Mechanisms of species invasion in apple snails: proteome of the egg perivitelline fluid, and proteomic responses of the adults to abiotic stressors

Huawei Mu

Follow this and additional works at: https://repository.hkbu.edu.hk/etd_oa

Recommended Citation
https://repository.hkbu.edu.hk/etd_oa/329
DATE: September 1, 2016

STUDENT'S NAME: MU Huawei

THESIS TITLE: Mechanisms of Species Invasion in Apple Snails: Proteome of the Egg Perivitelline Fluid, and Proteomic Responses of the Adults to Abiotic Stressors

This is to certify that the above student's thesis has been examined by the following panel members and has received full approval for acceptance in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairman: Dr. Shiu Kwok Keung
Associate Professor, Department of Chemistry, HKBU
(Designated by Dean of Faculty of Science)

Internal Members: Prof. Xia Yiji
Head and Professor, Department of Biology, HKBU

Prof. Wong Ricky N S
Chair Professor in Biology, Department of Biology, HKBU

External Members: Dr. Thiyagarajan Vengatesen
Associate Professor
School of Biological Sciences
The University of Hong Kong

Dr. Cheung Siu Gin
Associate Professor
Department of Biology and Chemistry
City University of Hong Kong

In-attendance: Dr. Qiu Jianwen
Associate Professor, Department of Biology, HKBU

Issued by Graduate School, HKBU
Mechanisms of Species Invasion in Apple Snails: Proteome of the Egg Perivitelline Fluid, and Proteomic Responses of the Adults to Abiotic Stressors

MU Huawei

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Principal supervisor: Dr. QIU Jian-Wen

Hong Kong Baptist University

September 2016
DECLARATION

I hereby declare that this thesis represents my own work which has been done after registration for the degree of PhD at Hong Kong Baptist University, and has not been previously included in a thesis, dissertation submitted to this or other institution for a degree, diploma or other qualification.

Signature: [Signature]

Date: September 2016
The *Pomacea* apple snails are amphibious operculate freshwater gastropods. They show a circum-tropical and subtropical distribution and exhibit different stress tolerant abilities as some of them are invasive (e.g., *Pomacea maculata* and *Pomacea canaliculata*) and others are non-invasive species (e.g., *Pomacea diffusa*). With the development of proteomic and transcriptomic techniques and their applications in the environmental studies on non-model organisms, I was able to explore the molecular mechanisms underlying species invasion and adaptive evolution in large scale. By using a comparative approach, I have profiled the perivitelline fluid of *P. maculata*, compared the physiological and proteomic responses of *P. canaliculata* and *P. diffusa* to thermal and hypoxic stressors, and examined their sequence divergences.

*Pomacea maculata* adopts aerial oviposition. Proteins inside the egg perivitelline fluid (PVF) are presumably important for the embryonic development, but little is known about their identities. Using SDS-PAGE coupled LC-MS/MS, I identified 74 proteins from the PVF of *P. maculata*, among which 32 are novel. This collection of egg PVF proteins is the largest among those of gastropods. The annotated proteins were classified into nine functional groups: three major perivitelline subunits, immune response, energy metabolism, protein degradation, oxidation-reduction, signaling and binding, transcription and translation, cytoskeleton and others. Comparison of the gene expression levels between albumen gland (the organ that secretes PVF) and other organs showed that 24 PVF genes were specifically expressed in albumen gland. Base substitution analysis of PVF and housekeeping orthologs between *P. maculata* and its closely related species *P. canaliculata* showed that the PVF genes have a higher mean nonsynonymous substitution rate (Ka), synonymous substitution rate (Ks), and Ka/Ks, indicating these reproductive genes have weaker selective constrains. Phylogenetic analysis of perivitellin subunits suggested an ancient gene duplication during the evolution of *Pomacea*.

Tolerance to extreme environmental conditions is a trait shared by many invasive species, but little is known about the molecular mechanisms behind such tolerance. I therefore compared the heat tolerance, proteomic response to heat stress, and adaptive sequence divergence in the invasive snail *P. canaliculata* and its non-invasive congener *P. diffusa*. The LT50 of *P. canaliculata* was significantly higher than that of *P. diffusa*. More than 3,350 proteins were identified from the hepatopancreas of the snails exposed to acute and chronic thermal stress using iTRAQ-coupled mass spectrometry. Acute exposure (3-h exposure at 37 °C with 25 °C as control) resulted in similar numbers (27 in *P. canaliculata* and 23 in *P. diffusa*) of differentially expressed proteins in the two species. Chronic exposure (3-wk exposure at 35 °C with 25 °C as control) caused differential expression of more proteins (58 in *P. canaliculata* and 118 in *P. diffusa*), with many of them related to restoration of
damaged molecules, ubiquitinating dysfunctional molecules, and utilization of energy reserves in both species; but only in *P. diffusa* there was a shift from carbohydrate to lipid catabolism. Analysis of orthologous genes encoding the differentially expressed proteins revealed two genes having clear evidence of positive selection (Ka/Ks > 1), and seven candidates for more detailed analysis of positive selection (Ka/Ks between 0.5 and 1). These nine genes are related to energy metabolism, cellular oxidative homeostasis, signaling and binding processes.

I also compared the responses of two congeners (*P. canaliculata* and *P. diffusa*) to hypoxia exposure by determining their mortality, analyzing their proteome and conducting an evolutionary analysis of genes encoding proteins that were differently expressed between the two species and between different hypoxia treatments. A 72 h experiment was conducted, which showed that *P. canaliculata* is much more tolerant to hypoxia than *P. diffusa*. The two species were then exposed to four levels of dissolved oxygen (6.7, 4.1, 2.0 and 1.0 mg O₂ L⁻¹) for 8 h, and their gill proteins were analyzed using iTRAQ-coupled LC-MS/MS. There were striking differences in protein expression profiles between the two species. Compared with *P. diffusa*, the more hypoxia tolerant *P. canaliculata* had more up-regulated signal transduction proteins and down-regulated proteins which are involved in glycolysis and the tricarboxylic acid cycle. Evolutionary analysis revealed three orthologous genes encoding the differentially expressed proteins having clear signal of positive selection, indicating selection has acted on some of the hypoxia responsive genes.

Overall, my study has revealed the considerable investment of *P. maculata* to ensure the development of its offspring in aerially deposited eggs. The proteomic and base substitution rate analyses indicate genetic basis of differential resistance to heat and hypoxia stress between *P. canaliculata* and *P. diffusa*. My combined proteomics and comparative evolutionary analysis provides a framework for studying the genetic basis of species invasion and predicting their further expansion in a changing climate for non-model species.
ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my advisor Dr. Qiu Jian-Wen. He gave me the opportunity to do the project and thus I have the chance to pursue science. During the past four years, Dr. Qiu has taught me how to do scientific research and how to think and write in a critical and logical way. His rigorous attitude to science and patience with supervising students have been influencing me greatly. I feel very lucky to have such a wonderful mentor at the beginning of my scientific research journey. I would also like to thank Dr. Sun Jin who is more like an elder brother for me. When I joined Dr. Qiu’s team in 2012, I almost knew nothing about molecular techniques. It is Dr. Sun who has spent plenty time on teaching me experimental techniques, background knowledge, and good laboratory habitats. He and his wife Mrs Sun Mingming also gave me much help on my daily life in Hong Kong. I really appreciate it. In addition, I would like to thank James Xie Yang who is one of my idols in science. In my view, he is the man of ‘pure interest in science’ and sets a good example for me. James has taught me how to rear snails which is very important for my projects. I learn how to fish from him rather than being given a fish.

I sincerely thank Dr. Cheung Siu Gin who offered me a lot help on several physiological experiments and hypoxia exposure experiment, and many useful comments on my projects. Many thanks for Professor Gray A. Williams who also helped me a lot on some physiological experiments, Professor Horacio Heras who kindly sent me snail samples and gave me many advices on the perivitelline fluid proteome analysis, Professor Luan Tiangang and Ms. Fang Ling on conducting the mass spectrometry analysis, and Professor Chris K. C. Wong on useful comments on my projects.

I would like to thank my labmates and friends who have given me much help on my projects and my daily life in Hong Kong: Dr. Liu Pei, Dr. Brent L. Lockwood, Ms. Zhang Yanjie, Ms. Xu Ting, Mr. Xu Zhonghu, Dr. Li Runsheng,
Ms. Yu Hiu-yan, Mr. Wang Zhi, Ms. Yin Zhao, Ms. Sun Yanan, Dr. Alice Ka Yee Chan, Dr. Stephen Robert Cartwright, Dr. Zhang Haoyu, Dr. Huang Minjuan, Dr. Wang Mingling etc. Many thanks for our technician Dr. Alice Law Yu Sheung who has helped me a lot on ordering consumables and dealing with other trivial things. I sincerely thank Hong Kong Baptist University and University Grants Committee for the financial support.

Last but not least, I would like to thank my family. I come from a big family in a small village, and my parents expected me or my siblings to become a medical doctor who will be rich once get a job. I let them down and chose the biological science. But now, I have realized half of their dreams: I become a ‘Doctor’ (of Philosophy) and I am rich in knowledge and spiritual world (: D). Without their understanding and support, I can’t reach so far. In the end, I would also like to thank my Mr. right-Mr. Wang Hao. Whenever I feel frustrated or depressed, he is the one who is always there. With his accompany and encouragement, I always feel energetic to cope with the problems. He is also one of my idols in science (o_O): Focus on one thing, devote all his mind to it, keep digging and try his best to explore scientific questions one by one. I am greatly influenced by him.

Too much help from my friends and family which can’t be listed here in details, but I will keep it in mind. Now I am close to the end of school life, but the scientific research journey just starts.

Stay hungry, stay foolish.
Table of contents

Chapter 1. General introduction ................................................................. 1
  1.1 Apple snails: the family Ampullariidae ................................................. 1
  1.2 Biology of Pomacea .............................................................................. 2
    1.2.1 Habitats and distribution ............................................................... 3
    1.2.2 Species invasion in Pomacea .......................................................... 5
    1.2.3 Physiological characteristics of Pomacea: amphibiousness and aerial
        oviposition ......................................................................................... 7
  1.3 Perivitelline fluid in Pomacea .............................................................. 9
    1.3.1 Origin of perivitelline fluid in Pomacea .......................................... 9
    1.3.2 Pomacea maculata ......................................................................... 11
  1.4 Stress resistance in Pomacea ............................................................... 13
    1.4.1 Pomacea canaliculata ..................................................................... 13
    1.4.2 Pomacea diffusa ............................................................................ 14
    1.4.3 Comparative study on stress tolerance ........................................... 15
  1.5 Overview of thesis .............................................................................. 17

Chapter 2 Identification and evolutionary analysis of egg perivitelline fluid
proteins ........................................................................................................ 20
  2.1 Introduction .......................................................................................... 20
  2.2 Materials and methods ......................................................................... 23
    2.2.1 Snail culture .................................................................................. 23
    2.2.2 Transcriptome sequencing, assembly and annotation .................... 23
    2.2.3 Egg mass collection and protein extraction .................................... 24
    2.2.4 SDS-PAGE and LC–MS/MS Analysis ........................................... 25
    2.2.5 Database search ............................................................................ 26
    2.2.6 Transcriptomic expression analysis .............................................. 27
    2.2.7 Bioinformatic analysis .................................................................. 27
    2.2.8 Evolutionary analysis ................................................................... 28
  2.3 Results .................................................................................................... 29
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4 Discussion</td>
<td>32</td>
</tr>
<tr>
<td>2.4.1 Functional characterization of PVF proteins</td>
<td>32</td>
</tr>
<tr>
<td>2.4.2 Evolutionary characterization of PVF</td>
<td>35</td>
</tr>
<tr>
<td>**Chapter 3 Genetic basis of differential heat resistance between two</td>
<td>46</td>
</tr>
<tr>
<td>congeneric freshwater snails</td>
<td></td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>46</td>
</tr>
<tr>
<td>3.2 Materials and methods</td>
<td>49</td>
</tr>
<tr>
<td>3.2.1 Animal collection and acclimation</td>
<td>49</td>
</tr>
<tr>
<td>3.2.2 Tolerance to acute thermal exposure</td>
<td>51</td>
</tr>
<tr>
<td>3.2.3 Protein expression</td>
<td>52</td>
</tr>
<tr>
<td>3.2.3.1 Exposure to acute and chronic thermal stress</td>
<td>52</td>
</tr>
<tr>
<td>3.2.3.2 Protein extraction, digestion and labelling</td>
<td>53</td>
</tr>
<tr>
<td>3.2.3.3 Sample fractionation and LC-MS/MS analysis</td>
<td>54</td>
</tr>
<tr>
<td>3.2.3.4 Mass spectrometry data analysis</td>
<td>56</td>
</tr>
<tr>
<td>3.2.3.5 Functional annotation</td>
<td>58</td>
</tr>
<tr>
<td>3.2.4 Base substitution rate analysis</td>
<td>58</td>
</tr>
<tr>
<td>3.2.4.1 Orthologous protein identification and Ka/Ks estimation</td>
<td>59</td>
</tr>
<tr>
<td>3.2.4.2 Differentially expressed orthologs between species and</td>
<td>60</td>
</tr>
<tr>
<td>treatments of the same species</td>
<td></td>
</tr>
<tr>
<td>3.3 Results</td>
<td>60</td>
</tr>
<tr>
<td>3.3.1 Tolerance to acute thermal exposure</td>
<td>61</td>
</tr>
<tr>
<td>3.3.2 Proteome analysis</td>
<td>61</td>
</tr>
<tr>
<td>3.3.2.1 Reproducibility of proteomic data</td>
<td>61</td>
</tr>
<tr>
<td>3.3.2.2 Comparison of protein expression among experimental treatments</td>
<td>62</td>
</tr>
<tr>
<td>3.3.3 Identification of orthologous proteins, and estimating Ka/Ks of</td>
<td>65</td>
</tr>
<tr>
<td>differentially expressed orthologs</td>
<td></td>
</tr>
<tr>
<td>3.3.4 Pathway enrichment analysis</td>
<td>66</td>
</tr>
<tr>
<td>3.4 Discussion</td>
<td>66</td>
</tr>
<tr>
<td>3.4.1 Molecular chaperones</td>
<td>67</td>
</tr>
</tbody>
</table>
Chapter 3 Functional proteomics of the mollusc Helix aspersa under hypoxic stress

3.4.2 Oxidative stress related proteins………………………………………69
3.4.3 Energy metabolism related proteins…………………………………….70
3.4.4 Immune responsive proteins…………………………………………………72
3.4.5 Comparison of the orthologous proteins between congener: challenges and insights………………………………………………………………………………….73

