protection of embryos within the egg capsule gland of muricids (Benkendorff *et al.* 2000, Benkendorff *et al.* 2001b). Furthermore the production of these bioactive compounds within the hypobranchial gland may facilitate the destruction and disposal of respiratory born pathogens trapped in epithelial secretions of adults (Westley *et al.* 2006). The cytotoxicity of these brominated indoles against a wide range of mammalian cell

lines (Benkendorff *et al.* 2011) suggests that they may also be effective against single celled eukaryotes, such as the intracellular ciliates. The use of indole compounds has been suggested in the treatment of the protozoan-caused human disease leishmaniasis (Mishra *et al.* 2009), and further research on the anti-protozoan activity of these brominated indoles would be of particular interest. Overall, the increase in Tyrian purple precursors may be either directly affecting the abundance of ciliate protozoans due to their toxicity, or indirectly affecting their abundance due to a reduction in available food sources as a result of the antibacterial activity associated with these compounds.

The abundance of ciliates in *D. orbita* individuals during and post-copulation is significantly higher than for pre-copulating individuals (Fig 5.3), which suggests ciliates may be transferred between individuals during copulation, or that copulation results in ideal conditions for ciliate colonization and reproduction. It is possible that any sperm that does not enter the ventral channel of the female gonoduct during copulation may get trapped on the surface of the hypobranchial gland, thus providing a food source to significantly boost ciliate numbers. However, sperm generally enters the ventral channel via the vaginal opening in *D. orbita*, where it is passed on to the albumen gland for fertilisation or storage in the seminal receptacles or the ingesting gland (Westley *et al.* 2010a). Post copulation, sperm can be stored for up to 10 months in *D. orbita* (Westely *et al.* 2010), which could sustain the slightly higher ciliate numbers post-copulation in comparison to the pre-copulatory numbers. The hypobranchial gland lies alongside the gonoduct in *D. orbita,* although there are no obvious ducts leading between the ingesting gland and the hypobranchial gland (Westley *et al.* 2010b). We are unaware of any previous studies that link the presence or abundance of intracellular ciliates to the reproductive cycles of marine gastropods. The infestation of the Mysid *Archaeomysis articulata* by the peretrich ciliate has been shown to be independent of seasonal changes (Hanamura 2000). A number of other freshwater studies have investigated the effects of seasonal variation, indicating that pH and salinity are more important determinants of ciliate populations (Fernandez-Leborans *et al.* 2002, Lei *et al.* 2009). The importance of reproductive cycles and environmental/seasonal cycles on the regulation of intracellular ciliate populations in marine hosts should be subjected to further investigation.

In conclusion, we have identified the presence of at least one, if not more, ciliate protozoans in the hypobranchial gland of the marine snail *Dicathais orbita*. SSH identified protein coding genes and ribosomal RNA sequences that belonged to ciliate protozoans. 18s ribosomal RNA

phylogenetic analysis confirmed the presence of a phyllopharangyean protozoan, whereas analysis of a second 18s ribosomal RNA sequence failed to identify which taxonomic class this ciliate sequence belonged to. Histological analysis confirmed the presence of ciliates intracellularly and on the epithelial surface of the hypobranchial gland. These results improve and advance our knowledge of host-endosymbiont relationships in gastropod molluscs, and show links between the reproductive cycle and the presence of intracellular ciliates in *D. orbita*. Furthermore this study highlights a potential new application of the SSH technique, toward the discovery and identification of multiple genomes in biological tissue samples.

5.6 Acknowledgements

This work was funded by an anonymous philanthropic foundation. Patrick Laffy was supported firstly by a Flinders University Faculty of Science and Engineering Research scholarship, followed by a Flinders University Postgraduate Research scholarship. All histological analysis and associated statistics in this manuscript was performed by Dr. Chantel Westley.

Table 5.1 Ciliate sequences identified within the hypobranchial gland of *D. orbita*. Full sequence annotation information can be obtained from GenBank using the corresponding Accession numbers

GenBank Accession Number	Gene Description	Genbank Accession Number	Gene Description
GU199594	18s ribosomal RNA	GD253714	L10A ribosomal protein
GU199595	18s ribosomal RNA	GD253689	Macronuclear PKG gene
GU199592	26s ribosomal RNA	GD253887	Melanocyte proliferating gene 1 MYG1
GU199593	26s ribosomal RNA	GD253865	Phosphotase 2A subunit
GD254014	Acetylcholinesterase	GD253775	Polyubiquitin
GD253832	Actin binding protein	GD253744	Proteasome beta 5 subunit
GD253915	Adenylyl cyclase	GD2537754	Ribonucleotide reductase
GD253857	Alphatubulin	GD253810	Ribosomal protein L11
GD253894	Alphatubulin	GD253858	Ribosomal protein L17A
GD253983	Alphatubulin	GD253682	Ribosomal protein L41
GD253713	ATPase	GD253937	Ribosomal protein S19
GD253693	Betatubulin	GD253952	Ribosomal protein S2
GD253801	Betatubulin	GD253987	Ribosomal protein L18A
GD253890	Betatubulin	GD253868	RNA polymerase 1
GD253913	Betatubulin	GD253806	T-complex protein 1
GD253991	Betatubulin	GD253728	Thymidine kinase
GD253755	Calcyclin binding protein	GD253818	Translation initiation factor IF-2
GD253989	cAMP dependant protein kinase	GD253745	Zinc finger protein delta transcription factor
GD253711	Cathepsin B	GD253678	Stress induced protein STI1
GD253966	DNA helicase	GD253720	Unknown
GD253871	Dynein	GD253730	Unknown
GD253778	Exonuclease	GD253776	Unknown
GD253850	Glucoamylase P GAMP	GD253803	Unknown
GD253813	GTP binding protein	GD256817	Unknown
GD253911	Heat shock protein 70 HSP70	GD253821	Unknown
GD253934	Heat shock protein 70 HSP70	GD253862	Unknown
GD253940	Heat shock protein 82 HSP82	GD253923	Unknown
GD253872	Heat shock protein cognate 5	GD253948	Unknown
GD253949	Heat shock protein cognate 5	GD253980	Unknown
GD254019	Histone H3	GE905837	Unknown
GD253784	Histone H4		

Table 5.2 Ciliate 18s ribosomal RNA sequences used in the phylogenetic analysis of GU199594

Class	Species	Accession number
Phyllopharangea	*Discophyra collini*	L26446
	Prodiscophryra sp.	AY331802
	Dysteria sp 1	AY331799
	Dysteria sp 2	AY331801
	Dysteria derouxi	AY378112
	Dysteria procera	DQ057347
	Isochona sp. 1	AY242116
	Chlamydodon triquetrus	AY331795
	Trithigmostoma steini	X71134
	Colepidae environmental sample	X76646
Litostomatea	*Macropodinium yalabense*	AF042486
	Epidinium ecaudatum caudatum	AM158474
	Metadinium medium	AM158464
	Eudiplodinium maggii	AM158452
	Diploplastron affine	AM158457
	Diplodinium dentatum	U57764
	Troglodytella abrassarti	EU680308
	Enchelydon sp.	U80313
	Chaenea vorax	DQ190461
	Amylovorax dogieli	AF298825
	Polycosta roundi	AF298819
	Ostracodinium gracile	AB535662
	Enoploplastron triloricatum	AM158462
Hererotrichia	*Condylostentor auriculatus*	DQ445605
	Condylstoma curva	EU379939
	Condylostoma spatiosum	DQ822483
	Blepharisma steini	AN713187
	Blepharisma americanum	AM713182
	Folliculina sp.	EU583992
	Stentor roeseli	AF357913
Prostomatea	*Coleps HBG2005 sp*.	DQ487194
	Coleps hirtus hirtus	AM292311
	Protocruzia sp.	AF194409
Oligohymenophorea	*Zoothamnium niveum*	DQ868350
	Urceolaria urechi	FJ499388
Spirotrichea	*Oxytricha longa*	AF508763
Outgroup (Yeast)	*Yarrowia lipolytica*	AF156969

Table 5.3 Ciliate 18s ribosomal RNA sequences used in the phylogenetic analysis of GU199595

Class	Species	Accession number
Spirotrichia	*Hypotrichidae sp.*	AF508778
	Stylonychia lemnaei	AF508773
	Oxytricha nova	X03948
	Stylongchia pustulata	X03947
	Pattersoniella vitiphila	AJ310495
	Cyrtongmena citrina	AF508755
	Neokeronopsis aurea	EU124669
	Paraurostyla weissei	AJ310485
	Hypotrichidae sp. 2	AF508763
	Oxytricha longa	AF508763
	Uroleptus gallina	AF508779
	Paraurostyla viridis	AF508766
	Urostyla grandis	AF508781
	Holosticha polystylata	AF508760
	Euplotidium arenarium	Y19166
	Gastrocirrhus monolifer	DQ864734
	Tintinnopsis beroidea	EF123709
Oligohymenophorea	*Trichodina sinonovaculae*	FJ499386
	Trichodina ruditapicis	FJ499385
	Trichodina nobilis	AY102172
	Trichodina hypsilepsis	EF524274
	Urceolaria urechi	FJ499388
	Frontonia sp.	AF255359
	Frontonia vernalis	U97110
	Paramecium putrinum	AF255360
	Paramecium bursaria	AB252002
	Frontonia didieri	DQ885986
	Anophytoides haemophila	U51554
	Pseudocohnilembus hargisi	AY212806
	Anoplophyra marylandensis	AY547546
	Aristerostoma marinum	EU264562
Nassophora	*Zosterodasys transverses*	EU286812
	Zosterodasys agamalievi	FJ008926
	Orthodonella apohamatus	DQ232761
	Orthodonella sp.	EU286809
Phyllopharangea	*Hartmannula derouxi*	AY378113
	Trichomodiella faurei	EU515792
Litostomatea	*Bitricha tasmaniensis*	AF298821
Armophorea	*Caenomorpha uniserialis*	U97108
Prostomatea	*Prorodon teres*	X71140
Outgroup (dinoflagellate)	*Prorocentrum lima*	EF025381

```
>gi|9756|emb|Z11763.1|OGALTUBG  O.granulifera gene for alpha-tubulin
Length=1672

 Score =  586 bits (1273),  Expect = 4e-179
 Identities = 243/261 (93%), Positives = 251/261 (96%), Gaps = 0/261 (0%)
 Frame = +3/+2

Query  84    LXKEIVDLCLDRIRKLADNCTGL*GFLVFNSVGGGTGSGLGSLLLERLSVDYGKKSKLGF  263
             + KEIVDLCLDRIRKLAD CTGL GFLVFNSVGGGTGSGLGSLLLERLSVDYGKKSKLGF
Sbjct  521   IGKEIVDLCLDRIRKLADQCTGLQGFLVFNSVGGGTGSGLGSLLLERLSVDYGKKSKLGF  700

Query  264   TVYPSP*ISNAIVEPYNSILSTHSLLEHTDVAVMLDNEAIYDICRRSLDIERPTYTNLNR  443
             TVYPSP*+S A+VEPYNS+LSTHSLLEHTDVAVMLDNEA+YDICRR+LDIERPTYTNLNR
Sbjct  701   TVYPSP*VSTAVVEPYNSVLSTHSLLEHTDVAVMLDNEAVYDICRRNLDIERPTYTNLNR  880

Query  444   LIS*VISSLTASLRFDGALNVDVTEF*TNLVPYPRIHFXLSSYGPIISAEKAYHE*LSVA  623
             LI+ VISSLTASLRFDGALNVDVTEF TNLVPYPRIHF LSSY P+ISAEKAYHE LSVA
Sbjct  881   LIAQVISSLTASLRFDGALNVDVTEFQTNLVPYPRIHFMLSSYAPVISAEKAYHEQLSVA  1060

Query  624   EITNSAFEPASMMVKCDPRHGKYMACCLMYRGDVVPKDVNAAVATIKTKRTIQFVDWCPT  803
             EITNSAFEPASMM KCDPRHGKYMACCLMYRGDVVPKDVNAAVATIKTKRTIQFVDWCPT
Sbjct  1061  EITNSAFEPASMMAKCDPRHGKYMACCLMYRGDVVPKDVNAAVATIKTKRTIQFVDWCPT  1240

Query  804   GFKCGINYQPPTVVPGGDLAK  866
             GFKCGINYQPPTVVPGGDLAK
Sbjct  1241  GFKCGINYQPPTVVPGGDLAK  1303
```

Figure 5.1 tBLASTx analysis of GD253983 EST sequence for alpha tubulin, using standard nuclear codon translation.

Sites where stop codons were identified within coding regions are highlighted in yellow. Note that a glutamine residue is present at the matching site in the subject sequence (Z11763 *Oxytrichia granulifera* sequence for alpha tubulin). The site in the matching subject sequence where a stop codon was present is highlighted in red. Further investigation into matching subject sequence (Z11763: *O. granulifera* sequence for alpha tubulin) identified the sequence is translated using Ciliate, Dasycladacean and Hexamita nuclear translation system (Translation code 6) and when correctly translated would code for a glutamine residue.

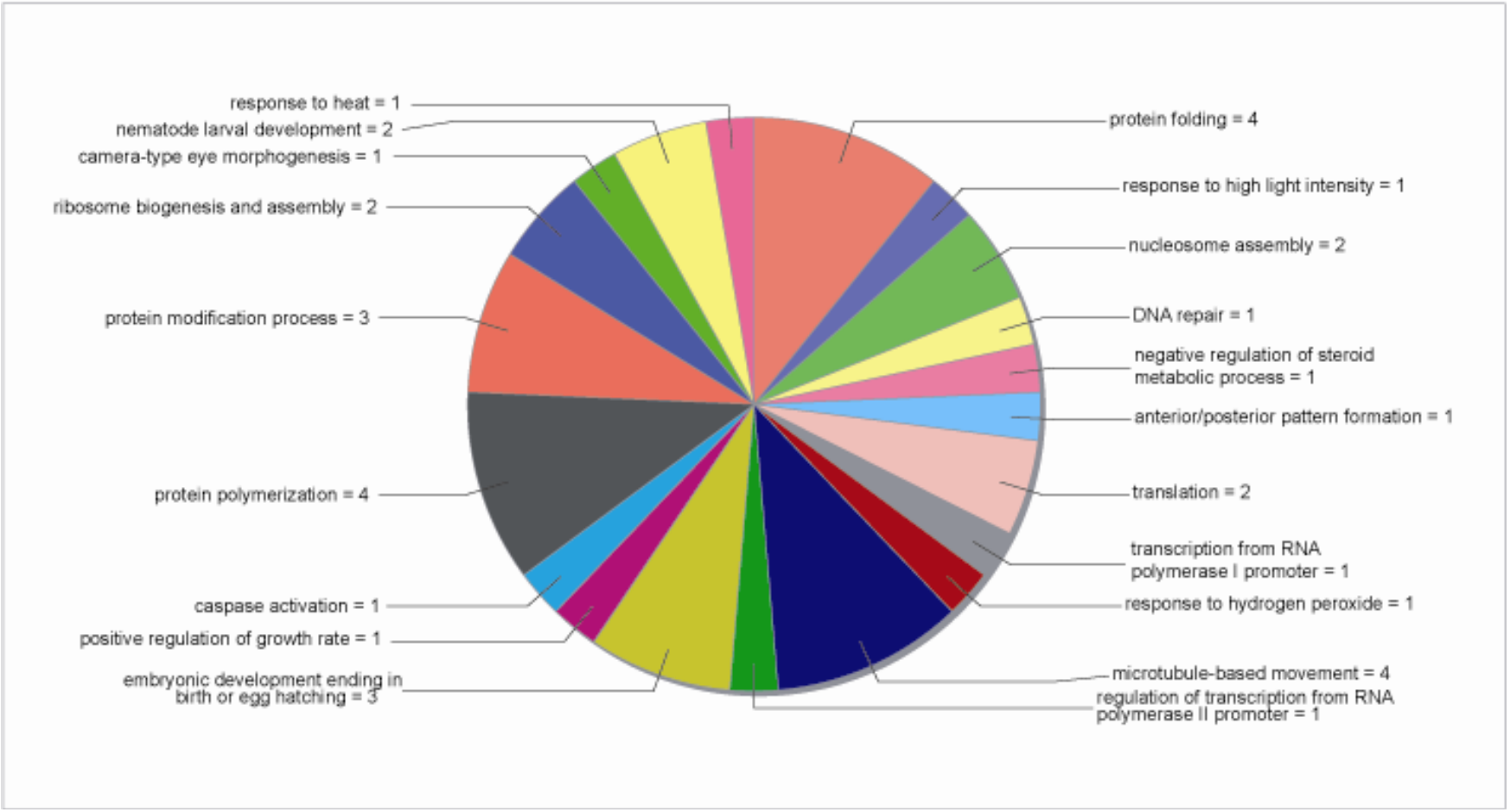


Figure 5.2 Biological process assignment (3rd level Gene ontology terms) of ciliate genes identified and expressed within the hypobranchial gland of *D. orbita.*

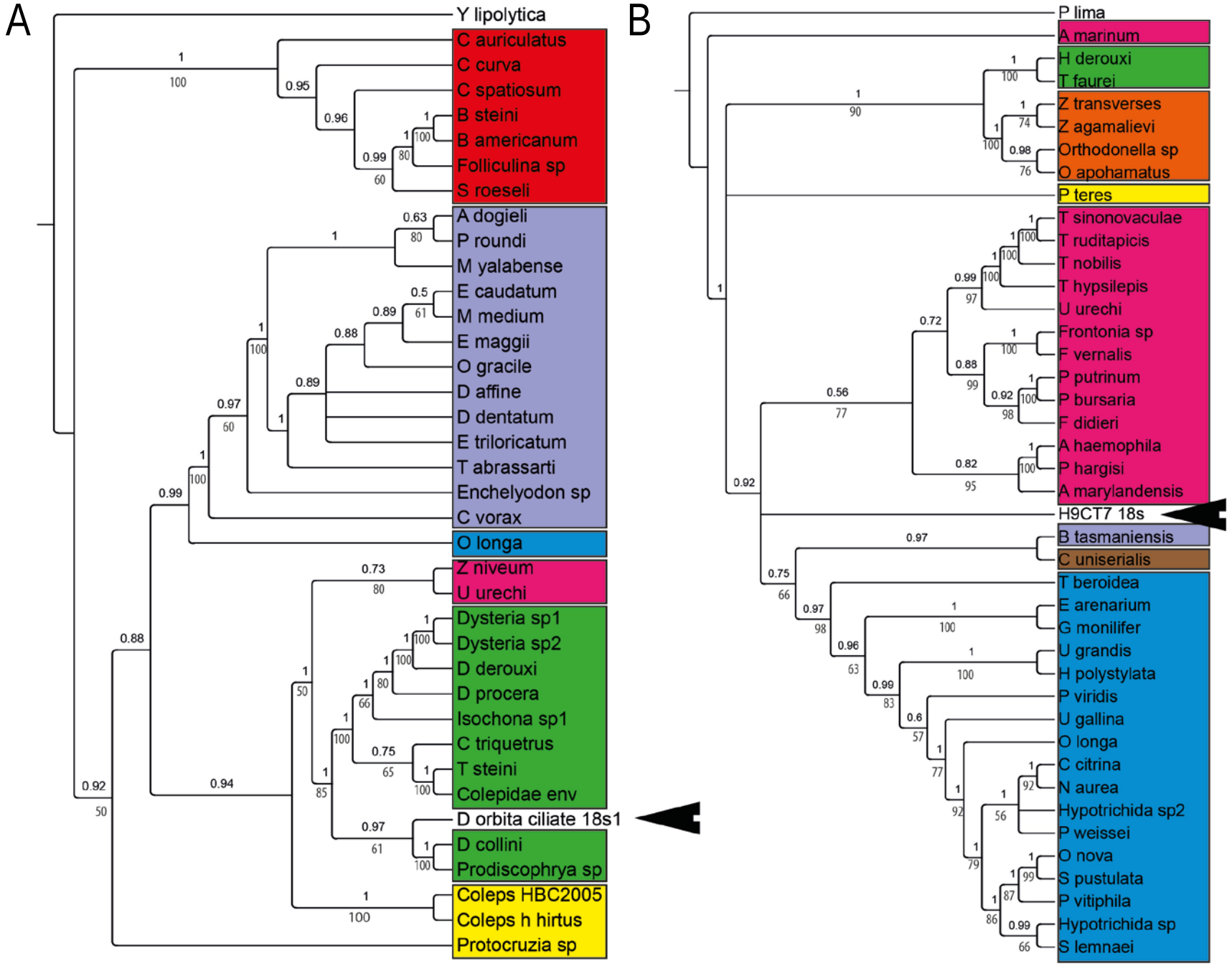

A
Y lipolytica
C auriculatus
C curva
C spatiosum
B steini
B americanum
Folliculina sp
S roeseli
A dogieli
P roundi
M yalabense
E caudatum
M medium
E maggii
O gracile
D affine
D dentatum
E triloricatum
T abrassarti
Enchelyodon sp
C vorax
O longa
Z niveum
U urechi
Dysteria sp1
Dysteria sp2
D derouxi
D procera
Isochona sp1
C triquetrus
T steini
Colepidae env
D orbita ciliate 18s1
D collini
Prodiscophrya sp
Coleps HBC2005
Coleps h hirtus
Protocruzia sp
B
P lima
A marinum
H derouxi
T faurei
Z transverses
Z agamalievi
Orthodonella sp
O apohamatus
P teres
T sinonovaculae
T ruditapicis
T nobilis
T hypsilepis
U urechi
Frontonia sp
F vernalis
P putrinum
P bursaria
F didieri
A haemophila
P hargisi
A marylandensis
H9CT7 18s
B tasmaniensis
C uniserialis
T beroidea
E arenarium
G monilifer
U grandis
H polystylata
P viridis
U gallina
O longa
C citrina
N aurea
Hypotrichida sp2
P weissei
O nova
S pustulata
P vitiphila
Hypotrichida sp
S lemnaei

Figure 5.3 Combined Bayesian and maximum parsimony phylogenetic analysis of ciliate protozoan 18s Ribosomal RNA sequences isolated from the hypobranchial gland of *D. orbita*.

Tree A represents phylogenetic analysis of *D. orbita* ciliate 18s #1 (GU199594). Tree B represents phylogenetic analysis of *D. orbita* ciliate 18s #2 (GU199595). *D. orbita* ciliate18s sequences are highlighted in each phylogenetic tree using a black arrowhead. Ciliate classes are labelled on each tree according to colour, including Phyllopharangea (Green), Oligohymenophorea (Pink), Litostomatea (Purple), Spirotrichea (Blue), Prostomatea (Yellow), Heterotrichea (Red), Amophorea (Brown), and Nassophora (Orange). Bayesian posterior probability values are listed above the line at each relevant node and maximum parsimony bootstrap values are included below the line at every site where they supported Bayesian topology. A 50% consensus majority rule was applied to both Bayesian and maximum parsimony analysis, and support values were only included when values were greater than 50%.

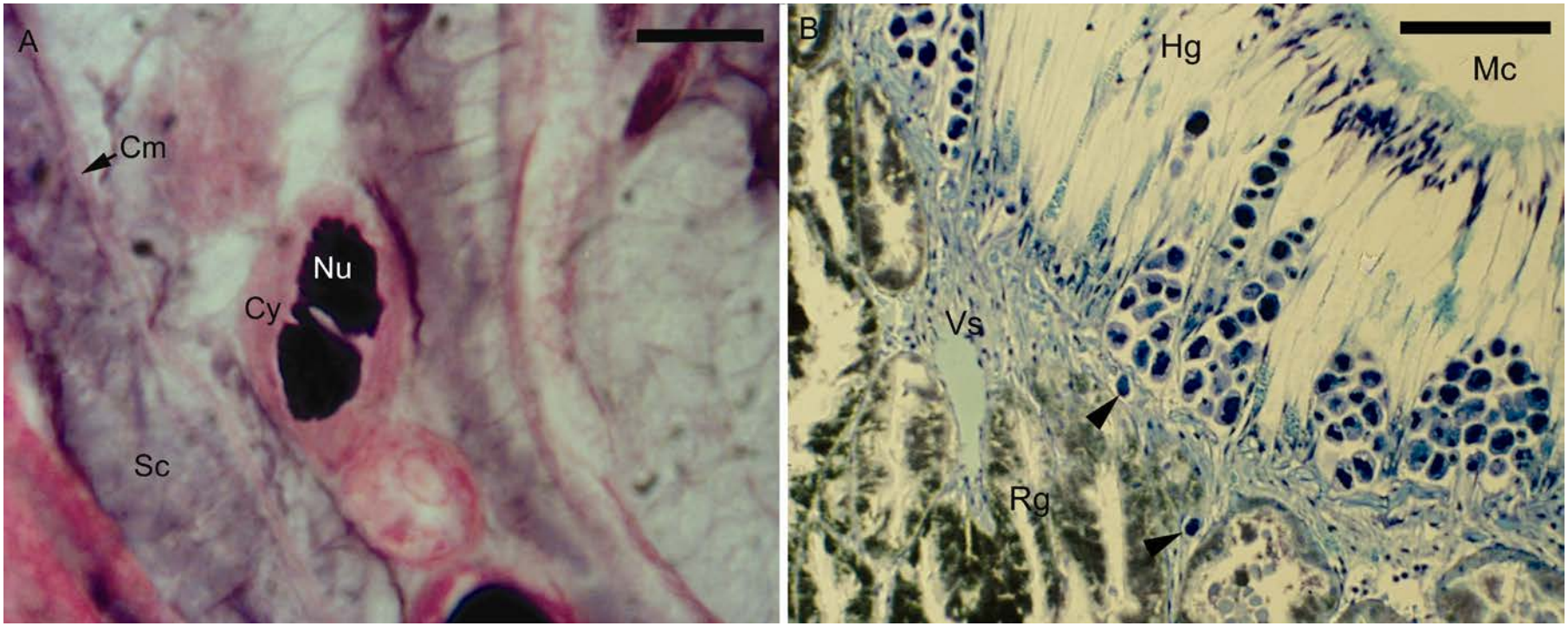


Figure 5.4 Histological analysis of the hypobranchial gland of *Dicathais orbita*. Transverse sections of (A) the hypobranchial gland (Hg) epithelium stained with H&E displaying intracellular ciliates undergoing mitotic division and (B) the subepithelium of the hypobranchial gland showing the presence of numerous ciliates in the subepithelium and a couple in the vascular sinus (Vs) (arrow heads) stained with Toluidine Blue.
Abbreviations: Cm, cell membrane; Cy, ciliate cytoplasm; Sc, secretion; MC, mantle cavity; Nu, ciliate nucleus; Rg, rectal gland. Scale bars = 10 μm (A) and 100 μm (B).

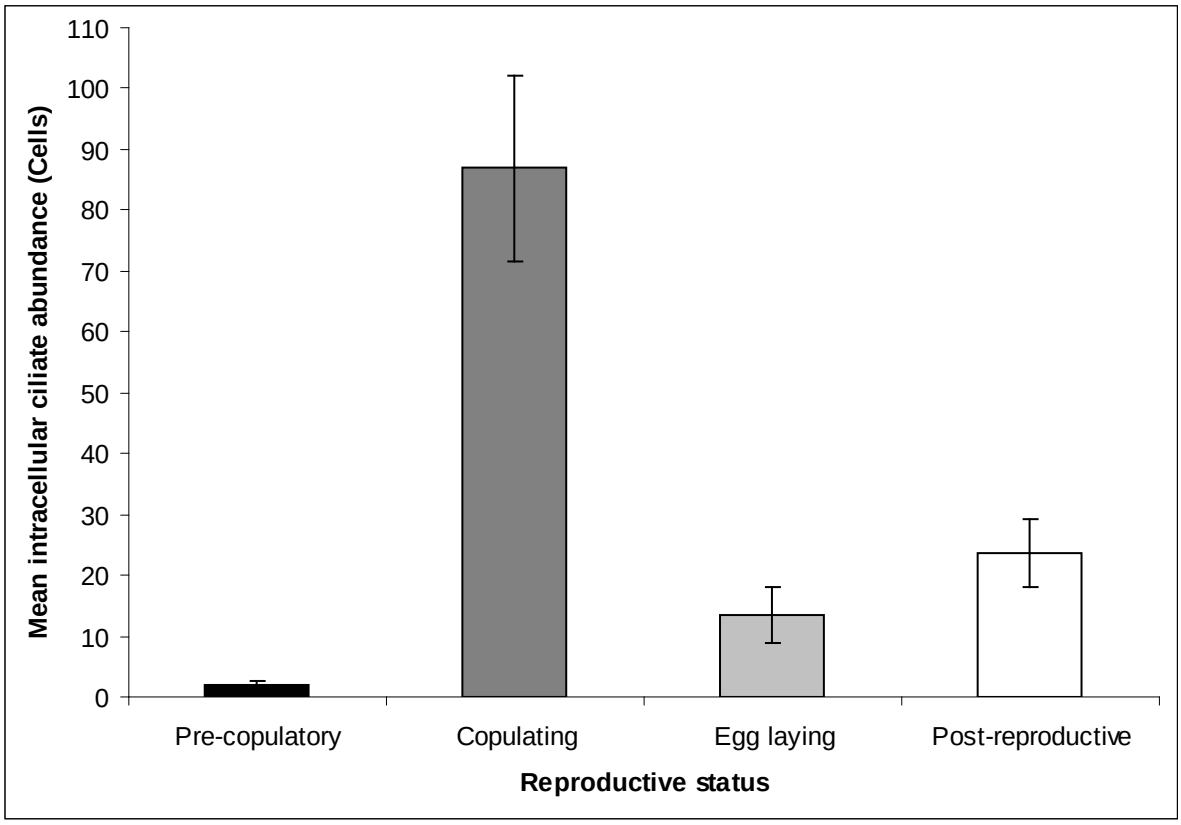


Figure 5.5 Mean total intracellular ciliate abundance (± S.E.) as a function of reproductive status within the hypobranchial gland's of female *D. orbita* (N=3).

Chapter 6. Characterisation and expression of recombinant arylsulfatase from the marine snail *Dicathais orbita*

Additional transfection experiments and western blot analysis presented in this chapter were completed by Dr Tong Chen at Flinders University, who will be credited as second author of this manuscript upon submission.

6.1 Abstract

The arylsulfatase enzyme is one of the key enzymes responsible for the production of bioactive secondary metabolites in muricid molluscs. A transcriptomics analysis of the hypobranchial gland of the Australian muricid *Dicathais orbita* revealed a partial arylsulfatase cDNA sequence, but a full length arylsulfatase enzyme has yet to be reported. Here, using 5' and 3' rapid amplification of cDNA ends, the full length *D. orbita* arylsulfatase gene was amplified. The arylsulfatase-short isoform (ARS_{562aa}) was cloned into a mammalian expression vector and transfected in HEK293T human embryonic kidney cells. This study successfully expressed ARS_{652aa}, however the resulting recombinant enzyme did not display enzyme activity. Soluble, membrane-soluble and secreted proteins were extracted, but no arylsulfatase enzyme activity was observed. Previous researchers have been unsuccessful at expressing invertebrate arylsulfatases in vertebrate expression systems suggesting that the C_{α}-formylglycine post-translational modification that is essential for enzymatic arylsulfatase activity may be a process that is specific to invertebrates. An alternate invertebrate expression system may be required to recombinantly express molluscan arylsulfatases in the future.

6.1 Introduction

Several tryptophan derived secondary metabolites, produced in the Muricidae family of marine molluscs, display potent anticancer and antibacterial activity (Benkendorff *et al.* 2000, Vine *et al.* 2007, Westley *et al.* 2010c, Benkendorff *et al.* 2011). These pharmacologically important indole products are the chemical precursors to the ancient dye Tyrian purple (Cooksey 2001, Benkendorff *et al.* 2004b). Several key enzymes are involved in the production of Tyrian purple and related compounds (Westley *et al.* 2006). The ultimate precursor to Tyrian purple is the indole tyrindoxyl sulfate. In order to produce tyrindoxyl sulfate, the amino acid tryptophan is converted to indoxyl via a tryptophanase enzyme (Miyazaki *et al.* 2000), before a methane thiol is added to the indole ring possibly involving sulfur transferase/sulfur reductase enzymes. A bromoperoxidase enzyme is responsible for the addition of a bromine group to this compound, which results in the formation of tyrindoxyl sulfate (Jannun & Coe 1987, Westley & Benkendorff 2009), the ultimate precursor to Tyrian purple. Tyrindoxyl sulfate is enzymatically oxidised by an arylsulfatase enzyme, which hydrolyses the sulfate ester (Baker & Duke 1973b). Following this, oxidation facilitates the production of the bioactive precursors, followed by dimerisation, then photolytic cleavage to produce the final product Tyrian purple (Cooksey 2001). The conversion of tyrindoxyl sulfate into Tyrian purple can also be triggered artificially by hydrochloric acid (Baker & Sutherland 1968, Westley & Benkendorff 2009). However, it is difficult to artificially control the subsequent reactions to obtain the bioactive precursors. Commercially available arylsulfatases have proven unsuccessful for hydrolysing tyrindoxyl sulfate (Westley & Benkendorff 2009). Due to the difficulties in chemically synthesizing the bioactive indoles from Muricidae (Vine *et al.* 2007, Benkendorff *et al.* 2011), further investigation of the Muricidae biosynthetic enzymes may facilitate their pharmacological development.

Sulfatases (E.C. 3.1.6.-) are a conserved family of enzymes responsible for the catalysis of hydrolytic desulfonation of sulfate esters and sulfamates (Hanson *et al.* 2004). They play numerous biological roles in different organisms, including degrading glycosaminoglycans in humans (Wilson *et al.* 1990) and being involved in estrogen biosynthesis in breast cancer (Pasqualini *et al.* 1997). They also display a digestive role in marine molluscs (Lloyd & Lloyd 1963), and process sugar sulfates and sulfated polysaccharides in molluscs (Hatanaka *et al.* 1976, Uzawa *et al.* 2003). The crystal structures of one bacterial (Boltes *et al.* 2001) and three

human arylsulfatase have been identified (Bond *et al.* 1997, Lukatela *et al.* 1998, Hernandez-Guzman *et al.* 2003), and two conserved regions of the proteins have been confirmed, both of which play a key role in the formation of the enzyme active site (Hanson *et al.* 2004). Using a post translational modification system, the conserved serine (a cysteine or serine in bacteria) found in the first position in the first conserved motif [CS]XP[SX]RXXX[LX][TX][GX][RX] is replaced with a Cα-formylglycine residue in enzymatically active sulfatases (Hanson *et al.* 2004). The second conserved motif G[YV]X[ST]XXGKXXH, in conjunction with the first conserved motif, is important in the creation of the enzymatic active site (Hanson *et al.* 2004). The Cα-formylglycine residue is essential for arylsulfatase activity in all sulfatase proteins (Schmidt *et al.* 1995). These two signature sequences have evolved under specific evolutionary constraints, whereby unique modification systems are present in individual species (Sardiello *et al.* 2005).

To date, limited research has been undertaken on arylsulfatase enzymes from molluscs. A 1557bp gene pertaining to an arylsulfatase enzyme from abalone was identified with a molecular weight of 42.5 kDa (Nikapitiya *et al.* 2007). Arylsulfatase enzyme activity has been identified in the marine gastropod *Littorina kurila* and the enzyme was isolated and found to have a molecular weight of around 45 kDa (Kusaykin *et al.* 2006). An arylsulfatase was extracted from another marine gastropod *Charonia lampas*, with specificity for the arylsulfatase enzyme substrates p-nitrophenyl sulphate and p-nitrocatechol sulphate (Hatanaka *et al.* 1976). This arylsulfatase had marked differences in substrate preference and enzymatic action compared to mammalian arylsulfatases (Hatanaka *et al.* 1976). Enzyme studies on the arylsulfatases from the roman land snail *Helix pomatia*, a non-specified abalone species and the common limpet *Patella vulgata* confirmed the unique substrate specificity and selectivity of molluscan arylsulfatases (Uzawa *et al.* 2003). Due to their varying substrate specificity molluscan sulfatases have been found to be useful in the synthesis of artificial glycopolymers for use in the treatment of inflammation, metastasis and autoimmune diseases (Bowman & Bertozzi 1999, Uzawa *et al.* 2003). Two full-length arylsulfatase cDNA sequences were cloned from the roman snail *H. pomatia* (Wittstock *et al.* 2000).