Chapter 4 Molecular mechanisms of tolerance to hypoxia stress in congeneric snails…………………………………………………………………………………………..82
4.1 Introduction………………………………………………………………………82
4.2 Material and methods…………………………………………………………..86
   4.2.1 Snail maintenance and hypoxia exposure system…………………..86
   4.2.2 Experiment 1. Differential mortality under hypoxia………………87
   4.2.3 Experiment 2. Responses of the proteome to acute hypoxia exposure……………………………………………………………………………………88
      4.2.3.1 Exposure to hypoxia, protein extraction and labelling…….88
      4.2.3.2 Fractionation and LC-MS/MS analysis………………………….89
      4.2.3.3 Protein identification…………………………………………………90
      4.2.3.4 Determination of differentially expressed proteins………..91
      4.2.3.5 Functional classification…………………………………………….92
      4.2.3.6 Evolutionary analysis of orthologous protein-coding genes93
4.3 Results……………………………………………………………………………94
   4.3.1 Differential tolerance to hypoxia……………………………………….94
   4.3.2 Proteome analysis…………………………………………………………..94
   4.3.3 Estimation of Ka/Ks…………………………………………………………97
4.4 Discussion……………………………………………………………………….97
   4.4.1 Hypoxia sensing and signal transduction……………………………99
   4.4.2 Protein synthesis arrest…………………………………………………..100
   4.4.3 Energy metabolism…………………………………………………………101
   4.4.4 Effects of different oxygen concentrations…………………………102
   4.4.5 Evolutionary analysis of orthologs…………………………………….104

Chapter 5 General conclusions and perspectives…………………………..114
Literature cited ........................................................................................................118
Appendixes .............................................................................................................143

Appendix 1. A python script used for filtering unpaired scans and separating data generated by the CID and HCD scans in mass spectrometry .......... 143
Appendix 2. Permission of reusing published work of Journal of Proteome Research in author’s thesis ................................................................. 148
CURRICULUM VITAE .......................................................................................... 149
List of Tables

Table 2.1. Proteins detected from the egg perivitelline fluid of *Pomacea maculata*…………………………………………………………………………………42

Table 2.2. Statistics of assembly quality of albumen gland (AG) and other tissues (OT) in *Pomacea maculata*………………………………………45

Table 3.1. The significantly enriched biological process gene ontology (GO) terms for differentially expressed orthologs between species and treatments…………………………………………………………………………75

Table 4.1. Log-likelihood values and parameter estimates for 3 orthologous protein-coding genes under two pairs of site models…………………………………………………………………………………113
List of Figures

**Figure 2.1.** An individual of *Pomacea maculata* (shell length = 63 mm) crawling near the water surface with extending siphon (A); a newly deposited clutch of eggs (scale bar = 0.5 cm) (B); a SDS-PAGE gel image showing proteins in the egg perivitelline fluid of *Pomacea maculata*. The left gel lane shows protein molecular weight markers (kDa), and the right lane is a representative replicate. Each replicate was cut into 10 slices and used for LC-MS/MS analysis (D); anatomy of *P. maculata* with or without shells showing the position of representative structures and albumen gland which was used for transcriptome sequencing..........................................................37

**Figure 2.2.** The correlated relationship between the protein abundance (emPAI) of 74 PVF proteins and their corresponding gene expression levels (expressed in TPM) in albumen gland (AG) and other tissues (OT)........................................................................................................38

**Figure 2.3.** Sequence alignment for four pairs of orthologous PmPV1 subunits between *P. maculata* (Pm) and *P. canaliculata* (Pc). * indicates a fully conserved amino acid site. There is one conserved motif: IXGGP........................................................................................................39
Figure 2.4. Phylogenetic tree showing the relationships of orthologous PmPV1 subunits between *P. maculata* (Pm) and *P. canaliculata* (Pc). Numbers on the branches are bootstrap values in the maximum likelihood analysis.

Figure 2.5. Distribution of Ka, Ks and Ka/Ks values in PVF protein coding and housekeeping genes.

Figure 3.1. A photograph of a harvested vegetable garden in Long Yuen, Hong Kong, showing *P. canaliculata* on the sediment surface and their pink egg clutches on a side wall, with an inset showing an enlarged individual (A); and a photograph of an individual of *P. diffusa* purchased from a local aquarium shop (B). Photograph courtesy of Jian-Wen Qiu, Jin Sun and Huawei Mu. Copyright 2015.

Figure 3.2. Percent mortality of *P. canaliculata* (*n* = 30) and *P. diffusa* (*n* = 35) after exposure to increasing temperatures. The mortality-temperature relationship was modelled using a logistic function.

Figure 3.3. Expression patterns of DEPs under acute and chronic heat exposure (37 °C/25 °C and 35 °C/25 °C) in *P. canaliculata* (Pc). The color scale bar on the left represents the fold change (red = up-regulation, green = down-regulation). On the left are level 2 biological processes GO terms.
Figure 3.4. Expression patterns of DEPs under acute and chronic heat exposure (37 °C/25 °C and 35 °C/25 °C) in *P. diffusa* (Pd). The color scale bar on the left represents the fold change (red = up-regulation, green = down-regulation. On the left are level 2 biological processes GO terms……………………………………………………………………………………………………………………………80

Figure 3.5. Ka/Ks values in different functional protein groups. Orthologs with Ka/Ks > 1 fall above the dotted line while orthologs with Ka/Ks between 0.5 and 1 fall between the dotted and solid lines.............81

Figure 4.1. Mortality rate of *P. canaliculata* (a) and *P. diffusa* (b) under control (6.7 mg O₂ L⁻¹) and three hypoxia conditions (4.1, 2.0 and 1.0 mg O₂ L⁻¹) in a 72-h experiment (n = 12 for both species)..................107

Figure 4.2. Expression patterns of 82 orthologous proteins that had significant oxygen concentration, species or their interaction effects (ANOVA, P < 0.02). Values showing significance are highlighted in red. Pc = *P. canaliculata*, Pd = *Pomacea diffusa*, 4.1, 2.0 and 1.0 represents the hypoxia treatment of 4.1, 2.0 and 1.0 mg O₂ L⁻¹, respectively. Color scale bar on the left represents the fold change (red = up-regulation, green = down-regulation).....................108

Figure 4.3. Expression patterns of differentially expressed proteins between control (6.7 mg O₂ L⁻¹) and three other treatment groups (4.1, 2.0 and 1.0 mg O₂ L⁻¹; ANOVA, P < 0.02) in *P. canaliculata* (a) and *P. diffusa* (b). Values showing significance are highlighted in red.
Color scale bar on the left represents the fold change (red = up-regulation, green = down-regulation). On the right of the heat map are protein annotation, functional classification, and $P$ value of post hoc test result………………………………………………109

Figure 4.4. Percentage composition of different functional groups of DEPs in $P$. canaliculata and $P$. diffusa under 4.1, 2.0 and 1.0 mg O$_2$ L$^{-1}$……..111

Figure 4.5. Possible pathways involved in differential hypoxia-related protein expression in $P$. canaliculata and $P$. diffusa. Protein names in red color represent differentially expressed proteins (including one-way and two-way ANOVA results) detected in our study. Enzymes in the TCA cycle: (1) Aconitase; (2) Succinyl-CoA synthetase; (3) Dihydrolipoyl dehydrogenase (FAD cofactor); (4) Malate dehydrogenase precursor; (5) Citrate synthase; NDUFV2: NADH dehydrogenase [ubiquinone] flavoprotein 2; NDUFA5: NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5; COX6B: cytochrome $c$ oxidase subunit VIb……………………………………112
Abbreviations

ACN: acetonitrile;
AG: albumen gland;
ANOVA: analysis of variance;
BP: biological process;
CC: cellular component;
CID: collision-induced dissociation;
COI: cytochrome c oxidase subunit I;
COX6B: cytochrome c oxidase subunit Vib;
DEOs: differentially expressed orthologs;
DEPs: differentially expressed proteins;
DO: dissolved oxygen;
EF1α: elongation factor 1-alpha;
emPAI: exponentially modified protein abundance index;
ETC: electron transport chain;
GO: gene ontologies;
GOEAST: gene ontology enrichment analysis software toolkit;
GST: glutathione S-transferase;
HCD: high-energy collision-induced dissociation;
HPPD: 4-hydroxyphenylpyruvate dioxygenase;
HSP: heat shock protein;
iTRAQ: isobaric tag for relative and absolute quantitation;

Ka: the number of nonsynonymous substitutions per non-synonymous site;

KEGG: Kyoto encyclopedia of genes and genomes;

Ks: the number of synonymous substitutions per synonymous site;

LC-MS/MS: liquid chromatography-mass spectrometry/mass spectrometry;

LRT: likelihood ratio test;

$L_{50}$: median lethal temperature after exposure of an organism to stressful condition;

LTQ: linear trap quadrupole;

MACPF: membrane attack complex/perforin

MF: molecular function;

MRD: metabolic rate depression;

NDUFV2: NADH dehydrogenase [ubiquinone] flavoprotein 2;

NDUFA5: NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5;

OT: other tissues;

PDI: protein disulfide isomerases.

PVF: perivitelline fluid;

RBH: reciprocal best hit method;

ROS: reactive oxygen species;

RPK: reads per kilobase;

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis;

SOD: superoxide dismutase;
TCA: tricarboxylic acid;

TPM: transcripts per million.
Chapter 1. General introduction

1.1 Apple snails: the family Ampullariidae

Apple snails refer to a family (Ampullariidae) of tropical and sub-tropical freshwater operculate snails. The name ‘apple snails’ comes from the fact that this family has members that could reach the size of an apple – the biggest among freshwater gastropods (Cowie, 2002; Hayes et al., 2015). For example, *Pomacea maculata* could attain a shell length of 15.5 cm. Ampullariidae belongs to Caenogastropoda, Gastropoda and contains more than 150 species in nine genera which include *Afropomus, Saulea, Pila, Lanistes, Asolene, Felipponea, Pomella, Marisa* and *Pomacea* (Ponder et al., 2008; Hayes et al., 2009a). They exhibit high diversities on aspects of behavior, morphology and physiological adaptation which make them a good ‘model’ to study the evolution of behavioral and physiological adaptation, speciation and historical biogeography (Hayes et al., 2009b).

The native distribution of Ampullariidae are mainly from South-East Asia, Africa, and South and Central America (Hayes et al., 2015). *Pomella* and *Felipponea* are present in South America; *Afropomus* and *Saulea* are present in Africa; *Asolene* and *Marisa* snails originate from Paraguay, Uruguay, parts of Bolivia, Brazil and Argentina as well as the northern part of South America; *Lanistes* is native to central Africa, eastern Madagascar, and the Nile river; Parts of *Pila’s* native origins overlap with *Lanistes*, and the other native range is
southern Asia; *Pomacea’s* native places are South and Central America (Ramakrishnan 2007; Jørgensen et al., 2008; Salzburger et al., 2014; Hayes et al., 2015).

During the past two decades, more and more researches had been conducted on apple snails partially due to the introduction of several species as secondary food source or aquarium pet world-widely, where they were released intentionally or unintentionally into natural filed and brought much damage to local ecology, economy and even human health. In particular, three genera in Ampullariidae, *Marisa*, *Pila* and *Pomacea*, have showed tenacious ability to adapt, thrive and spread quickly in the areas where they have been introduced (Rawlings et al., 2007). Among them, the genus *Pomacea* is the most researched one because it contains several invasive species (e.g., *Pomacea canaliculata*) which are listed among the 100 of the World's Worst Invasive Alien Species, and some non-invasive snails (e.g., *Pomacea diffusa*) which is an aquarium pet used to remove algae on the tank walls.

**1.2 Biology of Pomacea**

The *Pomacea* has around 96 valid species. It is the largest genus in the family Ampullariidae on the basis of species number (Hayes et al., 2003). This genus is often known as apple snails due to their large, round and sometimes greenish shells. They are also called ‘golden apple snail’, ‘channeled apple snails’ and ‘mystery snail’ (Cowie, 2002).
Species of *Pomacea* have native distributions ranging from Central and South America to the Caribbean. The environment they inhabit is mainly still or slow flowing waters in rivers, lakes, ponds, ditches, and swamps. For example, it was reported that *Pomacea maculata* (previously *Pomacea insularum*) preferred flowing water while *Pomacea canaliculata* was more frequently found in relatively standing waters (Ramakrishnan, 2007). The adaptation of living in stagnant waters for some *Pomacea* species allow them to thrive and become pests of local agriculture or ecological system; while living in flowing waters help them to spread through rivers and canals in which they float, drift and crawl (Cowie, 2002; EFSA PLH Panel, 2013). In addition, attached to animals (e.g., aquatic invertebrate) and association with human activities (e.g., aquarium trade and food import) also facilitate the spread of *Pomacea* snails.

Nowadays, several *Pomacea* snails have become invasive species in America, South Asia, Europe and Africa among which *P. canaliculata* and *P. maculata* are the most notorious ones. *Pomacea canaliculata* was deliberately introduced into Hawaii to provide food source and extra economic income for local farmers in the 1980s, but now it is regarded as a pest due to the great damages it has brought to agriculture (e.g., taro) and local ecosystem (Lach & Cowie, 1999; Tran *et al*., 2008). Later this species was found in the continental United States with increasing frequency. Rawlings *et al*. (2007) examined the
identity of introduced apple snail populations in continental United States with mtDNA and performed a phylogenetic analysis to clarify their geographical origin. Five non-native apple snails (i.e. *P. canaliculata*, *P. maculata*, *P. haustrum*, *P. diffusa* and *Marisa cornuarietis*) were found in their sampling, and two of them (*P. canaliculata* and *P. maculata*) were probably introduced from places near Buenos Aires, Argentina. In 2013, Panel on Plant Health of European Food Safety Authority (EFSA) conducted a thorough survey on the current status of apple snails in Europe (i.e. *P. canaliculata* and *P. maculata*) and assessed environmental risks and potential establishment by developing a population dynamics model (EFSA PLH Panel, 2013).

The initial origin of *Pomacea* in South-East Asia was from an illegal introduction from Argentina to Taiwan in 1980s because no prior marketing research was undertaken. The original purpose was for food consumption and restaurant trade with an expectation of profits. Subsequently, they were introduced to Japan, Philippines, Hong Kong, mainland China and some other regions of South-East Asia (Cowie, 2002). In 2008, Hayes *et al.* analyzed the cytochrome *c* oxidase subunit I (COI) sequences of 783 snails collected from 57 native South America places and 164 Asian locations using genealogical and phylogenetic methods in order to clarify their geographical origins and identities. They found that there were four non-native *Pomacea* species in Asia: *P. canaliculata*, *P. maculata*, *P. diffusa* and *P. scalaris*. The *P. canaliculata* and *P. maculata* may have multiple introductions which may partially explain the quick
and wide spread of these two species; while *P. diffusa* and *P. scalaris* may be native to Brazil and Argentina and spread through aquarium trade. Lv *et al.* (2012) identified the presence of *P. canaliculata* and *P. maculata* in mainland China. The high diversity and mosaic distribution of these two species also indicated that they may have multiple and secondary introductions. For the distribution of invasive *P. canaliculata* snails in Hong Kong, Kwong *et al.* (2008) conducted a survey of 61 sites to examine their abundance, and results showed that most of wetlands and farm lands of New Territories were occupied by *P. canaliculata* because these areas are abundant in plants and vegetables which are major food source for apple snails.

1.2.2 Species invasion in *Pomacea*

*Pomacea* species feed on various items such as macrophytes, biofilms, algae, living animals and even carrion (Hayes *et al.*, 2015). Several studies have been conducted on *P. canaliculata* to examine the effects different types of macrophytes (e.g., wild semi-aquatic and cultivated macrophytes, filamentous algae, phytoplankton, fresh and decaying leaves) on the feeding and life history of *Pomacea canaliculata*. These results indicated that *P. canaliculata* favored cultivated vegetables and macrophytes with high level of nitrogen content, low levels of phenolics and dry matter content because they have lower chemical/physical defenses and higher nutrition (Qiu & Kwong, 2009; Fang *et al.*, 2010; Kwong *et al.*, 2010; Wong *et al.*, 2010; Qiu *et al.*, 2011). This may be
parts of reasons that apple snails mainly invade agricultural areas in Asia. Laboratory experiments on *P. maculata*, an invasive species in continental United States, revealed that it consumed more of native aquatic plants (e.g., *Limnobium spongia*, *Ceratophyllum demersum*, and *Chara sp.*) than non-indigenous ones (e.g., *Panicum repens*, *Colocasia esculenta*, and *Hydrilla verticillata*) (Burlakova *et al.*, 2009; Baker *et al.*, 2010). Therefore, *P. maculata* could not be used as biological control agent for invasive plants; wetland restoration in areas which was invaded by *P. maculata* should also consider plants with low palatability. In addition, comparison between juvenile and adult of *P. maculata* indicated that the juveniles consumed more by biomass than adults (Boland *et al.*, 2008). Combined with the larger number and difficult detection of juveniles, these traits contributed to the success of global invasion of apple snails (Boland *et al.*, 2008).

The wide food spectrum has allowed *Pomacea* snails to live in various ecological environments, but they also cause many economic and ecological damage and health concern in their introduced areas. Several *Pomacea* species are listed among the most serious agricultural pests due to their voracious appetite for semi-aquatic vegetation. For example, in the invaded areas, *P. canaliculata* and *P. maculata* feed a lot on agricultural crops (e.g., taro in Hawaii islands and rice plants in South-East Asia) and bring great economic yield loss (e.g. US$ 30.9 million in Taiwan in 1986) (Cowie, 2002). In addition to the economic loss caused by introduced *Pomacea*, they also threaten local
ecological homeostasis by competing food and ecological niche with native species. Pointier et al. (1991) found that the expanding population of introduced *Pomacea glauca* led to a decreased population of *Biomphalaria glabrata*. Kwong et al. (2009) showed that the invasive *P. canaliculata* snails could predate on both juveniles and adults of five native freshwater snails, and cause significant mortality to all the early life-stages of five native species. Another concern on apple snails is their threat to human health because many *Pomacea* species are the vectors of parasites such as gnathostomiasis which is caused by *Gnathostoma spinigerum* whose intermediate hosts are *P. canaliculata* and *Pila ampullacea* (Komalamisra et al., 2009). The most ‘famous’ one is rat lungworm disease or angiostrongyliasis which is commonly caused by nematode *Angiostrongylus cantonensis*. According to the first national survey on *A. cantonensis* in China, Lv et al. (2009) revealed that *A. cantonensis* was widely distributed in southern China and there were two major intermediate hosts: *P. canaliculata* and *Achatina fulica*. Human get infected easily through eating raw or not well-cooked *P. canaliculata* infected with parasites (Hayes et al., 2015). Since the transmission of *A. cantonensis* was partially determined by the dormant period of intermediate host (e.g., *P. canaliculata*), global warming and climate change might cause a shorter dormant period of *P. canaliculata* and further lead to expansion of *A. cantonensis* (Lv et al., 2011).

1.2.3 Physiological characteristics of *Pomacea*: amphibiousness and aerial
oviposition

*Pomacea* snails have two unique characteristics: (1) amphibiousness which allows them to respire by lung and gill; (2) aerial oviposition. Eggs were deposited above water line on the stems of emergent plants, tree trunks, rocks and other hard structures (Hayes *et al*., 2015).

Species in *Pomacea* are usually considered as ‘amphibious’ because they have both a well-developed gill for aquatic respiration and fully functional lung (pulmonary sac) for aerial breathing. Juveniles of *P. canaliculata* which were hatched after three hours already can utilize aerial respiration. The lung ventilation frequency showed a positive correlation with snail’s shell size and water temperature (Seuffert & Martín, 2009 & 2010a). During lung ventilation, a flexible tube-shaped siphon is formed from the folding of nuchal lobe (mantle cavity) at the left side of their body. The siphon is responsible for conveying air into lung and allows snails to ventilate without exposing themselves above the water surface. The possession of lung and siphon helps *Pomacea* snails to survive even in low oxygen water systems (Hayes *et al*., 2015). Some species may even use air-filled lung to adjust their buoyancy, and float at the water surface when needed (Cowie, 2002). This strategy may allow the juveniles and adults to disperse.

*Pomacea* is gonochoristic, internally fertilizing, and oviparous (Cowie, 2002). There is evidence that in some species of *Pomacea*, females are bigger than males. Take *P. canaliculata* for example, males and females have different
shell morphology with the female being larger and having an oval shell aperture (Cazzaniga, 1990). The onset of breeding usually takes place after the snails reach 25 mm and become reproductively mature (Burela & Martín, 2011). The breeding is often seasonal and reproductive output varies with both environmental conditions (e.g., temperature and humidity) and food availability (Tamburi & Martín, 2011). For P. canaliculata, they are reproductively active all over the year in subtropical and tropical areas, while in temperate regions, they show high reproductive activity only in warm months (Estebenet & Martin, 2002). A characteristic feature for most Pomacea snails is aerial oviposition with snails laying eggs on the exposed parts of macrophytes, walls and rocks. This process usually lasts for several hours and take place at night or early morning in order to avoid aquatic/territorial predators, desiccation and high temperature stress on both eggs and adults (Burks et al., 2010; Kyle et al., 2011; Hayes et al., 2015).

1.3 Perivitelline fluid in Pomacea

1.3.1 Origin of perivitelline fluid in Pomacea

Eggs of most gastropods contain very limited egg yolk which is incorporated into primary oocytes during vitellogenesis, instead, they have larger amount of perivitelline fluid (PVF) (De Jong-Brink et al., 1983). PVF is a proteinaceous extracellular matrix surrounding the embryos in eggs. In Pomacea, when the fertilized oocytes entered into albumen gland and traverse its periphery, they
were coated with albumen by a ciliated tract (Hayes et al., 2012; Winik & Castro-Vazquez, 2015). Then round calcareous egg masses are deposited above water surface which avoids the aquatic predators but also exposes themselves to aerial stressors such as radiation, high temperature, terrestrial predators (i.e. snail kite and fire ant) and desiccation (Yusa, 2001; Hayes et al., 2009; Hayes et al., 2015). Therefore, there must have been some drastic changes in the perivitelline fluid composition to deal with the shift from aquatic to aerial oviposition.

Study on P. canaliculata indicated that the PVF is the primary energy store during embryo development consisting of 34.8% of carbohydrates, 13.0% of proteins, and 1.5% of lipids in dry weight at early stage of embryo development (Heras et al., 1998). Three major perivitellins of P. canaliculata (i.e., PcOvo, PcPV2 and PcPV3) were reported to form glyco-lipoproteic complexes and protect the embryos against predators and environmental stresses (Heras et al., 2007; Sun et al., 2012). The different major proteins, lipid compositions and glycosidic moieties of ovorubin, PV2 and PV3 have also been characterized (Cheesman 1958; Garin et al., 1996; Ituarte et al., 2010; Dreon et al., 2004). Ovorubin belonging to the small Kunitz type serine protease inhibitor family is responsible for conspicuous and presumably aposematic coloration of eggs as well as antidigestive property (Heras et al., 2007; Dreon et al., 2010). PcPV2 is a neurotoxin containing two subunit: tachylectin-like lectin and membrane attack complex/perforin (MACPF)-like toxin, which is the first evidence of
lectin-pore-forming toxin in animals, and they function in a way like the bacteria and plant seed defenses (Heras et al., 2008; Frassa et al., 2010; Dreon et al., 2002 & 2013). PV3 is also a glycolipoprotein with a high level of lipids (e.g., carotenoid pigment, free sterols and phospholipids) and is involved in transporting antioxidants from the membrane to developing embryo (Heras et al., 2007).

1.3.2 Pomacea maculata

Pomacea maculata (Perry, 1810), previously known as Pomacea insularum (d'Orbigny, 1835) (Hayes et al., 2012), is a good candidate for comparative study with *P. canaliculata* on the comparison and evolution of PVF proteins, because phylogenetic analysis of several ampullariid species indicated that *P. maculata* is the most closely related species to *P. canaliculata* (Hayes et al., 2008). Also a study comparing *P. maculata* and *P. canaliculata* with a nuclear gene (elongation factor 1-alpha) indicated the existence of genetic exchange between two species in both introduced area-Japan and their native ranges (Matsukura et al., 2008 & 2013). Despite their close phylogenetic relationship, most studies only focused on *P. canaliculata* as is the most notorious pest in the family Ampullariidae. In contrast, there have been only few studies on *P. maculata*. For example, several physiological parameters such as feeding and consumption rate, growth rate and efficiency, and tolerance to desiccation and low temperature were compared between *P. canaliculata* and *P. maculata*.
Results showed that *P. maculata* had a relatively higher feeding and growth rate than *P. canaliculata*; while *P. canaliculata* was more tolerant to cold temperature and desiccation than *P. maculata*. In order to win the war against invasive species, early detection of invasion (i.e. egg clutch) is required. Egg characteristics (i.e. egg clutch volume, mass, number, width, length, and depth) had been quantified by Kyle *et al.* (2013) to compare the differences between *P. maculata* and *P. canaliculata*. Recently, Pasquevich *et al.* (2014) characterized the major perivitelline fluid protein (PmPV1) in *P. maculata* such as the biochemical traits (e.g., molecular mass, pI, and lipid compositions) of five subunits.

Reproductive proteins usually showed signs of rapid divergence and interspecific difference in structural, molecular and functional aspects, thus, the mechanism driving the divergence and evolutionary diversity had not been well characterized due to limited studies and techniques (Swanson & Vacquier, 2002; Shu *et al.*, 2015). Gastropoda was the second largest class of Mollusca in total species number. But there were only a few studies conducted on one or a few compositions of egg proteins of Gastropoda: a freshwater snail *Biomphalaria glabrata* (Hathaway *et al.*, 2010), *Lymnaea stagnalis* snail (Nagle *et al.*, 2001), marine *Tegula* snail (Hellberg *et al.*, 2012) and abalone *Haliotis* spp. (Aagaard *et al.*, 2006; Aagaard *et al.*, 2010). Therefore, large suite of reproduction related proteins in gastropod is still needed. Due to the recent advances in
transcriptomic and proteomic techniques, it is now possible to identify reproductive proteins in large scale and thus facilitate inter-and intra-specific comparison (Dorus et al., 2006; Mann, 2007; Findlay & Swanson, 2010). In this study, I performed a proteomic profiling of the PVF in *P. maculata* and compared the sequence divergence between *P. maculata* and *P. canaliculata* aiming to better understand the evolution of reproductive proteins.

1.4 Stress resistance in *Pomacea*

1.4.1 *Pomacea canaliculata*

*Pomacea canaliculata* (Lamarck, 1822) receives majority of researches among the family Ampullariidae such as biogeography, phylogenetics and environmental impact assessment (Hayes et al., 2015). As an invasive species in East Asia and Southeast Asia, it has brought many damages to local ecology and economy. The tolerance of *P. canaliculata* to various environmental stresses has also been examined in several studies including both physiological and molecular levels. For example, the feeding and time spent active increased with temperature ranging from 10-30 °C (Seuffert et al., 2010b); the influences of high or low temperatures on adults or days estimated for embryonic development were also investigated (Matsukura et al., 2009; Seuffert et al., 2012); comparison between *P. canaliculata* and native snail *Pila scutata* in Thailand revealed that *P. canaliculata* inhabited a natural habit with an average dissolved oxygen value of 2.37 mg l⁻¹ which is quite lower than native snail
(8.03 mg O₂ l⁻¹) indicating the higher tolerance of invasive species (Chaichana & Sumpan, 2015); the changes of non-enzymatic antioxidants (e.g., uric acid and reduced glutathione), enzymatic antioxidants (e.g., catalase and superoxide dismutase), thiobarbituric acid reactive substances, and some heat shock proteins (e.g., Hsp70 and Hsp90) were studied in active, estivated and aroused *P. canaliculata* snails (Giraud-Billoud *et al*., 2011 & 2013). Sun *et al.* (2013) then compared the proteomic responses of *P. canaliculata* to estivation and arousal using a quantitative method which provided us with a global map of differentially expressed proteins (DEPs) under desiccation. The DEPs indicated that there was an energy shift from glucose to lipid in *P. canaliculata*, and pathways such as oxidative defense, prevention of protein degradation, and removal of toxic ammonia were also activated.

1.4.2 *Pomacea diffusa*

*Pomacea diffusa* (Blume, 1957), previously considered as subspecies of *Pomacea bridgesii*, is also known as spike-topped apple snail (Rawlings *et al*., 2007). This species is native to South America and transported world-wide as aquarium pet. Studies on *P. diffusa* are quite limited as it is usually considered as non-invasive species in the introduced ranges such as Sri Lanka, Hawaii, Florida and New Zealand (Nugaliyadde *et al*., 2001; Tran *et al*., 2008; Collier *et al*., 2011; Morrison & Hay, 2011). Unlike the invasive snail *P. canaliculata* which favors various macrophytes, *P. diffusa* prefer algae film on tank wall, organic
debris, and decomposed fish food, but rarely feed on rice crop (Nugaliyadde et al., 2001; Coelho et al., 2012). Combined with the shell color polymorphism of *P. diffusa*, this species was considered to be suitable for ornamental trade. Actually, the Department of Agriculture of United States lists the *P. diffusa* snail as innocuous species allowed to be transported interstate without a permit (Rawlings et al., 2007). Comparison of consumption and growth rate among *P. diffusa*, *P. canaliculata*, and *P. maculata* indicated that the non-invasive *P. diffusa* had a minimal rate for all tested macrophytes while the other two invasive snails consumed more and had higher conversion efficiency. The higher feeding and growth rates contributed to the greater expansion of invasive species (Morrison & Hay, 2011).