Arylsulfatase activity was first identified in muricid molluscs when Baker and Duke (1973) investigated ethanol extractions of the hypobranchial gland of *D. orbita* (Baker & Duke 1973b). Westley *et al* (2009b) used histochemistry to demonstrate the distribution of arylsulfatase activity throughout the hypobranchial gland and reproductive organs of *D. orbita*. Until the

identification of a partial arylsulfatase sequence from a transcriptomic study of the hypobranchial gland of *D. orbita* (Chapter 4), no arylsulfatase gene sequence had been identified in any muricid mollusc. To obtain the full length arylsulfatase sequence from *D. orbita*, 5' and 3' RACE was performed using hypobranchial gland cDNA. In order to determine whether the identified arylsulfatase sequences are responsible for the arylsulfatase activity from the hypobranchial gland of *D. orbita*, an expression vector containing the full length arylsulfatase gene was constructed using the pcDNA expression vector system and recombinant proteins were expressed transiently in mammalian cells.

6.3 Materials and Methods

6.3.1 Total RNA extraction

Total RNA was extracted from 100mg of branchial and medial hypobranchial gland's of *D. orbita* using the RNAqueous® RNA isolation kit (Ambion inc., Austin, TX, USA) according to the manufacturer's instructions with the modifications described in Chapter 3 (Laffy *et al.* 2009).

6.3.2 5' RACE of arylsulfatase sequence

To obtain the flanking 5' region of the partial *D. orbita* arylsulfatase gene (GD253910, Chapter 4), 5' RACE amplification was performed using the Invitrogen 5' RACE system for rapid amplification of cDNA ends version 2.0 (Invitrogen, Carlsbad, CA, USA) using the protocol described in the product manual. Template from hypobranchial gland RNA was used to produce first strand cDNA following the product manual. First strand cDNA was synthesized using the gene specific primer 5RACE_GSP1_ars. Three different specific 5' RACE primers, GSP_ARS2, 5_ARS_part2_GSP2 and 5ARS_part2_GSP3 (Table 6.1) were designed from GD253910 and used during the RACE protocol. Initial 5' RACE was performed using the kit forward primers and either GSP_ARS2 and 5_ARS_part2_GSP2 as the reverse primer under the following cycling conditions: An initial denaturation step was performed for 2 min at 94°C before 35 cycles were performed, first denaturing the template at 94°C for 1 min, facilitating primer annealing at 55°C for 1 min and allowing primer extension at 72°C for 2 min. A final extension step was performed at 72°C for 7 min. Products were visualized, gel purified using the Wizard SV gel and PCR cleanup system (Promega, Madison, WI, USA) and used as templates in nested PCR. Nested PCR was performed using initial 5' RACE PCR products as template using the nested primer 5_ARS_part2_GSP3 under the following conditions: An initial denaturation step was performed for 2 min at 94 °C before 35 cycles were performed, first

denaturing the template at 94°C for 1 min, facilitating primer annealing at 55 °C for 1 min and allowing primer extension at 72°C for 2 min. A final extension step was performed at 72°C for 7 min.

6.3.3 3' RACE of arylsulfatase gene

To obtain the 3' flanking region of known *D. orbita* arylsulfatase sequence (GD253910), 3' RACE amplification was performed using the Invitrogen 3' RACE system for rapid amplification of cDNA ends (Invitrogen) following the protocol described in the product manual. Template from hypobranchial gland RNA and polyT primer provided with the kit was used to produce first strand cDNA and arylsulfatase specific primers (Table 6.1) designed from GD253910 and used in the PCR reactions. Initial 3' RACE amplification was performed using ARS_3RACE_primer and cycled as follows: An initial denaturing step was performed at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 30 sec, annealing at 57°C for 30 sec and extension at 72°C for 2 min. To further enrich hypobranchial gland specific product formation, nested PCR was performed on the PCR products produced by the initial 3' RACE amplification. The nested primer ARS_3RACE_nestedPrimer with the nested polyT primer included in the kit were used in nested PCR reactions on PCR products (2 kb, 750 bp and 500 bp products) produced from initial 3' RACE PCR and cycled as follows: An initial denaturing step was performed at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 30 sec, annealing at 57°C for 30 sec and extension at 72°C for 2 min.

6.3.4 Cleanup, cloning and sequencing RACE products

Nested 5' RACE and 3' RACE PCR products were cleaned up using the Wizard PCR and Gel cleanup system (Promega) in accordance with the manufacturer's protocol. Ligation of products into pGEMt-easy vector and transformation of JM109 ultracompetent cells was performed following the protocol in the pGEMt-easy technical manual (Promega). White colonies were grown overnight at 37°C in Luria-Bertani media containing ampicillin and the resulting inoculant was used in plasmid DNA purification following the Wizard Miniprep plasmid Purification system (Promega). Plasmid DNA was digested to determine insert size, and vectors that were shown to contain insert DNA within the size ranges expected (500 bp-2000 bp), were sent for sequencing at the Australian Genome Research Sequencing Facility (AGRF, Brisbane, Australia) using M13F and M13R sequencing primers (Table 6.1). Sequencher 4.1.4 (Genecodes, Ann Arbour, MI, USA) was used to view sequenced plasmid DNA, clip vector sequence and assemble 5' and 3' flanking regions of GD253910 to obtain the full length

arylsulfatase sequence of *D. orbita*. When the potential translation of the full length arylsulfatase genes was investigated, it was unclear where in the sequence translation would be initiated due to the presence of two in-frame potential start codons.

6.3.5 Amplification and cloning of full length arylsulfatase sequences

In order to verify the full length of the assembled arylsulfatase sequence from *D. orbita*, two full length arylsulfatase sequences were amplified using primers (Table 6.1), designed from our assembled RACE sequences. To produce the longer ARS_{571AA} protein sequence, primers ARSfull5long_4 and ARS_rev_6_cloning; to produce the shorter ARS_{562aa} protein sequence, primers ARSfull5short_5 and ARS_rev_6_cloning were used. Full length sequences were amplified using Phusion High fidelity DNA polymerase (Finnzymes Oy, Finland). Reactions were cycled as follows; an initial denaturation step was performed at 98°C for 30 sec followed by 35 cycles of denaturation at 98°C for 10 sec, primer annealing at 55°C for 30 sec and primer extension at 72°C for 1 min. A final extension step was performed at 72°C for 10 mins at the end of the cycling. PCR products were visualized and purified using the Wizard SV gel cleanup kit (Promega) in accordance with the manufacturer's instructions. Restriction enzymes KpnI and ApaI were used to digest purified PCR products and the expression vector pcDNA™3.1/V5-His (Invitrogen). Digested vector was phosphatase treated using calf intestinal phosphatase (Promega) and inserts were ligated into the vector using T4 DNA ligase (Promega). XL10 Gold ultracompetent cells (Aligent, Wilmington, DE) were transformed with the ligated vector and inserts using the XL10 Gold transformation protocol. Colonies were selected, grown at 37°C overnight in Luria-bertani media and the resulting inoculants had their plasmid DNA purified using Wizard Miniprep plasmid purification system (Promega). Plasmid DNA was digested to determine insert size, and vectors that was shown to contain insert DNA within the size ranges expected (~1700 bp) were quantified using the Nanodrop spectrophotometer (Thermo Scientific, Rochester NY, USA) and sequenced in duplicate using sequencing primers T7 and BGH_reverse and gene specific primers ARS_3RACE_nestedPrimer and ARS3_3 (Table 6.1). This full length arylsulfutase sequence was submitted as a nucleotide entry to Genbank and can be found under the accession number HM246144. A full length expression vector identified as pcDNA-ARS_{562aa}_V5-His was found to contain the complete open reading frame (ORF) of the smaller arylsulfatase protein ARS_{562aa} and was used in subsequent transfection experiments.

6.3.6 Sequence comparisons of full length arylsulfatase sequences

BLASTP analysis was performed on the protein translation of our full length arylsulfatase gene (Genbank accession number ADK13094) and the top 10 BLAST results can be viewed in Figure 6.1. The results of this blast search confirmed sequence homology and shows the sequence conservation observed between HM246144 and other invertebrate arylsulfatase sequences. The full length arylsulfatase sequence was compared to known invertebrate arylsulfatase protein sequences obtained from Genbank: from the roman snail *Helix pomatia* two arylsulfatase sequences, AAF30402.1 and AAF30403.1; from the abalone *Haliotis discus discus*, AB026612.1; from the Florida lancelet *Branchiostoma floridae*, XP_002592251.1; from the red flour beetle *Tribolium castaneum*, XP972832.1; and from the marine tunicate *Ciona intestinalis* XM_002121052.1 using a ClustalX multiple sequence alignment (Thompson *et al.* 1997). A Gonnet protein weight matrix and a gap open penalty of 10 and a gap extension penalty of 0.2 was selected. Pairwise sequence analysis was also performed on all invertebrate sulfatase sequences included in multiple sequence alignment using Needle analysis as part of the EMBOSS package available at the European Molecular Biology Laboratory website (EMBL, www.embl.org).

Sequence analysis was performed on the full length arylsulfatase protein sequence from *D. orbita* using the EXPASY website (www.expasy.org/tools). The program ProtParam was used to identify the theoretical molecular weight and pI of the pcDNA-ARS$_{562aa}$_V5His protein, SignalP was used to identify whether the protein contained a signal peptide, NetPhos 2.0 was used to identify any posttranslational modification sites on the protein and protein sulfination sites were identified using Sulfinator (Gasteiger *et al.* 2005).

6.3.7 Recombinant expression of pcDNA-ARS562aa_V5-His

6.3.7.1 Mammalian cell culture and recombinant gene expression

Bacterial and yeast expression systems used in the recombinant expression of two molluscan arylsulfatases from *Helix pomatia* failed produce active arylsulfatse enzymes, and this has been accounted to an incompatibility in the expression hosts post-translational modification systems (Hanson *et al.* 2004). The same paper confirmed that invertebrates utilize the same posttranslational modification to the cysteine residue as humans do (Hanson *et al.* 2004), and as active human arylsulfatases have been successfully expressed in mammalian tissue culture systems (Lukatela *et al.* 1998), this expression system was chosen for use in our study.

pcDNA-ARS_{562aa}_V5-His plasmid DNA was amplified using the JETstar plasmid purification MAXI kit following the manufacturer's instructions (Genomed, Lohne, Germany). In order to investigate the activity of our pcDNA-ARS_{562aa}_V5-His construct, mammalian cell culture and transfection was performed. Human embryonic kidney cells HEK293T were grown in DMEM supplemented with 5% FCS, 100 U/ml penicillin and 100 μg/ml streptomycin at 37°C, 5% CO_2. Cells were cultured to approximately 75% confluence in T25 culture flasks, and transfection was carried out using methods previously described, with Fugene 6 (Roche, Indianapolis, IN, USA) or Lipofectamine (Invitrogen) (Abbott *et al.* 1999).

6.3.7.2 Preparation of arylsulfatase activity positive control

Dissected hypobranchial glands, snap frozen and stored at -80°C were homogenized under liquid N_2 with mortar and pestle, and then resuspended in 5 ml of 5 mM Tris-HCl (pH 7.0). The homogenate was pipetted thoroughly to mix and centrifuged at 100,000g for 30 min at 4°C in the TLX ultracentrifuge in a TLA 100.3 rotor (Beckman Coulter, Brea, CA, USA). The supernatant was removed and the protein concentration was estimated using Bradford assay (Biorad, Herculese, CA, USA) using bovine serum albumin as the standard, according to manufacturer's instructions. This was stored at 4°C and used as positive control for up to two weeks in enzyme assays.

6.3.7.4 Arylsulfatase enzyme activity assay

The presence of arylsulfatase activity was tested using the p-Nitrocatechol sulfate (pNCS) assay (Baum *et al.* 1959). Cell culture media (1 ml) was isolated from transfected cultures and centrifuged at 16,000 RPM for 30 mins at 4°C to investigate enzyme activity of secreted proteins. Harvested cells (Approx 1×10^6 cells) were suspended in 1 ml 5 mM Tris-HCl pH 7.0 with protease inhibitors 1 mM ETDA, 10 μM pepstatin A, 10 μM leupeptin and 10 μM E64. Cells were sonicated for 3 × 10 seconds on an Ultrasonic Processor XL cell Sonicator (Heat Systems inc., Farmingdale, NY, USA) at power level 3. Cells were centrifuged at 16,000 RPM for 30 min at 4°C and the supernatant was used as the soluble protein fraction. The resulting pellet was resuspended in 1 ml of 5 mM Tris-HCl pH 7.0 with protease inhibitors 1 mM ETDA, 10 μM pepstatin A, 10 μM leupeptin and 10 μM E64 and centrifuged at 100,000g in the Beckman Ultracentrifuge in a TLA100.3 rotor (Beckman Coulter) for 30 mins to produce a supernatant fraction containing membrane soluble proteins.

The assay substrate was made up with 10mM pNCS in 0.1 M 2-[N-morpholino] ethanesulfonic acid (MES) pH 6.2. The assays were carried out in 96 well plates, with the secreted protein, soluble protein and membrane soluble protein fractions all tested for arylsulfatase activity. Each sample was measured in triplicate, with 100 µl of protein extract (up to 2µg/µl) and 50 µl of substrate in each reaction. Arylsulfatase activity was measured at 510 nm every two minutes for 60 min using the Omega Fluorostar spectrophotometric plate reader (BMG Labtech, Offenburg, Germany).

6.3.7.5 Transfection control: DP enzyme activity

An expression construct of the human dipeptidyl peptidase IV (DPIV) was used as a positive control for our expression reactions. This expression vector had previously been shown to display DPIV activity when recombinantly expressed in HEK293T cells (Abbott *et al.* 2000), and produces a monomeric protein approximately 113 kDa in size, as well as a larger dimeric form of the protein. The DPIV expressing construct pcDNA-DPIV$_{776aa}$_V5-His was transfected simultaneously to pcDNA-ARS$_{562aa}$_V5-His, and the DPIV assay was performed on extracts from this transfection to confirm the success of the transfection process (Abbott *et al.* 2000). Assays were carried out in a 96 well plate with 1 mM H-Ala-Pro-p-nitroanilide (H-Ala-Pro-pNA) in 50 mM Tris-HCl pH 8.0, 100 mM NaCl buffer. Samples were measured in triplicate using 100µl of sample (approx 2µg/µl) and 50 µl of substrate per well (150 µl total). The production of pNA was measured for 60 min at 37°C, with absorbance readings taken every 10 min at 405 nm using the Omega Fluorostar spectrophotometric plate reader (BMG Labtech).

6.3.7.6 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Discontinuous polyacrylamide electrophoresis was performed according to the method of (Laemmli 1970) using Bio-Rad mini protean II electrophoresis apparatus. Gels (10% acrylamide, 375 mM Tris-HCl (pH 6.8), 0.1% (w/v) Sodium dodecyl sulfate (SDS), 10% (w/v) acrylamide/bis (29;1), 0.05% (w/v) ammonium persulfate and 0.01% (w/v) TEMED) were routinely run at 170 Volts (V) for 1 hr. Precision Plus Protein Standards Kaleidoscope markers (Cat# 161-0375) were included on every gel (Biorad). Protein samples (60-90 ug total protein) were solubilised using 3x SDS-PAGE loading buffer (62.5 mM Tris-HCl (pH 6.8), 20% (v/v) glycerol, 2% (w/v) SDS, 5% (v/v) β-mercaptoethanol) and boiled for 5 min. The running buffer contained 25 mM Tris, 192 mM glycine, 0.1 % (w/v) SDS, pH 8.3.

6.3.7.7 Western blotting

Transfer of protein to Polyvinylidene Fluoride (PVDF) membrane was undertaken following the method as previously described (Towbin *et al.* 1979). Transfer was carried out in a Tris/glycine buffer (25 mM Tris (pH 8.3), 152 mM glycine-without methanol) for 1 hour at 60 V using a Mini Trans-Blot Cell (BioRad).

6.3.7.8 Anti-V5 antibody immunoblotting

Protein samples were separated by SDS-PAGE and transferred to PVDF. Membranes were blocked with 5% skim milk powder blocking buffer (5% w/v skim milk powder (Diploma) in PBS pH 7.4, 1% tween-20 (v/v)) for 1 hr or overnight. The pcDNA expression system attaches a 14 amino acid V5 epitope to the C terminus of the recombinant protein which can be detected using immunoblotting. Monoclonal anti-V5 antibody raised in mouse (R930-CUS, Invitrogen) was diluted 1:5000 in blocking buffer and the membrane was incubated for 1 hr at room temperature (RT) with rocking. The membrane was washed for 3 x 10 min using blocking buffer (Sambrook *et al.* 1989). Polyclonal rabbit-anti mouse secondary antibody (P0161, Dako, Glostrup, Denmark), diluted 1:1000 in blocking buffer, was incubated for 1 hr at RT with rocking. The membrane was washed 3 x 10min with PBS with 0.1% (v/v) tween-20 (PBST) and incubated with chemiluminescence reagent (Thermo Scientific) for 5 min before exposure on X-ray film (X-omat, Kodak) for 5 min. Size of any resulting bands was calculated using the Versadoc imaging system and QualityOne software (Biorad)

6.4 Results

6.4.1 Full length sequence amplification of arylsulfatase genes

When 5'RACE and 3'RACE sequences were assembled with the known partial arylsulfatase sequence from *D. orbita*, we obtained a single contiguous 2210 bp sequence (HM246144). Translation of this sequence identified two methionine start codons that could initiate translation, at the start of a single open reading frame. The first initiator methionine generated a 571 aa protein ARS_{571aa}, and the second methionine would generate a 562 aa protein ARS_{562aa}, that is identical, lacking the first 8 aa. Two forward primers were designed to amplify DNA that would express a short and long form of arylsulfatase, both approximately 1700 bp long. Both forward primers amplified products of the appropriate size from *D. orbita* cDNA but only ARS_{562aa} was successfully ligated into the pcDNA-V5HisA expression vector after

repeated cloning experiments, where numerous variations of insert to vector ratios were attempted. Thus only ARS_{562aa} was investigated in transfection studies. Sequencing of the ARS sequence inserted into the expression vector, pcDNA-ARS_{562aa}V5-His confirmed that the sequence was in the correct frame and orientation to produce intact recombinant protein.

ARS_{562aa} had a predicted molecular weight of 66136.8 Daltons (Da) and a theoretical pI of 6.34 (ProtParam). An N-terminal signal peptide was identified on the ARS_{562aa} isoform, cleaved at the site VAA_{18}-Q_{19}N (SignalP) (Fig 6.2) that would be processed in the secretory pathway (TargetP, iPsort). Taking into account the presence of the N-terminal signal peptide identified from Signal P analysis, the predicted molecular weight of the cleaved ARS_{562aa} would be 64211.3 Da with a theoretical pI of 6.27 (ProtParam). Sixteen serine phosphorylation sites were predicted out of a total 30 serine residues, 11 threonine phosphorylation sites were predicted out of a total of 38 threonine residues and 8 tyrosine phosphorylation sites out of a total of 26 tyrosine residues were predicted (NetPhos 2.0). A single site for tyrosine sulfation was identified at the Tyr_{100} residue (Sulfinator). A multiple sequence alignment of the *D. orbita* ARS sequence with other arylsulfatase revealed a conserved sulfatase region. The consensus motif [CS]XP[SX]RXXX[LX][TX][GX][RX] was identified starting at residue 80 with a sequence of STPSRASYMTGY (Fig 6.2). Additionally, the second sulfatase conserved motif G[YV]X[ST]XXXGKXH was present in our *D. orbita* ARS sequence, starting at residue 126 of the peptide, with a sequence of GYVSHLVGKWH (Fig 6.2).

Table 6.2 details the amino acid identity and similarity percentage values for each pairwise comparison. ARS_{562aa} showed the highest level of amino acid percentage identity with the arylsulfatase from the lancelet *B. floridae* but the highest level of amino acid similarity to the gastropod *H. discus*.

6.4.2 Arylsulfatase enzyme activity

The pcDNA-ARS_{562aa}V5His expression construct was transfected into HEK293T cells and examined for its ability to produce arylsulfatase activity. All assays were performed using hypobranchial gland homogenate as a positive control for the arylsulfatase assay. A dipeptidyl peptidase IV (DPIV) expression construct, pcDNA-$DPIV_{766aa}$V5His construct (Abbott *et al.* 2000) was used as a transfection positive control. The arylsulfatase and DPIV assays were performed in triplicate on three separate transiently transfected cell extracts to confirm the accuracy of our transfection reaction. No significant arylsulfatase activity was identified from the

secreted, soluble or membrane soluble protein extracts from pcDNA-ARS_{562aa}V5-His positive control transfected cells compared to the cell only and vector only soluble and membrane soluble protein extracts (Fig 6.3A). In order to concentrate recombinant protein for arylsulfatase activity assays, nickel affinity purification was performed (see Appendix I). Purification failed to enrich arylsulfatase activity. The measurement of arylsulfatase activity from the hypobranchial gland homogenate confirmed the pNCS assay was working (Fig 6.3A). Significant DP activity was identified in both soluble and membrane soluble fractions of $DPIV_{766aa}$V5His transfected cell when compared to cell only and vector only fractions (Fig 6.3B).

6.4.3 Recombinant ARS_{563aa} is expressed in transiently transfected HEK293t cells

To determine whether the arylsulfatase expression construct was producing recombinant protein, a western blot was performed with an antibody which would detect the C-terminal V5 epitope. A prominent band was visible from the ARS_{562aa} total cell extract of approximately 63 kDa (Fig 6.4). Two bands were also observed from the $DPIV_{766aa}$ total cell extract, one with an approximate size of 113 kDa and an additional dimeric band larger than 250 kDa (Fig 6.4). No His-tagged protein was visible in the vector only control total cell extract.

6.5 Discussion

This work presents the isolation and characterization of the first full length arylsulfatase cDNA sequence and resulting protein sequence from any muricid mollusc, and is the first publication detailing a protein sequence responsible for enzyme activity involved in the formation of Tyrian purple and its bioactive precursors. Two possible translations from our full length *D. orbita* cDNA sequence were proposed in this study, however it is unclear from the cDNA sequence which methionine start codon would be used in natural expression of the *D. orbita* arylsulfatase gene.

The consensus sequence gccRccAUGG, found at the flanking 5' region of eukaryotic expressed genes (Kozak 1987), is known to be involved in the initiation of translation (Angioletti *et al.* 2004). The typical Kozak sequence is usually seen as GCCACCAUGG, where AUG refers to the initiating methionine residue, however variations on the consensus sequence are not uncommon. In fact, observations into translation initiation sequences have identified several different translation initiation consensus sequences for different phyla (Mankad *et al.* 1998).

While all vertebrates and invertebrates typically maintain a purine residue (usually A) at the -3 site, molluscs have been shown to exhibit the following Kozak sequence (a|t)(c|t)(a|c)Aa(a|c)ATGg. Both methionine codons in *D. orbita*'s arylsulfatase cDNA sequence are in an acceptable invertebrate Kozak context that could be used for translation. The cDNA sequence upstream of the met in ARS_{562aa} sequence contains a GACACCATGA Kozak region, showing the typical Kozak site in four of the six flanking nucleotides to the initiating ATG, including the conserved purine (A) at the -3 site, although the sequence did not contain the conserved G residue at the +4 position. The region upstream of the methionine that would encode an ARS_{571aa} product contains the TTGTTGATGG Kozak region, showing conservation in only two of the six flanking regions, not including the important purine residue at the -3 position, although displaying the conserved G in the +4 position. When comparing Kozak sequences, it remains unclear which of these sites of translation initiation is the likely start site for protein expression. Kozak's leaky scanning, first-AUG rule, identified that when two methionine residues are located in close proximity to each other at the 5' end of a sequence, it is unlikely that the second methionine residue will be preferentially chosen to initiate translation, even if its corresponding kozak sequence better conforms to the Kozak convention (Kozak 1995). Because of this, it is unlikely that the second downstream methionine residue is involved in translation initiation, which suggests that ARS_{562aa} does not represent the full length arylsulfatase gene expressed in *D. orbita*.

The inclusion of a signal cleavage site downstream of both possible start codons suggests that the N Terminus of either arylsulfatase translations would be cleaved upon processing, indicating that the signal peptide is unlikely to be involved in the activity of the arylsulfatase enzyme and regardless of the composition of the N-terminus of this signal peptide, active arylsulfatase enzyme should be produced. As we could only successfully clone the ARS_{562aa} construct into the expression vector, further investigation into the expression of the ARS_{571aa} construct is needed to confirm whether it has arylsulfatase enzyme activity. With such minimal difference in peptide sequence composition between both translation initiation sites, it was hoped that both sequences, ARS_{571aa} and ARS_{562aa}, would produce active arylsulfatase enzyme. Nevertheless, it is possible that this signal peptide could be critical for appropriate post-translational processing in the endoplasmic reticulum and further attempts at producing this larger recombinant protein should be made in the future to determine whether this signal peptide is involved in post-translational modification. An active human arylsulfatase has been expressed recombinantly in baby hamster kidney cells. This protein was secreted and purified

from the cell culture media indicating the signal peptide found on the N-terminus of eukaryotic arylsulfatases is not involved in enzyme activity (Lukatela *et al.* 1998).

The full length arylsulfatase sequence that was produced displayed key properties that would be expected from an eukaryotic arylsulfatase; both conserved motifs that are synonymous with sulfatase activity as described in Hanson *et al.* (2004) were present in the *D. orbita* arylsulfatase. When both of these conserved residues are compared to the known *H. pomatia* sulfatase sequences, none of the four variable residues in motif I are commonly expressed between *D. orbita* and *H. pomatia* (Fig 6.2). Interestingly, two of the four variable residues in motif I between *D. orbita* and *H discus* are identical (Fig 6.2). Conserved motif II has only two residues from its total 12 residues that display any sequence difference (Fig 6.2), and there are no similarities observed in these two residues in all four of the gastropod arylsulfatases reported. The cleavage site that was identified from SignalP analysis that is present in both ARS_{562aa} and ARS_{571AA} supports the finding that arylsulfatase is located in secretory vesicles within cells of the hypobranchial gland and is secreted into the mantle cavity when required (Westley & Benkendorff 2009, Westley *et al.* 2010b).

Analysis of the amino acid pairwise sequence similarity showed that this arylsulfatase sequence was highly conserved between other invertebrate and lancelet sequences (Table 6.1) ARS_{562aa} sequence shared the highest level of amino acid sequence similarity (46.9%) with the arylsulfatase sequence from the marine gastropod *H. discus* (Table 6.1). *Helix pomatia*, while also a gastropod mollusc, is a terrestrial snail, and the differences in habitat and phylogeny between the species is likely responsible for the lower level of sequence similarity and identidy observed between *H. pomatia* and *D. orbita* (43.75% and 21.4% respectively) (Table 6.1). Interestingly, the *D. orbita* arylsulfatase showed higher amino acid sequence similarity to the marine ascidian *Ciona intestinalis* and the lancet *Branchiostoma floridae* (Phyla Chordata) than the terrestrial gastropod *H. pomatia* (Table 6.1). However the range in values is between 45-46.9 % compaed to 43.7% so these differences may not be that significant. The C terminus of invertebrate sulfatases appears to display a high level of protein sequence homology (Fig 6.2). The location of conserved motifs I and II and the catalytic Cα-formylglycine residue within the first 150 aa of invertebrate sulfatases suggests that this region is highly specialized and essential for enzymatic activity. The N terminus of invertebrate arylsulfatases shows a much higher level of variability, even when comparing homologs within the same species (*H. pomatia* Fig 6.2). This increased variability may be important in determining the

different substrate specificities and functional divergences that so many invertebrate arylsulfatases display (Uzawa *et al.* 2003, Hanson *et al.* 2004, Kusaykin *et al.* 2006).

There was no significant difference in arylsulfatase activity between any of the ARS_{562aa} transfected cell extracts (membrane, soluble membrane or secreted media) and the corresponding cell only and vector only cell extracts. The presence of arylsulfatase activity in the hypobranchial gland homogenate confirmed that the p-NCS arylsulfatase assay was functioning correctly. The fact that significant increases in DP activity were observed in soluble and membrane soluble extracts also confirmed that the transfection protocol was effective at expressing pcDNA-V5-His recombinant proteins from the HEK293T cells. This would suggest that despite the lack of active arylsulfatase activity in our ARS_{562aa} cell extracts, this gene construct was successfully transfected into the cells. This is further supported by western blot analysis, where a 63 kDa His tagged protein was shown to be expressed in our ARS_{562aa} transfected cell extract (Fig 6.3A).

In order to identify the nucleotide sequence responsible for the arylsulfatase activity exhibited in the hypobranchial gland, further investigation is required on both the ARS_{562aa} and ARS_{571aa} gene sequences. There are many reasons why recombinant expression of an enzyme fails to yield active protein products. It has previously been reported that the presence of a histidine tag on recombinant enzymes can affect substrate specificity and activity (Lee *et al.* 1999). Consequently, the V5/His_6 tag which was confirmed via western blot analysis (Fig 6.4) may be responsible for the inactivity exhibited in our recombinant expression experiments. Its presence at the carboxyl terminus of the protein may interfere with the folding of the protein, inhibiting its arylsulfatase activity (Kao *et al.* 2009).

An alternative reason why arylsulfatase activity was not detected from our ARS_{562aa} expressed protein may be due to specific post-translational modification to the arylsulfatase protein that occurs *in vivo* in the hypobranchial gland of *D. orbita*. Differences in bacterial and lower eukaryotic sulfatases and sulfatase-modifying factors, compared to the relevant homologs in higher eukaryotes suggest that sulfatase modifying systems may not be compatible between mammalian and bacterial recombinant expression systems. All eukaryotic sulfatases display sequence conservation, and all contain a C_α-formylglycine residue that is present in their active site and essential for enzymatic activity (Hanson *et al.* 2004). This Cα-formylglycine modification is essential for the activation of sulfatase proteins and without the appropriate

sulfatase-modifying enzymes, sulfatase activity cannot be facilitated, as is demonstrated in the inherited human medical condition multiple sulfatase deficiency (Cameron *et al.* 2004). A full length arylsulfatase cDNA was identified from the sea urchin *Hemicentrotus pulcherrimus*, however, no arylsulfatase enzyme activity was reported in association with this recombinant enzyme (Sasaki *et al.* 1988). Similarly, the full length arylsulfatase gene has been identified from the disk abalone *H. discus* (Nikapitiya *et al.* 2007), which shares homology with ARS_{562aa}, however there have been no publications reporting on the successful expression of this enzyme in a recombinant system.

A recent paper has reviewed the identification of formylglycine-generating enzymes (FGE) in many eukaryotes including at least two gastropods and two bivalves (Gande *et al.* 2008). It is known that it is not possible to successfully express active sulfatase proteins from higher eukaryotes in prokaryotic or yeast expression system due to the compatibility of the C_{α}-formylglycine post-translational modification enzymes (Wittstock *et al.* 2000, Sardiello *et al.* 2005, Landazuri *et al.* 2009). However there have been several cases where higher eukaryote expression systems, such as tissue culture systems, have been used to recombinantly express arylsulfatases from other higher eukaryotes (Moro *et al.* , Nagamine *et al.* , Oshikawa *et al.* 2009). Our results suggest that the mechanisms involved in the post translational modification of higher eukaryotic arylsulfatases is so specific that molluscan arylsulfatases cannot be modified appropriately in a human tissue culture expression system. The prokaryotic and lower eukaryotic incompatibility that higher order eukaryotic sulfatases display has been documented in the literature (Moro *et al.* 2009, Nagamine *et al.* 2009), which prompted the selection of mammalian tissue culture system to express *D. orbita*'s arylsulfatase in this study. Sulfatase modifying factors have been identified that are responsible for the post-translational modification of the catalytic C_{α}-formylglycine residue in humans (Cosma *et al.* 2003, Dierks *et al.* 2003). Eukaryotic sulfatases show a high level of amino acid sequence similarity (Fig 6.2, table 6.1), but it is not known if conservation is also shared between human and *D. orbita* FGE amino acid sequences. If the C_{α}-formylglycine post-translational modifications that are essential for ARS_{562aa} activity are specific to invertebrates, or even specific to muricid molluscs, then this could account for why no arylsulfatase activity was detected in the mammalian expression system. A comprehensive review of FGE sequences and paralog FGE sequences in eukaryotes confirmed that a specific signal sequence responsible for transport to the endoplasmic reticulum is present in *A. californica* but absent in *Homo sapiens* (Gande *et al.* 2008). Without the correct signal peptide being present in the mammalian tissue culture

expression system, we are unlikely to facilitate the FGE to modify our recombinant arylsulfatase and gain enzyme activity. By expressing ARS_{562aa} in the baculovirus expression system (King & Possee 1992), which utilizes recombinant viral sequence expression within insect cells or insect larvae, it may be possible to determine whether fellow invertebrates can facilitate the post translational modification of ARS_{562aa}, or whether *D. orbita*'s arylsulfatase modification system is specific to molluscan taxa. If FGE sequences are identified from *D. orbita*, we should be able to determine what kind of signalling peptides are present and choose a suitable expression system that may produce active recombinant enzyme. Alternatively, identification and isolation of *D. orbita* FGE may allow us to express the corresponding FGE that specifically targets ARS_{562aa} and dual expression of these enzymes in a mammalian expression system may facilitate the production of functionally active enzymes.

This investigation identified the full length mRNA sequence for a muricid arylsulfatase and successfully produced a construct to express the recombinant ARS_{562aa} V5-His-tagged arylsulfatase protein in mammalian cells. While ARS_{562aa} V5-His-tagged protein was expressed recombinantly it did not exhibit arysulfatase activity in comparison to *D.orbita* HBG tissue. While the other variant of *D. orbita* arylsulfatase ARS_{571aa} should be investigated in addition to ARS_{562aa} in future work, any further investigations may need to utilize an alternate expression system in order to facilitate the correct post-translational modification system that is essential for arylsulfatase activity.

6.6 Acknowledgements

This work was funded by an anonymous philanthropic foundation. Patrick Laffy is supported firstly by a Flinders University Faculty of Science and Engineering Research scholarship, followed by a Flinders University Postgraduate Research scholarship. Dr Tong Chen was responsible for additional transfections and western blot analysis presented in this manuscript.

Table 6.1 Primers used in RACE, cDNA sequencing and cloning experiments. Describes the primer sequence and the primers' use for every primer used in the characterization and expression of the full length arylsulfatase sequence from *D. orbita*

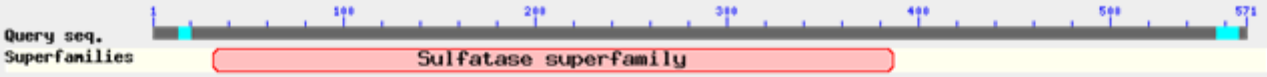
Primer name	Sequence 5'-3'	Type
5RACE_GSP1_ars	GGGGTCAAAGTCTTCATGGCTTGC	5'RACE first strand cDNA synthesis
GSP_ARS2	CCGTTGAGGAGACCCAGG	5' RACE
5_ARS_part2_GSP2	GCCATCTTTCCAAGGTTGGGGG	5' RACE
5ARS_part2_GSP3	CCCAGATCATCAGCCATGATGTA	5' RACE
ARS_3RACE_primer	CGAGCCGAGGGTTCGATACCTTCC	3' RACE
ARS_3RACE_nestedPrimer	CGAGCAAGCCATGAAGACTTTG	3' RACE/ Sequencing
M13F (universal)	TGTAAAACGACGGCCAGT	Sequencing
M13R (universal	CAGGAAACAGCTATGACC	Sequencing
ARSfull5long_4	CCGGGTACCTTGATGTCCACTGCAGCTCAC	Cloning
ARSfull5short_5	GCGGGTACCACCATGGGCAGAATGATGCTTC	Cloning
ARS_rev_6_cloning	GTCCGGGCCCGTCTGTTTCGCACCAAGTG	Cloning
T7	TAATACGACTCACTATAGGG	Sequencing
BGH_reverse (universal)	TAGAAGGCACAGTCGAGG	Sequencing
ARS3_3	GAAGGGCCCTGTAGGTGTCCACCTTAG	Sequencing

Table 6.2 Percentage identity and sequence similarity (bold) matrix between known eukaryotic sulfatase amino acid sequences. All pairwise sequence comparisons calculated using Needle in the EMBOSS software package available from the European Molecular Biology Laboratory website (EMBL, www.embl.org)

	Haliotis discus discus *481 aa*	***Ciona intestinalis*** *518 aa*	***Branchiostoma floridae*** *481 aa*	***Tribolium castaneum*** *646 aa*	***Helix pomatia 1*** *503 aa*	***Helix pomatia 2*** *266 aa*
Dicathais orbita	31.2%	32.1%	33.7%	26.5%	29.0%	14.1%
571 aa	**46.9%**	**45.0%**	**46.5%**	**41.6%**	**43.7%**	**21.4%**
Haliotis discus discus		36.6%	52.3%	30.9%	45.2%	20.4%
481 aa		**54.6%**	**67.4%**	**45.8%**	**60.6%**	**26.9%**
Ciona intestinalis			30.4%	41.5%	37.1%	12.8%
518 aa			**43.1%**	**55.4%**	**52.8%**	**17.9%**
Branchiostoma floridae				30.4%	43.3%	20.3%
481 aa				**43.1%**	**60.3%**	**28.2%**
Tribolium castaneum					31.5%	15.2%
646 aa					**45.6%**	**22.3%**
Helix pomatia 1						24.0%
503 aa						**29.3%**
Helix pomatia 2						
266 a						

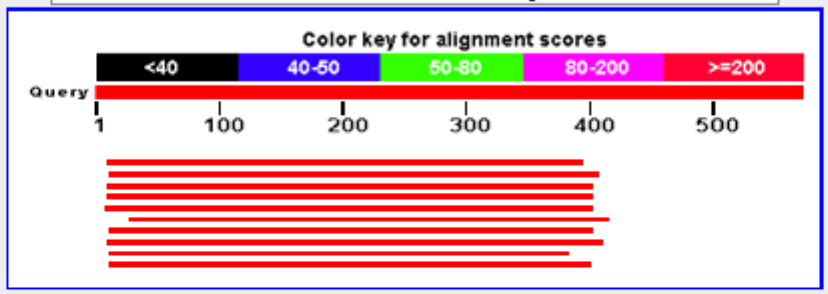
Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.



Distribution of 10 Blast Hits on the Query Sequence

Mouse-over to show defline and scores, click to show alignments



Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Links
ABO26695.1	sulfatase 1B precursor [Haliotis discus discus]	293	293	67%	2e-95	
ABO26612.1	arylsulfatase [Haliotis discus discus]	292	292	68%	9e-94	
XP_002121051.1	PREDICTED: similar to arylsulfatase B [Ciona intestinalis]	287	287	68%	2e-91	U G
ABO26692.1	sulfatase 1A precursor [Haliotis discus discus]	285	285	68%	4e-91	
XP_002127641.1	PREDICTED: similar to Arylsulfatase B precursor (ASB) (N-acet	284	284	69%	2e-90	U G
XP_003118422.1	CRE-SUL-3 protein [Caenorhabditis remanei] >gb|EFO83020.1	283	283	67%	4e-90	G
XP_002592276.1	hypothetical protein BRAFLDRAFT_206928 [Branchiostoma flor	276	276	68%	2e-87	U G
XP_002736785.1	PREDICTED: galactosamine (N-acetyl)-6-sulfate sulfatase-like	275	275	69%	3e-87	G M
XP_002610988.1	hypothetical protein BRAFLDRAFT_246447 [Branchiostoma flor	275	275	65%	4e-87	G
XP_002592251.1	hypothetical protein BRAFLDRAFT_206907 [Branchiostoma flor	275	275	68%	5e-87	U G

Figure 6.1 BLASTP analysis of translated full length arylsulfatase from *D. orbita*.
Protein sequence from the full length *D. orbita* arylsulfatase (Accession number HM246144) underwent BLASTP analysis on the NCBI blast server against the non-redundant protein sequence database (nr) using default parameters and a Blosum62 scoring matrix.

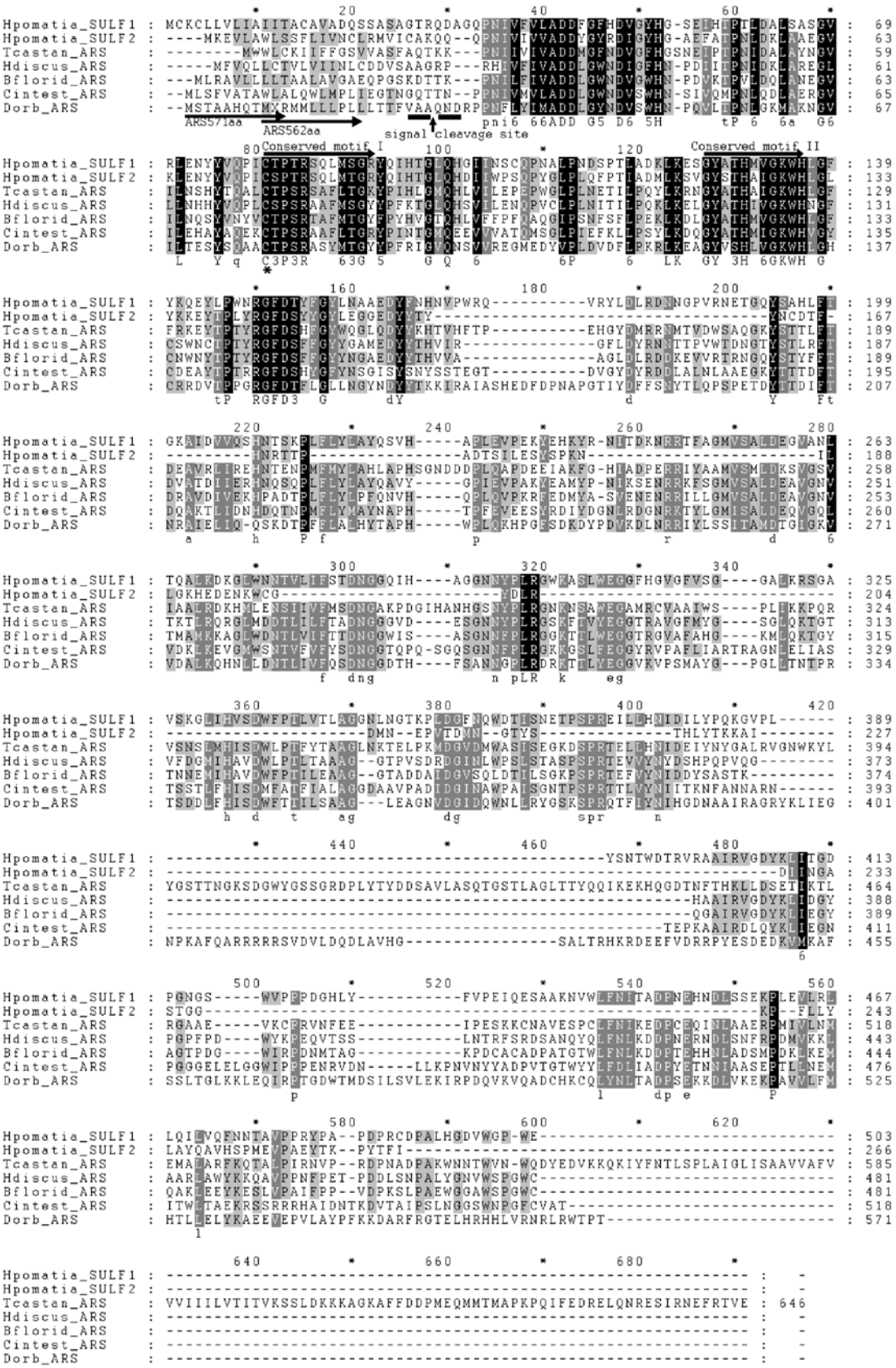
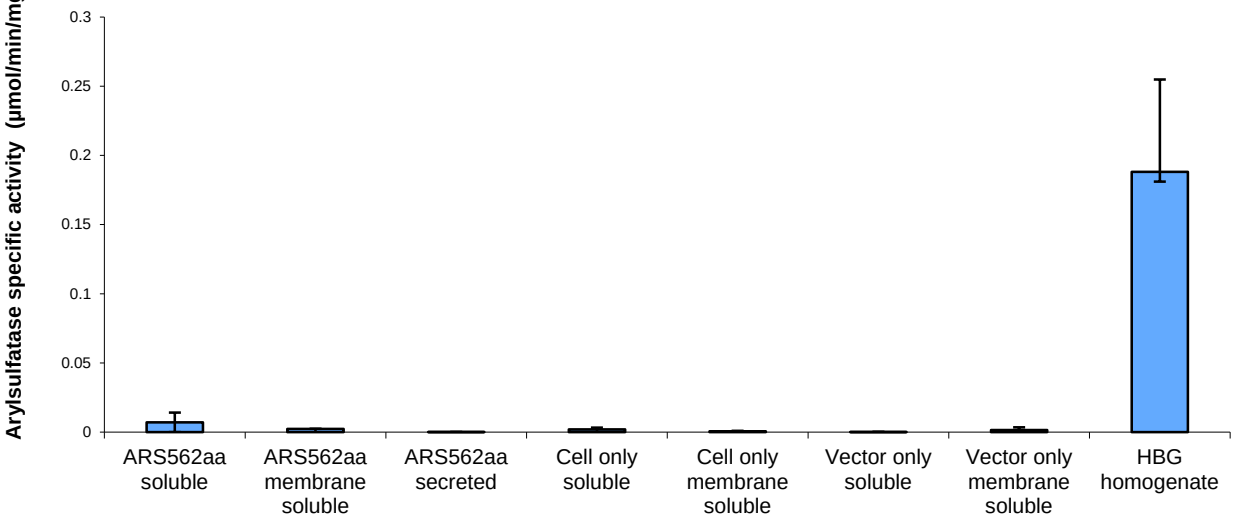

Hpomatia_SULF1
Hpomatia_SULF2
Tcastan_ARS
Hdiscus_ARS
Bflorid_ARS
Cintest_ARS
Dorb_ARS
ARS571aa
ARS562aa
signal cleavage site
Conserved motif I
Conserved motif II

Figure 6.2 Multiple sequence alignment showing sequence features of both ARS_{571aa} and ARS_{562aa} from *D. orbita*.
Hpomatia_SULF1 is sequence of sulfatase 1 from *Helix pomatia* AAF30402.1, Hpomatia_SULF2 is sequence of sulfatase 2 from *H. pomatia* AAF30403.1, Tcastan_ARS is the predicted arylsulfatase sequence from the red flour beetle *Tribolium castaneum* XP972832.1, Hdiscus_ARS is the arylsulfatase sequence from the disk abalone *Haliotis discus discus* AB026612.1, Bflorid_ARS is the arylsulfatase sequence from the hypothetical arylsulfatase protein from the Florida lancelet *Branchiostoma floridae* XP_002592251.1, Cintest_ARS is the arylsulfatase sequence from the *Ciona intestinalis* XM_002121051.1 and Dorb_ARS is the full length translation of the ARS protein sequences (Both ARS_{571aa} and ARS_{562aa}) produced from *D. orbita*. The predicted translation starting sites for ARS_{571aa} and ARS_{562aa} are labelled on the Dorb_ARS sequence, as is the signal cleavage site predicted by SignalP (expasy, www.expasy.com/tools). Both conserved arylsulfatase motifs as described in Hanson *et al.* (2004) (conserved motif I and conserved motif II) are labelled on the multiple sequence alignment. Shading indicates the level of sequence similarity. Black residues indicate where 100% sequence conservation is observed between sequences, dark grey indicates 80% sequence conservation between sequences and residues shaded in light grey indicate a conservation level between sequences greater than 60%. The first residue of conserved motif I is represented as a C, but will be modified to a Cα-formylglycine in active sulfatases (unmodified cysteine is denoted by an * underneath the consensus sequence).

A


Arylsulfatase specific activity (µmol/min/mg)
0.3
0.25
0.2
0.15
0.1
0.05
0
ARS562aa soluble
ARS562aa membrane soluble
ARS562aa secreted
Cell only soluble
Cell only membrane soluble
Vector only soluble
Vector only membrane soluble
HBG homogenate

B

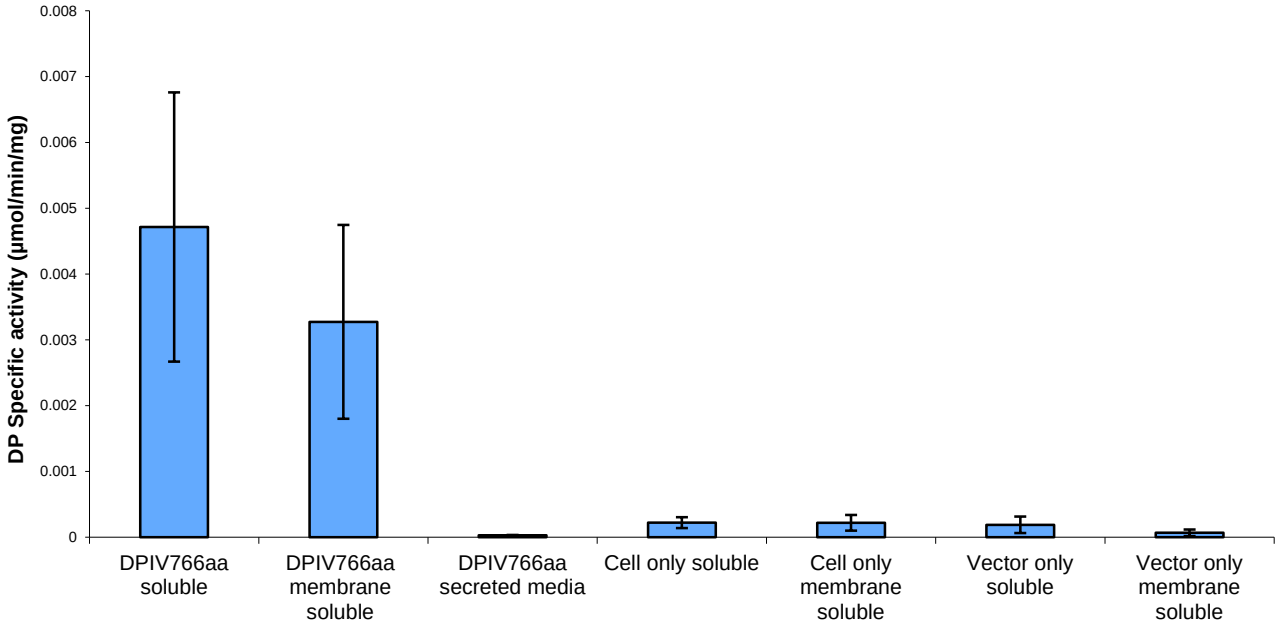

DP Specific activity (µmol/min/mg)
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
DPIV766aa soluble
DPIV766aa membrane soluble
DPIV766aa secreted media
Cell only soluble
Cell only membrane soluble
Vector only soluble
Vector only membrane soluble

Figure 6.3 Arylsulfatase (A) and dipeptidyl peptidase (B) specific activity of transiently transfected HEK293T cell extracts.

A) details the arylsulfatase activity of HEK293Tcells transiently transfected with ARS_{562aa} (N=3). Enzyme activity is expressed as μmol substrate cleaved per minute per mg of protein(μmol/min/mg), using an extinction coefficient of 12.6 mM-1.cm-1 for pNCS. B) details the Dipeptidyl Peptidase (DP) activity of HEK293T cells transiently transfected with $DPIV_{766aa}$ (N=3). Enzyme activity is expressed as μmol substrate cleaved per minute per mg of protein (μmol/min/mg), using an extinction coefficient of 9.45 mM-1.cm-1 for pNA (H-Ala-Pro-pNA) These results are a representative of three experiments.

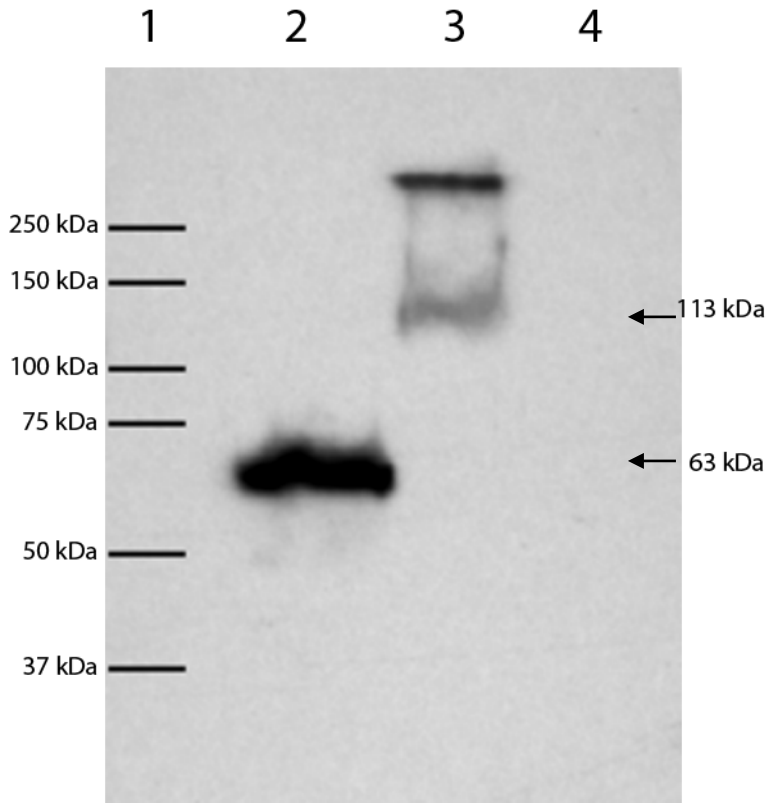


Figure 6.4 ARS562aa and DPIV766aa transfected cells produce His tagged proteins.
Total cell extracts were run on 10% SDS-PAGE gel and transferred to PVDF. Membrane was probed with monoclonal anti-V5 antibody raised in mouse. Rabbit anti-mouse was used as a secondary antibody and chemiluminescence reagent was applied before exposure for 5 min on X-ray film. Lane 1 contains Biorad precision plus protein standards; Lane 2 contains total cell extract from ARS_{562aa} transfected HEK293T cells; Lane 3 contains total cell extract from $DPIV_{766AA}$ transfected HEK293T cells; Lane 4 contains total cell extract from pcDNA vector-only transfected cells. Bands in Lane 1 are labelled with the molecular weight of the standard bands. Individual bands are labelled with sizes calculated using QualityOne quantitation software (Biorad).

Chapter 7. Final discussion and future directions

Brominated Tyrian purple precursors hold great potential for the development of future novel pharmaceutical and nutraceutical treatments. The distribution, prevalence and chemical composition of these brominated indoles has been historically documented throughout scientific literature. However, the enzymatic production and molecular processes involved in the formation of these compounds within the hypobranchial gland of muricid molluscs is still poorly understood. This thesis investigated the evolution of brominated indole investment within the egg capsules of the Muricidae, and identified this as a synapomorphic trait (Chapter two). In addition, a cDNA library from the hypobranchial gland of *D. orbita* identified 352 sequences, produced using SSH, of which, 110 sequences had their functions annotated (Chapter four). A further 188 novel sequences were found to have no homology to any known sequences in GenBank (Chapter four). This investigation also identified 57 genes with significant homology to ciliate protozoan gene sequences which appeared to use a ciliate alternate codon translation system (Chapters three and five). Ribosomal RNA sequences produced from this cDNA library were also investigated, and the presence of two different ciliates within the hypobranchial gland of *D. orbita* was identified, highlighting the potential SSH has for use in symbiosis studies (Chapter five). Finally, a full length arylsulfatase gene expressed in the hypobranchial gland of *D. orbita* was cloned and recombinantly expressed in mammalian tissue culture (Chapter six). Although this study has dramatically increased the knowledge of molecular mechanisms within the hypobranchial gland of *D. orbita*, further investigations are still needed in order to better understand Tyrian purple biogenesis and egg capsule investment.

7.1 The evolutionary significance of Tyrian purple investment in muricid egg capsules.

Molecular phylogenetic investigation into the investment of brominated indoles within the egg capsules of muricid molluscs has identified that this chemical investment is an evolutionary trait that has arisen multiple times within the Muricidae (Chapter two). Investigation into additional species within each subfamily will be required to better understand the evolution of this form of parental protection in muricid molluscs. A recent publication has reported an extensive molecular phylogenetic analysis of the Muricidae (Barco *et al.* 2010). In this study, nine of the ten widely accepted muricid subfamilies were included in a multigene molecular analysis of one nuclear (28s rRNA) and three mitochondrial genes (12s rRNA, 16s rRNA and Cytochrome oxidase I) for 72 different muricid molluscs. This study confirmed the monophyly of Rapaninae,

Ocenebrinae and Haustrinae families as we have suggested in chapter two, as well as Ergalataxinae, Coralliophilinae and Typhinae and reported polyphyly in the subfamilies Trophoninae, Muricinae and Muricopsinae (Barco *et al.* 2010). If egg capsule data can be collected from the additional species included within this study or any other family members, we will be able to further investigate the investment of Tyrian purple precursors within the egg capsules of the Muricidae, identifying how many times this trait has arisen in muricid evolution. The findings of this study also support the use of multiple genes in the molecular phylogenetic analysis of the Muricidae, and the inclusion of additional sequences within our dataset may facilitate more conclusive results.

7.2 Transcriptomic investigations of *D. orbita*

The investigation into gene expression within the hypobranchial gland of *D. orbita* uncovered several interesting findings of particular importance regarding the function and role of the hypobranchial gland in muricid molluscs. Gene expression patterns, identified using automated sequence annotation (Chapter four), identified several biological processes that are being undertaken within the gland. The expression of genes involved in mucus production and Tyrian purple biosynthesis support previous studies that identify the hypobranchial gland as the source of Tyrian purple precursors and mucus secretion within muricid (Baker 1974, Fretter & Graham 1994, Cooksey 2001, Westley & Benkendorff 2009). In order to confirm the capacity of SSH to isolate differentially expressed genes within the hypobranchial gland, we performed Quantitative realtime PCR and confirmed that sequences identified from our cDNA library were being differentially expressed in the hypobranchial gland compared to the mantle of *D. orbita*. Surveying gene expression from a larger cohort of individuals may facilitate a better understanding of hypobranchial gland gene expression which would strengthen our findings.

A paucity of genetic information is available on molluscan gene expression and our investigation into sequence homology assignments from genes expressed in the hypobranchial gland of *D. orbita* highlights the difficulty in identifying gene sequences when there is limited genetic information known from a closely related species (Chapter 3). Nevertheless, this study dramatically increased our knowledge of muricid gene expression patterns, particularly within the hypobranchial gland and provides further support for the role of the gland as a biosynthetic organ, vital for the production of secondary metabolites (Chapter four). In order to identify the

function of novel sequences expressed in the hypobranchial gland of *D. orbita*, reverse genetic studies are needed to determine the function of these transcripts. Techniques such as siRNA enable the silencing of genes in many model systems (Kuznetsov 2003), however their application *in vivo* can be problematic (Higuchi *et al.* 2010), and additional advances in delivery methods will be needed before they can be applied in marine molluscs. The localization and identification of where these genes are being expressed within the hypobranchial gland, using i*n situ* hybridization (Jackson *et al.* 2006), may enable characterization of these novel sequences and suggest possible functionality of these transcripts. The capacity to identify transcripts from non-model organisms using homologous sequence comparisons will improve in the future, due to advances in technology and bioinformatics techniques (Chapters three, four and five). The trace sequence data for the genome of the Californian sea hair *Aplysia californica* is currently available (NCBI 2010), and once this genome has been annotated, it may facilitate the annotation and classification of many of our unknown *D. orbita* sequences. A full length transcriptome of another muricid *Concholepas concholepas* has recently been published in an unassembled form (Cardenas *et al.* 2011), and if further analysis is completed on this transcriptome, we may be able to greatly increase the amount of sequence homology inferences that can be made from our data.

The development of next generation sequencing techniques in recent years has revolutionized the field of gene sequencing and transcriptomics throughout the scientific community (Linnarsson 2010, Miller *et al.* 2010). Not only has the cost of sequencing been dramatically reduced from these new technologies, but the scope of genetic information that is produced from these studies has dramatically increased the biological information that can be garnered from studies that use these methods. While large scale genomic and transcriptomic investigations within non-model organisms have traditionally been out of reach (Wolf *et al.* 2010), advances in sequencing techniques, analysis software and the increasing amount of genetic information available makes the use of these genomic techniques increasingly viable for the study of a variety of different biological problems, where genomic investigations were previously thought of as unfeasible (Buggs *et al.* 2010, Goetz *et al.* 2010). It has already been seen that there is a connection between the production of brominated indoles within the hypobranchial gland of muricid molluscs and the investment of these compounds within egg capsules (Benkendorff *et al.* 2000, Benkendorff *et al.* 2004b, Westley & Benkendorff 2008, Westley & Benkendorff 2009). Large scale sequence analysis on the reproductive organs and gonoduct of *D. orbita* may enable us to determine if there relationship between these organs

and hypobranchial gland. In addition, full sequence coverage of the cDNA library produced in this study may enable a deeper understanding of hypobranchial gland gene expression in the future, particularly regarding the discovered symbiotic relationship with ciliate protozoa (Chapters four and five).

7.3 Symbiosis within the hypobranchial gland of *D. orbita*

Our use of SSH to identify differential expression within the hypobranchial gland of *D. orbita* has facilitated the identification of intracellular ciliate protozoans within hypobranchial gland cells (Chapter five). Ciliate protozoa have been identified within the gills of several bivalves (Elston *et al.* 1999, Conn *et al.* 2008, Dias *et al.* 2008), but this is the first report of ciliates within the hypobranchial gland of any mollusc. Our work illustrated that ciliate abundance within the hypobranchial gland is at its highest in copulating individuals, but varies over the seasonal cycle (Chapter five). Because no ciliates have previously been identified to form a symbiotic relationship with the hypobranchial gland of molluscs, it remains to be seen what kind of symbiotic relationship exists in the gland.

This study of gene expression patterns within the hypobranchial gland of *D. orbita* identified a novel use of SSH in the identification of multiple genomes within tissue samples and highlights the potential of this method as a useful tool for future investigations of symbiotic relationships. In order to confirm its usefulness in the identification of symbiotic relationships, this method should be applied under biological conditions where symbiosis is well characterized. Many important biological systems are dependant on the successful interactions of symbionts and hosts, and studies into symbiosis have implications for human health (Ewaschuk & Dieleman 2006), agriculture (Zhao & Qi 2008, Aanen *et al.* 2009), bioremediation (Gohre & Paszkowski 2006) and the stability of ecosystems (Read & Perez-Moreno 2003, Seilacher *et al.* 2007). Any of a number of these biological systems could be investigated using SSH, allowing us to confirm its capacity to identify symbiosis. The freshwater snails of the genus *Biomphalaria* are known to play role in the lifecycle of the parasitic trematode symbionts responsible for the human disease schistosomiasis (Raghavan & Knight 2006, Guillou *et al.* 2007) and application of SSH in the study of these molluscs may facilitate a deeper understanding about the genetic interactions between host and parasite. Furthermore, the more that is known about symbionts

gene expression, the more useful the application of this method will become, as gene expression profiles may allow us to infer the type of symbiotic relationships under investigation.

Further characterization of these intracellular ciliates is needed in order to both identify their role in the hypobranchial gland and to determine if they are present throughout all populations of *D. orbita*, other muricids and even other molluscs. Ciliate specific 18s rRNA primers have been shown to be effective in identifying family classifications of many ciliate species within environmental samples (Dopheide *et al.* 2008), and similar investigations could be performed on hypobranchial gland samples from other communities of *D. orbita* in Australia and New Zealand, as well as other members of the Muricidae and Neogastropoda in order to identify the frequency and diversity of ciliates in the hypobranchial glands of other molluscs. Microscopic live observation, silver nitrate staining and silver carbonate impregmentation have been used in taxonomic studies of ciliates to identify ciliate features and classify individuals (Foissner 1991) and similar investigations will enable us to further characterize ciliates within the hypobranchial gland of *D. orbita*.

It is unclear whether ciliate protozoans are involved in Tyrian purple biosynthesis, but as their abundance within the hypobranchial gland varies seasonally and Tyrian purple precursors are constitutively expressed in many members of the Muricidae, it is unlikely that ciliates play a main part of the biogenesis of these bioactive compounds. Despite this, symbiosis has been shown to play a critical role in the biogenesis of several different secondary metabolites from symbiotic dinoflagellates within several coral and invertebrate species (Kita *et al.* 2010), as well as in the symbiotic microbial production of bioactive secondary metabolites in sea sponges (Hentschel *et al.* 2006, Thomas *et al.* 2010). As discussed above, large scale sequencing of our hypobranchial gland library using next generation sequence technology, particularly when snails are copulating when ciliate abundance is at a maximum, may identify a much larger transcriptome from ciliates, which may help to identify their role in hypobranchial gland physiology and identify if they play any role in Tyrian purple production in *D. orbita*.

7.4 The identification of the enzymes involved in the formation of Tyrian purple precursors in the hypobranchial gland *of D. orbita.*

In Chapter four, an arylsulfatase sequence (713 bp) was identified and found to be expressed in the hypobranchial gland of *D. orbita*. Arylsulfatase activity has been identified within the hypobranchial gland of *D. orbita*, and plays a role in the production of brominated indole precursors to Tyrian purple (Baker & Duke 1973b, Westley & Benkendorff 2009). Subsequently, cloning of a full length arylsulfatase gene from the hypobranchial gland *of D. orbita* identified a gene sequence 2210bp long containing an open reading frame (Chapter six). Despite the expression and detection of the full length tagged gene in mammalian cells, no corresponding measurable enzyme activity was observed. The post translational modification of a cysteine to formylglycine within a conserved domain of all arylsulfatases is essential for the activation of enzyme activity (Hanson *et al.* 2004). To our knowledge, no other study has successfully expressed any invertebrate arylsulfatase in a mammalian expression system despite many successful expressions of other vertebrate arylsulfatases within these systems, indicating that the compatibility of eukaryotic sulfatase-modifying factors is highly specific between different phyla.

A solution that may enable the expression of active recombinant arylsulfatase from *D. orbita* would be to express a tagged *D. orbita* gene within an invertebrate expression system. It has already been shown that the expression of molluscan sulfatases within yeast expression systems fails to yield active protein (Wittstock *et al.* 2000), however, the use of the baculovirus expression system may produce active recombinant arylsulfatase (Anderson *et al.* 1996). Baculoviral expression involves the transfection of insect cell lines or larvae with a viral expression vector capable of producing the desired gene product. Although it is unclear if sulfatase modifying factors are compatible between the Mollusca and the Insecta, it may be the most viable option we have of producing recombinant arylsulfatase from *D. orbita*. While primary cell culture of molluscan cells have been established *in vitro* (Poncet *et al.* 2000, Odoemelam *et al.* 2009, Odintsova *et al.* 2010), there is very little literature detailing recombinant gene expression within these cell lines and it remains to be seen whether these culture systems are suitable for recombinant expression.

Several other enzymes have been suggested to be involved in the formation of Tyrian purple precursors, including tryptophanase and bromoperoxidase, in addition to the aforementioned arylsulfatase. Despite the recent findings of Westley *et al.* (2009) who described enzymatic activity of a bromoperoxidase enzyme within the hypobranchial gland of *D. orbita,* PCR amplification using bromoperoxidase-designed primers failed to yield a bromoperoxidase product (Appendix III). Similarly, PCR amplification was used in an attempt to amplify tryptophanase sequence, but failed to yield enzyme sequence (Appendix III). Our findings indicate that no gene sequence could be amplified using the degenerate gene-specific primers we designed, and suggests that tryptophanase and bromoperoxidase share a low level of sequence conservations with known homologous sequences and may have unique sequence features. Approximately 54% of sequences produced in our cDNA library from the hypobranchial gland *of D. orbita* were novel sequences which showed no homology to any other sequence in GenBank (Chapter four), and it is possible that, although we haven't functionally identified or annotated a bromoperoxidase or tryptophanase sequence from our library, we may still have sequence data pertaining to these enzymes. Further investigations such as the synthesis of an expression library (Holz *et al.* 2001) containing hypobranchial gland expressed genes and screening of this library for bromoperoxidase and tryptophanase activity may help identify the source of enzyme activity, and may allow for the synthetic production of Tyrian purple precursors. Proteomics techniques such as two dimensional PAGE gel electrophoresis and peptide sequencing (Simonian *et al.* 2009) may be used to identify proteins specifically expressed in the hypobranchial gland in comparison to other tissues, uncovering the peptide sequences involved in Tyrian purple biogenesis.

Molecular investigations into the production of secondary metabolites in microbes has identified several gene clusters in marine invertebrates and bacteria that are responsible for the expression of polyketide synthase genes and non ribosomal peptide synthetases essential for secondary metabolite production (Salomon *et al.* 2004). These gene clusters facilitate the synthesis of many complex natural chemical products (Henning *et al.* 2002), and it is currently unknown whether the arylsulfatase involved in the production of Tyrian purple precursors is a component of a larger gene cluster involved in secondary metabolite synthesis. The identification of the chromosomal localization of the *D. orbita* arylsulfatase gene identified in this thesis and the sequencing of adjacent genes, may identify whether this gene is a part of a larger gene cluster, and may identify additional enzymes that are involved in the production of these bioactive Tyrian purple precursors.

7.5 Conclusions

The findings presented in this thesis clearly demonstrate the value of using molecular techniques to investigate the biosynthetic production of bioactive secondary metabolites. The genomics age is ever expanding. As more sequencing information is available and technologies become more accessible and affordable, there will be an increased demand for using molecular techniques to unlock the secrets of many biological problems. This study highlights the application of molecular techniques to study a species with very little known genetic information. From the findings of this thesis we have been able to uncover a wealth of knowledge about the enzymatic production of Tyrian purple, identified the evolutionary significance of these compounds within the Muricidae and helped to identify the main functional role of the hypobranchial gland as a biosynthetic organ in muricids. In addition, we have identified an uncharacterized symbiotic relationship within *D. orbita,* and a novel application for SSH methodology in the field of symbiosis has been proposed. Despite these promising findings, future investigations will be required in order to fully understand the investment of brominated Tyrian purple precursors within the egg capsules of the Muricidae. Additional studies will also be required on ciliate protozoans within the hypobranchial gland of *D. orbita*, in order to both taxonomically classify and further characterize their symbiotic relationships. Further molecular techniques may also be applied in order to elucidate remaining enzymes involved in Tyrian purple biogenesis within the hypobranchial gland of *D. orbita*. This will enable the development of a sustainable *D. orbita* based therapy to be used by the pharmaceutical and nutraceutical industry.

Appendix I. Additional molecular phylogenetic analysis of the Muricidae

Chapter two details the study of the evolution of Tyrian purple precursor investment within the egg capsules of the Muricidae and presents a molecular phylogenetic combined analysis using species for which we had both 18s and 28s ribosomal RNA sequences. The work presented in this appendix presents the Bayesian and maximum parsimony analysis based on either18s or 28s rRNA sequences, incorporating all additional muricid 18s and 28s rRNA sequences available in the GenBank

AI.1 Methods

Additional muricid 18s and 28s sequences were collected from GenBank. Sequences were aligned together with the sequences produced in chapter two and flanking 5' and 3' regions not displaying full coverage across all sequence were clipped and removed from alignment files. Methods for analysis were performed as described in chapter two for both Bayesian and maximum parsimony analysis. Sequences used in analyses are listed in Table AI.1, and include all sequences produced in chapter two as well as all homologous muricid sequences and representative sequences from other members of the Muricoidean superfamily, for each 18s and 28s ribosomal RNA sequence under investigation.