### 1.4.3 Comparative study on stress tolerance

In the genus *Pomacea*, several species are introduced out of their native ranges with some of them colonizing large habitats and becoming invasive (e.g., *P. canaliculata*) while others remaining non-invasive (e.g., *P. diffusa*). This may be partially due to their differential stress tolerant abilities and genetic divergence (Sexton et al., 2002). Comparative study especially congeneric comparison which integrates physiological, biochemical and high-throughput (e.g., proteomics and transcriptomics) analyses can be a powerful tool to understand the molecular mechanisms of differential stress responses and predict the impacts of global climate change on biota (Somero, 2011).
Comparative studies between congeners can be conducted to examine responses at different levels of organization. At the population level, studies can be conducted to examine the population structure, sexual size dimorphism, predator and prey interaction, etc. (Krecker, 1939; Wellborn, 1994; Zamudio, 1998). At the individual level, studies can be performed to investigate the behaviors, feeding rate, oxygen consumption rate, heart rate, molting frequency and so on (Das & Stickle, 1993; Qiu et al., 2002; Braby & Somero, 2006). At the biological level, characters of proteins and enzymes such as lactate concentration, expression patterns of heat shock proteins were also examined to illustrate possible explanations of discrete distribution of congeneric Petrolisthes crabs (Stillman & Somero, 1996), Tegula snails and Mytilus mussels (Tomanek & Somero, 2000 & 2002). Recently, high-throughput ‘omics’ techniques have been applied in various comparative studies. For example, transcriptomic and proteomic responses to hypoxia, thermal stress and salinity were well characterized in congeneric fish and Mytilus mussels (Boswell et al., 2009; Lockwood et al., 2010a & 2010b; Tomanek & Zuzow, 2010; Fields et al., 2012; Mandic et al., 2014). In my study, I combined state-of-the-art proteomic and transcriptomic techniques with measurements of traditional physiological parameters to understand the mechanisms of differential stress resistance (i.e. thermal stress and hypoxia) between the invasive P. canaliculata and non-invasive P. diffusa.
1.5 Overview of thesis

The *Pomacea* snails have a wide distribution showing their strong tolerant abilities to various environmental conditions; a characteristic reproductive strategy-aerial oviposition which may indicate some special biochemical compositions and changes of perivitelline fluid (PVF) in eggs; an amphibious behavior allowing them to escape from adverse (e.g., hypoxia, desiccation and thermal stress) environments. Besides, the *Pomacea* bring us with many ecological and economic concerns. Therefore, I selected three *Pomacea* species, *Pomacea maculata*, *P. canaliculata* and *P. diffusa*, as experimental targets to investigate the compositions of PVF, determined their sequence divergence at transcriptomic level and compare their different physiological and proteomic responses to thermal and hypoxic stressors.

Firstly, I conducted a proteomic profiling of PVF in *P. maculata*. Proteomics has been applied in many models animals (e.g., chicken and *Drosophila*) to uncover the constituents of reproduction-related proteins or discover previously unannotated genes (Mann, 2007; Findlay *et al.*, 2009). But researches on non-model organisms are quite limited, and hence hinders our understanding on their biogeographic distribution, phylogenetic evolution and further expansion. In our study, we used a combination of SDS-PAGE and LC-MS/MS to identify the protein components of PVF of *P. maculata* among which some were novel. We also constructed a transcriptome of *P. maculata* including both albumen gland (AG) and other tissues (OT). Gene expression
levels in AG and OT were compared to check whether all the proteins detected in the proteomic experiment were expressed by the AG. In addition, we compared the sequence divergence of PVF genes between *P. maculata* and *P. canaliculata* by calculating the nonsynonymous substitution rate (Ka), synonymous substitution rate (Ks), and Ka/Ks. Sequence alignment of several subunits of major perivitellins between *P. maculata* and *P. canaliculata* was also performed to detect the conserved regions.

Then I used a comparative approach to investigate the different physiological and proteomic responses of *P. canaliculata* and *P. diffusa* to acute and chronic thermal stresses and their adaptive sequence divergence. The *P. canaliculata* and *P. diffusa* were exposed to increasing temperatures ranging from 25 to 53 °C. The relationship between temperature and their mortality rate was expressed using a logistic model. The lethal temperatures at which 50% of snails died (*LT*<sub>50</sub>) of two species was compared using statistical method. The two snails were also exposed to acute (3 h exposure at 37 °C with 25 °C as control) and chronic (3 weeks of exposure at 35 °C with 25 °C as control) thermal stress. Using isobaric tags for relative and absolute quantitation (iTRAQ) coupled mass spectrometry, we were able to identify differentially expressed proteins under acute and chronic heat exposure in both species, and this helped us to uncover several functional protein groups or pathways which may play a vital role in coping with thermal stress. Furthermore, we calculated the base substitution rates of transcriptomic sequences between the two species and found several candidate genes showing a sign of
positive selection.

In the third part, we examined the responses of *P. canaliculata* and *P. diffusa* to hypoxic conditions (4.1, 2.0 and 1.0 mg O$_2$ L$^{-1}$). Firstly, we measured the mortality rate of two species under control and hypoxia in a 72-h experiment. Then the orthologs between the two species were identified by reciprocal best hit method using transcriptome data. In addition, using iTRAQ and LC-MS/MS, we determined the differentially expressed orthologs using a two-way analysis of variance. A separate one-way analysis of variance was also performed in order to determine the hypoxic treatment effects because some proteins without an ortholog in the other species may also respond to hypoxia. Evolutionary analysis of orthologous genes (encoding the differentially expressed proteins) of four ampullariid species (*P. canaliculata, P. diffusa, P. maculata, and Marisa cornuarietis*) provided us with some positively selected amino acid sites and revealed that several genes may undergo positive selection.
Chapter 2 Identification and evolutionary analysis of egg perivitelline fluid proteins

2.1 Introduction

There has been great interest over the last several decades in studying reproduction-related proteins as such studies have the potential to answer fundamental questions about speciation and evolution (Swanson & Vacquier, 2002; Shu et al., 2015), and solve practical problems of animal infertility and pest population outbreak (Thomas et al., 2000; Clark et al., 2006). Apple snails (Gastropoda: Ampullariidae) are as a model in evolutionary biology due to a number of characteristics including high diversity of reproductive strategies (Hayes et al., 2015). Members of Ampullariidae lay egg masses underwater, at the water surface or above the waterline on vegetation stems, rocks or other solid surfaces (Hayes et al., 2009).

Species of the South American Pomacea, the most species-rich and derived genus of the family (Hayes et al., 2008), deposit colorful egg masses (Fig. 2.1) well above the waterline. The shift to aerial oviposition can prevent the eggs from attack by aquatic predators, and is considered to be a critical innovation that has contributed to the diversification and range expansion of Pomacea (Hayes et al., 2009). Nevertheless, embryonic development in aerial eggs is challenging: the eggs are often exposed to stressful abiotic conditions such as desiccation, heat and UV radiation, as well as terrestrial predators (Heras et al.,...
Therefore drastic changes to the aerial eggs must have taken place to protect the embryos from physical stress and predation (Sun et al., 2012).

The ampullariid egg contains perivitelline fluid (PVF), which surrounds the embryo and provides energy and nutrients for embryonic development (Heras et al., 1998). Proteins, termed perivitellins, are an important component of organic matter in the PVF, which accounts for up to 13% of the total weight (Heras et al., 1998). Previous studies showed that Pomacea canaliculata (Lamarck, 1822) has three major perivitellins (termed PcOvo, PcPV2 and PcPV3) that serve not only as a source of energy and nutrient, but also several defense functions against abiotic stressors and predators (Dreon et al., 2004; Heras et al., 2008; Dreon et al., 2010; Frassa et al., 2010; Ituarte et al., 2010; Dreon et al., 2013). Moreover, a proteomic study led to the identification of 59 proteins in the PVF of P. canaliculata, showing that the PVF proteins are complex, and more studies are needed to better understand the maternal investment on the eggs (Sun et al., 2012).

Comparison between closely related species is an effective approach to study the evolution of reproductive proteins (Swanson et al., 2001; Schein et al., 2004; Findlay & Swanson, 2009). Pomacea maculata Perry, 1810 (previously Pomacea insularum (d'Orbigny, 1835) (Hayes et al., 2012) is a good candidate for comparative study with P. canaliculata for PVF protein evolution. The two species have been introduced out of South America to become invasive pests of
agricultural and natural wetlands (Rawlings et al., 2007, Hayes et al., 2008, Lv et al., 2011 & 2013, Hayes et al., 2015), therefore comparing their PVF proteins may reveal the mechanisms of range expansion. Their distribution ranges overlap in their native South American and invaded Asian wetlands (Rawlings et al., 2007, Hayes et al., 2008) and they can hybridize naturally (Matsukura et al., 2013), therefore studying PVF proteins can reveal their contribution to speciation in Pomacea. Nevertheless, little is known about the perivitellins of P. maculata. Recent studies have reported three major perivitellins of P. maculata, which are structurally and functionally similar to the corresponding PVF proteins of P. canaliculata (Pasquevich et al., 2014, Pasquevich et al., submitted). However, nothing is known about the presence and roles of less abundant PVF proteins in the embryonic development of P. maculata.

We therefore aimed to identify the PVF proteome of P. maculata. To enable such proteomic analysis in the non-model species, we sequenced the transcriptome albumen gland (AG), the organ that secretes PVF, to generate a translated protein database. We hypothesized that perivitellin genes be actively expressed in AG when compared to genes coding no perivitellin proteins. We therefore compared the expression level of PVF proteins in AG against other tissues. We then analyzed the substitution rates of protein-coding genes between P. maculata and P. canaliculata in order to gain insight into mechanisms of their divergence.
2.2 Materials and methods

2.2.1 Snail culture

Adults of *Pomacea maculata* (~ 20 mm shell length) were collected from National University of La Plate (Argentina) and transported to Hong Kong Baptist University. The identities of *P. maculata* were confirmed by amplifying fragments of the elongation factor 1-alpha (EF1α) and cytochrome c oxidase subunit I (COI) genes (Matsukura et al., 2013) and comparing them with the sequences deposited in the GenBank. Snails were cultured at 25 ± 1 °C in 250-L tanks filled with tap water. Aeration was provided by an air stone connected to an air pump, and the water was cleaned using an overflow filter. Water used in the snail culture was dechlorinated by aeration before use, and changed once a week. The snails were fed with lettuce, carrot and fish feed. Excessive leftovers were removed daily. Under these conditions, the snails could grow, mate and deposit egg clutches above the waterline on the tank wall.

2.2.2 Transcriptome sequencing, assembly and annotation

To obtain a database for protein identification following MS/MS analysis, and to detect genes specifically expressed in the albumen gland, we sequenced the transcriptomes of albumen gland (AG, Fig. 2.1), and other tissues (OT) pooled in equal amount from foot, mantle, testis, and digestive gland. Total RNA from AG and OT were extracted separately by using the TRIzol reagent (Invitrogen, Carlsbad, USA) following the manufacture’s manual. A high salt solution (1.2 M
sodium chloride and 0.8 M sodium citrate) and a lithium chloride solution (2 M) were added before and after the isopropanol precipitation step to remove polysaccharides. RNA quality was checked by agarose gel electrophoresis and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Messenger RNA was selected, reverse-transcribed into cDNA, and paired-end sequenced on an Illumina Hiseq 2000 to produce 100 base pair (bp) reads (Illumina, San Diego, CA, USA). Reads were assembled by Trinity (Grabherr et al., 2011). The generated sequences were annotated by searching against protein databases (NCBI nr, KEGG, COG, and Swissprot) using BLASTx with an E-value threshold of $1 \times 10^{-5}$. Protein-coding regions were translated into amino acid sequences and used for the following protein identification database.

### 2.2.3 Egg mass collection and protein extraction

Egg masses laid within 12 hours were removed from the culture tank wall carefully, rinsed several times with MilliQ water, and then dried in a fume hood. The egg shells were then cracked gently using a fine sterile needle and the perivitelline fluid (PVF) was extracted using 10 μL pipette tips. The PVF was dissolved in 8 M urea, and the mixed solution was homogenized thoroughly and centrifuged at 12000g for 10 min at 4 °C. Then the supernatant was transferred into new tubes. The protein solution was purified using the methanol/chloroform method (Friedman, 2007). Three biological samples were collected from different egg masses and the protein concentration in each sample was measured.
by using the RC-DC kit (Bio-Rad).

2.2.4 SDS-PAGE and LC-MS/MS Analysis

The purified protein solution was mixed with the SDS-PAGE buffer (50% glycerol, 0.2 M Tris-HCl pH = 6.8, 0.05% bromophenol blue, 10 mM dithiothreitol, and 10% SDS) at a ratio of 3:1 (v/v), and placed on a heating block at 105 °C for 5 min. Proteins in each sample was separated by SDS-PAGE and stained by Coomassie Brilliant Blue. 1% acetic acid was applied to destain the gel. Each sample was divided into 10 slices based on their band intensity and protein molecular weight (Fig. 2.1).

Gel slices were further destained with a solution of 50% methanol/ 50 mM NH₄HCO₃, washed with MilliQ, dried by 100% ACN, and re-hydrated with 100 mM NH₄HCO₃. 10 mM dithiothreitol and 55 mM iodoacetamide were sequentially applied to reduce the disulfide bonds of proteins and alkylate the exposed sulfhydryl (−SH) groups, respectively. The gel was then digested with sequencing grade trypsin (Promega, Madison, WI) in 50 mM NH₄HCO₃. Peptide solutions were recovered from the gel, desalted using Sep-Pak C18 cartridges (Waters, Milford, MA), and dried in a vacuum concentrator (Eppendorf, Hamburg, Germany).

The dried fraction in each biological sample was reconstituted with 0.1% formic acid and analyzed twice with a LTQ-Orbitrap Elite coupled to an Easy-nLC (Thermo Fisher, Bremen, Germany) as described (Mu et al., 2015).
Briefly, peptides were separated in a C18 capillary column (Michrom BioResources, CA) using a 90-min gradient. Mass spectrometry scans with a range of 350 to 1600 \( m/z \) were acquired with a resolution of 60000 and a positive charge mode. The five most abundant multiple-charged ions having a minimum signal threshold of 500.0 were selected for high-energy collision-induced dissociation (HCD) and fragmentation using collision-induced dissociation (CID). Both the HCD and CID scanning strategies adopted an isolation width of 2.0 \( m/z \). For HCD fragmentation, the activation time was 10 ms and the normalized collision energy was 45%. For CID fragmentation, the activation time was also 10 ms, but the normalized collision energy was 35%.

**2.2.5 Database search**

Raw LC-MS/MS data files were converted into .mgf files using Proteome Discovery 1.3.0.339 (Thermo Finnigan, CA), and submitted to Mascot version 2.3.2 (Matrix Sciences, London, U.K.) to search against the *P. maculata* database with 77584 protein sequences including both ‘target’ and ‘decoy’ sequences. Searching parameters were identical to those described in Mu *et al.* (2015) except that the maximum missed cleavage of trypsin was one, and the fixed medication was carbamidomethyl of cysteine. Matched peptides with an ion score of \( \geq 22 \) were retained to achieve a 95% confidence in identification. Only peptides longer than nine amino acids were retained because shorter peptides could easily match the decoy database. A 1% false discovery rate
threshold was applied in the final protein identification in each biological replicate. Proteins detected in at least two replicates and had at least three matched peptides were used for bioinformatic analysis. However, six proteins with at least 3 matched peptides and identified in one replicate were also kept because their presence could be confirmed transcriptomic analysis.

### 2.2.6 Transcriptomic expression analysis

To distinguish genes that were specifically expressed in albumen gland (AG) and other tissues (OT), we compared their expression levels quantified as transcripts per million (TPM) which takes gene length and sequence depth into account (Li & Dewey, 2011). TPM was calculated by dividing the read counts by the length of each gene in kilobases to get a reads per kilobase (RPK) value, summing up all the RPK values in AG and OT and dividing this number by 1000000 to get a “per million” scaling factor, then dividing the RPK values by the scaling factor. Genes which had an expression level (TPM) less than 0.5 were considered to be non-expressed (Grabherr et al., 2011).

### 2.2.7 Bioinformatic analysis

Protein abundance was quantified by Exponentially Modified Protein Abundance Index (emPAI) in Mascot (Ishihama et al., 2005). The PVF proteins were classified into several functional categories based on their annotation. SignalP 4.1 was applied to determine whether the PVF amino acid sequences
were secreted proteins (Petersen et al., 2011). Protein sequences of four subunits of PmPV1 (Pasquevich et al., 2014) were selected, and the corresponding orthologs in *P. canaliculata* were also picked up. The sequences were aligned using MUSCLE implemented in MEGA 7 (Kumar et al., 2016) in order to detect the conservative regions. They were also used to construct a phylogenetic tree to show the similarities of these protein homologues. RAxML1.5 beta (Silvestro & Michalak, 2012) was used to construct the tree with the following parameters: ML + rapid bootstrap with 1000 replicates, PROTGAMMA + GTR. The tree topology was visualized using FigTree v1.4.

### 2.2.8 Evolutionary analysis

A reciprocal best hit of the *P. maculata* and *P. canaliculata* (Sun et al., 2012) transcriptomes was used to identify orthologs between the two species (Mu et al., 2015). Only proteins with at least 67 amino acids were used in the orthologous analysis. Local BLASTp was performed with an *E*-value threshold of 1e-5. The paired orthologs between the two species were aligned by ParaAT1.0 (Zhang et al., 2012) and aligned codons with gaps were removed. KaKs_Calculator was applied to calculate the nonsynonymous substitutions per nonsynonymous site (Ka), synonymous substitutions per synonymous site (Ks), and the Ka/Ks ratio using the GY method (Zhang et al., 2006). Sequences with Ka > 2, Ks > 2, or Ks <0.01 were removed because they were either saturated for nonsynonymous and synonymous substitution or might result in high Ka/Ks
value due to the low Ks value (Villanueva-Cañas et al., 2013; Chen et al., 2014). To determine whether the genes coding PVF proteins evolved faster than housekeeping genes, the Ka, Ks, and Ka/Ks values of PVF gene orthologs were compared with housekeeping genes including actin, tubulin and ribosomal proteins (Zhang & Li, 2004) using the non-parametric Wilcoxon rank sum test (Zhang & Li, 2004; Lv et al., 2015). In order to make sure the expression levels of housekeeping genes were consistent in all tissues, expression levels (in TPM) of selected housekeeping genes in AG and OT were similar as an average ratio of AG to OT is 0.53.

2.3 Results
A total of 74 proteins were identified in the PVF of P. maculata on the basis of 6995 peptides, among which 364 were unique peptides (Table 2.1). They were classified into nine functional categories based on their annotation: three major perivitellin subunits, immune responsive proteins, oxidation-reduction, signaling and binding, cytoskeleton, protein degradation, energy metabolism, transcription and translation, and others. Abundance of these 74 proteins revealed by the emPAI method showed a broad range (from 0.02% to 24.41%), showing that our analytical approach was sensitive to detect the low abundance proteins. The five most abundant proteins in PVF were PmPV1-2 (24.41%), PmPV1-4a (18.75%), PmPV1-4b (14.15%), perivitellin ovorubin-3 (8.42%), and perivitellin ovorubin-2 (7.26%). This result is consistent with previous studies of P.
*canaliculata* which showed that ovorubin and two perivitellins were major components of PVF (Garin *et al.*, 1996; Sun *et al.*, 2012). Immune responsive proteins which played a vital role in egg defense also showed high abundant level as the 14 proteins accounted for 6.45% of total PVF abundance. Another three high abundant functional groups were signaling and binding (2.00%), protein degradation (1.27%) and energy metabolism (0.92%).

We got a total of 52,732,156 and 54,961,478 clean reads for albumen gland (AG) and other tissues (OT), respectively. The GC contents of AG and OT was 44.94% and 45.05%, respectively. The detailed assembly results were showed in Table 2.2. Briefly, there were 92,567, 130,305 and 105,349 assembled unigenes in AG, OT and combined AG & OT, respectively. The 105,349 sequences was used for protein identification database. These proteins were further annotated by searching against protein databases (i.e., NCBI nr, SwissProt, KEGG, GO and COG). After that, we got 38587 annotated unigenes.

Comparison of the transcriptome between albumen gland (AG) and other tissues (OT) showed that genes coding 24 of the PVF proteins were only highly expressed in AG such as PmPV1 subunits, PmPV2 and Kunitz-like protease inhibitor. Fig. 2.2 shows the correlation between the 74 PVF protein abundance and their corresponding gene expression levels in AG, and between the protein abundance and their expression levels in OT. Although six proteins were only detected only in one replicate, they were listed as putative PVF proteins because their TPM values in AG were at least five times higher than that of OT which
was a criteria used by a comparison of plant pollen and other organs (Huang et al., 2009). As shown in table 2.1, the five most highly expressed PVF genes in AG were PmPV1-4b, perivitellin ovorubin-2, PmPV1-4a, perivitellin protein, and perivitellin ovorubin-3, all with TPM value > 4300; while their expression levels in OT were less than 0.3, lower than the threshold of gene expression that is considered to be non-expressed (Grabherr et al., 2011). In particular, 24 genes were only expressed in AG among which most were perivitellins, indicating that AG organ is the source of many important PVF proteins. In addition, the expression levels of several PVF protein-coding genes were similar in AG and OT (i.e., tubulin, actin, and Ras-related proteins), indicating that these proteins are involved in general cellular functions.

The orthologous PmPV1 subunits between *P. maculata* and *P. canaliculata* were aligned together and a conserved motifs were detected: IXGGP (Fig. 2.3). Orthologous pairs between *P. maculata* and *P. canaliculata* showed a high similarity ranging from 89.7% to 94.4%; while the similarities between any two PmPV1 subunits was quite low with a range from 17.2% to 38.9%. Phylogenetic analysis of these sequences revealed that paired orthologs formed in a same clade and further grouped with other orthologs (Fig. 2.4).

All to all BLAST of the transcriptome between the two species resulted in 19072 pairs of orthologs. Base substitution analysis of the orthologs revealed the Ka/Ks values of 12231 pairs of genes. Among the genes coding the 74 PVF proteins, 19 had a Ka/Ks value. Three genes (i.e., perivitellin ovorubin-3,
15-hydroxyprostaglandin dehydrogenase [NAD+] and PmPV1-1) had a Ka/Ks value > 1, but the \( P \) value of Fisher’s exact test was larger than 0.05. Therefore, further evidences are needed to determine whether the PVF genes are undergoing positive selection. Comparison between PVF genes and housekeeping genes showed that the AG genes had significantly higher average Ka, Ks and Ka/Ks values (\( Ka = 0.014 \pm 0.012, \) Ks = \( 0.083 \pm 0.084, \) Ka/Ks = \( 0.208 \pm 0.134 \) in PVF genes, average ± SD; \( Ka = 0.0017 \pm 0.003, \) Ks = \( 0.030 \pm 0.022, \) Ka/Ks = \( 0.057 \pm 0.096 \) in housekeeping genes, average ± SD; Wilcoxon rank sum test, \( P < 0.01; \) Fig. 2.5) indicating that the PVF genes were under weaker selective constraints than housekeeping genes.

2.4 Discussion

2.4.1 Functional characterization of PVF proteins

The most abundant category of PVF proteins is the perivitellin group such as perivitellin ovorubin, PV2 MACPF and tachylectin subunits. Previous study on \( P. canaliculata \) indicated that perivitellins provided energy source and structural precursor during early embryo development (Heras \textit{et al.}, 1998). A recent study on the major perivitellin in \( P. maculata \) (i.e. PmPV1, representing ~52% of PVF proteins) characterized that it was a glyco-lipo-carotenoprotein and responsible for bright red coloration and protection from oxidation (Pasquevich \textit{et al.}, 2014). In the present study, 13 perivitellin homologs were detected in proteomic data, and the three most abundant proteins were ovorubin including its subunits,
perivitellin protein, and PV2 which is consistent with the composition of *P. canaliculata* (Sun et al., 2012).

Apart from the perivitellins that are involved in the conspicuous red coloration, antidigestion and antinutrition, immune responsive proteins form an important functional group in protecting the embryos from pathogens. Ferritin, which is the ninth abundant PVF protein (~1.38%) in *P. canaliculata*, is an iron storage protein and involved in the immunity (Bottke et al., 1988; Simonsen et al., 2011). It is a major exogenous yolk protein in the freshwater snails *Lymnaea stagnalis* and *Planorbarius corneus* (Bottke et al., 1988). Ferritin was overexpressed under 30-day estivation in *P. canaliculata* (Sun et al., 2013). Kunitz-like protease inhibitor also plays an essential role in innate immunity by serving as inhibitor of serine proteinase in the razor clam *Solen grandis* (Wei et al., 2012). Its protein abundance in our study was relatively low (~0.33%), but the transcript expression level in albumen gland (TPM = 1938.09) was much higher than that in other tissues (TPM = 0.30), indicating its high transcription rate in the AG of *P. maculata*. Serpin 1 and serine protease inhibitor 1 may protect the egg masses against foreign serine protease released by bacteria and viruses. They may also function in melanization by activating phenoloxidase (Hathaway et al., 2010). There are also several other immune-related proteins which were also detected in the PVF of *P. canaliculata* such as C1q domain-containing protein, peptidoglycan recognition protein S1L and thioester-containing protein-C (Sun et al., 2012). The thioester-containing
protein-C protein was almost equally transcribed in AG (TPM = 39.41) and OT (37.61) suggesting its essential role in the whole organism. But for C1q domain-containing protein and peptidoglycan recognition protein S1L, their transcripts abundance in AG (TPM = 19602.56 and 51.85, respectively) were much higher than that in OT (TPM = 0.06 and 0.00, respectively).

Perivitelline fluid is an energy source during the early development of the embryos. Study on *P. canaliculata* showed that the biochemical composition of PVF at stage I (from which stage it is almost impossible to separate perivitelline fluid and embryos) included 34.8% carbohydrates, 13.0% proteins, and 1.5% lipid in dry weight (Heras *et al*., 1998). Thus enzymes catalyzing carbohydrates, proteins and lipids are required. Our study detected several carbohydrate and protein metabolic enzymes, including 15-hydroxyprostaglandin dehydrogenase (NAD+), 4-hydroxyphenylpyruvate dioxygenase (HPPD), and cysteine proteases inhibitor. HPPD is an Fe(II)-dependent, non-heme oxygenase which catalyzes the 4-hydroxyphenylpyruvate into homogentisate. This enzyme functions in the second step of tyrosine catabolic pathway which is common to almost all aerobic life forms (Moran, 2005). Although the abundance of HPPD was low in protein and transcripts, its presence in the snail PVF further lends support to its wide distribution and conserved function. Cysteine proteases inhibitor played various roles in physiological and pathological processes such as inhibiting protein processing and degradation as well as MHC class II immune responses (Armstrong 2006; Vicik *et al*., 2006).
2.4.2 Evolutionary characterization of PVF

In Ampullariidae, different genera adopt various reproductive strategies including oviposition position, egg color and size, and shell texture (Hayes et al., 2009). For example, some genera laid gelatinous egg masses on the substrate in underwater (*Marisa* and *Lanistes*); while others laid calcareous eggs above the water such as *Pomacea* among which different species also differed in egg color and size. Since proteins in the perivitelline fluid of ampullariids played several vital roles including antioxidant, antinutrition and energy source (Heras et al., 2007), the PVF proteins may be associated with their different reproductive phenomena. Previous studies on the zona pellucida domain protein of abalone eggs suggested a positive selection of six of these reproductive genes (Aagaard et al., 2006). Therefore, we would like to examine whether the PVF gene are undergoing adaptive evolution.

Base substitution analysis of orthologs between *P. maculata* and *P. canaliculata* unveiled their sequence divergence. Although ten PVF genes had a Ka/Ks value > 0.5 or 1.0, the P values of Fisher’s exact text is larger than 0.05, thus further evidences (e.g., add additional sequences from other genera and use more sensitive method) were needed to determine potential positively selected amino acid sites. We further compared the Ka, Ks, and Ka/Ks values between PVF and a set of housekeeping genes and found that the PVF genes had significant higher values than housekeeping genes. This result is consistent with
previous study on mammalian tissue-specific (e.g., prostate, vulva, kidney, and liver) and housekeeping genes which showed that the former one had a lower level of evolutionary conservation (Zhang & Li, 2004).

Gene duplication was thought to play a vital role in the evolution of new genes (Findlay et al., 2009). Sun et al. (2012) analyzed the similarity of PVF proteins of *P. canaliculata* and identified 31 pairs of homologues which may be due to gene duplication. *Pomacea maculata* is a closely related congeneric species to *P. canaliculata*, and thus a good candidate to study the phylogenetic relationships of perivitellin homologs. As indicated in Figure 2.4, orthologs between *P. maculata* and *P. canaliculata* are roughly clustered in the same clade. Given the low similarities (21.78% to 45.23%) of perivitellin paralogs in *P. maculata*, we suggested that the PmPV1 subunit gene duplication may arise early during the speciation of the *Pomacea*. 
Figure 2.1. An individual of *Pomacea maculata* (shell length = 63 mm) crawling near the water surface with extending siphon (A); a newly deposited clutch of eggs (scale bar = 0.5 cm) (B); a SDS-PAGE gel image showing proteins in the egg perivitelline fluid of *Pomacea maculata*. The left gel lane shows protein molecular weight markers (kDa), and the right lane is a representative replicate. Each replicate was cut into 10 slices and used for LC-MS/MS analysis (D); anatomy of *P. maculata* with or without shells showing the position of representative structures and albumen gland which was used for transcriptome sequencing.
Figure 2.2. The correlated relationship between the protein abundance (emPAI) of 74 PVF proteins and their corresponding gene expression levels (expressed in TPM) in albumen gland (AG) and other tissues (OT).
Figure 2.3. Sequence alignment for four pairs of orthologous PmPV1 subunits between *P. maculata* (Pm) and *P. canaliculata* (Pc). * indicates a fully conserved amino acid site. There is one conserved motif: IXGGP.
Figure 2.4. Phylogenetic tree showing the relationships of orthologous PmPV1 subunits between *P. maculata* (Pm) and *P. canaliculata* (Pc). Numbers on the branches are bootstrap values in the maximum likelihood analysis.
Figure 2.5. Distribution of Ka, Ks and Ka/Ks values in PVF protein coding and housekeeping genes.
Table 2.1. Proteins detected from the egg perivitelline fluid of *Pomacea maculata*.