Table AI.1 Ribosomal RNA sequences used in molecular phylogenetic studies and their GenBank accession numbers

Superfamily FAMILY Subfamily	Species	18s rRNA accession number	28s rRNA accession number
Muricoidea			
MURICIDAE			
Rapaninae	*Dicathais orbita*	HM486916	HM486926
	Agnewia. tritoniformis	HM486912	HM486921
	Concholepas concholepas	HM486913	HM486923
	Thais clavigera	X91979	-
	Rapana vinosa	X98826	-
	Reishia bronni	X98827	-
Haustrinae	*Lepsiella flindersi*	HM486917	HM486928
	Lepsiella. vinosa	HM486918	HM486927
Ocenebrinae	*Acanthina monodon*	HM486911	HM486922
	Chorus giganteus	HM486914	HM486924
	Xanthochorus. cassidiformis	HM486920	HM486930
	Nucella lapillus	-	AJ420264.1
	Ocenebra erinacea	-	AF327546
Muricinae	*Pterynotus triformis*	HM486919	HM486929
	Bolinus brandaris	DQ279944	DQ279986
Buccinoidea			
Buccinidae	*Pisania striata*	X94272	-
	Buccinum undatum	-	AF327548
Nassaridae	*Ilyanassa obsoleta*	EU087578	-
Conoidea			
Conidae	*Raphitoma linearis*	DQ279945	-
Volutidae			
Mitridae	*Mitra cucumerina*	-	AY296899
Volutidae	*C hunteri*	-	DQ916584
Littorinimorpha: Ranellidae			
	Cabestana spengleri	HM486915	HM486930

- Sequences not used in analysis

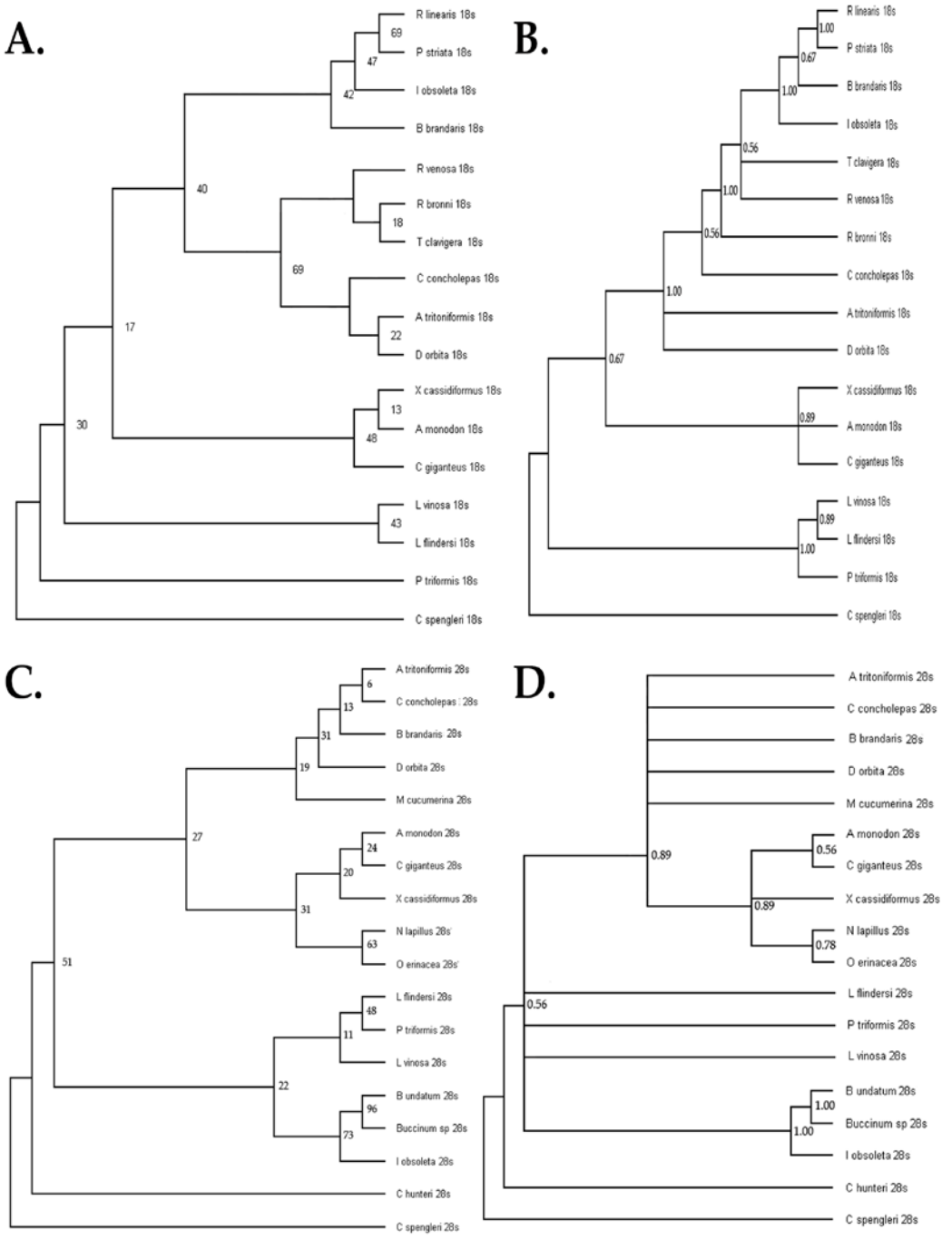


Figure AI.1 Maximum parsimony and Bayesian analysis of 18s and 28s ribosomal RNA fragments from muricid molluscs.

All trees rooted with *C. spengleri* sequence. A and B display analysis of 18s sequences. C and D display analysis of 28s sequences. A and C display bootstrapped maximum parsimony analysis with 100 replicates, bootstrap values as shown at tree nodes. B and D display Bayesian analysis with minimum mcmcp value of 3000000, and PP scores as shown at tree nodes.

AI.2 Results

AI.2.1 18s rRNA phylogenetic single gene analysis

Differing topologies were identified in 18s rRNA single gene analyses when comparing maximum parsimony and Bayesian analyses. Fig AI.1A shows maximum parsimony analysis of 18s ribosomal RNA, and Fig AI.1B shows the corresponding Bayesian analysis. Bootstrap support values for Fig AI.1A fail to support all node formation. In addition, multiple polytomies from Bayesian analysis were identified (Fig AI.1B)

AI.2.2 28s rRNA phylogenetic single gene analysis

Differing topologies were also observed from 28s rRNA single gene analysis when comparing maximum parsimony and Bayesian analyses. Fig AI.1C shows maximum parsimony analysis of 28s ribosomal RNA sequences from the Muricidae, and Fig AI.1D shows the corresponding Bayesian analysis. Bootstrap support failed to support all node formation, with the exception of members of the Buccinidae Family (Fig AI.1C). Bayesian analysis of the 28s rRNA sequences resulted in multiple polytomies at several nodes (Fig AI.1D).

AI.3 Conclusion

The lack of support for bootstrap values from 18s and 28s rRNA single gene analysis from the Muricidae indicated that the single gene analyses did not result in adequate parsimony informative sites within the multiple sequence alignment. Bayesian inferences also indicated that clear phylogenetic resolution between species was not possible for single 18s rRNA analyses. Similarly, 28s single gene analysis failed to clearly resolve the molecular phylogeny for either Bayesian or maximum parsimony analysis. Complex polytomial node formation and poor bootstrap support values also indicated that 28s single gene analysis did not provide enough sequence variation to resolve the phylogeny of our sequences. It was for this reason that sequence information was pooled together in order to effectively resolve the molecular phylogeny of the Muricidae in order to identify the evolutionary significance of Tyrian purple precursor egg capsule investment in chapter two. As 18s and 28s rRNA sequence information was not available for both genes for all species presented, the final combined phylogenetic analysis was performed only on sequences described in chapter two.

Appendix II. Arylsulfatase semi purification using nickel beads

In an attempt to purify and concentrate the amount of recombinant arylsulfatase enzyme produced in chapter six, Nickel affinity purification was performed. After purification was performed, a visible band was enriched from the ARS_{562aa} membrane soluble fraction. The enriched band was excised and underwent digestion and peptide sequencing, but it was found that the enriched band appeared to be a human GAPDH protein. As this process did not yield an enriched arylsulfatase protein, it was not reported in the manuscript. The raw peptide sequence results are included at the end of this Appendix.

AII.1 Methods

AII.1.1 Semi-purification of ARSshort_V5His purification

Protein samples were purified using ProBond Nickel-Chelating resin (Invitrogen) to enrich all Histidine tagged proteins within the sample. ProBond beads were washed with sterile water then centrifuged for 1 min at 5000 rpm to collect beads in the bottom of the tube. Supernatant was removed and beads were washed with 1ml bead washing buffer (20mM Tris-HCl pH 8.0, 10% glycerol, 0.1% DDM, 20mM imidazole). Beads were centrifuged for 1 min at 5000 rpm, then the supernatant was removed and the beads were washed with 1 ml bead washing buffer a further two times. Beads were centrifuged at 5000 RPM and resuspended in a 1:1 ratio of bead washing buffer. 100µl of ProBond beads were applied to each of the 6 protein samples; ARS_{562aa} soluble fraction, ARS_{562aa} membrane soluble fraction, ARS_{562aa} media (secreted fraction), $DPIV_{766aa}$ soluble fraction, cell only control soluble fraction, and pcDNA A vector only control soluble fraction. Samples were incubated on a rotating wheel overnight at 4°C. Beads were washed 3× with bead washing buffer, removed each time by centrifugation at 5000 RPM. His6 tagged proteins were eluted from nickel beads with denaturing elution buffer (8M urea, 20mM Sodium Phosphate pH 4.0, 500mM NaCl).

AII.1.2 Trypsin digestion of protein bands

Protein bands that were separated using SDS-PAGE agarose gel electrophoresis (Fig AII.2.1 underwent trypsin digestion and excision in order to identify the protein(s) in the band. After the

gel was photographed and the band for excision was identified, the gel was placed on clean glass plate and the band was excised using a new sterile scalpel blade. The gel band was placed in a fresh low binding eppendorff tube and washed with 50/50 (v/v) acetonitrile/water for 15 minutes. Tube was flick mixed to bring liquid to the bottom of the tube and liquid was removed via pipetting. This washing step was performed a further two times until the coomassie stained band was clear. The gel plug was then covered with acetonitrile, which was removed once the gel plug had shrunk to a small white plug. The gel plug was rehydrated in 100mM ammonium bicarbonate for 5 mins before an equal volume of acetonitrile was added to the gel plug. The reaction was incubated for 15 mins before all the liquid was removed and the plug was dried overnight at 37°C. The gel plug was rehydrated in 10mM DTT, 100mM ammonium bicarbonate. The mixture was incubated for 45 mins at 65°C in thermomixer. The solution was then removed and an equal volume of 100mM iodoacetamide in 100mM ammonium bicarbonate was added to the gel plugs. The reaction was incubated for 30 min at 30°C in thermomixer in the dark. Iodoacetamide solution was removed and gel pieces were washed with a 50:50 v/v acetonitrile:water solution. The solution was removed before being repeated an additional two times. Gel pieces were dried in a vacuum concentrator. Gel pieces were then covered in digestion solution (100mM ammonium bicarbonate, 0.5 mM $CaCl_2$ and 12.5 ng/µl Trypsin gold) and incubated on ice for 45 min. Excess digestion solution was removed from gel pieces and 15µl of 100mM ammonium bicarbonate was added, then incubated overnight at 37°C. More 100mM ammonium bicarbonate was then added to the gel pieces to ensure they were sufficiently covered and incubated for 15 min at room temperature. An equal volume of acetonitrile was then added to the gel pieces and incubated at room temperature for 15 min. Supernatant was removed and transferred to a fresh low binding eppendorff tube. Extraction was repeated twice using 5% formic acid instead of acetonitrile, and supernatant was transferred to the low binding eppendorff tube along with the acetonitrile extraction. Samples were concentrated down to a final volume of 1-2ul using the rotary evaporator and resuspended in 15µl of formic acid and stored at 4°C before mass spectrometry analysis was performed.

All mass spectrometry (MS) sequencing and database searching was performed under the direction of Dr. Tim Chataway at the Flinders Proteomics Facility (Flinders University, Adelaide, South Australia). Digested peptides were analysed with a Thermo LTQ XL linear ion trap mass spectrometer fitted with a nanospray source (Thermo Electron Corp, San Jose, CA). The samples were applied to a 300 µM i.d x 5 mm C18 PepMap 100 precolumn and separated

on a 75 µM x 150 mm C18 PepMap 100 column using a Dionex Ultimate 3000 HPLC (Dionex Corp, Sunnyvale, CA) with a 55 min gradient from 2% acetonitrile to 45% acetonitrile containing 0.1% formic acid at a flow rate of 200 nl/min followed by a step to 77% acetonitrile for 9 min. The mass spectrometer was operated in positive ion mode with one full scan of mass/charge (m/z) 300-2000 followed by product ion scans of the three most intense ions with dynamic exclusion of 30s and collision induced dissociation energy of 35%. Mass spectrometry spectra were searched with Bioworks v3.3 (Thermo Electron Corp, San Jose, CA) using the Sequest algorithm using the swissprot database containing all swissprot protein sequences (updated 25/09/2009).

AII.2 Results

AII.2.1 Purification of Histidine tagged proteins.

The ARS_{562aa} protein expressed in the HEK293t cells was concentrated using metal-affinity chromatography. A single band of approx 60 kDa proteins was enriched from the membrane fraction (Fig AII.1). No visible proteins were enriched from the ARS_{562aa} soluble fraction, the vector only soluble fraction or the cell only soluble fraction. Following the enrichment process several prominent were identified from the ARS_{562aa} media but none of these were the correct size for recombinant ARS_{562aa} or the size of the band enriched from ARS_{562aa} V5His transfected cells.

AII.2.2 Trypsin digestion of protein bands

The visibly enriched band from the ARS_{562aa} membrane soluble fraction was extracted, digested and sequenced via mass spectroscopy. When we look at the frequency of peptide matches for specific proteins (see Fig AII.2) we can see that 44% of our peptides match known Glyceraldehyde-3-Phosphate dehydrogenase (GAPDH) sequences (287 peptide matches). The second most dominant peptide match, with 26% of peptide matches was to keratin sequences (175 peptide matches), followed by matches to malate dehydrogenase with 7% of all peptide matches (48 matches). Fructose-biphosphate aldolase proteins matched 3% of our peptides (19 matches, 3% of our matches were also to NADPH quinone reductase (19 matches) and an additional 1% of our peptide matches were to trypsin (7 matches). The remaining 16 % of peptide matches were to other proteins (106 matches).

Table AII.1 Contents of lanes of coomassie gel containing nickel bead purified extracts. All samples that were purified were produced from our third transient transfection of HEK293t cells and were performed on fresh cell extracts. Each lane contained 20ul of purified cell extract.

Lane	contents	Lane	Contents
1	ARS_{562aa} media	11	Empty Lane
2	Purified ARS_{562aa} media	12	Empty lane
3	ARS_{562aa} soluble	13	Vector only soluble
4	Purified ARS_{562aa} soluble	14	Purified vector only soluble
5	ARS_{562aa} membrane soluble	15	Cell only soluble
6	Purified ARS_{562aa} membrane soluble	16	Purified cell only soluble
7	Protein Kaleidoscope ladder	17	Protein Kaleidoscope ladder
8	$DPIV_{766aa}$ soluble	18	ARS_{562aa} media
9	Purified $DPIV_{766aa}$ soluble	19	Purified ARS_{562aa} media
10	AbeM purified V5/His6 tagged protein	20	AbeM purified V5/His6 tagged protein

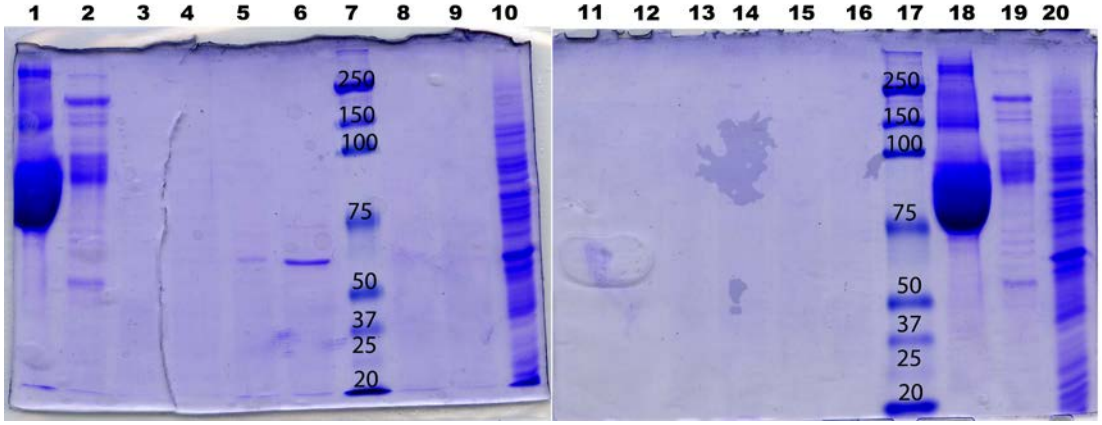


Figure AII.1 Total protein in samples before and after nickel bead purification.
Lane contents described in Table A.2.1 20 µl of protein extracts were run on 10% SDS PAGE gel electrophoresis at 170V for 1 hr and stained in Coomassie blue. Protein standard band sizes (kDa) is listed above the corresponding bands. A distinct band was observed in the ARS_{562aa} membrane soluble fraction (lane 5) which was concentrated upon purification (lane 6).

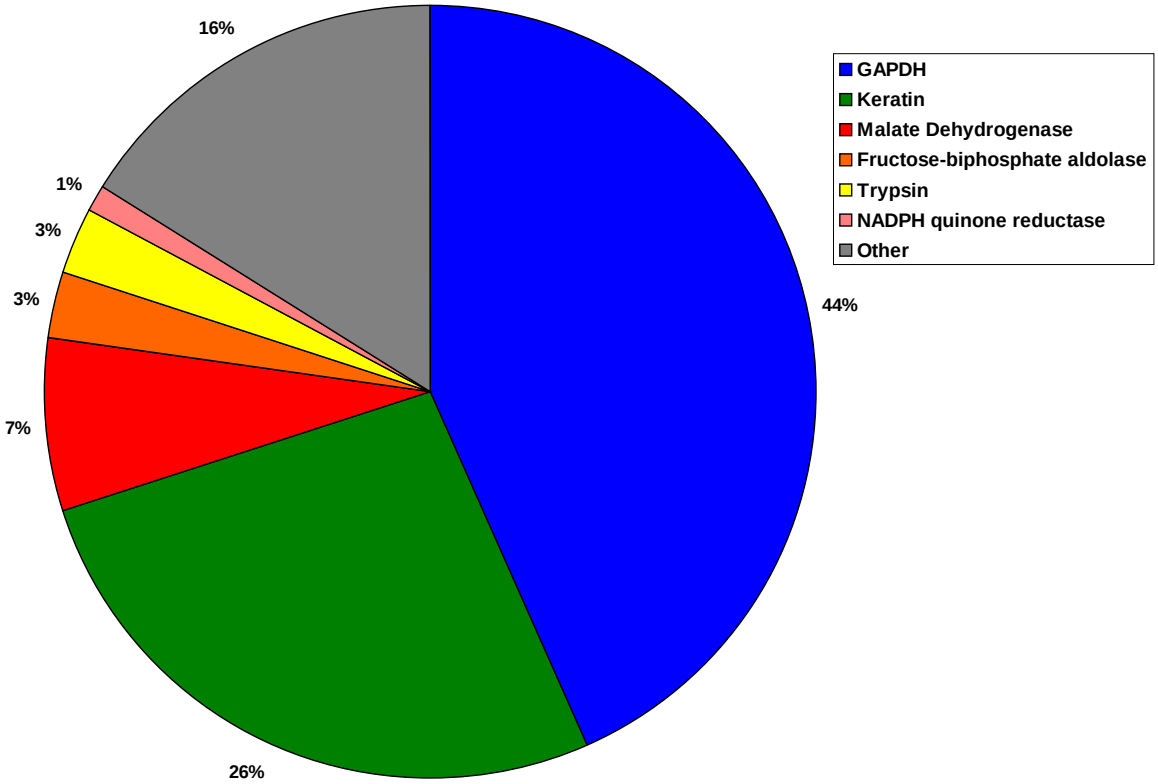


Figure AII.2 Mass spectroscopy peptide analysis of the enriched protein band identified from ARS_{562aa} membrane soluble fraction of transiently transfected 293t cell extracts.
A total of 661 peptides were sequenced from our trypsin digested gel slice. Mass spectrometry spectra were searched with Bioworks v3.3 (Thermo Electron Corp, San Jose, CA) using the Sequest algorithm using the swissprot database containing all swissprot protein sequences (updated 25/09/2009).

AII.3 Conclusion

Despite nickel affinity purification being performed to enrich for proteins with Histidine tags, we failed to purify or enrich the recombinantly expressed arylsulfatase enzyme. Western blot analysis confirms the expression of a 63 kDa recombinant protein corresponding to the *D. orbita* ARS_{562aa} protein. It is unclear why this recombinant arylsulfatseprotein was not enriched from nickel affinity purification, but possible competition with other proteins containing histidine residues may have outcompeted the binding of tagged recombinant protein. The enriched band visible in the membrane soluble purified fraction (Lane 6, Fig AII.1) appeared to contain GAPDH expressed in the mammalian cells. No arylsulfatase or sulfatase sequence was identified from the band digestion and subsequent peptide sequencing, but 44% of the resulting peptides corresponded with GAPDH sequences. GAPDH, a protein important in glycolysis, forms complexes with other glycolytic enzymes (Campanella *et al.* 2005). Human GAPDH (NP_002037) does not contain long stretches of histidine residues that could interact and outcompete histidine tagged recombinant proteins in binding to Nickel beads, however, if GAPDH has formed complexes with other proteins, this may explain its enrichment in nickel purification. GAPDH NP_002937 is approximately 38 kDa in size, and the presence of a much larger peptide band in our purified membrane soluble fraction indicates this NADPH is forming either a dimer or some kind of protein complex. The second most frequent peptide identified by mass spectrometry was for human keratin. Typically, the identification of human keratin is indicative of human hair contamination during the electrophoresis or peptide digestion and is not necessarily indicative of the presence of keratin protein within gel bands (Shevchenko *et al.* 2006).

Future investigations will investigate whether any bands produced from Nickel purification can be detected via western blot analysis before peptide sequencing is performed. At the time Nickel purification was performed, western blot analysis using an Anti-His antibody was failing to detect both the expressed arylsulfatase and the DPIV positive control, and it wasn't until after purification and peptide sequencing was performed that a fresh anti-V5 antibody (chapter six) successfully detected recombinant protein.

SRF File: E:\data\Patrick\09-12-15\PL09A1swissprot.srf
Database... indexed - swissprot_08_10_03.fasta.hdr (9/25/2008)
Filter(s)... xc (± 1,2,3)=1.50,2.00,2.50 ; peptide probability<=1e-003
Mods: (STY* +79.96633) (M# +15.99492) (C@ +58.00548)

Reference Scan(s)	Sequence	MH+	z	P P	Score XC	Coverage DeltaCn	Sp	Accession RSp	Ions
G3P_RAT Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (38 k				1e-015	106.3	20.1		122065190	
8869	K.RVIISAPSADAPM#FVM#GVNHEK.Y	2401.20	3	3e-006	4.139	0.257	778.4	1	31/84
8878	K.RVIISAPSADAPM#FVM#GVNHEK.Y	2401.20	3	2e-006	3.814	0.233	646.2	3	30/84
9187	K.RVIISAPSADAPMFVM#GVNHEK.Y	2385.21	3	0.0009	4.676	0.223	1011.9	1	32/84
9316	R.VIISAPSADAPM#FVM#GVNHEK.Y	2245.10	2	1e-007	3.849	0.373	353.9	1	17/40
9352	R.VIISAPSADAPM#FVM#GVNHEK.Y	2245.10	2	6e-008	3.387	0.429	535.9	1	20/40
9558	K.RVIISAPSADAPMFVM#GVNHEK.Y	2385.21	3	1e-005	4.046	0.028	909.8	2	30/84
9558	K.RVIISAPSADAPM#FVMGVNHEK.Y	2385.21	3	0.0001	3.933	0.126	969.4	1	31/84
9649	R.VIISAPSADAPMFVM#GVNHEK.Y	2229.10	2	5e-007	3.518	0.029	778.8	1	22/40
9649	R.VIISAPSADAPM#FVMGVNHEK.Y	2229.10	2	4e-008	3.416	0.362	697.0	2	21/40
9655	R.VIISAPSADAPMFVM#GVNHEK.Y	2229.10	2	6e-008	4.451	0.014	832.3	1	22/40
9655	R.VIISAPSADAPM#FVMGVNHEK.Y	2229.10	2	4e-009	4.390	0.301	659.9	2	20/40
9719	K.RVIISAPSADAPMFVMGVNHEK.Y	2369.21	2	2e-005	5.072	0.416	662.1	1	21/42
9769	K.RVIISAPSADAPMFVMGVNHEK.Y	2369.21	2	1e-012	5.286	0.486	1037.1	1	25/42
10033	K.RVIISAPSADAPMFVMGVNHEK.Y	2369.21	3	7e-014	5.340	0.172	2376.0	1	39/84
10166	R.VIISAPSADAPMFVMGVNHEK.Y	2213.11	2	3e-008	3.145	0.289	549.1	11	18/40
10173	R.VIISAPSADAPMFVMGVNHEK.Y	2213.11	2	2e-007	3.370	0.122	545.4	6	18/40
10299	R.VIISAPSADAPMFVMGVNHEK.Y	2213.11	2	7e-011	4.656	0.126	966.7	1	24/40
10306	R.VIISAPSADAPMFVMGVNHEK.Y	2213.11	2	4e-010	4.563	0.146	826.3	1	23/40
10780	K.WGDAGAEYVVESTGVFTTMEK.A	2277.04	2	0.0003	3.756	0.284	549.7	1	17/40
11353	K.WGDAGAEYVVESTGVFTTMEK.A	2277.04	2	1e-009	5.942	0.612	2180.7	1	27/40
11379	K.VIHDNFGIVEGLM#TTVHAITATQK.T	2611.35	2	0.0002	4.502	0.520	612.0	1	22/46
11399	K.VIHDNFGIVEGLM#TTVHAITATQK.T	2611.35	3	7e-010	5.009	0.179	1059.3	1	35/92
11428	K.WGDAGAEYVVESTGVFTTMEK.A	2277.04	2	1e-010	5.586	0.602	1901.9	1	25/40
11851	K.VIHDNFGIVEGLMTTVHAITATQK.T	2595.36	3	2e-008	6.639	0.246	1291.9	1	38/92
11878	K.VIHDNFGIVEGLMTTVHAITATQK.T	2595.36	3	1e-015	6.753	0.235	1701.4	1	43/92
25 of 46 peptide matches reported, 21 removed due to filtering									
K1C9_HUMAN RecName: Full=Keratin, type I cytoskeletal 9; AltNa				4e-012	90.3	0.0		81175178	
7093	R.SGGGGGGGLGSGGSIR.S	1232.60	2	0.0002	4.184	0.446	1076.1	1	22/30
7181	R.FSSSSGYGGGSSR.V	1235.53	2	4e-005	2.824	0.311	439.4	20	14/24
7203	K.STMQELNSR.L	1065.50	2	0.0003	2.187	0.130	755.1	2	13/16
7522	R.GGSGGSHGGGSGFGGESGGSYGGGEEASGSGGGYG	3223.28	3	0.0001	5.607	0.430	1362.6	1	37/156
8305	K.VQALEEANNDLENK.I	1586.77	2	2e-007	4.671	0.380	971.9	1	20/26
8705	K.VQALEEANNDLENK.I	1586.77	2	2e-006	4.021	0.343	1089.4	1	20/26
8825	R.IKFEMEQNLR.Q	1307.68	2	0.0002	2.491	0.051	464.2	4	13/18
8836	R.IKFEMEQNLR.Q	1307.68	2	9e-005	2.350	0.115	383.8	10	12/18
9695	K.EIETYHNLLEGGQEDFESSGAGK.I	2510.13	3	2e-005	4.404	0.305	667.2	30	28/88
10150	K.EIETYHNLLEGGQEDFESSGAGK.I	2510.13	3	5e-005	3.726	0.075	420.1	43	24/88
10381	R.HGVQELEIELQSQLSK.K	1837.97	2	1e-007	5.355	0.503	1854.2	1	23/30
10429	R.HGVQELEIELQSQLSK.K	1837.97	2	4e-012	5.575	0.511	2322.1	1	23/30
11095	K.DIENQYETQITQIEHEVSSSGQEVQSSAK.E	3264.51	3	1e-006	4.635	0.310	1061.6	1	35/112
11107	K.DIENQYETQITQIEHEVSSSGQEVQSSAK.E	3264.51	3	1e-009	4.892	0.386	1214.7	1	34/112
14 of 28 peptide matches reported, 14 removed due to filtering									
K2C1_HUMAN RecName: Full=Keratin, type II cytoskeletal 1; AltN				5e-013	80.3	0.0		1346343	
6895	K.SKAEAESLYQSK.Y	1340.67	2	2e-006	3.952	0.356	1232.8	1	18/22
6916	K.SKAEAESLYQSK.Y	1340.67	2	1e-007	4.193	0.405	1528.6	1	19/22
7234	R.GGGGGGYGSGGSSYGSGGGSYGSGGGGGGGR.G	2383.95	2	5e-013	5.754	0.621	1092.9	1	24/60
7252	R.GGGGGGYGSGGSSYGSGGGSYGSGGGGGGGR.G	2383.95	2	1e-009	6.016	0.656	1359.7	1	27/60
7291	K.SKAEAESLYQSK.Y	1340.67	2	2e-006	4.157	0.323	1041.6	1	17/22
7306	K.NM#QDM#VEDYR.N	1332.52	2	0.0001	2.441	0.151	433.6	4	12/18
7329	K.AEAESLYQSK.Y	1125.54	2	6e-005	2.900	0.448	709.0	1	14/18
7343	K.AEAESLYQSK.Y	1125.54	2	7e-005	2.986	0.464	823.1	2	14/18
7514	R.GSYGSGGSSYGSGGGSYGSGGGGGGHGSYGSGSSS	3312.31	3	1e-012	4.655	0.587	857.8	1	36/152
7524	R.GSYGSGGSSYGSGGGSYGSGGGGGGHGSYGSGSSS	3312.31	3	6e-010	4.327	0.359	582.8	8	29/152
9130	R.SGGGFSSGSAGIINYQR.R	1657.79	2	2e-007	4.162	0.379	1928.9	1	22/32
9139	R.SGGGFSSGSAGIINYQR.R	1657.79	2	1e-006	3.773	0.359	1160.1	1	20/32
9149	K.NMQDMVEDYR.N	1300.53	2	0.0004	2.249	0.143	373.0	112	12/18
9196	K.NMQDMVEDYR.N	1300.53	2	0.0003	2.460	0.164	684.3	2	14/18
9875	K.LNDLEDALQQAK.E	1357.70	2	0.0004	4.379	0.324	2285.5	1	19/22
9888	K.LNDLEDALQQAK.E	1357.70	2	0.0006	3.934	0.268	1539.7	1	17/22
16 of 28 peptide matches reported, 12 removed due to filtering									
G3P_FELCA Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)				4e-009	72.2	3.6		62286684	
6808	R.VVDLM#AHM#ASKE	1362.64	2	3e-005	3.016	0.308	455.9	1	16/22
6814	R.VVDLM#AHM#ASKE	1362.64	2	1e-006	2.673	0.280	374.1	1	14/22
7184	R.VVDLM#AHM#ASKE	1362.64	2	6e-005	2.493	0.094	411.4	1	15/22
7193	R.VVDLM#AHM#ASKE	1362.64	2	1e-005	2.935	0.228	469.6	1	16/22
7829	R.VVDLMAHM#ASK.-	1217.60	2	0.0008	3.009	0.026	931.8	2	17/20
7835	R.VVDLMAHM#ASK.-	1217.60	2	0.0002	2.547	0.059	757.0	2	16/20
7875	R.VVDLMAHM#ASKE.-	1346.64	2	6e-005	2.638	0.013	919.9	2	18/22
7875	R.VVDLM#AHMASKE	1346.64	2	0.0008	2.431	0.230	1119.9	1	19/22
8626	R.VVDLMAHMASK.E	1201.61	1	0.0002	2.299	0.179	656.4	1	13/20
8636	R.VVDLMAHMASK.E	1201.61	1	0.0001	2.509	0.126	713.4	1	14/20
8682	R.VVDLMAHMASK.E	1201.61	2	8e-007	3.734	0.128	1613.0	1	18/20
8689	R.VVDLMAHMASK.E	1201.61	2	1e-006	3.825	0.113	1644.9	1	18/20
8706	R.VVDLMAHM#ASKE.-	1346.64	2	0.0002	2.222	0.023	380.7	35	13/22
8706	R.VVDLM#AHMASKE.-	1346.64	2	1e-005	2.145	0.068	408.0	24	13/22
8920	R.VVDLMAHMASKE	1330.65	2	4e-009	4.027	0.108	1942.2	1	19/22
8930	R.VVDLMAHMASKE	1330.65	2	2e-008	3.392	0.097	1069.4	1	19/22

9060	R.VVDLMAHMASK.E	1201.61	2	3e-007	3.094	0.115	1402.9	1	18/20
9067	R.VVDLMAHMASK.E	1201.61	2	4e-008	3.247	0.109	1450.0	1	18/20
9296	R.VVDLMAHMASKE	1330.65	2	4e-007	3.306	0.135	1152.5	1	16/22
9310	R.VVDLMAHMASKE.-	1330.65	2	3e-006	2.187	0.012	355.0	29	13/22

20 of 43 peptide matches reported, 23 removed due to filtering

K1C10_HUMAN Keratin, type I cytoskeletal 10 (Cytokeratin-10) (8e-011 50.3 0.0 147744568

7056	R.SQYEQLAEQNRK.D	1493.73	2	0.0004	3.002	0.298	497.3	5	13/22
7062	R.SQYEQLAEQNRK.D	1493.73	2	0.0009	3.089	0.263	459.2	3	13/22
7496	R.SLLEGEGSSGGGGR.G	1262.60	2	1e-005	2.822	0.396	878.4	1	20/26
7872	R.SQYEQLAEQNR.K	1365.64	2	3e-005	3.597	0.000	1328.9	1	16/20
9663	K.GSLGGGFSSGGFSGGSFSR.G	1707.77	2	8e-011	5.353	0.000	1720.8	1	23/36
9951	K.ELTTEIDNNIEQISSYK.S	1996.97	2	3e-006	3.373	0.206	791.3	2	18/32
9959	K.ELTTEIDNNIEQISSYK.S	1996.97	2	1e-005	3.456	0.255	972.0	1	19/32
10511	K.ELTTEIDNNIEQISSYK.S	1996.97	2	2e-005	3.396	0.111	411.5	1	16/32

8 of 11 peptide matches reported, 3 removed due to filtering

MDHM_PIG Malate dehydrogenase, mitochondrial precursor 1e-010 40.3 13.6 2506849

8029	K.AGAGSATLSM#AYAGAR.F	1470.70	2	3e-005	2.913	0.355	504.7	6	15/30
9096	K.AGAGSATLSMAYAGAR.F	1454.71	2	4e-005	4.006	0.397	1419.5	1	21/30
9103	K.AGAGSATLSMAYAGAR.F	1454.71	2	0.0008	3.677	0.328	1299.3	1	20/30
10761	K.IFGVTTLDIVR.A	1233.72	2	0.0002	2.651	0.041	551.1	32	14/20
11133	K.IFGVTTLDIVR.A	1233.72	2	3e-005	2.835	0.043	769.2	5	15/20
11571	K.VAVLGASGGIGQPLSLLLK.N	1793.09	2	1e-010	4.882	0.184	1339.4	1	23/36
11598	K.VAVLGASGGIGQPLSLLLK.N	1793.09	2	2e-008	5.532	0.239	1705.3	1	24/36
11976	K.VAVLGASGGIGQPLSLLLK.N	1793.09	2	4e-008	5.561	0.237	1478.1	1	24/36
12144	K.VAVLGASGGIGQPLSLLLK.N	1793.09	2	6e-006	4.845	0.229	1147.9	1	20/36

9 of 13 peptide matches reported, 4 removed due to filtering

K22E_HUMAN Keratin, type II cytoskeletal 2 epidermal (Cytokera 5e-008 40.2 0.0 547754

6712	R.SKEEAEALYHSK.Y	1391.68	2	1e-006	4.365	0.380	1327.9	1	19/22
6733	R.SKEEAEALYHSK.Y	1391.68	2	8e-008	3.919	0.416	1100.0	1	18/22
6761	R.GSSSGGGYSSGSSSYGSGGR.Q	1740.71	2	5e-008	3.742	0.001	1175.5	1	21/38
6799	R.GSSSGGGYSSGSSSYGSGGR.Q	1740.71	2	2e-007	4.049	0.454	1043.6	1	23/38
7671	R.HGGGGGGFGGGGFGSR.S	1320.58	2	4e-005	3.371	0.456	964.2	1	22/30
7676	R.HGGGGGGFGGGGFGSR.S	1320.58	2	2e-007	3.251	0.424	361.8	1	16/30
11247	K.VDLLNQEIEFLK.V	1460.80	2	3e-006	3.672	0.242	1222.1	1	18/22
11276	K.VDLLNQEIEFLK.V	1460.80	2	3e-005	4.199	0.248	1525.8	1	19/22

8 of 17 peptide matches reported, 9 removed due to filtering

G3P_HUMAN Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) 6e-012 40.2 11.3 120649

9193	R.GALQNIIPASTGAAK.A	1411.79	2	0.0003	4.253	0.277	817.0	1	21/28
9204	R.GALQNIIPASTGAAK.A	1411.79	1	2e-005	2.477	0.383	487.9	1	15/28
9571	R.GALQNIIPASTGAAK.A	1411.79	2	0.0007	3.904	0.221	755.6	1	20/28
9952	R.GALQNIIPASTGAAK.A	1411.79	2	0.0008	2.367	0.213	393.7	43	13/28
10435	K.LVINGNPITIFQERDPSK.I	2041.11	2	6e-012	3.404	0.237	760.2	1	19/34
10449	K.LVINGNPITIFQERDPSK.I	2041.11	2	2e-009	2.897	0.209	569.6	1	18/34
10587	K.AENGKLVINGNPITIFQER.D	2113.14	2	1e-005	2.943	0.262	585.9	1	16/36
10810	K.LVINGNPITIFQERDPSK.I	2041.11	2	0.0003	2.044	0.139	256.4	2	14/34
10985	K.LVINGNPITIFQER.D	1613.90	2	2e-008	3.081	0.206	1493.8	1	21/26
10992	K.LVINGNPITIFQER.D	1613.90	2	8e-006	3.220	0.179	1580.6	1	21/26
11363	K.LVINGNPITIFQER.D	1613.90	2	7e-009	3.196	0.105	1547.4	1	21/26
11734	K.LVINGNPITIFQER.D	1613.90	2	3e-008	3.340	0.190	1427.7	1	21/26
11741	K.LVINGNPITIFQER.D	1613.90	2	4e-008	3.228	0.151	1299.1	1	20/26
12282	K.LVINGNPITIFQER.D	1613.90	2	2e-006	2.434	0.230	1119.9	1	18/26
12462	K.LVINGNPITIFQER.D	1613.90	2	1e-006	2.914	0.105	1112.6	1	18/26
12631	R.GALQNIIPASTGAAK.A	1411.79	2	0.001	2.652	0.149	402.7	6	15/28
12641	R.GALQNIIPASTGAAK.A	1411.79	2	0.0004	3.205	0.303	662.4	1	18/28
12996	R.GALQNIIPASTGAAK.A	1411.79	2	0.0002	2.785	0.033	307.8	6	14/28
13541	K.LVINGNPITIFQER.D	1613.90	2	3e-007	3.110	0.257	1476.1	1	21/26
13898	K.LVINGNPITIFQER.D	1613.90	2	1e-005	2.610	0.120	975.7	1	18/26
14080	R.GALQNIIPASTGAAK.A	1411.79	2	0.0003	3.132	0.221	571.6	1	17/28
14266	K.LVINGNPITIFQER.D	1613.90	2	3e-007	2.969	0.234	1757.2	1	21/26
14286	K.LVINGNPITIFQER.D	1613.90	2	1e-005	2.506	0.079	1158.7	1	18/26
14289	K.LVINGNPITIFQER.D	1613.90	2	0.0009	2.445	0.101	1001.5	1	17/26
14657	K.LVINGNPITIFQER.D	1613.90	2	4e-006	2.252	0.134	790.4	1	15/26
15456	K.LVINGNPITIFQER.D	1613.90	2	3e-005	2.595	0.003	924.1	1	17/26
17303	R.GALQNIIPASTGAAK.A	1411.79	2	0.0003	2.808	0.211	312.7	35	13/28

27 of 52 peptide matches reported, 25 removed due to filtering

K1C10_BOVIN Keratin, type I cytoskeletal 10 (Cytokeratin-10) (1e-005 36.2 0.0 125074

7872	R.SQYEQLAEKNR.R	1365.68	2	3e-005	3.597	0.300	1328.9	1	16/20
8536	K.IRLENEIQTYR.S	1434.77	2	2e-005	2.802	0.217	864.7	2	16/20
8544	K.IRLENEIQTYR.S	1434.77	2	1e-005	2.820	0.196	736.1	1	14/20
8936	K.QSLEASLAETEGR.Y	1390.68	2	3e-005	2.652	0.271	398.2	3	15/24
9509	R.DAEAWFNEK.S	1109.49	2	1e-005	2.236	0.029	340.5	109	10/16

5 of 11 peptide matches reported, 6 removed due to filtering

G3P_MESAU Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) 1e-015 34.3 14.7 1730158

9721	R.VIISAPSADAPM#FVM#GVNHDK.Y	2231.08	2	0.0007	3.020	0.173	317.0	11	14/40
10033	K.RVIISAPSADAPMFVM#GVNHDK.Y	2371.19	3	0.0003	4.422	0.363	1406.3	2	33/84
11399	K.VIHDNFGIVKGLM#TTVHAITATQK.T	2610.41	3	2e-007	4.115	0.367	807.4	2	32/92
11851	K.VIHDNFGIVKGLMTTVHAITATQK.T	2594.41	3	3e-007	5.003	0.593	865.8	2	33/92

11878	K.VIHDNFGIVKGLMTTVHAITATQK.T	2594.41	3	1e-015	5.165	0.563	1420.1	2	41/92

5 of 22 peptide matches reported, 17 removed due to filtering

MDHM_HUMAN Malate dehydrogenase, mitochondrial precursor — 1e-015 30.3 14.2 6648067

8115	R.VNVPVIGGHAGK.T	1147.66	2	0.0007	2.790	0.187	563.9	30	16/22
10530	K.VDFPQDQLTALTGR.I	1560.80	2	0.0002	3.444	0.122	437.3	1	18/26
10537	K.VDFPQDQLTALTGR.I	1560.80	2	4e-005	3.877	0.162	373.4	1	18/26
10733	R.LTLYDIAHTPGVAADLSHIETK.A	2365.24	3	1e-014	6.820	0.410	3020.8	1	44/84
10740	R.LTLYDIAHTPGVAADLSHIETK.A	2365.24	3	1e-015	6.778	0.399	3613.4	1	46/84
10903	K.VDFPQDQLTALTGR.I	1560.80	2	0.0008	2.790	0.393	170.9	28	16/26

6 of 9 peptide matches reported, 3 removed due to filtering

MDHC_HUMAN Malate dehydrogenase, cytoplasmic (Cytosolic malate — 2e-009 28.2 11.4 1708967

7830	K.GEFVTTVQQR.G	1164.60	2	0.0006	3.202	0.302	1336.2	1	16/18
7856	K.GEFVTTVQQR.G	1164.60	2	3e-005	3.297	0.313	1368.8	1	16/18
8467	K.DVIATDKEDVAFK.D	1450.74	2	2e-008	4.179	0.384	1603.2	1	19/24
8473	K.DVIATDKEDVAFK.D	1450.74	2	3e-009	4.244	0.352	1829.9	1	19/24
8668	K.DVIATDKEDVAFK.D	1450.74	2	2e-009	3.893	0.429	1200.8	1	18/24
10553	K.EVGVYEALKDDSWLK.G	1751.89	2	3e-009	3.937	0.312	880.1	1	18/28
10595	K.EVGVYEALKDDSWLK.G	1751.89	2	2e-006	4.219	0.337	952.1	1	19/28

7 of 10 peptide matches reported, 3 removed due to filtering

G3P_XENLA Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) — 2e-007 24.2 3.6 1730164

7875	R.VVDLVCHM#ASKE.-	1346.64	2	0.0008	2.604	0.078	738.7	3	17/22
8682	R.VVDLVCHMASK.-	1201.61	2	2e-005	3.254	0.394	1054.6	2	16/20
8689	R.VVDLVCHMASK.-	1201.61	2	3e-005	3.392	0.323	1073.8	2	16/20
8920	R.VVDLVCHMASKE.-	1330.65	2	2e-007	3.590	0.360	1327.4	2	17/22
8930	R.VVDLVCHMASKE.-	1330.65	2	7e-007	3.063	0.429	1355.1	3	17/22
9060	R.VVDLVCHMASK.-	1201.61	2	9e-006	2.737	0.392	916.8	2	16/20
9067	R.VVDLVCHMASK.-	1201.61	2	2e-006	2.894	0.299	946.1	2	16/20
9296	R.VVDLVCHMASKE.-	1330.65	2	9e-006	2.859	0.283	919.5	2	15/22
9310	R.VVDLVCHMASKE.-	1330.65	2	6e-005	2.162	0.293	328.6	56	13/22

9 of 26 peptide matches reported, 17 removed due to filtering

TRYP_PIG Trypsin precursor — 6e-008 20.3 17.3 136429

9768	R.LGEHNIDVLEGNEQFINAAK.I	2211.10	2	2e-006	4.602	0.037	2100.5	1	24/38
9795	R.LGEHNIDVLEGNEQFINAAK.I	2211.10	2	6e-008	5.366	0.085	2242.5	1	26/38
10110	R.LGEHNIDVLEGNEQFINAAK.I	2211.10	3	0.0001	4.672	0.092	1285.4	1	31/76
10166	R.LGEHNIDVLEGNEQFINAAK.I	2211.10	2	0.0003	4.128	0.126	1596.7	1	24/38
10425	K.IITHPNFNGNTLDNDIMLIK.L	2283.18	2	4e-006	4.030	0.336	1537.9	1	24/38
10825	K.IITHPNFNGNTLDNDIMLIK.L	2283.18	2	3e-006	3.369	0.297	1375.7	1	23/38

6 of 19 peptide matches reported, 13 removed due to filtering

LDHB_RAT L-lactate dehydrogenase B chain (LDH-B) (LDH heart su — 5e-011 20.2 8.7 1170739

7172	K.IVADKDYSVTANSK.I	1510.77	2	9e-011	4.101	0.496	1501.5	1	20/26
7177	K.IVADKDYSVTANSK.I	1510.77	2	5e-011	4.059	0.385	1313.3	1	19/26
12178	K.SLADELALVDVLEDK.L	1629.86	2	2e-008	4.961	0.428	1785.2	1	20/28
12191	K.SLADELALVDVLEDK.L	1629.86	2	3e-006	4.412	0.373	1766.0	1	20/28

4 of 6 peptide matches reported, 2 removed due to filtering

ALDOA_HUMAN Fructose-bisphosphate aldolase A (Muscle-type aldo — 6e-010 20.2 4.1 113606

9363	K.FSHEEIAM#ATVTALR.R	1691.84	2	3e-008	2.879	0.155	717.2	1	16/28
9897	K.FSHEEIAMATVTALR.R	1675.85	2	6e-010	3.911	0.422	1456.9	1	21/28
9915	K.FSHEEIAMATVTALR.R	1675.85	2	5e-009	4.446	0.435	1591.4	1	21/28

3 of 4 peptide matches reported, 1 removed due to filtering

QOR_HUMAN Quinone oxidoreductase (NADPH:quinone reductase) (Ze — 8e-011 20.2 8.8 585013

6654	K.VAEAHENIIHGSGATGK.M	1690.85	2	8e-011	4.273	0.504	1166.3	1	21/32
6667	K.VAEAHENIIHGSGATGK.M	1690.85	2	6e-009	4.399	0.472	1262.3	1	21/32
7081	K.ILGTAGTEEGQK.I	1203.62	2	0.0001	2.618	0.302	654.4	1	16/22

3 of 5 peptide matches reported, 2 removed due to filtering

CASA1_BOVIN Alpha-S1-casein precursor [Contains: Antioxidant p — 4e-005 20.2 14.5 115646

9735	K.YKVPQLEIVPNS*AEER.L	1951.95	2	0.0005	4.351	0.354	433.5	1	22/45
9746	K.HQGLPQEVLNENLLR.F	1759.94	2	4e-005	3.366	0.390	615.7	1	17/28

2 of 8 peptide matches reported, 6 removed due to filtering

ALDOA_RAT Fructose-bisphosphate aldolase A (Muscle-type aldola — 5e-007 20.2 7.4 113609

7369	R.LQSIGTENTEENR.R	1490.71	2	0.0002	3.293	0.352	1222.2	1	19/24
7388	R.LQSIGTENTEENR.R	1490.71	2	0.0005	3.251	0.294	757.7	1	17/24
8456	K.GILAADESTGSIAK.R	1332.70	2	4e-006	3.934	0.178	1428.1	1	20/26
8490	K.GILAADESTGSIAK.R	1332.70	2	5e-007	4.142	0.145	1294.2	1	19/26

4 of 6 peptide matches reported, 2 removed due to filtering

K2C6A_RAT Keratin, type II cytoskeletal 6A (Cytokeratin-6A) (C — 0.0001 20.2 0.0 123781839

7394	K.AQYEEIAKR.S	1107.58	2	0.0004	2.301	0.000	448.3	16	12/16

7408	K.AQYEEIAKR.S	1107.58	2	0.0005	2.567	0.017	700.2	1	13/16
8766	K.YEELQITAGR.H	1179.60	2	0.0001	3.728	0.349	1860.1	1	17/18

3 of 8 peptide matches reported, 5 removed due to filtering

K1C17_HUMAN Keratin, type I cytoskeletal 17 (Cytokeratin-17) (2e-005 20.2 0.0 547751

7987	R.LEQEIATYR.R	1122.58	2	0.0004	2.252	0.089	503.1	1	12/16
8553	K.TRLEQEIATYR.R	1379.73	2	2e-005	3.281	0.282	929.6	1	16/20

2 of 7 peptide matches reported, 5 removed due to filtering

DCD_HUMAN Dermcidin precursor (Preproteolysin) [Contains: Surv 4e-005 20.1 0.0 20141302

6998	K.ENAGEDPGLAR.Q	1128.53	2	0.0003	2.862	0.244	783.8	1	15/20
9217	K.LGKDAVEDLESVGK.G	1459.76	2	4e-005	2.955	0.269	726.2	1	16/26

2 of 3 peptide matches reported, 1 removed due to filtering

G3PC_HORVU Glyceraldehyde-3-phosphate dehydrogenase, cytosolic 1e-006 20.1 0.0 120668

10639	K.LVSWYDNEWGYSTR.V	1775.80	2	1e-006	2.248	0.106	254.0	14	12/26
10756	R.VPTVDVSVVDLTVR.L	1498.85	2	2e-005	2.740	0.296	417.0	1	17/26

2 of 2 peptide matches reported, 0 removed due to filtering

K2C4_HUMAN Keratin, type II cytoskeletal 4 (Cytokeratin-4) (CK 2e-006 18.2 0.0 82654947

7394	R.AQYEEIAQR.S	1107.54	2	0.0004	2.301	0.079	448.3	16	12/16
7408	R.AQYEEIAQR.S	1107.54	2	0.0005	2.567	0.202	700.2	1	13/16
8489	K.VQQLQISVDQHGDNLK.N	1821.95	2	2e-006	4.180	0.293	937.9	1	19/30
8525	K.VQQLQISVDQHGDNLK.N	1821.95	2	6e-005	4.724	0.332	1175.5	1	21/30

4 of 9 peptide matches reported, 5 removed due to filtering

K2C73_HUMAN Keratin, type II cytoskeletal 73 (Cytokeratin-73) 0.0002 18.2 0.0 74750553

7286	K.SKAEAEALYQTK.F	1338.69	2	0.0002	3.751	0.180	1402.9	1	18/22
8815	R.FLEQQNQVLETK.W	1476.77	2	0.0002	3.691	0.024	1401.9	2	18/22

2 of 15 peptide matches reported, 13 removed due to filtering

WAKLE_ARATH Wall-associated receptor kinase-like 5 precursor 3e-006 18.2 0.0 116256120

10743	R.IRS*SPEDLEAHIENDDEEDQVMEIS*RE.-	3345.35	3	3e-006	3.538	0.066	596.3	29	38/208
10743	R.IRSS*PEDLEAHIENDDEEDQVMEIS*RE.-	3345.35	3	0.0002	3.303	0.148	566.1	51	37/208

2 of 2 peptide matches reported, 0 removed due to filtering

CISY2_BACSU RecName: Full=Citrate synthase 2; AltName: Full=Ci 0.0001 14.2 0.0 3123196

9460	K.RLT*NLTGES*KWYEMS*IR.I	2323.96	3	0.001	3.332	0.035	511.0	89	39/160
9460	K.RLT*NLTGES*KWY*EMSIR.I	2323.96	3	0.0001	3.109	0.138	528.1	66	39/160

2 of 2 peptide matches reported, 0 removed due to filtering

ALDOA_RABIT Fructose-bisphosphate aldolase A (Muscle-type aldo 2e-005 10.2 0.0 113608

10575	K.IGEHTPSALAIMENANVLAR.Y	2107.10	2	2e-005	4.881	0.240	1228.9	1	23/38
10768	K.IGEHTPSALAIMENANVLAR.Y	2107.10	3	0.0004	3.863	0.151	894.1	1	27/76

2 of 2 peptide matches reported, 0 removed due to filtering

G3P_BUCAP Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) 6e-008 10.2 0.0 1169790

9830	R.VPTANVSVVDLT*VR.Y	1549.80	2	1e-005	4.611	0.355	1751.7	1	27/39
9837	R.VPTANVSVVDLT*VR.Y	1549.80	2	3e-006	4.829	0.318	1337.9	2	24/39
10208	R.VPTANVSVVDLT*VR.Y	1549.80	2	4e-006	2.914	0.208	817.9	3	21/39
10215	R.VPTANVSVVDLT*VR.Y	1549.80	2	6e-008	3.279	0.278	1001.4	1	21/39

4 of 11 peptide matches reported, 7 removed due to filtering

G3P2_DROPS Glyceraldehyde-3-phosphate dehydrogenase 2 (Glycera 3e-009 10.2 0.0 115312169

10558	K.LISWYDNEFGYSNR.V	1763.80	2	1e-008	4.453	0.479	1744.8	1	20/26
10635	K.LISWYDNEFGYSNR.V	1763.80	2	5e-009	4.580	0.441	2200.5	1	21/26
10773	K.LISWYDNEFGYSNR.V	1763.80	1	2e-005	2.040	0.218	194.1	8	12/26
10999	K.LISWYDNEFGYSNR.V	1763.80	2	5e-007	4.587	0.413	1816.0	1	20/26
11007	K.LISWYDNEFGYSNR.V	1763.80	2	3e-009	4.583	0.448	1873.7	1	20/26
11387	K.LISWYDNEFGYSNR.V	1763.80	2	3e-007	4.388	0.370	1708.8	1	19/26
11394	K.LISWYDNEFGYSNR.V	1763.80	2	4e-008	4.392	0.458	1391.3	1	18/26
11767	K.LISWYDNEFGYSNR.V	1763.80	2	2e-005	4.525	0.450	1814.2	1	20/26
11793	K.LISWYDNEFGYSNR.V	1763.80	2	4e-008	4.509	0.507	1507.4	1	19/26
12159	K.LISWYDNEFGYSNR.V	1763.80	2	1e-007	4.509	0.450	1889.1	1	19/26
12198	K.LISWYDNEFGYSNR.V	1763.80	2	8e-007	4.210	0.472	1240.3	1	17/26
12606	K.LISWYDNEFGYSNR.V	1763.80	2	1e-006	4.342	0.447	1763.5	1	19/26
12613	K.LISWYDNEFGYSNR.V	1763.80	2	4e-009	4.148	0.437	1486.2	1	18/26
12962	K.LISWYDNEFGYSNR.V	1763.80	2	8e-006	4.479	0.453	1440.9	1	18/26
12971	K.LISWYDNEFGYSNR.V	1763.80	2	2e-005	4.128	0.385	1219.2	1	17/26
13552	K.LISWYDNEFGYSNR.V	1763.80	2	0.0002	2.801	0.324	525.5	2	16/26
13557	K.LISWYDNEFGYSNR.V	1763.80	2	6e-008	3.175	0.502	668.3	1	16/26
13916	K.LISWYDNEFGYSNR.V	1763.80	2	2e-005	4.339	0.496	1195.1	1	18/26
14315	K.LISWYDNEFGYSNR.V	1763.80	2	3e-006	3.304	0.459	491.8	1	17/26
14336	K.LISWYDNEFGYSNR.V	1763.80	2	3e-005	3.229	0.246	873.4	1	15/26

20 of 21 peptide matches reported, 1 removed due to filtering

ALDOC_MACFA Fructose-bisphosphate aldolase C (Brain-type aldol 1e-010 10.2 0.0 56748614

9973	K.GVVPLAGTDGETTTQGLDGLSER.C	2273.13	2	1e-010	4.562	0.302	446.6	1	20/44

1 of 1 peptide matches reported, 0 removed due to filtering

G3P_PARBR Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) 2e-005 10.2 0.0 30580398

9599	R.VPTANVSVVDLTC@R.T	1531.78	2	0.0007	4.094	0.387	1130.2	1	19/26
9606	R.VPTANVSVVDLTC@R.T	1531.78	2	2e-005	4.455	0.389	1189.4	1	20/26
10028	R.VPTANVSVVDLTC@R.T	1531.78	2	0.0005	3.234	0.318	772.7	4	16/26
11704	R.VPTANVSVVDLTC@R.T	1531.78	2	0.001	2.371	0.175	395.0	5	14/26
13039	R.VPTANVSVVDLTC@R.T	1531.78	2	0.0002	2.743	0.246	543.8	4	15/26

5 of 21 peptide matches reported, 16 removed due to filtering

K2C71_MOUSE Keratin, type II cytoskeletal 71 (Cytokeratin-71) 9e-008 10.2 0.0 60390219

8438	R.FLEQQNQVLQTK.W	1475.79	2	8e-007	4.364	0.046	2618.1	1	20/22
8446	R.FLEQQNQVLQTK.W	1475.79	2	8e-007	4.403	0.055	2801.5	1	20/22
8808	R.FLEQQNQVLQTK.W	1475.79	2	9e-008	3.733	0.053	1828.3	1	19/22
8815	R.FLEQQNQVLQTK.W	1475.79	2	4e-007	3.781	0.024	1401.9	2	18/22

4 of 6 peptide matches reported, 2 removed due to filtering

MDHC_CHICK Malate dehydrogenase, cytoplasmic (Cytosolic malate 3e-009 10.2 0.0 82082933

10553	K.EVGVYEAIKDDSWLK.G	1751.89	2	3e-009	3.937	0.000	880.1	1	18/28
10595	K.EVGVYEAIKDDSWLK.G	1751.89	2	2e-006	4.219	0.000	952.1	1	19/28

2 of 2 peptide matches reported, 0 removed due to filtering

LYSC_CHRAM Lysozyme C (1,4-beta-N-acetylmuramidase C) 7e-005 10.2 0.0 126609

9614	R.NTDGSTDYGILQINSR.W	1753.84	2	0.0003	3.569	0.281	436.4	7	15/30
9629	R.NTDGSTDYGILQINSR.W	1753.84	2	7e-005	4.211	0.075	1071.0	1	19/30

2 of 2 peptide matches reported, 0 removed due to filtering

K2C75_HUMAN Keratin, type II cytoskeletal 75 (Cytokeratin-75) 0.0002 10.2 0.0 74739699

11553	R.NLDLDSIIAEVK.A	1329.73	2	0.0002	3.836	0.246	1100.5	1	16/22

1 of 9 peptide matches reported, 8 removed due to filtering

K1C15_SHEEP Keratin, type I cytoskeletal 15 (Cytokeratin-15) (0.0001 10.2 0.0 75058787

8326	R.ALEEANADLEVK.I	1301.66	2	0.0001	3.920	0.299	1329.9	1	18/22

1 of 1 peptide matches reported, 0 removed due to filtering

CAS2_BOVIN Alpha-S2-casein precursor [Contains: Casocidin-1 (C 0.0002 10.2 0.0 115654

8564	K.TVDM#ES*TEVFTK.K	1482.61	2	0.0002	3.687	0.057	1600.5	2	24/33

1 of 4 peptide matches reported, 3 removed due to filtering

G3P_PICAN Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) 0.0003 10.2 0.0 6016084

9361	K.ILSNASCTTNCLAPLAK.V	1719.88	2	0.0005	3.218	0.000	988.7	1	18/32
9733	K.ILSNASCTTNCLAPLAK.V	1719.88	2	0.0005	3.570	0.000	1182.9	1	20/32
9740	K.ILSNASCTTNCLAPLAK.V	1719.88	2	0.0003	2.840	0.000	1012.0	1	18/32
11038	K.ILSNASCTTNCLAPLAK.V	1719.88	2	0.0005	2.304	0.000	552.2	1	15/32

4 of 10 peptide matches reported, 6 removed due to filtering

ACT_AJECA Actin 2e-006 10.2 0.0 1703149

9153	R.VAPEEHPVLLTEAPINPK.S	1954.06	2	2e-006	2.850	0.000	235.8	69	14/34
9216	R.VAPEEHPVLLTEAPINPK.S	1954.06	2	0.0001	3.524	0.000	941.1	1	21/34

2 of 2 peptide matches reported, 0 removed due to filtering

MDHC_ECHGR Malate dehydrogenase, cytoplasmic 4e-005 10.2 0.0 6016537

8831	K.VLVVGNPANTNCLIMSK.Y	1772.94	2	4e-005	3.214	0.121	965.5	1	19/32

1 of 3 peptide matches reported, 2 removed due to filtering

G3P_COLGL Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) 3e-006 10.2 0.0 462142

10299	K.VIISAPSADAPMY*VMVVNEK.S	2214.06	2	4e-005	3.371	0.343	261.9	29	16/57
10306	K.VIISAPSADAPMY*VMVVNEK.S	2214.06	2	0.0001	3.348	0.302	342.5	4	18/57
10671	K.VIISAPSADAPMY*VMVVNEK.S	2214.06	2	3e-006	3.173	0.071	221.5	32	15/57
10679	K.VIISAPSADAPMY*VMVVNEK.S	2214.06	2	4e-006	3.430	0.133	342.2	8	18/57

4 of 15 peptide matches reported, 11 removed due to filtering

IAA1_WHEAT Alpha-amylase inhibitor 0.19 (0.19 alpha-AI) (0.19 3e-006 10.2 0.0 123963

7104	K.EHGAQEGQAGTGAFPR.C	1612.75	2	3e-006	3.422	0.513	632.7	1	20/30
7108	K.EHGAQEGQAGTGAFPR.C	1612.75	2	3e-005	3.051	0.346	603.4	1	20/30

2 of 2 peptide matches reported, 0 removed due to filtering

```
ALF_MAIZE Fructose-bisphosphate aldolase, cytoplasmic isozyme     0.0001  10.2     0.0       113621
  8741        R.LSSINVENVEENR.R                    1502.74   2  0.0003  3.256  0.364   773.3  1   17/24
  8844        R.LSSINVENVEENR.R                    1502.74   2  0.0001  3.091  0.276   771.3  1   16/24

  2 of 2 peptide matches reported, 0 removed due to filtering


SSG1_ORYSJ RecName: Full=Granule-bound starch synthase 1, chlo    6e-006  10.2     0.0       122168648
  11996       R.VLTVSPYYAEELISGIAR.G                1981.06   2  6e-006  3.216  0.281   287.9  22  13/34

  1 of 2 peptide matches reported, 1 removed due to filtering


KIF27_MOUSE Kinesin-like protein KIF27                            0.0003  10.2     0.0       81894342
  6500        R.QELEMLEEDLKK.R                     1504.76   2  0.0003  3.027  0.113   886.3  8   15/22
  6505        R.QELEMLEEDLKK.R                     1504.76   2  0.0003  3.159  0.148   1016.4 5   15/22

  2 of 5 peptide matches reported, 3 removed due to filtering


K2C1_CANFA Keratin, type II cytoskeletal 1 (Cytokeratin-1) (CK    5e-005  10.2     0.0       75062693
  7882        R.TNAENEFVTIKK.D                     1393.73   2  5e-005  2.598  0.159   263.4  132 12/22

  1 of 7 peptide matches reported, 6 removed due to filtering


TOM1_ASHGO Probable E3 ubiquitin-protein ligase TOM1              0.0003  10.1     0.0       50401471
  7490        R.ELDAGFLNT*IPET*LR.Q                1848.82   3  0.0003  2.928  0.120   365.7  53  29/112

  1 of 3 peptide matches reported, 2 removed due to filtering


G3PC_LEIME Glyceraldehyde-3-phosphate dehydrogenase, cytosolic    0.0008  10.1     0.0       232117
  8008        K.TVDGPSLKDWR.G                      1273.65   2  0.001   2.914  0.142   544.0  4   13/20
  8022        K.TVDGPSLKDWR.G                      1273.65   2  0.0008  2.769  0.109   486.5  3   13/20

  2 of 2 peptide matches reported, 0 removed due to filtering


K1C15_HUMAN Keratin, type I cytoskeletal 15 (Cytokeratin-15) (    9e-005  10.1     0.0       125081
  8462        R.LKYENELALR.Q                       1248.69   2  9e-005  2.852  0.296   491.1  6   14/18

  1 of 3 peptide matches reported, 2 removed due to filtering


Y333_BACAN Uncharacterized RNA methyltransferase BA_0333/GBAA0    0.0007  10.1     0.0       50401540
  7172        K.VKGY*PIFVKNGLPGEEAQIK.I            2267.18   3  0.0007  2.790  0.060   605.8  7   28/114

  1 of 2 peptide matches reported, 1 removed due to filtering


ORF1A_TAV Replication protein 1a [Includes: ATP-dependent heli    0.0009  10.1     0.0       137257
  8250        -.MAASAFNIHKLVAS*HGDK.G              1976.94   3  0.0009  2.773  0.044   434.5  79  26/102

  1 of 2 peptide matches reported, 1 removed due to filtering


K2C5_BOVIN RecName: Full=Keratin, type II cytoskeletal 5; AltN    0.0002  10.1     0.0       75062316
  7201        K.YEELQQTAGR.H                       1194.57   2  0.0002  2.767  0.276   486.4  4   13/18

  1 of 4 peptide matches reported, 3 removed due to filtering


NU2M_RHISA NADH-ubiquinone oxidoreductase chain 2 (NADH dehydr    0.0002  10.1     0.0       8473787
  9959        K.Y*KLIKQILIFSSISHQGWILCLIAK.K        2995.66   3  0.0002  2.752  0.037   413.5  90  27/144

  1 of 1 peptide matches reported, 0 removed due to filtering


QOR_CAVPO Quinone oxidoreductase (NADPH:quinone reductase) (Ze    0.0002  10.1     0.0       117549
  8783        K.VHAC@GINPVETYIR.S                  1629.81   2  0.0002  2.691  0.115   314.4  3   13/26

  1 of 2 peptide matches reported, 1 removed due to filtering


DDB2_CHICK RecName: Full=DNA damage-binding protein 2; AltName    0.0004  10.1     0.0       82233793
  10813       K.LLSWEVEEM#GT*KTK.D                 1746.80   2  0.0004  2.593  0.132   365.2  74  15/39

  1 of 4 peptide matches reported, 3 removed due to filtering


Y856_MIMIV Putative TPR repeat-containing protein R856            9e-006  10.1     0.0       82000252
  7605        K.Y*PNKLFNIAFTIS*R.L                 1843.85   3  9e-006  2.596  0.062   665.8  1   29/104

  1 of 3 peptide matches reported, 2 removed due to filtering


SUCD_ECOLI RecName: Full=Succinyl-CoA ligase [ADP-forming] sub    6e-007  10.1     0.0       84027802
  7102        R.MGHAGAIIAGGK.G                     1082.58   2  6e-007  2.586  0.403   1128.4 1   19/22

  1 of 2 peptide matches reported, 1 removed due to filtering


RL14_CLOB1 50S ribosomal protein L14                              0.0001  10.1     0.0       166232505
  10335       R.VLGGSHRKWGNIGDVIVAS*VK.S           2272.20   3  0.0001  2.566  0.076   361.6  47  31/120
```

1 of 1 peptide matches reported, 0 removed due to filtering

RF3_SHEPA RecName: Full=Peptide chain release factor 3; Short= 0.0008 10.1 0.0 189040019

9795	R.KCSM#NLALDGGDNLTYIAPTMVNLNLSM#ER.Y	3316.57	3	0.0008	2.551	0.002	339.6	102	22/116

1 of 2 peptide matches reported, 1 removed due to filtering

ENO1_MAIZE Enolase 1 (2-phosphoglycerate dehydratase 1) (2-pho 9e-007 10.1 0.0 119355

9174	R.IEEELGDAAVYAGAK.F	1535.76	2	9e-007	2.531	0.309	306.9	1	14/28

1 of 1 peptide matches reported, 0 removed due to filtering

FBX5_XENTR F-box only protein 5 0.0007 10.1 0.0 118573326

9069	R.FEEAM#CST*LKK.M	1382.57	2	0.0007	2.454	0.024	363.8	7	16/30

1 of 1 peptide matches reported, 0 removed due to filtering

LACB_BOVIN Beta-lactoglobulin precursor (Beta-LG) (Allergen Bo 0.0002 10.1 0.0 125910

7878	R.TPEVDDEALEK.F	1245.58	2	0.0002	2.396	0.082	504.3	39	13/20

1 of 1 peptide matches reported, 0 removed due to filtering

GPD1_ZYGRO Glycerol-3-phosphate dehydrogenase [NAD+] 1 (ZrGPD1 0.0004 10.1 0.0 14285460

9240	K.KTDS*SMSVVTAENPY*K.V	1916.77	2	0.0004	2.347	0.132	331.2	3	21/60
9240	K.KTDSS*MSVVTAENPY*K.V	1916.77	2	0.0004	2.261	0.178	331.2	3	21/60

2 of 2 peptide matches reported, 0 removed due to filtering

DDAH1_RAT RecName: Full=N(G),N(G)-dimethylarginine dimethylami 0.0004 10.1 0.0 6831527

10310	K.DENATLDGGDVLFTGR.E	1679.79	2	0.0004	2.259	0.149	245.5	9	13/30

1 of 1 peptide matches reported, 0 removed due to filtering

K1C13_HUMAN RecName: Full=Keratin, type I cytoskeletal 13; Alt 2e-006 10.1 0.0 6016411

7958	K.ILTATIENNR.V	1144.63	2	2e-006	2.216	0.186	322.5	11	12/18

1 of 1 peptide matches reported, 0 removed due to filtering

UHPT_CHLTR RecName: Full=Probable hexose phosphate transport p 0.0008 10.1 0.0 9979133

9707	K.FVS*GVMS*DQS*NPR.Y	1663.56	2	0.0008	2.076	0.079	336.0	5	25/60

1 of 1 peptide matches reported, 0 removed due to filtering

AK1A1_PIG RecName: Full=Alcohol dehydrogenase [NADP+]; AltName 0.0005 10.1 0.0 1703236

10912	K.MPLIGLGTWK.S	1115.63	2	0.0005	2.032	0.000	765.3	2	13/18

1 of 1 peptide matches reported, 0 removed due to filtering

K1C13_PROAT Keratin, type I cytoskeletal 13 (Cytokeratin-13) (0.0007 10.1 0.0 82176030

7473	K.EAMQNLNDR.L	1090.49	2	0.0007	2.028	0.008	545.4	5	13/16

1 of 3 peptide matches reported, 2 removed due to filtering

HIS3_HUMAN Histatin-3 precursor (Histidine-rich protein 3) (PB 0.0007 10.1 0.0 123143

8640	R.SNYLYDN.-	888.37	1	0.0007	1.870	0.346	336.6	4	8/12

1 of 2 peptide matches reported, 1 removed due to filtering

K1C10_CANFA Keratin, type I cytoskeletal 10 (Cytokeratin-10) (8e-011 8.3 0.0 75043394

9663	K.GSIGGGFSSGGFSGGSFSR.G	1707.77	2	8e-011	5.353	0.594	1720.8	1	23/36

1 of 5 peptide matches reported, 4 removed due to filtering

TRY1_RAT Anionic trypsin-1 precursor (Anionic trypsin I) (Pret 6e-008 8.2 0.0 136409

9768	R.LGEHNINVLEGDEQFINAAK.I	2211.10	2	2e-006	4.431	0.348	1850.8	2	23/38
9795	R.LGEHNINVLEGDEQFINAAK.I	2211.10	2	6e-008	4.909	0.325	1875.5	2	24/38

2 of 7 peptide matches reported, 5 removed due to filtering

ENGA_STAS1 GTP-binding protein engA 0.0006 8.2 0.0 123642627

9938	K.EFQFLDY*AEIAFVSAKEK.Q	2215.04	2	0.0006	4.516	0.108	599.2	51	17/51

1 of 2 peptide matches reported, 1 removed due to filtering

MDHM_YEAST Malate dehydrogenase, mitochondrial precursor 8e-005 8.2 0.0 547901

11571	K.VTVLGAGGGIGQPLSLLLK.L	1793.09	2	0.0008	3.983	0.478	809.6	2	19/36
11598	K.VTVLGAGGGIGQPLSLLLK.L	1793.09	2	0.0003	4.212	0.450	1162.2	2	21/36
11976	K.VTVLGAGGGIGQPLSLLLK.L	1793.09	2	8e-005	4.242	0.415	913.9	2	20/36