<table>
<thead>
<tr>
<th>Name</th>
<th>Annotation</th>
<th>MW (KDa)</th>
<th>No. of peptides</th>
<th>Average % (average ± SD)</th>
<th>Pm_TPM(^b)</th>
<th>AG</th>
<th>OT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Three major perivitellin subunits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL7293.C3</td>
<td>perivitellin protein(^5)</td>
<td>22.499</td>
<td>770</td>
<td>24.41 ± 3.58</td>
<td>46075.76</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Unigene35350</td>
<td>PmPV1-4a(^5)</td>
<td>21.944</td>
<td>1148</td>
<td>18.75 ± 8.93</td>
<td>66379.16</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Unigene40233</td>
<td>PmPV1-4b(^5)</td>
<td>22.368</td>
<td>800</td>
<td>14.15 ± 3.43</td>
<td>92663.06</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>Unigene34885</td>
<td>perivitellin ovarubin-3(^5)</td>
<td>22.68</td>
<td>226</td>
<td>8.42 ± 1.47</td>
<td>43994.87</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>Unigene34956</td>
<td>perivitellin ovarubin-2(^5)</td>
<td>23.748</td>
<td>385</td>
<td>7.26 ± 0.75</td>
<td>80798.28</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>CL2290.C1</td>
<td>PmPV1-1b(^5)</td>
<td>23.226</td>
<td>238</td>
<td>3.80 ± 1.34</td>
<td>3889.26</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>CL6258.C4</td>
<td>PV2 MACPF subunit(^5)</td>
<td>63.171</td>
<td>674</td>
<td>1.75 ± 0.37</td>
<td>4747.31</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>CL7293.C1</td>
<td>PmPV1-2</td>
<td>8.951</td>
<td>52</td>
<td>0.76 ± 0.12</td>
<td>15012.89</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>CL4382.C1</td>
<td>PV2 tachylectin subunit(^5)</td>
<td>31.728</td>
<td>180</td>
<td>0.63 ± 0.37</td>
<td>1979.86</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>CL4382.C2</td>
<td>PV2 tachylectin subunit(^5)</td>
<td>32.16</td>
<td>191</td>
<td>0.67 ± 0.22</td>
<td>1465.04</td>
<td>1.25</td>
<td></td>
</tr>
<tr>
<td>CL5816.C3</td>
<td>perivitellin protein(^5)</td>
<td>22.491</td>
<td>95</td>
<td>0.40 ± 0.16</td>
<td>40.45</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Unigene34835</td>
<td>tachylectin-like protein</td>
<td>29.536</td>
<td>80</td>
<td>0.59 ± 0.27</td>
<td>2580.39</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>CL5816.C1</td>
<td>perivitellin protein(^5)</td>
<td>22.505</td>
<td>80</td>
<td>0.40 ± 0.16</td>
<td>40.45</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td><strong>Immune responsive proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unigene40434</td>
<td>ferritin, partial</td>
<td>13.144</td>
<td>24</td>
<td>1.38 ± 1.24</td>
<td>1090.93</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>CL9552.C2</td>
<td>transferrin-like protein</td>
<td>36.762</td>
<td>79</td>
<td>0.83 ± 0.09</td>
<td>53.68</td>
<td>1.19</td>
<td></td>
</tr>
<tr>
<td>Unigene40228</td>
<td>C1q domain-containing protein(^5)</td>
<td>18.338</td>
<td>53</td>
<td>0.63 ± 0.22</td>
<td>19602.56</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>CL9552.C3</td>
<td>transferrin-like protein inhibitor-1(^5)</td>
<td>81.962</td>
<td>95</td>
<td>0.54 ± 0.14</td>
<td>56.52</td>
<td>1.64</td>
<td></td>
</tr>
<tr>
<td>Unigene38276</td>
<td>serine protease inhibitor-1(^5)</td>
<td>8.144</td>
<td>23</td>
<td>0.48 ± 0.24</td>
<td>122.04</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>CL6429.C2</td>
<td>serpin 1</td>
<td>39.27</td>
<td>95</td>
<td>0.45 ± 0.05</td>
<td>4856.93</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>CL8110.C1</td>
<td>peptidoglycan recognition protein S1L(^5)</td>
<td>11.343</td>
<td>24</td>
<td>0.42 ± 0.17</td>
<td>51.85</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>CL3024.C1</td>
<td>universal stress protein Stl1101-like</td>
<td>14.563</td>
<td>3</td>
<td>0.40 ± 0.27</td>
<td>6.70</td>
<td>8.26</td>
<td></td>
</tr>
<tr>
<td>CL539.C1</td>
<td>deleted in malignant brain tumors 1 protein</td>
<td>11.324</td>
<td>183</td>
<td>0.35 ± 0.18</td>
<td>796.74</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>CL1118.C4</td>
<td>Kunitz-like protease inhibitor(^5)</td>
<td>21.582</td>
<td>107</td>
<td>0.33 ± 0.09</td>
<td>1938.09</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>CL3601.C1</td>
<td>thioester-containing protein-C</td>
<td>157.143</td>
<td>127</td>
<td>0.27 ± 0.10</td>
<td>39.41</td>
<td>37.61</td>
<td></td>
</tr>
<tr>
<td>CL4196.C2</td>
<td>alpha-2-macroglobulin</td>
<td>31.204</td>
<td>18</td>
<td>0.20 ± 0.05</td>
<td>2.89</td>
<td>3.01</td>
<td></td>
</tr>
<tr>
<td>Unigene28932</td>
<td>ferritin, partial</td>
<td>19.276</td>
<td>3</td>
<td>0.14 ± 0.15</td>
<td>134.29</td>
<td>174.65</td>
<td></td>
</tr>
<tr>
<td>CL237.C1</td>
<td>putative chitinase 3</td>
<td>10.649</td>
<td>3</td>
<td>0.03 ± 0.06</td>
<td>334.51</td>
<td>35.15</td>
<td></td>
</tr>
<tr>
<td><strong>Oxidation-reduction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL9797.C1</td>
<td>pi-class glutathione</td>
<td>24.322</td>
<td>8</td>
<td>0.15 ± 0.10</td>
<td>41.03</td>
<td>12.27</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Average \% calculated from the total number of peptides detected.

\(^b\) Pm_TPM: Perivitellin transcript per million.
<table>
<thead>
<tr>
<th>Unigene/Clin</th>
<th>Description</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL4528.C1</td>
<td>S-transferase putative ferric-chelate reductase 1 homolog</td>
<td>62.479</td>
<td>30</td>
<td>0.10 ± 0.01</td>
<td>7.21</td>
</tr>
<tr>
<td>Unigene7542</td>
<td>Superoxide dismutase [Cu-Zn]</td>
<td>103.473</td>
<td>19</td>
<td>0.07 ± 0.01</td>
<td>88.14</td>
</tr>
<tr>
<td><strong>Signaling and binding</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL762.C1</td>
<td>calcium-binding protein, partial §</td>
<td>17.683</td>
<td>169</td>
<td>0.62 ± 0.14</td>
<td>10530.26</td>
</tr>
<tr>
<td>Unigene31436</td>
<td>PREDICTED: ras-related protein Rab-10-like</td>
<td>22.096</td>
<td>32</td>
<td>0.27 ± 0.13</td>
<td>14.19</td>
</tr>
<tr>
<td>CL762.C2</td>
<td>calcium-binding protein, partial</td>
<td>14.856</td>
<td>5</td>
<td>0.31 ± 0.29</td>
<td>15846.13</td>
</tr>
<tr>
<td>Unigene14225</td>
<td>Ras-like protein</td>
<td>21.046</td>
<td>14</td>
<td>0.22 ± 0.04</td>
<td>26.77</td>
</tr>
<tr>
<td>Unigene33406</td>
<td>RAB protein</td>
<td>23.054</td>
<td>19</td>
<td>0.16 ± 0.08</td>
<td>9.29</td>
</tr>
<tr>
<td>CL1720.C2</td>
<td>Rab3</td>
<td>20.738</td>
<td>8</td>
<td>0.14 ± 0.14</td>
<td>2.33</td>
</tr>
<tr>
<td>Unigene6988</td>
<td>Calcium-dependent protein kinase 31</td>
<td>16.928</td>
<td>5</td>
<td>0.12 ± 0.20</td>
<td>34.42</td>
</tr>
<tr>
<td>Unigene3919</td>
<td>Ras-related protein Rab-33B</td>
<td>26.752</td>
<td>16</td>
<td>0.06 ± 0.03</td>
<td>4.02</td>
</tr>
<tr>
<td>CL5295.C3</td>
<td>Synaptotagmin-like protein 5</td>
<td>45.449</td>
<td>21</td>
<td>0.06 ± 0.05</td>
<td>0.44</td>
</tr>
<tr>
<td>CL10768.C1</td>
<td>Choline transporter-like protein 2, partial</td>
<td>78.431</td>
<td>9</td>
<td>0.04 ± 0.04</td>
<td>7.38</td>
</tr>
<tr>
<td><strong>cytoskeleton</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL1298.C2</td>
<td>actin</td>
<td>42.119</td>
<td>30</td>
<td>0.21 ± 0.06</td>
<td>1.15</td>
</tr>
<tr>
<td>Unigene35143</td>
<td>CD109 antigen-like</td>
<td>43.83</td>
<td>21</td>
<td>0.14 ± 0.13</td>
<td>3.95</td>
</tr>
<tr>
<td>CL421.C7</td>
<td>Tubulin beta chain</td>
<td>50.26</td>
<td>12</td>
<td>0.08 ± 0.08</td>
<td>51.85</td>
</tr>
<tr>
<td>CL421.C2</td>
<td>beta-tubulin</td>
<td>50.25</td>
<td>12</td>
<td>0.08 ± 0.08</td>
<td>20.76</td>
</tr>
<tr>
<td>Unigene14616</td>
<td>ezrin/radixin/moesin</td>
<td>69.078</td>
<td>5</td>
<td>0.04 ± 0.04</td>
<td>32.09</td>
</tr>
<tr>
<td>Unigene434</td>
<td>Myosin-Id</td>
<td>104.954</td>
<td>6</td>
<td>0.03 ± 0.03</td>
<td>3.51</td>
</tr>
<tr>
<td><strong>Protein degradation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL1290.C1</td>
<td>N-acetylated-alpha-linked acid dipeptidase-like protein §</td>
<td>17.286</td>
<td>16</td>
<td>0.36 ± 0.18</td>
<td>4853.39</td>
</tr>
<tr>
<td>CL4546.C1</td>
<td>cysteine proteases inhibitor</td>
<td>21.102</td>
<td>10</td>
<td>0.29 ± 0.30</td>
<td>138.64</td>
</tr>
<tr>
<td>CL14.C1</td>
<td>leukocyte elastase inhibitor-like</td>
<td>40.886</td>
<td>12</td>
<td>0.18 ± 0.10</td>
<td>3.61</td>
</tr>
<tr>
<td>CL1656.C1</td>
<td>Amiloride-sensitive amine oxidase [copper-containing]§</td>
<td>84.207</td>
<td>47</td>
<td>0.17 ± 0.06</td>
<td>131.34</td>
</tr>
<tr>
<td>Unigene35321</td>
<td>Transmembrane protease, serine 2</td>
<td>25.376</td>
<td>20</td>
<td>0.10 ± 0.01</td>
<td>115.88</td>
</tr>
<tr>
<td>CL1819.C1</td>
<td>Transmembrane protease, serine 2</td>
<td>29.842</td>
<td>7</td>
<td>0.09 ± 0.09</td>
<td>182.26</td>
</tr>
</tbody>
</table>

<p>| § Indicates partial sequence. |</p>
<table>
<thead>
<tr>
<th>Unigene/Unigene ID</th>
<th>Gene Description</th>
<th>EmPAI (mean ± SD)</th>
<th>Log2 fold change</th>
<th>Unigene/Unigene ID</th>
<th>Gene Description</th>
<th>EmPAI (mean ± SD)</th>
<th>Log2 fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL7499.C1</td>
<td>alpha-1-inhibitor 3-like</td>
<td>71.366</td>
<td>9</td>
<td>0.08 ± 0.01</td>
<td>5.00</td>
<td>9.94</td>
<td></td>
</tr>
<tr>
<td><strong>Energy metabolism</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unigene6216</td>
<td>ADP-ribosylation factor 2-like</td>
<td>21.048</td>
<td>15</td>
<td>0.39 ± 0.19</td>
<td>30.90</td>
<td>31.66</td>
<td></td>
</tr>
<tr>
<td>CL9967.C1</td>
<td>15-hydroxyprostaglandin dehydrogenase [NAD+]</td>
<td>25.919</td>
<td>32</td>
<td>0.24 ± 0.09</td>
<td>52.09</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>CL10187.C1</td>
<td>4-hydroxyphenylpyruvate dioxygenase</td>
<td>43.541</td>
<td>17</td>
<td>0.17 ± 0.06</td>
<td>2.96</td>
<td>6.68</td>
<td></td>
</tr>
<tr>
<td>CL6535.C1</td>
<td>aldehyde dehydrogenase family 3 member B1-like isoform X1</td>
<td>54.884</td>
<td>24</td>
<td>0.09 ± 0.08</td>
<td>0.99</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Unigene34578</td>
<td>folate receptor beta-like</td>
<td>23.34</td>
<td>5</td>
<td>0.03 ± 0.06</td>
<td>1230.04</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td><strong>Transcription and translation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL2610.C3</td>
<td>apoptosis-inducing factor 3-like</td>
<td>58.429</td>
<td>108</td>
<td>0.43 ± 0.09</td>
<td>19.15</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>Unigene28387</td>
<td>ribosomal protein S14</td>
<td>16.193</td>
<td>37</td>
<td>0.11 ± 0.05</td>
<td>117.25</td>
<td>141.78</td>
<td></td>
</tr>
<tr>
<td>CL7340.C2</td>
<td>elongation factor 1 alpha</td>
<td>50.033</td>
<td>55</td>
<td>0.09 ± 0.02</td>
<td>1263.38</td>
<td>1555.68</td>
<td></td>
</tr>
<tr>
<td>Unigene18923</td>
<td>40S ribosomal protein S6</td>
<td>14.733</td>
<td>12</td>
<td>0.09 ± 0.09</td>
<td>1.28</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td><strong>Others</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unigene34633</td>
<td>BRAFLDRAFT_214672</td>
<td>15.939</td>
<td>38</td>
<td>1.53 ± 0.67</td>
<td>2351.89</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>CL2247.C1</td>
<td>Novel protein</td>
<td>29.305</td>
<td>130</td>
<td>1.19 ± 0.78</td>
<td>184.50</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Unigene40414</td>
<td>LOTGIDRAFT_234762S</td>
<td>24.551</td>
<td>33</td>
<td>0.72 ± 0.37</td>
<td>1172.28</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Unigene28826</td>
<td>Lustrin A, partial</td>
<td>16.64</td>
<td>34</td>
<td>0.46 ± 0.12</td>
<td>593.08</td>
<td>64.75</td>
<td></td>
</tr>
<tr>
<td>Unigene34623</td>
<td>Novel protein</td>
<td>34.624</td>
<td>34</td>
<td>0.35 ± 0.12</td>
<td>1537.44</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Unigene6438</td>
<td>BRAFLDRAFT_113974</td>
<td>53.817</td>
<td>49</td>
<td>0.25 ± 0.05</td>
<td>138.29</td>
<td>26.28</td>
<td></td>
</tr>
<tr>
<td>CL677.C1</td>
<td>tetraspanin, partial</td>
<td>25.157</td>
<td>12</td>
<td>0.07 ± 0.06</td>
<td>13.34</td>
<td>4.67</td>
<td></td>
</tr>
<tr>
<td>Unigene37522</td>
<td>Novel protein</td>
<td>12.038</td>
<td>3</td>
<td>0.07 ± 0.12</td>
<td>3.86</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Unigene7330</td>
<td>CGL_10006272</td>
<td>47.604</td>
<td>4</td>
<td>0.06 ± 0.06</td>
<td>15.80</td>
<td>20.58</td>
<td></td>
</tr>
<tr>
<td>Unigene34632</td>
<td>BRAFLDRAFT_92818</td>
<td>8.262</td>
<td>5</td>
<td>0.05 ± 0.08</td>
<td>8.59</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Unigene7050</td>
<td>BRAFLDRAFT_187602</td>
<td>23.366</td>
<td>3</td>
<td>0.04 ± 0.07</td>
<td>53.73</td>
<td>5.99</td>
<td></td>
</tr>
<tr>
<td>CL8811.C1</td>
<td>LOC101850990</td>
<td>122.888</td>
<td>27</td>
<td>0.02 ± 0.01</td>
<td>8.46</td>
<td>9.56</td>
<td></td>
</tr>
</tbody>
</table>

S indicates a predicted signal peptide.

aThis column shows the normalized emPAI values in the proteomic experiment. Data are expressed as average ± standard deviation which were calculated from three biological replicates.

bThis column shows the transcriptomic expression levels of genes detected in albumen gland (AG) and other tissues (OT). Data are expressed in transcripts per million (TPM).
Table 2.2. Statistics of assembly quality of albumen gland (AG) and other tissues (OT) in *Pomacea maculata*.