3 of 4 peptide matches reported, 1 removed due to filtering

```
ACT22_DICDI Putative actin-22                                    2e-006  8.2      0.0       122057287
  9153        R.VAPEEHPVLLTEAPLNPK.A              1954.06   2  2e-006  2.850  0.229   235.8  69  14/34
  9216        R.VAPEEHPVLLTEAPLNPK.A              1954.06   2  0.0001  3.524  0.155   941.1  1   21/34

  2 of 6 peptide matches reported, 4 removed due to filtering


G3P_CANAL Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)       0.0003  8.2      0.0       2494636
  9361        K.IISNASCTTNCLAPLAK.V               1719.88   2  0.0005  3.218  0.039   988.7  1   18/32
  9733        K.IISNASCTTNCLAPLAK.V               1719.88   2  0.0005  3.570  0.191   1182.9 1   20/32
  9740        K.IISNASCTTNCLAPLAK.V               1719.88   2  0.0003  2.840  0.117   1012.0 1   18/32
  11038       K.IISNASCTTNCLAPLAK.V               1719.88   2  0.0005  2.304  0.124   552.2  1   15/32

  4 of 9 peptide matches reported, 5 removed due to filtering


ALF2_PEA Fructose-bisphosphate aldolase, cytoplasmic isozyme 2   0.0001  8.2      0.0       1168410
  8456        K.GILAADESTGTIGK.R                  1332.70   2  0.0007  3.234  0.446   941.8  3   17/26
  8490        K.GILAADESTGTIGK.R                  1332.70   2  0.0001  3.542  0.392   1051.5 2   18/26

  2 of 2 peptide matches reported, 0 removed due to filtering


UN13A_RAT Protein unc-13 homolog A (Munc13-1)                    0.0004  8.2      0.0       51316551
  10287       R.ILS*QRSNDEVAKEFVK.L               1942.96   2  0.0004  3.325  0.110   1094.3 1   23/45

  1 of 2 peptide matches reported, 1 removed due to filtering


MDHM_RAT Malate dehydrogenase, mitochondrial precursor           0.0005  8.2      0.0       122065494
  10537       K.VDFPQDQLATLTGR.I                  1560.80   2  0.0005  3.249  0.335   256.9  5   16/26

  1 of 3 peptide matches reported, 2 removed due to filtering


ALDOC_CARAU Fructose-bisphosphate aldolase C (Brain-type aldol   6e-007  8.2      0.0       1703244
  9973        K.GVVPLAGTNGETTTQGLDGLSER.C         2272.14   2  6e-007  3.182  0.440   355.0  2   18/44

  1 of 2 peptide matches reported, 1 removed due to filtering


HMCS_SCHPO Hydroxymethylglutaryl-CoA synthase (HMG-CoA synthas   0.0005  8.2      0.0       1708240
  9954        K.RVSPSVYAPTNCGNMY*TAS*IFSCLTALLSR.V 3382.51   3  0.0005  3.033  0.155   289.7  65  37/232

  1 of 1 peptide matches reported, 0 removed due to filtering


MDH_HAES1 Malate dehydrogenase                                   0.0005  8.1      0.0       123031394
  11133       K.LFGVTTLDVLR.S                     1233.72   2  0.0005  2.714  0.043   636.0  14  14/20

  1 of 4 peptide matches reported, 3 removed due to filtering


G3P2_TRIKO Glyceraldehyde-3-phosphate dehydrogenase 2 (GAPDH2)   3e-005  8.1      0.0       462132
  10215       K.LTGM#SIRVPTANVSVVDLT*VR.I         2324.20   3  3e-005  2.609  0.047   585.9  44  29/120

  1 of 1 peptide matches reported, 0 removed due to filtering


K1C27_MOUSE Keratin, type I cytoskeletal 27 (Cytokeratin-27) (   0.0007  8.1      0.0       81882110
  7473        K.VTMQNLNDR.L                       1090.53   2  0.0007  2.012  0.089   547.4  4   13/16

  1 of 4 peptide matches reported, 3 removed due to filtering


TIP30_MOUSE Oxidoreductase HTATIP2                               0.0009  8.1      0.0       78099267
  10961       K.EILGQNLFS*KVTLIGR.R               1868.00   2  0.0009  2.198  0.022   220.0  167 14/45

  1 of 1 peptide matches reported, 0 removed due to filtering


CATB_CHICK Cathepsin B precursor (Cathepsin B1) [Contains: Cat   0.0005  8.1      0.0       1168790
  9610        R.GEDHC@GIESEIVAGVPR.M              1825.84   2  0.0005  2.150  0.030   466.4  1   14/32

  1 of 1 peptide matches reported, 0 removed due to filtering


ALDR_HUMAN Aldose reductase (AR) (Aldehyde reductase)            0.0005  8.1      0.0       113596
  10912       K.MPILGLGTWK.S                      1115.63   2  0.0005  2.032  0.215   765.3  2   13/18

  1 of 3 peptide matches reported, 2 removed due to filtering


ZCH11_HUMAN Zinc finger CCHC domain-containing protein 11        0.0007  6.2      0.0       116242850
  10188       R.M#DDFQLKGIVEEKFVK.W               1942.00   2  0.0007  2.316  0.056   609.1  1   15/30

  1 of 6 peptide matches reported, 5 removed due to filtering


LDHB_PELSJ L-lactate dehydrogenase B chain (LDH-B)               4e-005  6.1      0.0       17368592
  12191       K.GLCDELALVDVLEDK.L                 1631.82   2  4e-005  2.642  0.419   1101.7 2   16/28

  1 of 2 peptide matches reported, 1 removed due to filtering
```

Y3709_CLOAB Uncharacterized protein CA_C3709				0.0007	6.1	0.0		19924238			
8510	K.LLTETVT*ALVEY*K.N	1639.76	2	0.0007	2.361	0.089	352.1	2	19/48		

1 of 2 peptide matches reported, 1 removed due to filtering

G3P1_CAEEL Glyceraldehyde-3-phosphate dehydrogenase 1 (GAPDH-1				7e-008	4.2	0.0		120639			
10166	K.VIISAPSADAPM#YVVGVNHEK.Y	2213.13	2	7e-008	2.934	0.325	522.5	13	17/40		
10173	K.VIISAPSADAPM#YVVGVNHEK.Y	2213.13	2	5e-005	3.170	0.340	585.2	3	18/40		
10299	K.VIISAPSADAPM#YVVGVNHEK.Y	2213.13	2	7e-008	4.067	0.276	536.0	3	19/40		
10306	K.VIISAPSADAPM#YVVGVNHEK.Y	2213.13	2	3e-006	3.897	0.266	445.3	3	18/40		

4 of 4 peptide matches reported, 0 removed due to filtering

IML1_CANGA Vacuolar membrane-associated protein IML1				0.0002	2.2	0.0		74608471			
10442	R.LLVGFQIC@TSSTIEEVESAR.K	2240.11	3	0.0002	3.004	0.131	744.3	31	27/76		

1 of 7 peptide matches reported, 6 removed due to filtering

ATM1_YARLI RecName: Full=Iron-sulfur clusters transporter ATM1				0.0004	2.1	0.0		74689505			
10999	K.YALVTGAT*M#VSYAIFTITTTSWR.T	2649.27	3	0.0004	2.620	0.138	517.3	66	28/132		

1 of 2 peptide matches reported, 1 removed due to filtering

FTSH2_SYNY3 Cell division protease ftsH homolog 2				0.0009	2.1	0.0		2492511			
9776	R.RSANASGQAM#SFGKS*K.A	1722.76	2	0.0009	2.509	0.116	426.7	8	18/45		

1 of 3 peptide matches reported, 2 removed due to filtering

Appendix III. PCR amplification of arylsulfatase, tryptophanase and bromoperoxidase sequences from the hypobranchial gland of *D. orbita*

In order to amplify PCR products corresponding to expressed arylsulfatase, bromoperoxidase and tryptophanase genes from *D. orbita*, PCR primers were designed from multiple sequence alignments from known enzyme sequences.

AIII.1 Methods

AIII.1.1 Primer design

PCR primers were designed for arylsulfatase, bromoperoxidase and tryptophanase PCR products. Table AIII.1 lists sequence information for all primers used. Primers were designed from nucleotide multiple sequence alignments of arylsulfatase, bromoperoxidase and tryptophanase sequences obtained from Genbank. Sequences used in multiple sequence alignments are listed in Table AIII.1. Due to the size of multiple sequence files, and the multiple number of alignments that were formed, alignments haven't been presented in this appendix. Multiple combinations of enzyme sequences were aligned together until consistant conserved multiple sequence alignments were produced. Primers were designed on consensus sequences of regions were conserved homology was identified. Final gene specific primers are listed in Table AIII.2 which also lists the estimated PCR product size.

Table AIII.1 Enzyme gene sequences used in the primer design for arylsulfatase, bromoperoxidase and tryptophanase sequences in *D. orbita*.

Sequence	Species	Accession #	length (bp)
Arylsulfatase	*Bos taurus* (cow)	BT030685	2088
	Canis lupus familiaris (dog)	NM_001048116	1797
	Danio rerio (zebrafish)	XM_687145	1987
	Helix pomatia (roman snail)	AF109924	2805
		AF109925	853
	Homo sapiens (human)	NM001085427	1873
	Macaca mulatta (rhesus monkey)	XM_001096903	2459
	Monodelphis domestica (opposum)	XM_001365962	1824
	Mus musculus (mouse)	BC098075	3522
		BC141169	1842
	Ornithorhynchus anatinus (playtpus)	XM_001514402	1794
	Pan troglodytes (chimpanzee)	XM_527073	2022
	Rattus norvegicus (rat)	NM_001047887	1794
	Strongylocentrotus purpuratus (purple sea urchin)	XM_001186405	2174
Bromoperoxidase	*Corallina officinalis* (algae)	AF218810	1908
	Corallina pilulifera (algae)	D87657	2035
		D87658	2029
	Fucus distichus (algae)	AF053411	2931
	Fucus vesiculosus (algae)	AF368992	773
	Laminaria digitata (algae)	AJ491786	3379
		AJ491787	3420
		AJ491787	3420
	Pseudomonas putida (bacteria)	AB034986	847
	Streptomyces aureofaciens (bacteria)	M84990	1711
		U01096	2112
	Streptomyces violaceus (bacteria)	X74791	2216
Tryptophanase	*Aedes aegypti (yellow fever mosquito)*	AF325458	1405
	Anopheles gambiae (African malarial mosquito)	L76432	1940
	Bos taurus (cow)	NM_001046313	1613
	Canis lupus familiaris (dog)	XM_532700	1494
	Chlamys farreri (scallop)	AY965263	1292
	Danio rerio (zebrafish)	NM_199856	1799
	Enterobacter aerogenes (bacteria)	D14297	2230
	Equus caballus (horse)	XM_001500725	1221
	Escherichia coli (bacteria)	X15974	1416
	Marmota monax (woodchuck)	AY253726	440
	Monodelphis domestica (opposum)	XM_001375244	1221
	Mus musculus (mouse)	U24493	1249
	Ornithorhynchus anatinus (playtpus)	XM_001520587	341
	Pan troglodytes (chimpanzee)	XM_001140187	1703
	Plodia interpunctella (Indianmeal moth)	AY427951	1450
	Proteus inconstans (bacteria)	AB019704	2285
	Proteus vulgaris (bacteria)	M93277	3148
	Symbiobacterium thermophilium (bacteria)	AB010832	7896

Table AIII.2 Primers for arylsulfatase, bromoperoxidase and tryptophanase products.

Primer name	5'-3' primer sequence	Approx PCR product size
ARS_for	gtttatcctagctgatgactagggc	~220 bp
ARS_rev	aatcccaggtgccacttgcc	
BPO_for	tacgaggagcctccctcccgagcc	~1900 bp
BPO_rev	gagctcggtgtgtccgtggag	
Tryp_for	gcttatgaactctggtttaagc	~660 bp
Tryp_rev	caatgctccctgaagtgc	

AII.1.2Total RNA extraction and cDNA synthesis

Total RNA was extracted from hypobranchial glands, mantle, egg capsule glands and prostate glands of *D. orbita* using the RNAqueous® RNA isolation kit (Ambion inc., Austin, TX, USA) according to the manufacturer's instructions with the modifications described in chapter three (Laffy *et al.* 2009). First strand cDNA was performed using Superscript®-II reverse transcriptase in accordance with manufacturers protocol (Invitrogen, Carlsbad, CA, USA). *D. orbita* genomic DNA was used as a positive control for PCR reactions using methods described in chapter two. Additional positive control Southern Bluefin Tuna (SBT) *Thunnus maccoyii* genomic DNA was obtained from Melissa Gregory at Flinders University.

AIII.1.3 PCR reactions

One to ten nanograms of cDNA template was used in each PCR reaction consisting of 2.5 mM dNTPs, 1.5 mM MgCl2, 200 nM each primer, 0.1 units Platinum taq polymerase (Invitrogen) in 10× PCR buffer. Initial cycling conditions for PCR included gradient PCR with an annealing temperature ranging from 40-55°C, however no products were visible for any of these PCR reactions (data not shown). As no PCR amplification using the primers designed in this appendix, only one representative experiment is presented Cycling conditions varied for each reaction and are listed below. A list of individual PCR reaction conditions is listed in Table AIII.3

AIII.1.3.1 Arylsulfatase cycling conditions

Reactions were cycled as follows; an initial denaturation step was performed at 94°C for 2 min followed by 35 cycles of denaturation at 94°C for 30 sec, primer annealing at 51°C for 30 sec and primer extension at 72°C for 2 min (7-12 in Table AIII.3).

AIII.1.3.2 Bromoperoxidase cycling conditions

Reactions were cycled as follows; an initial denaturation step was performed at 94°C for 2 min followed by 35 cycles of denaturation at 94°C for 30 sec, primer annealing at 55°C for 30 sec and primer extension at 72°C for 2 min (13-18 in Table AIII.3).

AIII.1.3.3 Tryptophanase cycling conditions

Reactions were cycled as follows; an initial denaturation step was performed at 94°C for 2 min followed by 35 cycles of denaturation at 94°C for 30 sec, primer annealing at 50°C for 30 sec and primer extension at 72°C for 2 min (1-6 in Table AIII.3).

28s ribosomal RNA primers used in phylogenetic analysis were used as a positive control for PCR cycling following the methods described in Chapter 2 and using genomic DNA from *D. orbita* and SBT (19-22 in Table AIII.3). Products were run on a 1% agarose gel as shown in Fig AIII.1.

Table AIII.3 PCR reaction setup for the enzymes involved in the formation of Tyrian purple.

Lane	Template	Primer set
1	HBG cDNA1	tryptophanase primers
2	HBG cDNA2	tryptophanase primers
3	mantle cDNA	tryptophanase primers
4	egg capsule cDNA	tryptophanase primers
5	prostate cDNA	tryptophanase primers
6	No template control	tryptophanase primers
7	HBG cDNA1	arylsulfatase primers
8	HBG cDNA2	arylsulfatase primers
9	mantle cDNA	arylsulfatase primers
10	egg capsule cDNA	arylsulfatase primers
11	prostate cDNA	arylsulfatase primers
12	No template control	arylsulfatase primers
13	HBG cDNA1	bromoperoxidase primers
14	HBG cDNA2	bromoperoxidase primers
15	mantle cDNA	bromoperoxidase primers
16	egg capsule cDNA	bromoperoxidase primers
17	prostate cDNA	bromoperoxidase primers
18	No template control	bromoperoxidase primers
19	SBT liver DNA	28s ribosomal RNA primers
20	SBT gonad DNA	28s ribosomal RNA primers
21	*D. orbita* DNA	28s ribosomal RNA primers
22	No template control	28s ribosomal RNA primers

AIII.2 Results

All PCR reactions performed attempting to amplify arylsulfatase, bromoperoxidase and tryptophanase sequences failed to produce PCR products. Fig AIII.1 is a representative result from these PCR reactions. Positive control reactions, run to verify the efficacy of polymerase and PCR related using 28s rRNA primers (see chapter two) resulted in the amplification of PCR products approximate size of 300 bp as well as other non specific PCR products. All gradient PCR reactions, performed to test the optimal annealing temperature of primers failed to amplify products even at the lowest annealing temperature (data not shown). All PCR reactions were repeated to confirm the lack of amplification from these primers, but no products were visible (data not shown).

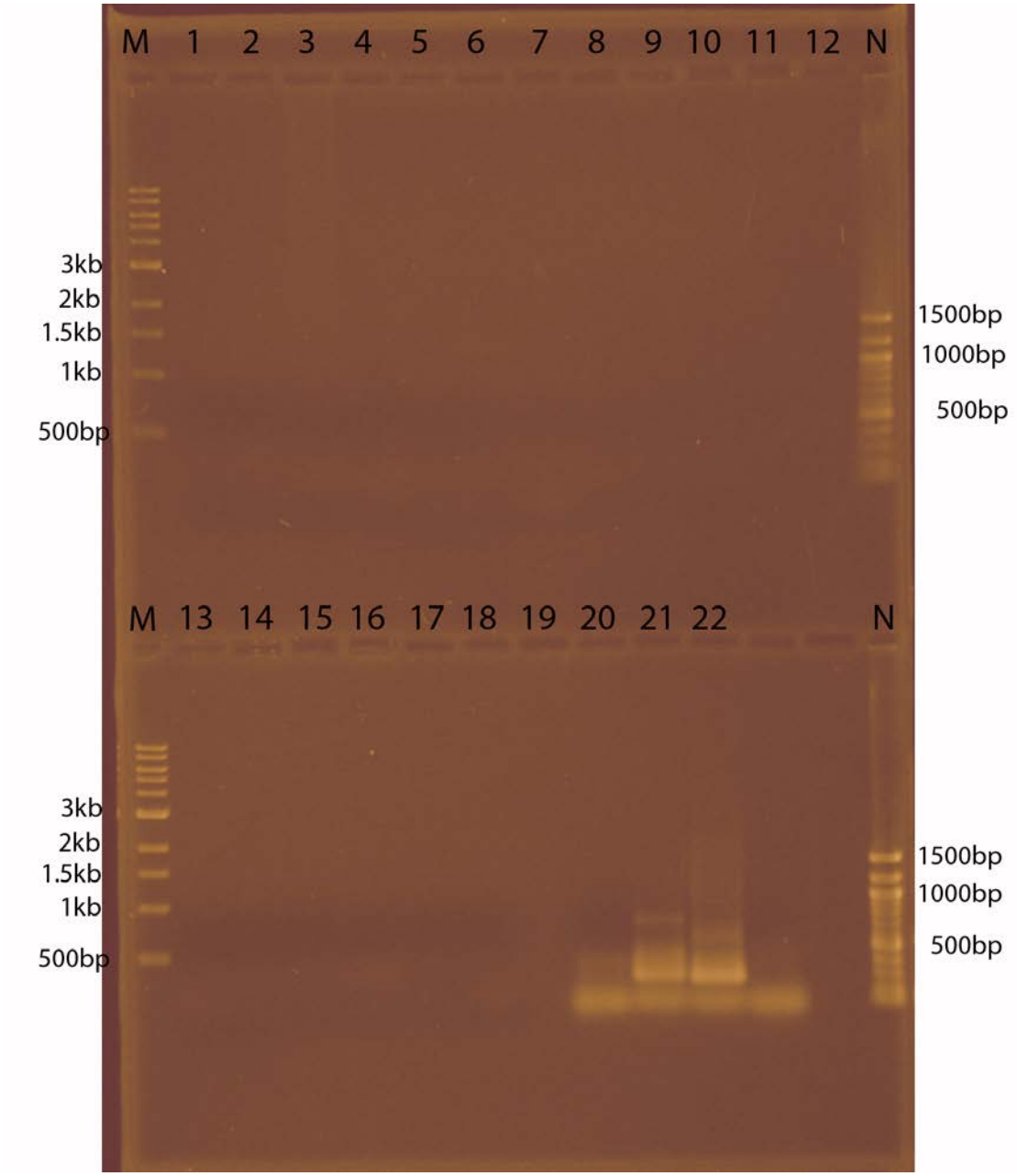


Figure AIII.1 Absence of tryptophanase, arylsulfatase and bromoperoxidase PCR products on a 1% TAE agarose gel.
5 µl of PCR products ran on gel for 1hr at 100V. Lane contents are listed in Table A.3.2.

AIII.3 Conclusion

No PCR products were amplified using gene specific arylsulfatase, bromoperoxidase and tryptophanase primers as can be observed in Fig AIII.1. The presence of bands in positive control reactions confirms that all generic PCR components are able to amplify PCR products. This means that the PCR primers designed were unable to find suitable template sequence in which to bind and amplify. As these primers were designed from a variety of different organisms, primer regions chosen may not have shared adequate homology with *D. orbit*a genes to allow amplification. While no *D. orbita* sequences have been identified for tryptophanase or bromoperoxidase enzymes, a full length arylsulfatase sequence has been identified (chapter six) and the C terminal region of this sequence has shown the greatest amount of variability when compared to other invertebrate arylsulfatase sequences (Fig 6.1). As arylsulfatase primers designed in this study were located in the C-terminal end of the sequence, it is likely that this increased variability between sequences is responsible for the lack of PCR product synthesis when these primers were used. Similarly, as bromoperoxidase and tryptophanase primers were designed based on sequence information from a a wide range of species, it is unclear whether *D. orbita* sequences show conservation or increased variability at primer sites, which may be the reason no products were amplified. Alternatively, due to the specific role these enzymes play in Tyrian purple synthesis, their sequences in *D. orbita* may show little or no sequence conservation with known bromoperoxidases and tryptophanases and may be the product of divergent evolution.

References

AANEN, DK, SLIPPERS, B, and WINGFIELD, MJ. (2009) Biological pest control in beetle agriculture *Trends in Microbiology* 17, 179-182

ABBOTT, CA, MCCAUGHAN, GW, LEVY, MT, CHURCH, WB, and GORRELL, MD. (1999) Binding to human dipeptidyl peptidase IV by adenosine deaminase and antibodies that inhibit ligand binding involves overlapping, discontinuous sites on a predicted beta propeller domain *European Journal of Biochemistry* 266, 798-810

ABBOTT, CA, YU, DMT, WOOLLATT, E, SUTHERLAND, GR, MCCAUGHAN, GW, and GORRELL, MD. (2000) Cloning, expression and chromosomal localization of a novel human dipeptidyl peptidase (DPP) IV homolog, DPP8 *European Journal of Biochemistry* 267, 6140-6150

AIHW (AUSTRALIAN INSTITUTE OF HEALTH AND WELFARE), and AACR (AUSTRALASIAN ASSOCIATION OF CANCER REGISTRIES). (2008) Cancer in Australia: an overview, 2008. AIHW, Canberra, Australia

ALTINCICEK, B, KNORR, E, and VILCINSKAS, A. (2008) Beetle immunity: Identification of immune-inducible genes from the model insect *Tribolium castaneum Dev. Comp. Immunol.* 32, 585-595

ANDERSON, D, HARRIS, R, POLYAYES, D, CICCARONE, V, DONAHUE, R, GERARD, G, and JESSEE, J. (1996) Rapid generation of Recombinant Baculoviruses and Expression of Foreign Genes Using the Bac-To-Bac® Baculovirus Expression System *Focus* 17, 53-58

ANDREWS, EB, and LITTLE, C. (1972) Structure and function in the excretory system of some terrestrial prosobranch snails (Cyclophoridae) *Journal of Zoology (London)* 168, 395-422

ANG, KKH, HOLMES, MJ, HIGA, T, HAMANN, MT, and KARA, UAK. (2000) *In vivo* antimalarial activity of the beta-carboline alkaloid manzamine A *Antimicrob. Agents Chemother.* 44, 1645-1649

ANGIOLETTI, MD, LACERRA, G, SABATO, V, and CARESTIA, C. (2004) beta+45 G -> C: a novel silent beta-thalassaemia mutation, the first in the Kozak sequence *Br. J. Haematol.* 124, 224-231

ATLAN, G, BALMAIN, N, BERLAND, S, VIDAL, B, and LOPEZ, E. (1997) Reconstruction of human maxillary defects with nacre powder: histological evidence for bone regeneration *Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences* 320, 253-258

AUSLANDER, M, YUDKOVSKI, Y, CHALIFA-CASPI, V, HERUT, B, OPHIR, R, REINHARDT, R, NEUMANN, PM, and TOM, M. (2008) Pollution-affected fish hepatic transcriptome and its expression patterns on exposure to cadmium *Mar. Biotechnol.* 10, 250-261

AVILA, C. (2006) Molluscan natural products as biological models: Chemical ecology, histology and laboratory culture. in *Molluscs: Progress in Molecular and Subcellular Biology Subseries Marine Molecular Biochemistry* (CIMINO, G, and GAVAGNIN, M eds.), Springer Berlin Heidelburg. pp

BAEK, JH, WOO, TH, KIM, CB, PARK, JH, KIM, H, LEE, S, and LEE, SH. (2009) Differential Gene Expression Profiles in the Venom Gland/sac of *Orancistrocerus drewseni* (Hymenoptera: Eumenidae) *Archives of Insect Biochemistry and Physiology* 71, 205-222

BAI, Z, YIN, Y, HU, S, WANG, G, ZHANG, X, and LI, J. (2010) Identification of genes potentially involved in pearl formation by expressed sequence tag analysis of mantle from freshwater pearl mussel (*Hyriopsis Cumingii Lea*) *Journal Of Shellfish Research* 29, 527-534

BAKER, J, and DUKE, C. (1973a) Chemistry of the Indoleninones II. Isolation from the hypobranchial glands of marine molluscs of 6-bromo-2,2-dimethylthioindolin-3-one and 6-bromo-2-methythioindoleninone as alternative precursors to tyrian purple *Australian Journal of Chemistry* 26, 2153-2157

BAKER, J, and DUKE, C. (1973b) Isolation from the hypobranchial glands of marine molluscs of 6-bromo-2,2-dimethylthioindolin-3-one and 6-bromo-2-methylthioindoleninone as alternative precursors to tyrian purple *Tetrahedron Letters*, 2481-2482

BAKER, J, and DUKE, C. (1976) Isolation of choline and choline ester salts of tyrindoxyl sulphate from the marine molluscs *Dicathais orbita* and *Mancinella keineri Tetrahedron Letters*, 1233-1234

BAKER, J, and SUTHERLAND, M. (1968) Pigments of marine animals VIII. Precursors of 6,6'-dibromoindigotin (Tyrian Purple) from the mollusc *Dicathais orbita gmelin Tetrahedron Letters*, 43-46

BAKER, JT. (1974) Tyrian purple: an ancient dye, a modern problem *Endeavour (Oxford)* 33, 11-17

BALABAN, N, and DELL'ACQUA, G. (2005) Barriers on the road to new antibiotics *Scientist* 19, 42-43

BALFOURS-PAUL, J. (1998) *Indigo*, British Museum Press, London

BARCO, A, CLAREMONT, M, REID, DG, HOUART, R, BOUCHET, P, WILLIAMS, ST, CRUAUD, C, COULOUX, A, and OLIVERIO, M. (2010) A molecular phylogenetic framework for the Muricidae, a diverse family of carnivorous gastropods *Molecular Phylogenetics and Evolution* 56, 1025-1039

BAROINTOURANCHEAU, A, DELGADO, P, PERASSO, R, and ADOUTTE, A. (1992) A broad molecular phylogeny of ciliates- Identification of major evolutionary trends and radiations within the phylum *Proc. Natl. Acad. Sci. U. S. A.* 89, 9764-9768

BAUM, H, DOGSON, KS, and SPENCER, B. (1959) The assay of the arylsulfatase A and B in human urine. *Clinica Chimica Acta* 4, 453-455

BENKENDORFF, K. (1999) Bioactive molluscan resources and their conservation: Biological and chemical studies on the egg masses of marine molluscs. in *School of Biological Sciences*, University of Woolongong, Woolongong

BENKENDORFF, K. (2009) Aquaculture and the production of pharmaceuticals and nutraceuticals. in *New Technologies in Aquaculture: Improving production efficiency, quality and environmental management* (BURNELL, G, and ALLAN, G eds.), Woodhead Publishing Limited, Cambridge. pp 866-892

BENKENDORFF, K. (2010) Molluscan biological and chemical diversity: secondary metabolites and medicinal resources produced by marine molluscs *Biological Reviews (Cambridge)* doi 10.1111/j.1469-185X.2010.00124.x

BENKENDORFF, K, BREMNER, JB, and DAVIS, AR. (2000) Tyrian purple precursors in the egg masses of the Australian muricid, *Dicathais orbita*: A possible defensive role *J. Chem. Ecol.* 26, 1037-1050

BENKENDORFF, K, BREMNER, JB, and DAVIS, AR. (2001a) Indole derivatives from the egg masses of Muricid molluscs *Molecules* 6, 70-78

BENKENDORFF, K, DAVIS, AR, and BREMNER, J. (2001b) Chemical defense in the egg masses of benthic invertebrates: An assessment of antibacterial activity in 39 mollusks and 4 polychaetes *J. Invertebr. Pathol.* 78, 109-118

BENKENDORFF, K, MCIVER, CM, and ABBOTT, CA. (2011) Bioactivity of the Murex Homeopathic Remedy and of Extracts from an Australian Muricid Mollusc against Human Cancer Cells *Evid.-based Complement Altern. Med.*, 1-12

BENKENDORFF, K, PILLAI, R, and BREMNER, JB. (2004a) 2,4,5-tribromo-1H-imidazole in the egg masses of three muricid molluscs *Natural Product Research* 18, 427-431

BENKENDORFF, K, WESTLEY, CB, and GALLARDO, CS. (2004b) Observations on the production of purple pigments in the egg capsules, hypobranchial and reproductive glands from seven species of Muricidae (Gastropoda : Mollusca) *Invertebr. Reprod. Dev.* 46, 93-102

BERLAND, S, MARIE, A, DUPLAT, D, MILET, C, SIRE, JY, and BEDOUET, L. (2011) Coupling Proteomics and Transcriptomics for the Identification of Novel and Variant Forms of Mollusk Shell Proteins: A Study with *P. margaritifera Chembiochem* 12, 950-961

BOERICKE, J. (1999) Murex. in *Homeopathic Materia Medica*, 9th Ed., Presented by Medi-T http://www.homeoint.org/books/boericmm/index.htm. pp

BOLTES, I, CZAPINSKA, H, KAHNERT, A, VON BULOW, R, DIERKS, T, SCHMIDT, B, VON FIGURA, K, KERTESZ, MA, and USON, I. (2001) 1.3 angstrom structure of

arylsulfatase from *Pseudomonas aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family *Structure* 9, 483-491

BOND, CS, CLEMENTS, PR, ASHBY, SJ, COLLYER, CA, HARROP, SJ, HOPWOOD, JJ, and GUSS, JM. (1997) Structure of a human lysosomal sulfatase *Structure* 5, 277-289

BOWMAN, KG, and BERTOZZI, CR. (1999) Carbohydrate sulfotransferases: mediators of extracellular communication *Chemistry & Biology* 6, R9-R22

BRIEN, S, PRESCOTT, P, COGHLAN, B, BASHIR, N, and LEWITH, G. (2008) Systematic review of the nutritional supplement *Perna Canaliculus* (green-lipped mussel) in the treatment of osteoarthritis *Qjm-an International Journal of Medicine* 101, 167-179

BROEMER, M, and MEIER, P. (2009) Ubiquitin-mediated regulation of apoptosis *Trends in Cell Biology* 19, 130-140

BUGGS, RJA, CHAMALA, S, WU, W, GAO, L, MAY, GD, SCHNABLE, PS, SOLTIS, DE, SOLTIS, PS, and BARBAZUK, WB. (2010) Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping *Molecular Ecology* 19, 132-146

BYCHENKO, OS, SUKHANOVA, LV, UKOLOVA, SS, SKVORTSOV, TA, POTAPOV, VK, AZHIKINA, TL, and SVERDLOV, ED. (2009) Genome similarity of Baikal omul and sig *Russ. J. Bioorg. Chem.* 35, 86-93

CAMERON, DM, GREGORY, ST, THOMPSON, J, SUH, MJ, LIMBACH, PA, and DAHLBERG, AE. (2004) Thermus thermophilus L11 methyltransferase, PrmA, is dispensable for growth and preferentially modifies free ribosomal protein L11 prior to ribosome assembly *J. Bacteriol.* 186, 5819-5825

CAMPANELLA, ME, CHU, HY, and LOW, PS. (2005) Assembly and regulation of a glycolytic enzyme complex on the human erythrocyte membrane *Proc. Natl. Acad. Sci. U. S. A.* 102, 2402-2407

CARDENAS, L, SANCHEZ, R, GOMEZ, D, FUENZALIDA, G, GALLARDO-ESCARATE, C, and TANGUY, A. (2011) Transcriptome analysis in *Concholepas concholepas*

(Gastropoda, Muricidae): Mining and characterization of new genomic and molecular markers *Mar. Genom.* 4, 197-205

CARIO-TOUMANIANTZ, C, BOULARAN, C, SCHURGERS, LJ, HEYMANN, MF, LE CUNFF, M, LEGER, J, LOIRAND, G, and PACAUD, P. (2007) Identification of differentially expressed genes in human varicose veins: Involvement of matrix GLA protein in extracellular matrix remodeling *J. Vasc. Res.* 44, 444-459

CARTER, NJ, and KEAM, SJ. (2007) Trabectedin- A review of its use in the management of soft tissue sarcoma and ovarian cancer *Drugs* 67, 2257-2276

CASSIDY-HANLEY, DM, CORDONNIER-PRATT, MM, PRATT, LH, DEVINE, C, HOSSAIN, MM, DICKERSON, HW, and CLARK, TG. (2011) Transcriptional profiling of stage specific gene expression in the parasitic ciliate *Ichthyophthirius multifiliis Mol. Biochem. Parasitol.* 178, 29-39

CELLURA, C, TOUBIANA, M, PARRINELLO, N, and ROCH, P. (2007) Specific expression of antimicrobial peptide and HSP70 genes in response to heat-shock and several bacterial challenges in mussels *Fish Shellfish Immunol.* 22, 340-350

CHANG, HCJ, NATHAN, DF, and LINDQUIST, S. (1997) *In vivo* analysis of the Hsp90 cochaperone Sti1 (p60) *Molecular and Cellular Biology* 17, 318-325

CHRISTOPHERSON, C. (1983) Marine Indoles. in *Marine Natural Products, Chemical and Biological Perspectives* (SCHEUER, P ed.), Academic Press, New York. pp 259-285

CHRISTOPHERSON, C, WATJEN, F, BUCHARDT, O, and ANTHONI, U. (1978) A Revised Structure of Tyriverdin The precursor of Tyrian Purple *Tetrahedron Letters* 34, 2779-2781

CLAREMONT, M, REID, DG, and WILLIAMS, ST. (2008) A molecular phylogeny of the Rapaninae and Ergalataxinae (Neogastropoda : Muricidae) *J. Molluscan Stud.* 74, 215-221

CLARK, MS, THORNE, MAS, VIEIRA, FA, CARDOSO, JCR, POWER, DM, and PECK, LS. (2010) Insights into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing *BMC Genomics* 11

COBB, G, and WILLAN, RC. (2006) *Undersea Jewels: A Colour Guide to Nudibranchs*, First Edition ed., Advance Press Pty Ltd, Canberra

COLGAN, D, PONDER, W, BEACHAM, E, and MACARANAS, J. (2003) Gastropod phylogeny based on six segments from four genes representing coding or non-coding and mitochondrial or nuclear DNA *Molluscan Research*, 123-148

COLGAN, DJ, PONDER, WF, BEACHAM, E, and MACARANAS, J. (2007) Molecular phylogenetics of Caenogastropoda (Gastropoda: Mollusca) *Molecular Phylogenetics and Evolution* 42, 717-737

COLGAN, DJ, PONDER, WF, and EGGLER, PE. (2000) Gastropod evolutionary rates and phylogenetic relationships assessed using partial 28S rDNA and histone H3 sequences *Zool. Scr.* 29, 29-63

CONESA, A, GOTZ, S, GARCIA-GOMEZ, JM, TEROL, J, TALON, M, and ROBLES, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research *Bioinformatics* 21, 3674-3676

CONN, DB, SIMPSON, SE, MINCHIN, D, and LUCY, FE. (2008) Occurrence of *Conchophthirus acuminatus* (Protista : Ciliophora) in *Dreissena polymorpha* (Mollusca : Bivalvia) along the River Shannon, Ireland *Biological Invasions* 10, 149-156

CONNOR, VM. (1986) The use of mucous trails by intertidal limpets to enhance food resources *Biological Bulletin* 171, 548-564

COOKSEY, C, and SINCLAIR, R. (2005) Colour Variations in Tyrian Purple Dyeing *Dyes in History and Archaeology* 20, 127-135

COOKSEY, CJ. (2001) Tyrian purple: 6,6 '-dibromoindigo and related compounds *Molecules* 6, 736-769

COSMA, MP, PEPE, S, ANNUNZIATA, I, NEWBOLD, RF, GROMPE, M, PARENTI, G, and BALLABIO, A. (2003) The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases *Cell* 113, 445-456

CROFT, AC, D'ANTONI, AV, and TERZULLI, SL. (2007) Update on the antibacterial resistance crisis *Medical Science Monitor* 13, RA103-RA118

CUADRADO, A, GARCIA-FERNANDEZ, LF, GONZALEZ, L, SUAREZ, Y, LOSADA, A, ALCAIDE, V, MARTINEZ, T, FERNANDEZ-SOUSA, JM, SANCHEZ-PUELLES, JM, and MUNOZ, A. (2003) Aplidin (TM) induces apoptosis in human cancer cells via glutathione depletion and sustained activation of the epidermal growth factor receptor, Src, JNK, and p38 MAPK *J. Biol. Chem.* 278, 241-250

CUEVAS, C, PEREZ, M, MARTIN, MJ, CHICHARRO, JL, FERNANDEZ-RIVAS, C, FLORES, M, FRANCESCH, A, GALLEGO, P, ZARZUELO, M, DE LA CALLE, F, GARCIA, J, POLANCO, C, RODRIGUEZ, I, and MANZANARES, I. (2000) *Org. Lett.* 2, 2545-2548

CULBERT, AA, BROWN, MJ, FRAME, S, HAGEN, T, CROSS, DAE, BAX, B, and REITH, AD. (2001) GSK-3 inhibition by adenoviral FRAT1 overexpression is neuroprotective and induces Tau dephosphorylation and beta-catenin stabilisation without elevation of glycogen synthase activity *FEBS Lett.* 507, 288-294

DE CARALT, S, AGELL, G, and URIZ, MJ. (2003) Long-term culture of sponge explants: conditions enhancing survival and growth, and assessment of bioactivity *Biomolecular Engineering* 20, 339-347

DEFER, D, BOURGOUGNON, N, and FLEURY, Y. (2009) Screening for antibacterial and antiviral activities in three bivalve and two gastropod marine molluscs *Aquaculture* 293, 1-7

DEMAIN, AL. (2009) Antibiotics: Natural Products Essential to Human Health *Med. Res. Rev.* 29, 821-842

DIAS, RJP, D'AVILA, S, and D'AGOSTO, M. (2006) First record of epibionts Peritrichids and Suctorians (Protozoa, Ciliophora) on *Pomacea lineata* (Spix, 1827) *Brazilian Archives of Biology and Technology* 49, 807-812

DIAS, RJP, D'AVILA, S, WIELOCH, AH, and D'AGOSTO, M. (2008) Protozoan ciliate epibionts on the freshwater apple snail *Pomacea figulina* (Spix, 1827) (Gastropoda, Ampullariidae) in an urban stream of south-east Brazil *Journal of Natural History* 42, 1409-1420

DIATCHENKO, L, LUKYANOV, S, LAU, YFC, and SIEBERT, PD. (1999) Suppression subtractive hybridization: A versatile method for identifying differentially expressed

genes. in *cDNA Preparation And Characterization*, Academic Press Inc, San Diego. pp 349-380

DIERKS, T, SCHMIDT, B, BORISSENKO, LV, PENG, JH, PREUSSER, A, MARIAPPAN, M, and VON FIGURA, K. (2003) Multiple sulfatase deficiency is caused by mutations in the gene encoding the human C-alpha-formylglycine generating enzyme *Cell* 113, 435-444

DIMASI, JA, HANSEN, RW, and GRABOWSKI, HG. (2003) The price of innovation: new estimates of drug development costs *Journal of Health Economics* 22, 151-185

DISALVO, LH. (1994) Adios, Locos *Nat. Hist.* 103, 14-18

DO, Y, HEGDE, VL, NAGARKATTI, PS, and NAGARKATTI, M. (2004) Bryostatin-1 enhances the maturation and antigen-presenting ability of murine and human dendritic cells *Cancer Res.* 64, 6756-6765

DOPHEIDE, A, LEAR, G, STOTT, R, and LEWIS, G. (2008) Molecular characterization of ciliate diversity in stream biofilms *Applied and Environmental Microbiology* 74, 1740-1747

ELSTON, RA, CHENEY, D, FRELIER, P, and LYNN, D. (1999) Invasive orchitophryid ciliate infections in juvenile Pacific and Kumomoto oysters, *Crassostrea gigas* and *Crassostrea sikamea Aquaculture* 174, 1-14

ERBA, E, BASSANO, L, DI LIBERTI, G, MURADORE, I, CHIORINO, G, UBEZIO, P, VIGNATI, S, CODEGONI, A, DESIDERIO, MA, FAIRCLOTH, G, JIMENO, J, and D'INCALCI, M. (2002) Cell cycle phase perturbations and apoptosis in tumour cells induced by aplidine *Br. J. Cancer* 86, 1510-1517

ERSPAMER, V. (1946) Ricerche chimiche e farmacolgiche sugli estratti di ghiandola ipobranchiale di *Murex (Truncularia) trunculus (L.)*, *Murex (Biolinus) brandaris (L.) e Tritonalia erinacei (L.) Pub. Stax, Zool, Napoli* 20, 91-101

EWASCHUK, JB, and DIELEMAN, LA. (2006) Probiotics and prebiotics in chronic inflammatory bowel diseases *World Journal of Gastroenterology* 12, 5941-5950

FAGERHOLM, AE, HABRANT, D, and KOSKINEN, AMP. (2010) Calyculins and Related Marine Natural Products as Serine-Threonine Protein Phosphatase PP1 and PP2A Inhibitors and Total Syntheses of Calyculin A, B, and C *Marine Drugs* 8, 122-172

FAIRCLOTH, G, and CUEVAS, C. (2006) Kahalalide F and ES285; Potent anticancer agents from marine molluscs *Progress in Molecular and Subcellular Biology* 43, 363-379

FANG, D, XU, G, HU, Y, PAN, C, XIE, L, and ZHANG, R. (2011) Identification of Genes Directly Involved in Shell Formation and Their Functions in Pearl Oyster, *Pinctada fucata Plos One* 6

FAULKER, DJ. (2000) Marine Pharmacology *Antonie van Leeuwenhoek.* 77, 135-145

FELDMEYER, B, WHEAT, CW, KREZDORN, N, ROTTER, B, and PFENNINGER, M. (2011) Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (*Radix balthica*, Basommatophora, Pulmonata), and a comparison of assembler performance *BMC Genomics* 12

FENG, ZP, ZHANG, Z, VAN KESTEREN, RE, STRAUB, VA, VAN NIEROP, P, JIN, K, NEJATBAKHSH, N, GOLDBERG, JI, SPENCER, GE, YEOMAN, MS, WILDERING, W, COORSSEN, JR, CROLL, RP, BUCK, LT, SYED, NI, and SMIT, AB. (2009) Transcriptome analysis of the central nervous system of the mollusc *Lymnaea stagnalis BMC Genomics* 10, 15

FERNANDEZ-LEBORANS, G, HANAMURA, Y, and NAGASAKI, K. (2002) A new suctorian, *Flectacineta isopodensis* (Protozoa : Ciliophora) epibiont on marine isopods from Hokkaido (Northern Japan) *Acta Protozoologica* 41, 79-84

FERNANDEZ-LEBORANS, G, HANAMURA, Y, and NAGASAKI, K. (2003) Presence of two new species of suctorian (Protozoa : Ciliophora) epibiont on the marine crustacean isopod *Excirolana chiltoni Journal of the Marine Biological Association of the United Kingdom* 83, 447-452

FERNANDEZ, LS, HODGES, LD, MACRIDES T.A., WRIGHT P.F.A., and P.M., W. (2005) Bioactivity of Secondary Metabolites from Northern Pacific Sea Stars (*Asterias amurensis*). in *Combio 2005*, Australian Society for Biochemistry and Molecular Biology Inc., Adelaide, South Australia

FLIK, G, RIJS SJOERD, JH, and WENDELAAR BONGA, E. (1984) Evidence for the presence of calmodulin in fish mucus *European Journal of Biochemistry* 138, 651-654

FOISSNER, W. (1991) Basic light and scanning electron-microscope methods for taxonomic studies of ciliated protozoa *European Journal of Protistology* 27, 313-330

FORTMAN, JL, and SHERMAN, DH. (2005) Utilizing the power of microbial genetics to bridge the gap between the promise and the application of marine natural products *Chembiochem* 6, 960-978

FOX, DL. (1974) Biochromes. Occurrence, Distribution and Comparative Biochemistry of Prominent Natural Pigments in the Marine World. in *Biochemical and Biophysical Perspectives in Marine Biology* (MALINS, DC, and SARGENT, JR eds.), Academic Press, New York. pp 170-209

FRETTER, V, and GRAHAM, A. (1994) *British Prosobranch Molluscs*, The Ray Society, Dorset

FRIEDLANDER, P. (1909) Uber den Farbstoff des antiken Purpurs aus *murex brandaris Berichte Der Deutschen Chemischen Gesellschaft*, 765

FURUYA, H, OTA, M, KIMURA, R, and TSUNEKI, K. (2004) Renal organs of cephalopods: A habitat for dicyemids and chromidinids *Journal of Morphology* 262, 629-643

GANDE, SL, MARIAPPAN, M, SCHMIDT, B, PRINGLE, TH, VON FIGURA, K, and DIERKS, T. (2008) Paralog of the formylglycine-generating enzyme - retention in the endoplasmic reticulum by canonical and noncanonical signals *Febs J.* 275, 1118-1130

GARBER, K. (2005) Peptide leads new class of chronic pain drugs *Nature Biotechnology* 23, 399-399

GASTEIGER, E, HOOGLAND, C, GATTIKER, A, DUVAUD, S, WILKINS, MR, APPEL, RD, and BAIROCH, A. (2005) Protein identification and analysis tools on the ExPASy Server. in *The Proteomics Protocols Handbook* (WALKER, JM ed.), Humana Press. pp 571-607

GESTAL, C, COSTA, MM, FIGUERAS, A, and NOVOA, B. (2007) Analysis of differentially expressed genes in response to bacterial stimulation in hemocytes of the carpet-shell clam *Ruditapes decussatus*: Identification of new antimicrobial peptides *Gene* 406, 134-143

GOETZ, F, ROSAUER, D, SITAR, S, GOETZ, G, SIMCHICK, C, ROBERTS, S, JOHNSON, R, MURPHY, C, BRONTE, CR, and MACKENZIE, S. (2010) A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*) *Molecular Ecology* 19, 176-196

GOHRE, V, and PASZKOWSKI, U. (2006) Contribution of the arbuscular mycorrhizal symbiosis to heavy metal phytoremediation *Planta* 223, 1115-1122

GONG, J, STOECK, T, YI, ZZ, MIAO, M, ZHANG, QQ, ROBERTS, DM, WARREN, A, and SONG, WB. (2009) Small Subunit rRNA Phylogenies Show that the Class Nassophorea is Not Monophyletic (Phylum Ciliophora) *Journal of Eukaryotic Microbiology* 56, 339-347

GONZALEZ, IL, and SYLVESTER, JE. (1997) Incognito rRNA and rDNA in databases and libraries *Genome Res.* 7, 65-70

GOTZ, S, GARCIA-GOMEZ, JM, TEROL, J, WILLIAMS, TD, NAGARAJ, SH, NUEDA, MJ, ROBLES, M, TALON, M, DOPAZO, J, and CONESA, A. (2008) High-throughput functional annotation and data mining with the Blast2GO suite *Nucleic Acids Research* 36, 3420-3435

GUILLOU, F, MITTA, G, GALINIER, R, and COUSTAU, C. (2007) Identification and expression of gene transcripts generated during an anti-parasitic response in *Biomphalaria glabrata Dev. Comp. Immunol.* 31, 657-671

GULL, K. (2001) Protist tubulins: new arrivals, evolutionary relationships and insights to cytoskeletal function *Current Opinion in Microbiology* 4, 427-432

HACKSTEIN, JHP, and STUMM, CK. (1994) Methane production in terrestrial arthropods *Proc. Natl. Acad. Sci. U. S. A.* 91, 5441-5445

HAEFNER, B. (2003) Drugs from the deep: marine natural products as drug candidates *Drug Discov. Today* 8, 536-544

HANAMURA, Y. (2000) Seasonality and infestation pattern of epibiosis in the beach mysid *Archaeomysis articulata Hydrobiologia* 427, 121-127

HANSEN, BH, ALTIN, D, NORDTUG, T, and OLSEN, AJ. (2007) Suppression subtractive hybridization library prepared from the copepod *Calanus finmarchicus* exposed to a sublethal mixture of environmental stressors *Comparative Biochemistry and Physiology D-Genomics & Proteomics* 2, 250-256

HANSON, SR, BEST, MD, and WONG, CH. (2004) Sulfatases: Structure, mechanism, biological activity, inhibition, and synthetic utility *Angew. Chem.-Int. Edit.* 43, 5736-5763

HARASEWYCH, MG, ADAMKEWICZ, SL, BLAKE, JA, SAUDEK, D, SPRIGGS, T, and BULT, CJ. (1997) Neogastropod phylogeny: A molecular perspective *J. Molluscan Stud.* 63, 327-351

HARRY, OG. (1977) Observation on *Lichnophora auerbachii* A Ciliate from Eyes of Queen Scallop *Chlamys opercularis Journal of Protozoology* 24, A48-A48

HATANAKA, H, OGAWA, Y, and EGAMI, F. (1976) Arylsulphatase and Glycosulphatase of *Charonia lampas*: Substrate Specificity Towards Sugar Sulphate Derivatives *Biochemical Journal* 159, 445-448

HAUG, T, KJUUL, AK, STENSVAG, K, SANDSDALEN, E, and STYRVOLD, OB. (2002a) Antibacterial activity in four marine crustacean decapods *Fish Shellfish Immunol.* 12, 371-385

HAUG, T, KJUUL, AK, STYRVOLD, OB, SANDSDALEN, E, OLSEN, OM, and STENSVAG, K. (2002b) Antibacterial activity in *Strongylocentrotus droebachiensis* (Echinoidea), *Cucumaria frondosa* (Holothuroidea), and *Asterias rubens* (Asteroidea) *J. Invertebr. Pathol.* 81, 94-102

HAUG, T, STENSVAG, K, OLSEN, OM, SANDSDALEN, E, and STYRVOLD, OB. (2004) Antibacterial activities in various tissues of the horse mussel, *Modiolus modiolus J. Invertebr. Pathol.* 85, 112-119

HEDGECOCK, D, GAFFNEY, PM, GOULLETQUER, P, GUO, XM, REECE, K, and WARR, GW. (2005) The case for sequencing the Pacific oyster genome *Journal of Shellfish Research* 24, 429-441

HENNING, DM, DIRK, S, and MOHAMED, AM. (2002) Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases13 *Chembiochem* 3, 490-504

HENRY, JJ, PERRY, KJ, FUKUI, L, and ALVI, N. (2010) Differential Localization of mRNAs During Early Development in the Mollusc, *Crepidula fornicata Integrative and Comparative Biology* 50, 720-733

HENTSCHEL, U, USHER, KM, and TAYLOR, MW. (2006) Marine sponges as microbial fermenters *FEMS Microbiology Ecology* 55, 167-177

HERNANDEZ-GUZMAN, FG, HIGASHIYAMA, T, PANGBORN, W, OSAWA, Y, and GHOSH, D. (2003) Structure of human estrone sulfatase suggests functional roles of membrane association *J. Biol. Chem.* 278, 22989-22997

HERSHKO, A, and CIECHANOVER, A. (1998) The ubiquitin system *Annual Review of Biochemistry* 67, 425-479

HEYLAND, A, VUE, Z, VOOLSTRA, CR, MEDINA, M, and MOROZ, LL. (2011) Developmental Transcriptome of *Aplysia californica Journal of Experimental Zoology Part B-Molecular and Developmental Evolution* 316B, 113-134

HIGUCHI, Y, KAWAKAMI, S, and HASHIDA, M. (2010) Strategies for In Vivo Delivery of siRNAs Recent Progress *BioDrugs* 24, 195-205

HOCHACHKA, P. (1994) Solving the common problem: matching ATP synthesis to ATP demand during exercise. *Advances in Veterinary Science and Comparative Medicine* 38, 41-56

HOFER, R, SALVENMOSER, W, and FRIED, J. (2005) Population dynamics of *Capriniana piscium* (Ciliophora, Suctoria) on the gill surface of Arctic char (*Salvelinus alpinus*) from high mountain lakes *Archiv Fur Hydrobiologie* 162, 99-109

HOFFMAN, DC, ANDERSON, RC, DUBOIS, ML, and PRESCOTT, DM. (1995) Macronuclear gene-sized molecules of hypotrichs *Nucleic Acids Research* 23, 1279-1283

HOLZ, C, LUEKING, A, BOVEKAMP, L, GUTJAHR, C, BOLOTINA, N, LEHRACH, H, and CAHILL, DJ. (2001) A human cDNA expression library in yeast enriched for open reading frames *Genome Res.* 11, 1730-1735

HOLZBAUR, ELF, and VALLEE, RB. (1994) DYNEINS - MOLECULAR-STRUCTURE AND CELLULAR FUNCTION *Annual Review of Cell Biology* 10, 339-372

HUNT, S. (1967) Secretion of Mucopolysaccharide in the Whelk *Buccinum undatum L Nature* 214, 395-396

HUNT, S. (1973) Fine structure of the secretory epithelium in the hypobranchial gland of the prosobranch gastropod mollusc *Buccinum undatum L. Journal of the Marine Biological Association of the United Kingdom* 53, 59-71

HUNT, S, and JEVONS, FR. (1965) The hypobranchial mucin of the whelk *Buccinum undatum L*. Properties of the mucin and of the glycoprotein component *Biochemical Journal* 97, 701-709

HUSSEIN, AS, HAREL, M, and SELKIRK, ME. (2002) A distinct family of acetylcholinesterases is secreted by *Nippostrongylus brasiliensis Molecular & Biochemical Parasitology* 123, 125-134

HWANG, IH, KIM, DW, KIM, SJ, MIN, BS, LEE, SH, SON, JK, KIM, C-H, CHANG, HW, and NA, M. (2011) Asterosaponins Isolated from the Starfish *Asterias amurensis Chem. Pharm. Bull.* 59, 78-83

IDE, S, MIYAZAKI, T, MAKI, H, and KOBAYASHI, T. (2010) Abundance of Ribosomal RNA Gene Copies Maintains Genome Integrity *Science* 327, 693-696

IN, O, BERBERICH, T, ROMDHANE, S, and FEIERABEND, J. (2005) Changes in gene expression during dehardening of cold-hardened winter rye (*Secale cereale L.*) leaves and potential role of a peptide methionine sulfoxide reductase in cold-acclimation *Planta* 220, 941-950

ISUPOV, MN, DALBY, AR, BRINDLEY, AA, IZUMI, Y, TANABE, T, MURSHUDOV, GN, and LITTLECHILD, JA. (2000) Crystal structure of dodecameric vanadium-dependent bromoperoxidase from the red algae *Corallina officinalis Journal of Molecular Biology* 299, 1035-1049

JACKSON, D, ELLEMOR, N, and DEGNAN, B. (2005) Correlating gene expression with larval competence, and the effect of age and parentage on metamorphosis in the tropical abalone *Haliotis asinina Marine Biology* 147, 681-697

JACKSON, DJ, and DEGNAN, BM. (2006) Expressed sequence tag analysis of genes expressed during development of the tropical abalone *Haliotis asinina Journal of Shellfish Research* 25, 225-231

JACKSON, DJ, MCDOUGALL, C, GREEN, K, SIMPSON, F, WORHEIDE, G, and DEGNAN, BM. (2006) A rapidly evolving secretome builds and patterns a sea shell *BMC Biology* 4

JACKSON, DJ, MCDOUGALL, C, WOODCROFT, B, MOASE, P, ROSE, RA, KUBE, M, REINHARDT, R, ROKHSAR, DS, MONTAGNANI, C, JOUBERT, C, PIQUEMAL, D, and DEGNAN, BM. (2010) Parallel Evolution of Nacre Building Gene Sets in Molluscs *Mol. Biol. Evol.* 27, 591-608

JACOB, J, MITREVA, M, VANHOLME, B, and GHEYSEN, G. (2008) Exploring the transcriptome of the burrowing nematode *Radopholus similis Mol. Genet. Genomics* 280, 1-17

JANNUN, R, and COE, EL. (1987) Bromoperoxidase from the marine snail *Murex trunculus Comparative Biochemistry and Physiology* 88B, 917-922

JIMENEZ, EC, CRAIG, AG, WATKINS, M, HILLYARD, DR, GRAY, WR, GULYAS, J, RIVIER, JE, CRUZ, LJ, and OLIVERA, BM. (1997) Bromocontryphan: Post-translational bromination of tryptophan *Biochemistry* 36, 989-994

JIMENEZ, EC, OLIVERA, BM, GRAY, WR, and CRUZ, LJ. (1996) Contryphan Is a D-Tryptophan-containing Conus Peptide. in *J. Biol. Chem.*

JIMENEZ, EC, WATKINS, M, JUSZCZAK, LJ, CRUZ, LJ, and OLIVERA, BM. (2001) Contryphans from Conus textile venom ducts *Toxicon* 39, 803-808

JONCKHEERE, N, PERRAIS, M, MARIETTE, C, BATRA, SK, AUBERT, JP, PIGNY, P, and VAN SEUNINGEN, I. (2004) A role for human MUC4 mucin gene, the ErbB2 ligand, as a target of TGF-beta in pancreatic carcinogenesis *Oncogene* 23, 5729-5738

JOUBERT, C, PIQUEMAL, D, MARIE, B, MANCHON, L, PIERRAT, F, ZANELLA-CLEON, I, COCHENNEC-LAUREAU, N, GUEGUEN, Y, and MONTAGNANI, C. (2010) Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: focus on biomineralization *BMC Genomics* 11

KALUJNAIA, S, MCWILLIAM, IS, ZAGUINAIKO, VA, FEILEN, AL, NICHOLSON, J, HAZON, N, CUTLER, CP, and CRAMB, G. (2007) Transcriptomic approach to the study of osmoregulation in the European eel *Anguilla anguilla Physiol. Genomics* 31, 385-401

KANEHISA, M, GOTO, S, KAWASHIMA, S, OKUNO, Y, and HATTORI, M. (2004) The KEGG resource for deciphering the genome *Nucleic Acids Research* 32, 277-280

KAO, T-T, CHANG, W-N, WU, H-L, SHI, G-Y, and FU, T-F. (2009) Recombinant Zebrafish Î³-Glutamyl Hydrolase Exhibits Properties and Catalytic Activities Comparable with Those of Mammalian Enzyme *Drug Metabolism and Disposition* 37, 302-309

KAY, E, WELLS, F, and PONDER, W. (1998) *Mollusca: The Southern Synthesis Part B*, CSIRO Publishing

KELLEY, WP, WOLTERS, AM, SACK, JT, JOCKUSCH, RA, JURCHEN, JC, WILLIAMS, ER, SWEEDLER, JV, and GILLY, WF. (2003) Characterization of a novel gastropod toxin (6-bromo-2-mercaptotryptamine) that inhibits shaker K channel activity *J. Biol. Chem.* 278, 34934-34942

KIM, K, KIM, PD, ALKER, AP, and HARVELL, CD. (2000) Chemical resistance of gorgonian corals against fungal infections *Marine Biology* 137, 393-401

KING, LA, and POSSEE, RD. (1992) *The Baculovirus Expression System: A Laboratory Guide*, Chapman Hall, New York, NY

KINOSHITA, S, WANG, N, INOUE, H, MAEYAMA, K, OKAMOTO, K, NAGAI, K, KONDO, H, HIRONO, I, ASAKAWA, S, and WATABE, S. (2011) Deep Sequencing of ESTs from Nacreous and Prismatic Layer Producing Tissues and a Screen for Novel Shell Formation-Related Genes in the Pearl Oyster *Plos One* 6

KIRKIN, V, and DIKIC, I. (2007) Role of ubiquitin- and Ubl-binding proteins in cell signaling *Current Opinion in Cell Biology* 19, 199-205

KITA, M, OHNO, O, HAN, CG, and UEMURA, D. (2010) Bioactive Secondary Metabolites from Symbiotic Marine Dinoflagellates: Symbiodinolide and Durinskiols *Chemical Record* 10, 57-69

KOCOT, KM, CANNON, JT, TODT, C, CITARELLA, MR, KOHN, AB, MEYER, A, SANTOS, SR, SCHANDER, C, MOROZ, LL, LIEB, B, and HALANYCH, KM. (2011) Phylogenomics reveals deep molluscan relationships *Nature* 477, 452-U101

KOOL, SP. (1993a) Phylogenetic Analysis of the Rapaninae (Neogastropoda, Muricidae) *Malacologia* 35, 155-259

KOOL, SP. (1993b) The Systematic Position of the Genus Nucella (Prosobranchia, Muricidae, Ocenebrinae) *Nautilus* 107, 43-57

KOZAK, M. (1987) An Analysis of 5'-Noncoding Sequences from 699 Vertebrate Messenger-Rnas *Nucleic Acids Research* 15, 8125-8148

KOZAK, M. (1995) Adherence to the First-Aug Rule When a Second Aug Codon Follows Closely Upon the First *Proc. Natl. Acad. Sci. U. S. A.* 92, 2662-2666

KRAMER, H, and WINDRUM, G. (1954) The metachromic staining reaction *Journal of Histochemistry and Cytochemistry* 3, 227-237

KUSAYKIN, MI, PESENTSEVA, MS, SIL'CHENKO, AS, AVILOV, SA, SOVA, VV, ZVYAGINTSEVA, TN, and STONIK, VA. (2006) Aryl sulfatase of unusual specificity from the liver of marine mollusk *Littorina kutila Russ. J. Bioorg. Chem.* 32, 63-70

KUZNETSOV, VV. (2003) RNA interference. An approach to produce knockout organisms and cell lines *Biochem.-Moscow* 68, 1063-1076

LAEMMLI, UK. (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4 *Nature* 227, 680-685

LAFFY, PW, BENKENDORFF, K, and ABBOTT, CA. (2009) Trends in molluscan gene sequence similarity: An observation from genes expressed within the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae) *Nautilus* 123, 154-158

LANDAZURI, P, POUTOU-PINALES, RA, ACERO-GODOY, J, CORDOBA-RUIZ, HA, ECHEVERRI-PENA, OY, SAENZ, H, DELGADO, JM, and BARRERA-AVELLANEDA, LA. (2009) Cloning and shake flask expression of hrIDS-Like in *Pichia pastoris African Journal of Biotechnology* 8, 2871-2877

LARUELLE, F, MOLLOY, DP, FOKIN, SI, and OVCHARENKO, MA. (1999) Histological analysis of mantle-cavity ciliates in *Dreissena polymorpha*: Their location, symbiotic relationship, and distinguishing morphological characteristics *Journal of Shellfish Research* 18, 251-257

LEBEAU, T, and ROBERT, JM. (2003) Diatom cultivation and biotechnologically relevant products. Part II: Current and putative products *Applied Microbiology and Biotechnology* 60, 624-632

LEE, YL, SU, MS, HUANG, TH, and SHAW, JF. (1999) C-terminal His-tagging results in substrate specificity changes of the thioesterase I from *Escherichia coli Journal of the American Oil Chemists Society* 76, 1113-1118

LEI, YL, XU, KD, CHOI, JK, HONG, HP, and WICKHAM, SA. (2009) Community structure and seasonal dynamics of planktonic ciliates along salinity gradients *European Journal of Protistology* 45, 305-319

LEVINTON, JS. (2001) *Marine Biology: Function, biodiversity and ecology*, 2nd ed., Oxford University Press, New York

LI, CH, ZHAO, JM, and SONG, LS. (2009) A review of advances in research on marine molluscan antimicrobial peptides and their potential application in aquaculture *Molluscan Research* 29, 17-26

LI, ZY, and LIU, Y. (2006) Marine sponge *Craniella austrialiensis*-associated bacterial diversity revelation based on 16S rDNA library and biologically active Actinomycetes screening, phylogenetic analysis *Letters in Applied Microbiology* 43, 410-416

LIN, A, and MEYERS, MA. (2005) Growth and structure in abalone shell *Materials Science and Engineering a-Structural Materials Properties Microstructure and Processing* 390, 27-41

LINNARSSON, S. (2010) Recent advances in DNA sequencing methods - general principles of sample preparation *Experimental Cell Research* 316, 1339-1343

LIPSCOMB, D, PLATNICK, N, and WHEELER, Q. (2003) The intellectual content of taxonomy: a comment on DNA taxonomy *Trends in Ecology & Evolution* 18, 65-66

LLOYD, PF, and LLOYD, KO. (1963) Sulfatases and sulfated polysaccharides in the viscera of marine molluscs *Nature* 199, 287

LOCKYER, AE, SPINKS, JN, WALKER, AJ, KANE, RA, NOBLE, LR, ROLLINSON, D, DIAS-NETO, E, and JONES, CS. (2007) *Biomphalaria glabrata* transcriptome: Identification of cell-signalling, transcriptional control and immune-related genes from open reading frame expressed sequence tags (ORESTES) *Dev. Comp. Immunol.* 31, 763-782

LUCKENBACH, JA, ILIEV, DB, GOETZ, FW, and SWANSON, P. (2008) Identification of differentially expressed ovarian genes during primary and early secondary oocyte growth in coho salmon, *Oncorhynchus kisutch Reprod. Biol. Endocrinol.* 6, 15

LUKATELA, G, KRAUSS, N, THEIS, K, SELMER, T, GIESELMANN, V, VON FIGURA, K, and SAENGER, W. (1998) Crystal Structure of Human Arylsulfatase A: The Aldehyde Function and the Metal Ion at the Active Site Suggest a Novel Mechanism for Sulfate Ester Hydrolysis *Biochemistry (Moscow)* 37, 3654-3664

LYNN, DH. (2003) Morphology or molecules: How do we identify the major lineages of ciliates (Phylum Ciliophora)? *European Journal of Protistology* 39, 356-364

LYNN, DH. (2008) *The ciliated protozoa: characterization, classification, and guide to the literature*, Third ed., Springer

MA, XH, and WANG, ZW. (2009) Anticancer drug discovery in the future: an evolutionary perspective *Drug Discov. Today* 14, 1136-1142

MANKAD, RV, GIMELBRANT, AA, and MCCLINTOCK, TS. (1998) Consensus translational initiation sites of marine invertebrate phyla *Biological Bulletin* 195, 251-254

MANRIQUEZ, PH, DELGADO, AP, JARA, ME, and CASTILLA, JC. (2008) Field and laboratory pilot rearing experiments with early ontogenic stages of *Concholepas concholepas* (Gastropoda : Muricidae) *Aquaculture* 279, 99-107

MARIA, H, BRIAN, C, JOHN, S, and WOLFRAM, B. (2008) Mining marine shellfish wastes for bioactive molecules: Chitin and chitosan ndash; Part A: extraction methods *Biotechnology Journal* 3, 871-877

MARKO, PB, and VERMEIJ, GJ. (1999) Molecular Phylogenetifs and the Evolution of Labral spines among eastern pacific Ocenebrinae gatropods *Molecular Phylogenetics and Evolution* 13, 275-288

MARTEL, C, VIARD, F, BOURGUET, D, and GARCIA-MEUNIER, P. (2004) Invasion by the marine gastropod *Ocinebrellus inornatus* in France. II. Expansion along the Atlantic coast *Mar. Ecol.-Prog. Ser.* 273, 163-172

MASSILIA, GR, SCHININÀ, ME, ASCENZI, P, and POLTICELLI, F. (2001) Contryphan-Vn: A Novel Peptide from the Venom of the Mediterranean Snail *Conus ventricosus Biochemical and Biophysical Research Communications* 288, 908-913

MAX, B. (1989) This and that on colour and catecholamines *Trends in Pharmacological Sciences* 10, 60-63

MCMANUS, J. (1946) Histological Demonstration of Mucin after periodic acid *Nature (London)* 158, 202

MEIJER, L, SKALDSOUNIS, AL, MAGIATIS, P, POLYCHRONOPOULOS, P, KNOCKAERT, M, LEOST, M, RYAN, XP, VONICA, CA, BRIVANLOU, A, DAJANI, R, CROVACE, C, TARRICONE, C, MUSACCHIO, A, ROE. S.M., PEARL, L, and GREENGARD, P. (2003) GSK-3-Selective Inhibitors Derived from Tyrian Purple Indirubins *Chemistry & Biology* 10, 1255-1266

MENG, X, CHANG, Y, QIU, X, and WANG, X. (2010) Generation and analysis of expressed sequence tags from adductor muscle of Japanese scallop *Mizuhopecten yessoensis Comparative Biochemistry and Physiology D-Genomics & Proteomics* 5, 288-294

MILLER, JR, KOREN, S, and SUTTON, G. (2010) Assembly algorithms for next-generation sequencing data *Genomics* 95, 315-327

MISHRA, BB, SINGH, RK, SRIVASTAVA, A, TRIPATHI, VJ, and TIWARI, VK. (2009) Fighting Against Leishmaniasis: Search of Alkaloids as Future True Potential Anti-Leishmanial Agents *Mini-Reviews in Medicinal Chemistry* 9, 107-123

MITTA, G, VANDENBULCKE, F, and ROCH, P. (2000) Original involvement of antimicrobial peptides in mussel innate immunity *FEBS Lett.* 486, 185-190

MIYAZAKI, T, AOYAMA, I, ISE, M, SEO, H, and NIWA, T. (2000) An oral sorbent reduces overload of indoxyl sulphate and gene expression of TGF-{beta}1 in uraemic rat kidneys *Nephrology Dialysis Transplantation* 15, 1773-1781

MOLINSKI, TF, DALISAY, DS, LIEVENS, SL, and SALUDES, JP. (2009) Drug development from marine natural products *Nature Reviews Drug Discovery* 8, 69-85

MORO, E, TOMANIN, R, FRISO, A, MODENA, N, TISO, N, SCARPA, M, and ARGENTON, F. (2009) A novel functional role of iduronate-2-sulfatase in zebrafish early development *Matrix Biology* 29, 43-50

MOROZ, LL, EDWARDS, JR, PUTHANVEETTIL, SV, KOHN, AB, HLA, T, HEYLAND, A, KNUDSEN, L, SAHNI, A, YU, FH, LIU, L, JEZZINI, S, LOVELL, P, IANNUCCULLI, W, CHEN, MC, NGUYEN, T, SHENG, HT, SHAW, R, KALACHIKOV, S, PANCHIN, YV, FARMERIE, W, RUSSO, JJ, JU, JY, and KANDEL, ER. (2006) Neuronal transcriptome of *Aplysia*: neuronal compartments and circuitry *Cell* 127, 1453-1467

MOROZ, LL, JU, J, RUSSO, JJ, PUTHANVEETTI, S, KOHN, A, MEDINA, M, WALSH, PJ, BIRREN, B, LANDER, ES, and KANDEL, ER. (2004) Sequencing the *Aplysia* Genome: a model for single cell, real-time and comparative genomics. in *Submission of intent to complete sequencing project from the Aplysia genome project* *http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/AplysiaSeq.pdf*

MURILLO, I, RAVENTOS, D, JAECK, E, and SAN SEGUNDO, B. (1995) Isolation of Total RNA and mRNA from Plant Tissues. in *Promega Notes Magazine*

NAEGEL, LCA. (2005) The effect of periodically "milking" to obtain tyrian purple from *Plicopurpura pansa* (Gould, 1853) on the frequency of expulsion and mortality *Journal of Shellfish Research* 24, 85-90

NAEGEL, LCA, and AGUILAR-CRUZ, CA. (2006) Hypobranchial gland from the purple snail *Plicopurpura pansa* (Gould, 1853) (Prosobranchia : Muricidae) *Journal of Shellfish Research* 25, 391-394

NAEGEL, LCA, and ALVAREZ, JIM. (2005) Biological and chemical properties of the secretion from the hypobranchial gland of the purple snail *Plicopurpura pansa* (Gould, 1853) *Journal Of Shellfish Research* 24, 421-428

NAEGEL, LCA, and COOKSEY, C. (2002) Tyrian purple from marine muricids, especially *Plicopurpura pansa* (Gould, 1853) *Journal Of Shellfish Research* 21, 193-200

NAGAMINE, S, KEINO-MASU, K, SHIOMI, K, and MASU, M. (2009) Proteolytic cleavage of the rat heparan sulfate 6-O-endosulfatase SulfFP2 by furin-type proprotein convertases *Biochemical and Biophysical Research Communications* 391, 107-112

NCBI. (2009) Submitting sequences to dbEST and GenBank http://www.ncbi.nlm.nih.gov/dbEST/how_to_submit.html.