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Total Number</th>
<th>Mean Length(nt)</th>
<th>N50</th>
<th>Total Consensus Sequences&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Distinct Clusters&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Distinct Singletons&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>179,342</td>
<td>292</td>
<td>427</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>OT</td>
<td>211,148</td>
<td>343</td>
<td>557</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Unigene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>92,567</td>
<td>565</td>
<td>878</td>
<td>92,567</td>
<td>13,896</td>
<td>78,671</td>
</tr>
<tr>
<td>OT</td>
<td>130,305</td>
<td>629</td>
<td>1109</td>
<td>130,305</td>
<td>27,438</td>
<td>102,867</td>
</tr>
<tr>
<td>AG&amp;OT</td>
<td>105,349</td>
<td>785</td>
<td>1332</td>
<td>105,349</td>
<td>26,056</td>
<td>79,293</td>
</tr>
</tbody>
</table>

<sup>a</sup>All the assembled unigenes.

<sup>b</sup>The cluster unigenes; The same cluster contains many high similar unigenes which have more than 70% of similarity, and these unigenes may come from homologous or same gene.

<sup>c</sup>Distinct Singletons represents this unigene come from a single gene.
Chapter 3 Genetic basis of differential heat resistance between two congeneric freshwater snails

3.1 Introduction

The introduction of non-native species has become a serious problem around the world in recent decades due to the rapid increase in travel and trade (Strayer, 2010). Many non-native species which have successfully established, so called invasive species, have threatened ecosystems worldwide by reducing biodiversity, altering food web structure and changing ecosystem services (Keller et al., 2011). Whether an introduced species becomes invasive is dependent on both the characteristics of the host ecosystem and the biological traits of the species. Ecosystem characteristics include competition and facilitation by indigenous species, resource abundance, diversity, and frequency of disturbance (Blumenthal, 2006). Species traits include species' life span, trophic spectrum, reproductive output, dispersal ability, genetic adaptation, and tolerance to physical conditions (Kolar & Lodge, 2001; Franks & Munshi-South, 2014).

Tolerance to adverse physical conditions is considered to be one of the most important traits of many invasive species (Kolar & Lodge, 2001; Weir & Salice, 2012). A high degree of tolerance may allow invasive species to survive through temporarily adverse conditions that are lethal to other species, thereby allowing them to occupy environments that are otherwise uninhabited in the invaded ecosystem. This ability may also allow invasive species to develop new
adaptations through additive genetic variance or directional selection (Prentis & Wilson, 2008; Vandepitte et al., 2014).

Research on stress responses is one of the focuses of invasive species biology, as the results can help not only in understanding why some species become invasive, but also in predicting species’ range expansion and impact (Hochachka & Somero, 2002). The comparative study of congeners has been promoted as an approach to understand the differential stress responses of invasive vs. noninvasive species, as it allows one to distinguish effects of adaptive variation from those of phylogeny (Cooper, 1999; Somero, 2011). In this context, a variety of responses can be measured at different biological organization levels spanning from the molecule to the population (Rahel & 2008; Dong & Somero, 2009; Zerebecki & 2011; Lejeusne et al., 2014). Comparative studies integrating conventional physiological (e.g. activity pattern, heart rate, oxygen consumption, survivorship and growth) and high-throughput “omics” analyses, can reveal the molecular basis of physiological changes, and have been considered as a “crystal ball” that can be used to predict the impact of environmental changes on biota (Somero, 2011). For example, a comparison of the native blue mussel Mytilus trossulus and its invasive congener Mytilus galloprovincialis in the eastern Pacific using heart rate (Braby & Somero, 2006), transcriptomics (Lockwood & Somero, 2011) and proteomics (Tomanek & Zuzow, 2010; Tomanek, 2011; Fields et al., 2012) uncovered the causal mechanisms underlying the shift in their biogeographic limits in response to long term changes in seawater temperature.
The present study examined the responses of two congeneric apple snails of the family Ampullariidae (Gastropoda) to thermal stress (Fig. 3.1). *Pomacea canaliculata*, an indigenous species of South America, is now widely spread in Asia, Hawaii and southern USA (Joshi & Sebastian, 2006; Rawlings et al., 2007; Kwong et al., 2008; Hayes et al., 2008 & 2015; Yoshida et al., 2013). This species is also a threat to human health as it hosts a number of human parasites (Lv et al., 2009). A high tolerance of environmental stresses has been suggested to contribute to the invasion success of several species of *Pomacea* including *P. canaliculata*, as this may determine whether these snails can establish in an ecosystem once introduced, and also affects a number of life history traits (survival, growth, reproduction) that are critical to the population expansion of the invasive species (Cowie, 2002; Heras et al., 2007; Hayes et al., 2008; Yamanishi et al., 2012; Yoshida et al., 2013). Tolerance to cold has been investigated to understand the seasonal survival of *P. canaliculata* (Matsukura et al., 2009), and applied in climate models to predict its spread in Australia, New Zealand, Southeast Asia, East Asia, North America (Baker, 1998), and Europe (EFSA PLH Panel, 2013). *Pomacea diffusa*, another native of South America, is considered a non-invasive species. It is a common species in the aquarium trade around the world and has been introduced into the wetlands of Florida (Rawlings et al., 2007), Australia and Sri Lanka (Hayes et al., 2008; Tran et al., 2008), and New Zealand (Collier et al., 2011), but it has never been reported to form large populations or become a pest. In fact, *P. diffusa* is the only member of the family Ampullariidae
that is allowed to be transported interstate without a permit by the United States Department of Agriculture (Rawlings et al., 2007). The two Pomacea congeners are both indigenous to South America but their native distribution ranges do not overlap (Kenneth Hayes, pers. comm.). Pomacea canaliculata likely evolved in a region that experiences greater temperature extremes than P. diffusa, and the two species might have differential tolerances and physiological responses to temperature stresses.

Two experiments were thus conducted to compare the responses of the two apple snails to heat stress in the present study. The first experiment determined the lethal temperature by exposing the snails to acute heat stress. The second experiment compared the protein expression profiles of the two species exposed to acute and chronic thermal stress using iTRAQ-coupled LC-MS/MS, in order to detect protein biomarkers for assessing the graded responses of the two congeners to different types of thermal stress. Genes encoding the differentially expressed proteins between different thermal treatments of the same species, as well as the same thermal treatment between the two species, were analyzed for sequence divergence in order to identify genes that may have undergone rapid evolution. These analyses should allow us to gain insight into the genetic basis of differential thermal tolerance in the two congeners.

3.2 Materials and methods

3.2.1 Animal collection and acclimation
Adult *P. canaliculata* were collected from drainage channels and ponds in Yuen Long (22.505° N, 114.114° E), the only vegetable producing area in Hong Kong, whereas adult *P. diffusa* were purchased from an aquarium shop in Hong Kong. Their identities were confirmed by sequencing a fragment of the cytochrome *c* oxidase subunit I gene (GenBank accession numbers: KT313032, KT313033, KT313034, and KT313035) after amplifying the DNA using the LCO 1490 and HCO2198 primers (Folmer *et al.*, 1994), and comparing the sequences with those deposited in GenBank. The two species were cultured separately in 250-L aquaria at Hong Kong Baptist University in tap water that was dechlorinated by aeration with an air pump for at least two days before use. Snails were acclimated at 25 ± 1 °C which is an optimal temperature for reproduction and growth of laboratory cultured *Pomacea* snails (Estebenet & Cazzaniga, 1992). In each aquarium, an air pump was used to provide aeration, a submerged pump was used to recirculate the water, and a canister filter was used to remove solid wastes. About one third of water was changed every three days. The snails were fed with lettuce leaves, carrot and Hikari Lionhead fish feed pellets (Himeji, Japan), and any excess was removed daily. As both species grazed on the biofilm developed on the aquarium walls, the aquaria were illuminated to encourage algal growth with a 14 h light: 10 h dark photoperiod. Both species fed on the four food items (lettuce, carrot, fish feed and encrusting algae), although *P. diffusa* consumed less lettuce and carrot than *P. canaliculata*. All snails were acclimated for at least four weeks before being used in experiment.
3.2.2 Tolerance to acute thermal exposure

The upper thermal tolerance limit, indicated by the death of the test animal, was determined using an experimental protocol similar to that of Somero & Stillman (1996). Snails (*P. canaliculata* shell length = 19.56 ± 2.61 mm, mean ± SD, *n* = 30; *P. diffusa* = 18.71 ± 2.04 mm, *n* = 35) were starved for 24 h and then placed into different glass aquaria inside a JULABO TW20 water bath (JULABO Labortechnik GmbH, Seelbach, Germany). The water inside the aquaria was continuously aerated using an air pump. A grid was placed near the water surface to prevent snails from crawling out of water. The snails were transferred from a stock culture maintained at 25 °C to the exposure aquaria at the same temperature. The water used in the experiment was also taken from the stock tank. The snails were allowed to acclimatize for 30 min. Afterwards, the temperature was increased at a rate of 1 °C per 20 min. At every 20 min interval, the snails were checked for survival. The foot of the non-motile snails was prodded using a stick to determine whether they reacted to physical stimuli. Those without response were all confirmed to be dead after the snails were individually labelled and transferred back to 25 °C water for a 24-h recovery. The relationship between percent mortality (*Y*) and temperature (*X*) was expressed using a logistic model (Motulsky & Christopoulos, 2003; Qiu *et al*., 2007):

\[
Y = \frac{100}{1 + 10^S(LT_{50} - X)}
\]

where *S* is the hillslope of the regression, and *LT*\(_{50}\) the lethal temperature at which
50% of snails died. $F$ test was conducted to compare the hillslope and $LT_{50}$ between the two species using GraphPad Prism 4.

### 3.2.3 Protein expression

#### 3.2.3.1 Exposure to acute and chronic thermal stress

Acute and chronic heat exposure was conducted at 37 °C and 35 °C, respectively. Of the two temperatures, 37 °C is stressful for the two species as, in a preliminary experiment, all individuals of *P. diffusa* cultured at this temperature died within 24 h of exposure, and all individuals of *P. canaliculata* were inactive at day 1 of exposure to 37 °C, and approximately half of them died at day 10; HM, personal observation). Previous studies have shown that *P. canaliculata* cultured at 35 °C had reduced crawling and feeding activities (Seuffert *et al.*, 2010) as well as reduced growth (Seuffert & Martín, 2013), therefore this temperature was considered chronically stressful. Adults of *P. canaliculata* and *P. diffusa* (shell length 20-30 mm, $n = 9$) were used in the experiment. For each species, the acute heat treatment involved transferring the snails to an aquarium where the water temperature was gradually increased (0.3 °C/min) to 37 °C over 30 min, and maintained at this temperature for 3 h (Acute37 °C), followed by a 24-h recovery period at 25 °C before sampling to account for the lag between stress exposure and protein expression (Dowd *et al.*, 2010; Tomanek & Zuzow, 2010). Snails maintained at 25 °C served as control (Acute 25 °C). For the chronic heat exposure, snails were maintained at 35 °C for 21 days (Chronic 35 °C) and at
25 °C for 21 days as control (Chronic 25 °C). This chronic exposure control was used to eliminate the potential effects of subtle differences in experimental conditions on protein expression. Similar to the diet used during the acclimation period, both species were provided with sufficient lettuce, carrot and fish feed daily, and the aquaria had biofilm on the walls. Excess food and waste were removed daily. At the end of the exposure, snails were dissected and their hepatopancreas placed in 8 M urea and stored at -80 °C. The hepatopancreas was selected for analysis as this organ is important for energy metabolism and waste storage, and its protein expression has been shown to be responsive to another environmental stressor (desiccation) (Giraud-Billoud et al., 2011 & 2013; Sun et al., 2013).

3.2.3.2 Protein extraction, digestion and labelling

Samples were placed on ice to reduce the influence of heat. Hepatopancreas tissues were homogenized using a plastic pestle. Sonication (Sonic Dismembrator 300, Thermo Fisher Scientific, Waltham, MA) was then applied to break cell membranes with the following settings: 1 min at a power setting of 28, followed by 3 cycles of 15-s burst with 30-s pause intervals at a power setting of 35. The protein solutions were then centrifuged for 15 min at 4 °C and 15,000 g. The supernatant was collected, purified using a 2D-cleanup kit (Bio-Rad, Hercules, CA), and reconstituted using 8 M urea. Each sample was quantified using a RC-DC kit (Bio-Rad). An aliquot containing 200 μg protein was reduced by 5
mM triscarboxyethyl phosphine hydrochloride at 60 °C for 1 hour. Methylethanethiosulfonate (10 mM) was used to alkylate the exposed sulfhydryl (–SH) groups for 20 min at room temperature. Sample solutions were diluted 8-fold with 50 mM triethylammonium bicarbonate. Afterwards, protein samples were digested for 16 h at 37 °C using sequencing-grade trypsin (Promega, Madison, WI) with an enzyme-to-protein ratio of 1:50 (w/w). Digested solutions were desalted using Sep-Pak C18 cartridges (Waters, Milford, MA) and dried in a vacuum concentrator (Eppendorf, Hamburg, Germany). For each species and each biological replicate, samples from Chronic 25 °C, Acute 25 °C, Chronic 35 °C and Acute 37 °C were labeled with iTRAQ reagents 114, 115, 116 and 117 (AB Sciex, Framingham, MA), respectively. Labeled samples (including control and treatments labeled with 114, 115, 116, and 117 reagents) from one biological replicate were then pooled together and dried, and a total of three biological replicates for each species were used for the following analysis.