NCBI. (2010) ENTREZ Genome Project Home http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj.

NIKAPITIYA, C, KANG, H-S, OH, C, and LEE, J. (2007) Cloning, Expression and Purification of Arylsulfatase from Disk Abalone *haliotis discus discus* cDNA. in *Asian-Pacific Aquaculture 2007*, Hanoi, Vietnam

NOREN, M, and JONDELIUS, U. (1999) Phylogeny of the Prolecithophora (Platyhelminthes) Inferred from 18s rRNA sequences *Cladistics* 15, 103-112

ODINTSOVA, NA, DYACHUK, VA, and NEZLIN, LP. (2010) Muscle and neuronal differentiation in primary cell culture of larval *Mytilus trossulus* (Mollusca: Bivalvia) *Cell and Tissue Research* 339, 625-637

ODOEMELAM, E, RAGHAVAN, N, MILLER, A, BRIDGER, JM, and KNIGHT, M. (2009) Revised karyotyping and gene mapping of the *Biomphalaria glabrata* embryonic (Bge) cell line *International Journal for Parasitology* 39, 675-681

OLIVEIRA, G, RODRIGUES, NB, ROMANHA, AJ, and BAHIA, D. (2004) Genome and genomics of schistosomes *Can. J. Zool.-Rev. Can. Zool.* 82, 375-390

OLIVERIO, M, CERVELLI, M, and MARIOTTINI, P. (2002) ITS2 rRNA evolution and its congruence with the phylogeny of muricid neogastropods (Caenogastropoda, Muricoidea) *Molecular Phylogenetics and Evolution* 25, 63-69

OLIVERIO, M, and MARIOTTINI, P. (2001) A molecular framework for the phylogeny of Coralliophila and related muricoids *J. Molluscan Stud.* 67, 215-224

OSHIKAWA, M, USAMI, R, and KATO, S. (2009) Characterization of the arylsulfatase I (ARSI) gene preferentially expressed in the human retinal pigment epithelium cell line ARPE-19 *Molecular Vision* 15, 482-494

PAGEL, M, and MEADE, A. (2006) Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo *American Naturalist* 167, 808-825

PAGEL, M, MEADE, A, and BARKER, D. (2004) Bayesian estimation of ancestral character states on phylogenies *Systematic Biology* 53, 673-684

PANTZARTZI, C, DROSOPOULOU, E, YIANGOU, M, DROZDOV, I, TSOKA, S, OUZOUNIS, CA, and SCOURAS, ZG. (2010) Promoter Complexity and Tissue-Specific Expression of Stress Response Components in Mytilus galloprovincialis, a Sessile Marine Invertebrate Species *Plos Computational Biology* 6

PASCUAL, L, BLANCA, JM, CANIZARES, J, and NUEZ, F. (2007) Analysis of gene expression during the fruit set of tomato: A comparative approach *Plant Sci.* 173, 609-620

PASQUALINI, JR, CORTESPRIETO, J, CHETRITE, G, TALBI, M, and RUIZ, A. (1997) Concentrations of estrone, estradiol and their sulfates, and evaluation of sulfatase and aromatase activities in patients with breast fibroadenoma *Int. J. Cancer* 70, 639-643

PELLETIER, I, PFEIFER, O, ALTENBUCHNER, J, and VAN PEE, KH. (1994) Cloning of a second non-haem bromoperoxidase gene from *Streptomyces aureofaciens* ATCC 10762: sequence analysis, expression in *Streptomyces lividans* and enzyme purification *Microbiology* 140, 509-516

PETTIT, GR, KAMANO, Y, HERALD, CL, TUINMAN, AA, BOETTNER, FE, KIZU, H, SCHMIDT, JM, BACZYNSKYJ, L, TOMER, KB, and BONTEMS, RJ. (1987)

Antineoplastic agents. 136. The isolation and structure of a remarkable marine animal antineoplastic constituent - Dolastatin 10 *J. Am. Chem. Soc.* 109, 6883-6885

PICKART, CM. (2002) DNA repair - Right on target with ubiquitin *Nature* 419, 120-121

PITOT, HC, MCELROY, EA, REID, JR, WINDEBANK, AJ, SLOAN, JA, ERLICHMAN, C, BAGNIEWSKI, PG, WALKER, DL, RUBIN, J, GOLDBERG, RM, ADJEI, AA, and AMES, MM. (1999) Phase I trial of dolastatin-10 (NSC 376128) in patients with advanced solid tumors *Clin. Cancer Res.* 5, 525-531

PONCET, JM, SERPENTINI, A, THIEBOT, B, VILLERS, C, BOCQUET, J, BOUCAUD-CAMOU, E, and LABEL, JM. (2000) *In vitro* synthesis of proteoglycans and collagen in primary cultures of mantle cells from the nacreous mollusk, *Haliotis tuberculata*: A new model for study of molluscan extracellular matrix *Mar. Biotechnol.* 2, 387-398

PONDER, WF, and LINDBERG, DR. (2008) *Phylogeny and evolution of the Mollusc*, University of California Press, Berkely

PROKSCH, P, EDRADA-EBEL, R, and EBEL, R. (2003) Drugs from the Sea - Opportunities and Obstacles *Marine Drugs* 1, 5-17

PROTA, G. (1980) Nitrogenous pigments in marine invertebrates. in *Marine Natural Products: Chemical and Biological Perspectives* (SCHEUER, PJ ed.), Academic Press, New York. pp 141-178

QUILANG, J, WANG, SL, LI, P, ABERNATHY, J, PEATMAN, E, WANG, YP, WANG, LL, SHI, YH, WALLACE, R, GUO, XM, and LIU, ZJ. (2007) Generation and analysis of ESTs from the eastern oyster, *Crassostrea virginica* Gmelin and identification of microsatellite and SNP markers *BMC Genomics* 8, 11

RAGHAVAN, N, and KNIGHT, M. (2006) The snail (*Biomphalaria glabrata*) genome project *Trends in Parasitology* 22, 148-151

RAMBAUT, A, and DRUMMOND, AJ. (2007) Tracer v4.1, Available from http://beast.bio.ed.ac.uk/Tracer. v1.4 Ed.

RAYYAN, A, DAMIANIDIS, P, ANTONIADOU, C, and CHINTIROGLOU, CC. (2006) Protozoan parasites in cultured mussels *Mytilus galloprovincialis* in the Thermaikos Gulf (North Aegean Sea, Greece) *Diseases of Aquatic Organisms* 70, 251-254

READ, DJ, and PEREZ-MORENO, J. (2003) Mycorrhizas and nutrient cycling in ecosystems - a journey towards relevance? *New Phytologist* 157, 475-492

REEVES, WK, DILLON, RT, JR., and DASCH, GA. (2008) Freshwater snails (Mollusca : Gastropoda) from the Commonwealth of Dominica with a discussion of their roles in the transmission of parasites. in *American Malacological Bulletin*. pp 59-63

REGENSBOGENOVA, M, MCEWAN, NR, JAVORSKY, P, KISIDAYOVA, S, MICHALOWSKI, T, NEWBOLD, CJ, HACKSTEIN, JHP, and PRISTAS, P. (2004) A re-appraisal of the diversity of the methanogens associated with the rumen ciliates *FEMS Microbiol. Lett.* 238, 307-313

RIFAI, S, FASSOUANE, A, KIJJOA, A, and VAN SOEST, R. (2004) Antimicrobial activity of utenospongin B, a metabolite from the marine sponge *Hippospongia communis* collected from the atlantic coast of morocco *marine drugs* 1, 147-153

ROLLER, RA, RICKETT, JD, and STICKLE, WB. (1995) The hypobranchial gland of the estuarine snail *Stramonita haemastoma* canaliculata (Gray) (Prosobranchia: Muricidae): a light and electron microscopical study *American Malacological Bulletin* 11, 177-190

RONQUIST, F, and HUELSENBECK, JP. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models *Bioinformatics* 19, 1572-1574

ROSEGHINI, M, SEVERINI, C, ERSPAMER, G, and ERSPAMER, V. (1995) Choline esters and biogenic amines in the hypobranchial gland of 55 molluscan species of the neogastrapod muricoidea superfamily *Toxicon* 34, 33-55

ROUSSIS, V, CHINOU, IB, TSITSIMPIKOU, C, VAGIAS, C, and PETRAKIS, PV. (2001) Antibacterial activity of volatile secondary metabolites from Caribbean soft corals of the genus Gorgonia *Flavour Frag. J.* 16, 364-366

SAAVEDRA, C, and BACHERE, E. (2006) Bivalve genomics *Aquaculture* 256, 1-14

SALOMON, CE, MAGARVEY, NA, and SHERMAN, DH. (2004) Merging the potential of microbial genetics with biological and chemical diversity: an even brighter future for marine natural product drug discovery. *Nat. Prod. Rep.* 21, 105-121

SALZET, M, DELOFFRE, L, BRETON, C, VIEAU, D, and SCHOOFS, L. (2001) The angiotensin system elements in invertebrates *Brain Research Reviews* 36, 35-45

SAMBROOK, J, FRITSCH, EF, and MANIATIS, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbour Laboratory Press

SANDSDALEN, E, HAUG, T, STENSVAG, K, and STYRVOLD, OB. (2003) The antibacterial effect of a polyhydroxylated fucophlorethol from the marine brown alga, *Fucus vesiculosus World J. Microbiol. Biotechnol.* 19, 777-782

SANFORD, E, ROTH, MS, JOHNS, GC, WARES, JP, and SOMERO, GN. (2003) Local Selection and Latitudinal Variation in a Marine Predator-Prey Interaction *Science* 300, 1135-1137

SARDIELLO, M, ANNUNZIATA, I, ROMA, G, and BALLABIO, A. (2005) Sulfatases and sulfatase modifying factors: an exclusive and promiscuous relationship.

SASAKI, H, YAMADA, K, AKASAKA, K, KAWASAKI, H, SUZUKI, K, SAITO, A, SATO, M, and SHIMADA, H. (1988) cDNA cloning, nucleotide sequence and expression of the gene for arylsulfatase in the sea urchin (*Hemicentrotus pulcherrimus*) embryo *European Journal of Biochemistry* 177, 9-13

SATTELLE, DB, and BUCKINGHAM, SD. (2006) Invertebrate studies and their ongoing contributions to neuroscience *Invertebrate neuroscience : IN* 6, 1-3

SAWYERS, C. (2004) Targeted cancer therapy *Nature* 432, 294-297

SCHAUFELBERGER, DE, KOLECK, MP, BEUTLER, JA, VATAKIS, AM, ALVARADO, AB, ANDREWS, P, MARZO, LV, MUSCHIK, GM, ROACH, J, ROSS, JT, LEBHERZ, WB, REEVES, MP, EBERWEIN, RM, RODGERS, LL, TESTERMAN, RP, SNADER, KM, and FORENZA, S. (1991) The large-scale isolation of Bryostatin-1 from *Bugula nerita* following current good manufacturing practices *J. Nat. Prod.* 54, 1265-1270

SCHMIDT, B, SELMER, T, INGENDOH, A, and VON FIGURA, K. (1995) A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency *Cell* 82, 1-8

SCHOFFSKI, P, THATE, B, BEUTEL, G, BOLTE, O, OTTO, D, HOFMANN, M, GANSER, A, JENNER, A, CHEVERTON, P, WANDERS, J, OGUMA, T, ATSUMI, R, and SATOMI, M. (2004) Phase I and pharmacokinetic study of TZT-1027, a novel synthetic dolastatin 10 derivative, administered as a 1-hour intravenous infusion every 3 weeks in patients with advanced refractory cancer *Ann. Oncol.* 15, 671-679

SCHUMACHER, M, CERELLA, C, EIFES, S, CHATEAUVIEUX, S, MORCEAU, F, JASPARS, M, DICATO, M, and DIEDERICH, M. (2010) Heteronemin, a spongean sesterterpene, inhibits TNF alpha-induced NF-kappa B activation through proteasome inhibition and induces apoptotic cell death *Biochemical Pharmacology* 79, 610-622

SCHUSTER, FL, and RAMIREZ-AVILA, L. (2008) Current World Status of *Balantidium coli* *Clinical Microbiology Reviews* 21, 626-638

SEBERG, O, HUMPHRIES, CJ, KNAPP, S, STEVENSON, DW, PETERSEN, G, SCHARFF, N, and ANDERSEN, NM. (2003) Shortcuts in systematics? A commentary on DNA-based taxonomy *Trends in Ecology & Evolution* 18, 63-65

SEGADE, P, KHER, CP, LYNN, DH, and IGLESIAS, R. (2009) Morphological and molecular characterization of renal ciliates infecting farmed snails in Spain *Parasitology* 136, 771-782

SEILACHER, A, REIF, WE, and WENK, P. (2007) The parasite connection in ecosystems and macroevolution *Naturwissenschaften* 94, 155-169

SHACKEL, NA, MCGUINNESS, PH, ABBOTT, CA, GORRELL, MD, and MCCAUGHAN, GW. (2003) Novel differential gene expression in human cirrhosis detected by suppression subtractive hybridization *Hepatology* 38, 577-588

SHAW, P, MCCLURE, W, VAN BLARICOM, G, SIMS, J, FENICAL, W, and RUDE, J. (1974) Antimicrobial activities from marine organisms. in *From the Sea* (WEBBER, M, and RUGGIERI, G eds.), Marine Technology Society, Washington D.C

SHEVCHENKO, A, TOMAS, H, HAVLIS, J, OLSEN, JV, and MANN, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes *Nature Protocols* 1, 2856-2860

SIMONIAN, M, NAIR, SV, NELL, JA, and RAFTOS, DA. (2009) Proteomic clues to the identification of QX disease-resistance biomarkers in selectively bred Sydney rock oysters, *Saccostrea glomerata Journal of Proteomics* 73, 209-217

SIPKEMA, D, OSINGA, R, SCHATTON, W, MENDOLA, D, TRAMPER, J, and WIJFFELS, R. (2005) Large-Scale Production of Pharmaceuticals by Marine Sponges: Sea, Cell, or Synthesis? *Biotechnology and Bioengineering* 90, 201-222

SMITH, AM, and MORIN, MC. (2002) Biochemical differences between trail mucus and adhesive mucus from marsh periwinkle snails *Biological Bulletin* 203, 338-346

SMITH, AM, QUICK, TJ, and PETER, RLS. (1999) Differences in the composition of adhesive and non-adhesive mucus from the limpet *Lottia limatula Biological Bulletin* 196, 34-44

SNOEYENBOS-WEST, OLO, COLE, J, CAMPBELL, A, COATS, DW, and KATZ, LA. (2004) Molecular phylogeny of phyllopharyngean ciliates and their group I introns *Journal of Eukaryotic Microbiology* 51, 441-450

SNOEYENBOS-WEST, OLO, SALCEDO, T, MCMANUS, GB, and KATZ, LA. (2002) Insights into the diversity of choreotrich and oligotrich ciliates (Class : Spirotrichea) based on genealogical analyses of multiple loci *International Journal of Systematic and Evolutionary Microbiology* 52, 1901-1913

SOROKIN, AV, KIM, ER, and OVCHINNIKOV, LP. (2009) Proteasome system of protein degradation and processing *Biochem.-Moscow* 74, 1411-1442

SPIERS, ZB, BEARHAM, D, JONES, JB, O'HARA, AJ, and RAIDAL, SR. (2008) Intracellular ciliated protozoal infection in silverlip pearl oysters, *Pinctada maxima* (Jameson, 1901) *J. Invertebr. Pathol.* 99, 247-253

SRIVASTAVA, P. (2002) Roles of heat-shock proteins in innate and adaptive immunity *Nat. Rev. Immunol.* 2, 185-194

SUNAGAWA, S, WILSON, EC, THALER, M, SMITH, ML, CARUSO, C, PRINGLE, JR, WEIS, VM, MEDINA, M, and SCHWARZ, JA. (2009) Generation and analysis of transcriptomic resources for a model system on the rise: the sea anemone *Aiptasia pallida* and its dinoflagellate endosymbiont *BMC Genomics* 10, 10

SUNDERMANN, CA, and PAULIN, JJ. (1981) Ultrastructural Features of *Allantosoma intestinalis*, a Suctorian Ciliate Isolated from the Large Intestine of the Horse *Journal of Eurkaryotic Microbiology* 28, 400-405

SWOFFORD, DL. (1999) PAUP. Phylogenetics analysis using parsimony (and other methods) Version 4.0b Sinauer Associates, Sunderland, MA

TAN, KS. (2003) Phylogenetic analysis and taxonomy of some southern Australian and New Zealand Muricidae (Mollusca: Neogastropoda) *Journal of Natural History* 37, 911-1028

TANGUY, A, BIERNE, N, SAAVEDRA, C, PINA, B, BACHERE, E, KUBE, M, BAZIN, E, BONHOMME, F, BOUDRY, P, BOULO, V, BOUTET, I, CANCELA, L, DOSSAT, C, FAVREL, P, HUVET, A, JARQUE, S, JOLLIVET, D, KLAGES, S, LAPEGUE, S, LEITE, R, MOAL, J, MORAGA, D, REINHARDT, R, SAMAIN, J-F, ZOUROS, E, and CANARIO, A. (2008) Increasing genomic information in bivalves through new EST collections in four species: Development of new genetic markers for environmental studies and genome evolution *Gene* 408, 27-36

TAYLOR, JD, MORRIS, NJ, and TAYLOR, CN. (1980) Food specialization and the evolution of predatory prosobranch gastropods *Paleontology* 23

TAYLOR, MW, RADAX, R, STEGER, D, and WAGNER, M. (2007) Sponge-associated microorganisms: Evolution, ecology, and biotechnological potential *Microbiology and Molecular Biology Reviews* 71, 295-+

TENIKOFF, D, MURPHY, KJ, LE, M, HOWE, PR, and HOWARTH, GS. (2005) Lyprinol (stabilised lipid extract of New Zealand green-lipped mussel): a potential preventative treatment modality for inflammatory bowel disease *Journal of Gastroenterology* 40, 361-365

THOMAS, TRA, KAVLEKAR, DP, and LOKABHARATHI, PA. (2010) Marine Drugs from Sponge-Microbe Association - A Review *Marine Drugs* 8, 1417-1468

THOMPSON, JD, GIBSON, TJ, PLEWNIAK, F, JEANMOUGIN, F, and HIGGINS, DG. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools *Nucleic Acids Research* 25, 4876-4882

THOMPSON, S. (1966) *Selected Histological and Histopathological Methods*, Thomas Books, Springfield, Illinois

TOWBIN, H, STAEHELIN, T, and GORDON, J. (1979) Electrophoretic transfer of protein from plyacrylamide gens to nitrocellulose sheets- Procedure and some applications *Proc. Natl. Acad. Sci. U. S. A.* 76, 4350-4354

TOWHEED, TE, MAXWELL, L, ANASTASSIADES, TP, SHEA, B, HOUPT, J, ROBINSON, V, HOCHBERG, MC, and WELLS, G. (2005) Glucosamine therapy for treating osteoarthritis *Cochrane Database of Systematic Reviews*, 55

UZAWA, H, NISHIDA, Y, SASAKI, K, MINOURA, N, and KOBAYASHI, K. (2003) Synthetic Potential of Molluscan Sulfatases for the Library Synthesis of Regioselectively *O*-Sulfonated D-*Galacto*-Sugars *Chembiochem* 4, 640-647

VAN AS, LL, VAN AS, JG, and BASSON, L. (1999) *Licnophora limgetae* sp n. (Ciliophora : heterotrichea) ectosymbiont of South African limpets *Acta Protozoologica* 38, 249-253

VAN HOEK, A, VAN ALEN, TA, SPRAKEL, VSI, LEUNISSEN, JAM, BRIGGE, T, VOGELS, GD, and HACKSTEIN, JHP. (2000) Multiple acquisition of methanogenic archaeal symbionts by anaerobic ciliates *Mol. Biol. Evol.* 17, 251-258

VAN TREECK, P, EISINGER, M, MULLER, J, PASTER, M, and SCHUHMACHER, H. (2003) Mariculture trials with Mediterranean sponge species - The exploitation of an old natural resource with sustainable and novel methods *Aquaculture* 218, 439-455

VASCONCELOS, P, GASPAR, MB, JOAQUIM, S, MATIAS, D, and CASTRO, M. (2004) Spawning of *Hexaplex (Trunculariopsis) trunculus* (Gastropoda : Muricidae) in the laboratory: description of spawning behaviour, egg masses, embryonic development, hatchling and juvenile growth rates *Invertebr. Reprod. Dev.* 46, 125-138

VENTER, JC, ADAMS, MD, MYERS, EW, LI, PW, MURAL, RJ, SUTTON, GG, SMITH, HO, YANDELL, M, EVANS, CA, HOLT, RA, GOCAYNE, JD, AMANATIDES, P, BALLEW,

RM, HUSON, DH, WORTMAN, JR, ZHANG, Q, KODIRA, CD, ZHENG, XH, CHEN, L, SKUPSKI, M, SUBRAMANIAN, G, THOMAS, PD, ZHANG, J, GABOR MIKLOS, GL, NELSON, C, BRODER, S, CLARK, AG, NADEAU, J, MCKUSICK, VA, ZINDER, N, LEVINE, AJ, ROBERTS, RJ, SIMON, M, SLAYMAN, C, HUNKAPILLER, M, BOLANOS, R, DELCHER, A, DEW, I, FASULO, D, FLANIGAN, M, FLOREA, L, HALPERN, A, HANNENHALLI, S, KRAVITZ, S, LEVY, S, MOBARRY, C, REINERT, K, REMINGTON, K, ABU-THREIDEH, J, BEASLEY, E, BIDDICK, K, BONAZZI, V, BRANDON, R, CARGILL, M, CHANDRAMOULISWARAN, I, CHARLAB, R, CHATURVEDI, K, DENG, Z, FRANCESCO, VD, DUNN, P, EILBECK, K, EVANGELISTA, C, GABRIELIAN, AE, GAN, W, GE, W, GONG, F, GU, Z, GUAN, P, HEIMAN, TJ, HIGGINS, ME, JI, R-R, KE, Z, KETCHUM, KA, LAI, Z, LEI, Y, LI, Z, LI, J, LIANG, Y, LIN, X, LU, F, MERKULOV, GV, MILSHINA, N, MOORE, HM, NAIK, AK, NARAYAN, VA, NEELAM, B, NUSSKERN, D, RUSCH, DB, SALZBERG, S, SHAO, W, SHUE, B, SUN, J, WANG, ZY, WANG, A, WANG, X, WANG, J, WEI, M-H, WIDES, R, XIAO, C, YAN, C, YAO, A, YE, J, ZHAN, M, ZHANG, W, ZHANG, H, ZHAO, Q, ZHENG, L, ZHONG, F, ZHONG, W, ZHU, SC, ZHAO, S, GILBERT, D, BAUMHUETER, S, SPIER, G, CARTER, C, CRAVCHIK, A, WOODAGE, T, ALI, F, AN, H, AWE, A, BALDWIN, D, BADEN, H, BARNSTEAD, M, BARROW, I, BEESON, K, BUSAM, D, CARVER, A, CENTER, A, CHENG, ML, CURRY, L, DANAHER, S, DAVENPORT, L, DESILETS, R, DIETZ, S, DODSON, K, DOUP, L, FERRIERA, S, GARG, N, GLUECKSMANN, A, HART, B, HAYNES, J, HAYNES, C, HEINER, C, HLADUN, S, HOSTIN, D, HOUCK, J, HOWLAND, T, IBEGWAM, C, JOHNSON, J, KALUSH, F, KLINE, L, KODURU, S, LOVE, A, MANN, F, MAY, D, MCCAWLEY, S, MCINTOSH, T, MCMULLEN, I, MOY, M, MOY, L, MURPHY, B, NELSON, K, PFANNKOCH, C, PRATTS, E, PURI, V, QURESHI, H, REARDON, M, RODRIGUEZ, R, ROGERS, Y-H, ROMBLAD, D, RUHFEL, B, SCOTT, R, SITTER, C, SMALLWOOD, M, STEWART, E, STRONG, R, SUH, E, THOMAS, R, TINT, NN, TSE, S, VECH, C, WANG, G, WETTER, J, WILLIAMS, S, WILLIAMS, M, WINDSOR, S, WINN-DEEN, E, WOLFE, K, ZAVERI, J, ZAVERI, K, ABRIL, JF, GUIGO, R, CAMPBELL, MJ, SJOLANDER, KV, KARLAK, B, KEJARIWAL, A, MI, H, LAZAREVA, B, HATTON, T, NARECHANIA, A, DIEMER, K, MURUGANUJAN, A, GUO, N, SATO, S, BAFNA, V, ISTRAIL, S, LIPPERT, R, SCHWARTZ, R, WALENZ, B, YOOSEPH, S, ALLEN, D, BASU, A, BAXENDALE, J, BLICK, L, CAMINHA, M, CARNES-STINE, J, CAULK, P, CHIANG, Y-H, COYNE, M, DAHLKE, C, MAYS, AD, DOMBROSKI, M, DONNELLY, M,

ELY, D, ESPARHAM, S, FOSLER, C, GIRE, H, GLANOWSKI, S, GLASSER, K, GLODEK, A, GOROKHOV, M, GRAHAM, K, GROPMAN, B, HARRIS, M, HEIL, J, HENDERSON, S, HOOVER, J, JENNINGS, D, JORDAN, C, JORDAN, J, KASHA, J, KAGAN, L, KRAFT, C, LEVITSKY, A, LEWIS, M, LIU, X, LOPEZ, J, MA, D, MAJOROS, W, MCDANIEL, J, MURPHY, S, NEWMAN, M, NGUYEN, T, NGUYEN, N, NODELL, M, PAN, S, PECK, J, PETERSON, M, ROWE, W, SANDERS, R, SCOTT, J, SIMPSON, M, SMITH, T, SPRAGUE, A, STOCKWELL, T, TURNER, R, VENTER, E, WANG, M, WEN, M, WU, D, WU, M, XIA, A, ZANDIEH, A, and ZHU, X. (2001) The Sequence of the Human Genome *Science* 291, 1304-1351

VERMEIJ, GJ, and CARLSON, SJ. (2000) The muricid gastropod subfamily Rapaninae: phylogeny and ecological history *Paleobiology* 26, 19-46

VINE, KL, LOCKE, JM, RANSON, M, BENKENDORFF, K, PYNE, SG, and BREMNER, JB. (2007) In vitro cytotoxicity evaluation of some substituted isatin derivatives *Bioorganic & Medicinal Chemistry Letters* 15, 931-938

VOIGT, J, and NAGEL, K. (1993) Regulation of elongation factor G GTPase activity by the ribosomal state. The effects of initiation factors and differentially bound tRNA, aminoacyl-tRNA, and peptidyl-tRNA *J. Biol. Chem.* 268, 100-106

WAEGELE, H, DEUSCH, O, HAENDELER, K, MARTIN, R, SCHMITT, V, CHRISTA, G, PINZGER, B, GOULD, SB, DAGAN, T, KLUSSMANN-KOLB, A, and MARTIN, W. (2011) Transcriptomic Evidence That Longevity of Acquired Plastids in the Photosynthetic Slugs *Elysia timida* and *Plakobranchus ocellatus* Does Not Entail Lateral Transfer of Algal Nuclear Genes *Mol. Biol. Evol.* 28, 699-706

WALTERS, ET, and MOROZ, LL. (2009) Molluscan Memory of Injury: Evolutionary Insights into Chronic Pain and Neurological Disorders *Brain Behavior and Evolution* 74, 206-218

WANG, L, SONG, L, ZHAO, J, QIU, L, ZHANG, H, XU, W, LI, H, LI, C, WU, L, and GUO, X. (2009) Expressed sequence tags from the zhikong scallop (*Chlamys farreri*): Discovery and annotation of host-defense genes *Fish Shellfish Immunol.* 26, 744-750

WESTBROEK, P, and MARIN, F. (1998) A marriage of bone and nacre *Nature* 392, 861-862

WESTLEY, C. (2008) The distribution, biosynthetic origin and functional significance of Tyrian purple precursors in the Australian muricid *Dicathais orbita* (Neogastropoda: Muricidae). in *School of Biological Sciences*, Flinders University, Adelaide

WESTLEY, C, and BENKENDORFF, K. (2008) Sex-Specific Tyrian Purple Genesis: Precursor and Pigment Distribution in the Reproductive System of the Marine Mollusc, *Dicathais orbita J. Chem. Ecol.* 34, 44-56

WESTLEY, C, and BENKENDORFF, K. (2009) The distribution of precursors and biosynthetic enzymes required for Tyrian purple genesis in the hypobranchial gland, gonoduct, an egg masses of *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae) *Nautilus* 123, 148-153

WESTLEY, C, BENKENDORFF, K, and VINE, KL. (2006) A proposed functional role for indole derivatives in reproduction and the defense system of Muricidae (Neogastropoda:Mollusca) (MEIJER, L, GUYARD, N, SKALDSOUNIS, L, and EISENBRAND, G eds.), Life in Progress editions, Roscoff, France. pp 31-44

WESTLEY, C, LEWIS, MC, and BENKENDORFF, K. (2010a) Histomorphology of the female pallial gonoduct in *Dicathais orbita* (Neogastropoda, Muricidae): sperm passage, fertilization and sperm storage potential *Invertebrate Biology* DOI: 10.1111/j.1744-7410.2010.00190.x

WESTLEY, CB, LEWIS, MC, and BENKENDORFF, K. (2010b) Histomorphology of the hypobranchial gland in *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae) *J. Molluscan Stud.* doi:10.1093/mollus/eyp056

WESTLEY, CB, MCIVER, CM, ABBOTT, CA, LE LEU, RK, and BENKENDORFF, K. (2010c) Enhanced acute apoptotic response to axoxymethane-induced DNA damage in the rodent colonic epithelium by Tyrian purple precursors: A potential colorectal cancer chemopreventative *Cancer Biology & Therapy* 9, 371-379

WEYAND, M, HECHT, HJ, KIESS, M, LIAUD, MF, VILTER, H, and SCHOMBURG, D. (1999) X-ray structure determination of a vanadium-dependent haloperoxidase from *Ascophyllum nodosum* at 2.0 angstrom resolution *Journal of Molecular Biology* 293, 595-611

WHITMAN, WB, COLEMAN, DC, and WIEBE, WJ. (1998) Prokaryotes: The unseen majority *Proc. Natl. Acad. Sci. U. S. A.* 95, 6578-6583

WHITTAKER, VP. (1960) Pharmacologically active choline esters in marine gastropods *Annals of the New York Academy of Sciences* 90, 695-705

WIESER, W, and KRUMSCHNABEL, G. (2001) Hierarchies of ATP-consuming processes: direct compared with indirect measurements, and comparative aspects *Biochemical Journal* 355, 389-395

WILL, KW, and RUBINOFF, D. (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification *Cladistics-the International Journal of the Willi Hennig Society* 20, 47-55

WILSON, PJ, MORRIS, CP, ANSON, DS, OCCHIODORO, T, BIELICKI, J, CLEMENTS, PR, and HOPWOOD, JJ. (1990) Hunter Syndrome: Isolation of an Iduronate-2-Sulfatase cDNA Clone and Analysis of Patient DNA *Proc. Natl. Acad. Sci. U. S. A.* 87, 8531-8535

WINSTEAD, JT, VOLETY, AK, and TOLLEY, SG. (2004) Parasitic and symbiotic fauna in oysters (*Crassostrea virginica*) collected from the Caloosahatchee River and Estuary in Florida *Journal of Shellfish Research* 23, 831-840

WITTSTOCK, U, FISCHER, M, SVENDSEN, I, and HALKIER, BA. (2000) Cloning and Characterization of Two cDNAs Encoding Sulfatases in the Roman Snail, *Helix pomatia IUBMB Life* 49, 71-76

WOLF, JBW, LINDELL, J, and BACKSTROM, N. (2010) Speciation genetics: current status and evolving approaches Introduction *Philosophical Transactions of the Royal Society B-Biological Sciences* 365, 1717-1733

WOLTERS, AM, JAYAWICKRAMA, DA, and SWEEDLER, JV. (2005) Comparative analysis of a neurotoxin from *Calliostoma canaliculatum* by on-line capillary isotachophoresis/H-1 NMR and diffusion H-1 NMR *J. Nat. Prod.* 68, 162-167

WOODCOCK, SH, and BENKENDORFF, K. (2008) The impact of diet on the growth and proximate composition of juvenile whelks, *Dicathais orbita* (Gastropoda : Mollusca) *Aquaculture* 276, 162-170

WRIGHT, DJ, MATTHEWS, CL, DONALDSON, I, MARTIN, AP, and HENDERSON, IF. (1997) Immunocytochemical localisation of glycosaminoglycan-like activity in the mucus and associated tissues of terrestrial slug species (Gastropoda : Pulmonata) *Comp. Biochem. Physiol. B-Biochem. Mol. Biol.* 118, 825-828

WWW.NEWSWISE.COM. (2005) Scientists hopes to pry open part of the clam genome. in *Newswise*, MarineBiologicalLaboratory

XIAO, ZJ, HAO, YS, LIU, BC, and QIAN, LS. (2002) Indirubin and meisoindigo in the treatment of chronic myelogenous leukemia in China *Leukemia & Lymphoma* 43, 1763-1768

YIN, JL, SHACKEL, NA, ZEKRY, A, MCGUINNESS, PH, RICHARDS, C, VAN DER PUTTEN, K, MCCAUGHAN, GW, ERIS, JM, and BISHOP, GA. (2001) Real-time reverse transcriptase-polymerase chain reaction (RT-PCR) for measurement of cytokine and growth factor mRNA expression with fluorogenic probes or SYBR Green I *Immunol Cell Biol* 79, 213-221

YORK, PS, CUMMINS, SF, DEGNAN, SM, WOODCROFT, BJ, and DEGNAN, BM. (2010) Identification of genes differentially expressed in the ganglia of growing *Haliotis asinina* *Journal of Shellfish Research* 29, 741-752

YUAN, LH, ZUO, XG, CHEN, JP, LIN, BF, ZHANG, JP, SUN, M, and ZHANG, SY. (2008) Differential expression analysis of Liprin-alpha 2 in hibernating bat (*Rhinolophus ferrumequinum*) *Prog. Nat. Sci.* 18, 27-32

ZELLER, D, and PAULY, D. (2005) Good news, bad news: global fisheries discards are declining, but so are total catches *Fish and Fisheries* 6, 156-159

ZHAN, ZF, XU, KD, WARREN, A, and GONG, YC. (2009) Reconsideration of Phylogenetic Relationships of the Subclass Peritrichia (Ciliophora, Oligohymenophorea) Based on Small Subunit Ribosomal RNA Gene Sequences, with the Establishment of a New Subclass Mobilia Kahl, 1933 *Journal of Eukaryotic Microbiology* 56, 552-558

ZHAO, JM, SONG, LS, LI, CH, NI, DJ, WU, LT, ZHU, L, WANG, H, and XU, W. (2007) Molecular cloning, expression of a big defensin gene from bay scallop *Argopecten irradians* and the antimicrobial activity of its recombinant protein *Mol. Immunol.* 44, 360-368

ZHAO, SZ, and QI, XQ. (2008) Signaling in plant disease resistance and symbiosis *Journal of Integrative Plant Biology* 50, 799-807

ZIDERMAN, II. (1990) An exotic tale of shellfish, antiquity and prayer shawls. Tyrian and hyacinth purples *Trends in Pharmacological Sciences* 11, 60