3.2.3.3 Sample fractionation and LC-MS/MS analysis

In each biological replicate, dried samples were reconstituted using Buffer A (10 mM KH₂PO₄ and 20% acetonitrile (ACN), pH 3.0) and fractionated using a PolySULFOETHYL strong cation-exchange column (200x4.6 mm, 5 μm particle size, 200-Å pore size, PolyLC, Columbia, MD) on a Waters 2695 HPLC. Sample solutions were fractionated using a 50 min fractionation gradient at a flow rate of 1 mL/min: 100% Buffer A for the initial 5 min, 0-30% (linear changing gradient)
Buffer B (10 mM KH$_2$PO$_4$, 0.5 M KCl and 20% ACN, pH 3.0) for 28 min, 30-100% (linear changing gradient) Buffer B for 5 min, 100% Buffer B for 5 min and a final 7 min of 100% Buffer A. Fourteen fractions were collected and dried in a vacuum concentrator, then desalted using Sep-Pak C18 cartridges.

Dried fractions were reconstituted with 0.1% formic acid. Each sample was analyzed twice with LTQ-Orbitrap Elite coupled with an Easy-nLC (Thermo Fisher, Bremen, Germany). A C18 capillary column (Michrom BioResources, CA) was used to separate peptides with a 90 min gradient: 100% Solution A (0.1% formic acid in Milli-Q water) for 5 min, 0-30% (linear changing gradient) Solution B (0.1% formic acid in ACN) for 55 min, 30-98% Solution B for 10 min, maintained at 98% Solution B for 10 min, and finally re-equilibrated at 100% Solution A for 10 min. Mass spectrometry scans ranging from 350 to 1,600 m/z were acquired with a resolution of 60,000 in the positive charge mode. Five multiple-charged ions with the highest intensities and a minimum signal threshold of 500.0 were selected for fragmentation using collision-induced dissociation (CID) and high-energy collision-induced dissociation (HCD). The CID spectra were used for identifying peptides in LTQ, while the HCD spectra were used for iTRAQ quantification in C-trap. Both CID and HCD scanning approaches which had activated dynamic exclusion adopted an isolation width of 2.0 m/z. The dynamic exclusion time was 30s and mass window for precursor ion selection was 2.0Th. For the CID fragmentation, normalized collision energy of 35%, activation Q of 0.25 and activation time of 10 ms were used. For the HCD fragmentation,
parameters were set as follows: full scan with FTMS at a resolution of 15,000 in a centroid mode, activation time of 10 ms and normalized collision energy of 45%.

3.2.3.4 Mass spectrometry data analysis

Raw data were converted to .mgf files using Proteome Discoverer 1.3.0.339 (Thermo Finnigan, CA). A Python script (Appendix 1) was applied to filter unpaired scans, separate data generated by the CID and HCD scans, replace reporter groups (i.e. 114-117.5 Da) in CID scans by those of HCD scans from the same parent ion, and normalize the intensity of ions in HCD. The newly generated CID and HCD data files were separately submitted to Mascot version 2.3.2 (Matrix Sciences, London, UK) to search against a *P. canaliculata* database (Sun *et al.*, 2012) and a *P. diffusa* protein database (81,992 protein sequences which were translated from transcriptomic sequences including both ‘target’ and ‘decoy’ sequences). The search criteria for CID scans were set as monoisotopic experimental mass values, no restriction on protein mass, peptide charges of 2+, 3+ and 4+, a maximum number of 2 missed cleavages for trypsin-digested peptides, fixed modification for methylthio of cysteine, variable modification for deamidated asparagine and glutamine, variable modification for oxidation of methionine, ± 5 ppm for precursor ions, MS/MS ions search with fragment ions tolerance of ± 0.6 Da, iTRAQ 4 plex quantification method, and ESI-TRAP instrument. Parameter settings for HCD were similar to those of CID except that fragment ions tolerance was set to ± 20 mmu. The results were filtered to remove
data with ion scores less than 28, which indicated that the observed match was a random event (5% probability). Peptides which contained less than 7 amino acids were deleted as short peptides might easily match decoy sequences. The false discovery rate was set as 1% in the target-decoy database search (Sun et al., 2013). A Python script was applied to remove peptides which matched more than one protein in order to reduce quantitative bias caused by conserved regions within the same protein group. Both unlabeled peptides and peptides with erratic ratio between any two iTRAQ reporters (e.g. 116/114 more than 100 or less than 0) were removed. The median ratios between different iTRAQ reporters (i.e. 116/114 and 117/115) were normalized to 1. Proteins that were detected in three biological replicates and had at least 4 summed peptides in all of the three replicates were used for quantification analysis. Protein fold change in each replicate was quantified based on the summed intensity of matched peptides. In order to determine the criteria for significant differential expression, data from two replicate runs (Set 1 and Set 2) of the same biological replicate were analyzed (Yang et al., 2012). The log2-transformed fold-change ratios between Set 1 and Set 2 in 116/114 and 117/115 followed a normal distribution, respectively. The cutoff values were defined as the log2 transformed data at which 95% of all target proteins did not deviate from each other. The cutoff values were 0.45 and -0.47 for 116/114, and 0.32 and -0.32 for 117/115 which corresponded to 1.4 and 0.71-fold for 116/114, and 1.25 and 0.80-fold for 117/115. These fold-values were then used as thresholds for determining significantly up- or down-regulated proteins.
proteins between the control and treatment groups of each species.

3.2.3.5 Functional annotation

Unigenes in the transcriptome of each species were searched against NCBI nr, SwissProt, KEGG and COG databases using BLASTx with an \( E \)-value threshold of 1.00e-5. Protein functions were predicted from the annotation of the most similar protein among those databases. Blast2GO annotation resulted in the assignment of proteins into the Gene Ontologies (GOs) of Biological Process (BP), Molecular Function (MF) or Cellular Component (CC). Differentially expressed proteins (DEPs) which were annotated by GO were further classified by WEGO(Ye et al., 2006) and assigned to several level 2 BP terms. DEPs without GO annotations were labeled as “Others”. The expression levels of DEPs under each GO term were compared using log2-transformed data to produce a Spearman rank correlation similarity matrix, and clustered using the centroid linkage method and visualized by Java Treeview (Sun et al., 2013).

3.2.4 Base substitution rate analysis

Like many other “non-model” organisms, the scarcity of genomic resources has limited the application of the high-throughput “omics” approach, which is a major obstacle for revealing molecular mechanisms underlying responses to various stressors (Elmer et al., 2010). To enable proteomic study in \( P. \) diffusa, we conducted \textit{de novo} transcriptome sequencing of its pooled tissues using an
Illumina HiSeq 2000, which generated 108 million clean reads with 9.8 billion nucleotides. Among the assembled sequences, 39,074 unigenes were annotated and translated to protein sequences. Coding regions of another 7,208 sequences were predicted by ESTscan software. The database of *P. canaliculata* has been reported, and applied to understand the snail’s responses to desiccation (Sun *et al.*, 2012 & 2013). With these species-specific transcriptomic databases and the newly acquired MS data, we were able to identify a set of orthologous proteins between these two congeners.

### 3.2.4.1 Orthologous protein identification and Ka/Ks estimation

A reciprocal best hit method (Elmer *et al.*, 2010) was adopted to identify orthologous proteins between *P. canaliculata* and *P. diffusa*. Proteins with less than 67 amino acids which correspond to 200 bp nucleotides were removed (Wang *et al.*, 2011), and then local BLASTp (NCBI_blast_2.2.26) was performed with an *E*-value threshold of 1.00e-05. Paired orthologous protein-coding sequences were aligned by ParaAT1.0 (Zhang *et al.*, 2012). To examine the sequence divergence pattern of the orthologous proteins, nonsynonymous substitutions per nonsynonymous site (Ka), synonymous substitutions per synonymous site (Ks) and the ratio between Ka and Ks were calculated using the MLWDL method implemented in KaKs_Calculator (Zhang *et al.*, 2006). Protein-coding sequences with Ks > 0.1 were excluded in further analysis in order to avoid the inclusion of paralogs (Bustamante *et al.*, 2005).
3.2.4.2 Differentially expressed orthologs between species and treatments of the same species

The expression levels of the orthologs were compared to determine the differentially expressed proteins between the two species and different thermal treatments of the same species. For the same thermal exposure, protein fold-change ratios between the two species were calculated and Student $t$-test was then applied with a $P$-value of 0.02 instead of using multiple comparison corrections to limit the number of false positives (Tomanek & Zuzow, 2010). For the different thermal treatments of the same species (e.g. Chronic 35 °C/Chronic 25 °C, Acute 37 °C/ Acute 25 °C), protein fold-change values are also compared using the same test and significance criterion.

The differentially expressed orthologs were classified into nine categories on the basis of their GO functional annotation. Proteins without annotation were classified into the ‘others’ group. To determine the functional relationships of the DEPs, GO enrichment was performed using the Gene Ontology Enrichment Analysis Software Toolkit (GOEAST) (Zheng & Wang, 2008). Data were analyzed with Fisher’s exact test and Adrian Alexa’s improved weighted scoring algorithm, with the significance level of enrichment set as 0.01 as there was no multi-test adjustment (Morris et al., 2014).

3.3 Results
3.3.1 Tolerance to acute thermal exposure

With the increase in temperature, the two species of *Pomacea* exhibited similar behavioral patterns: normal behaviors (attached to wall or slowly crawling) near acclimation temperature; increasingly active (actively crawling and trying to crawl out of water) at higher temperatures until reaching roughly 38 to 39 °C, and reduced activities to ceased movement at higher temperatures. An analysis of the mortality-temperature curves showed that the hillslopes (Fig. 3.2) did not differ significantly between the two species (1.12 in *P. canaliculata* vs. 0.88 in *P. diffusa*, $P = 0.1208$). However, the $LT_{50}$ of *P. canaliculata* was significantly higher than that of *P. diffusa* (47.0 °C for vs. 44.8 °C, $P < 0.0001$). The mean lethal temperature of *P. canaliculata* was also significantly higher (47.3 °C) than that of *P. diffusa* (44.9 °C) ($P < 0.00001$).

3.3.2 Proteome analysis

3.3.2.1 Reproducibility of proteomic data

Data reproducibility among technical and biological replicates is a critical issue in quantitative proteomics (Gan *et al*., 2007). Data from two replicate runs (Set 1 and Set 2) of the same biological replicate were used to determine the technical variance of our experimental workflow (Yang *et al*., 2012). Fold-change ratios of 116/114 and 117/115 generated by the two separate experimental runs were log₂-transformed. Linear correlation analysis revealed significant positive correlation between the two runs (116/114: $R^2 = 0.7250$, $P <$
0.0001; 117/115: \( R^2 = 0.6701, P < 0.0001 \). This analysis has allowed us to
determine the cut-off values for differential expression.

In *P. canaliculata*, 2,389, 2,644, and 2,459 quantified proteins were detected
in the three biological replicates (each replicate was a pooled sample from 3
individuals), respectively; and 1,677 were shared among the three replicates. In *P.
diffusa*, 2,415, 2,524 and 2,648 quantified proteins were detected in the three
replicates, respectively; and 1,802 were shared among the three replicates. Our
examination of the data showed that the ~ 1000 proteins that did not occur in all
the three replicates were usually low abundance proteins. For example, in the one
biological replicate of *P. canaliculata*, 1,677 quantified proteins which were
detected among three replicates account for 96.4% of total protein abundance in
this replicate, and the remaining 712 quantified proteins which were not shared by
three replicates only account for 3.6%. Even though they are low-abundant, they
were still detected probably because we have used a high-end LC-MS/MS system.
Compared with several recent proteomic studies, the percentages of shared
proteins among replicates were similar (Pütz et al., 2012; Effertz et al., 2014), but
the absolute numbers of shared identified proteins among replicates were much
higher, indicating the high-throughput of our experimental workflow without
compromising data reproducibility.

3.3.2.2 Comparison of protein expression among experimental treatments
in the same species
A total of 3,395 and 3,368 proteins were identified from the three biological replicates of *P. canaliculata* and *P. diffusa*, respectively. Among them 1,677 and 1,802 proteins (intersection of three biological replicates, with at least four peptides matched in all these three replicates) were quantified for *P. canaliculata* and *P. diffusa*, respectively. Acute and chronic thermal exposure resulted in 27 and 58 DEPs in *P. canaliculata*, and 22 and 118 DEPs in *P. diffusa*, respectively.

The DEPs were classified into five level 2 Biological Process (BP) GO terms: localization, immune system process, response to stimulus, cellular process, and metabolic process (Figs. 3.3 & 3.4). The “localization” GO term included proteins that are responsible for transporting molecules such as fatty acid-binding protein (adipocyte), which transports lipids between organelles. The “response to stimulus” GO term mainly included molecular chaperones such as heat shock proteins (HSPs). Based on their molecular weight and similarity in function, HSPs are divided into 6 families: HSP110, HSP90, HSP70, HSP40, small HSP and chaperonin (Kampinga *et al.*, 2009). Among the differentially expressed chaperones found in *P. diffusa*, 13 HSPs (1 HSP90; 6 HSP70; 1 HSP40; 4 small HSP; 1 chaperonin) were induced by acute heat exposure, and 14 HSPs (3 HSP90; 5 HSP70; 1 HSP40; 5 small HSP) and 4 protein disulfide isomerases (PDIs) were induced by chronic heat exposure. In *P. canaliculata*, acute heat exposure resulted in the differential expression of 4 HSPs (1 HSP70; 3 small HSPs); chronic heat exposure resulted in the
differential expression of 8 HSPs (4 HSP70; 4 HSP90) and 2 PDIs. Both “cellular process” and “metabolic process” GO terms contained many enzymes and some other proteins related to carbohydrate and fatty acid metabolism.

The expression patterns of DEPs under each GO term of *P. canaliculata* and *P. diffusa* are shown in Figs. 3.3 & 3.4, respectively. Most of the acute exposure induced DEPs are heat shock proteins and they were all significantly up-regulated. Under chronic heat exposure, roughly half of the DEPs classified under “others”, “cellular process”, “metabolic process” and “localization” terms were up-regulated and half were down-regulated in both species. The “immune system process” included several immune responsive proteins such as peptidoglycan recognition protein, C-type lectin D1, vesicular integral-membrane protein VIP36 and haemocyanin in *P. canaliculata*. C-type lectin plays a crucial role in the innate immune system against pathogen infection in the scallop *Chamys farreri* (Zhang et al., 2009). Vesicular integral-membrane protein VIP36 containing a lectin-like domain acts as pattern recognition receptors in the Chinese mitten crab *Eriocheir sinensis* (Huang et al., 2014). Peptidoglycan recognition proteins function as innate immunity molecules in mollusks (Dziarski & Gupta, 2006). Haemocyanin, an oxygen transporting protein, also has phenoloxidase (PO) activity. Crustaceans can convert haemocyanin to PO-like enzyme. One by-product of the PO cascade is phenol, which plays a vital role in immune reaction of crustaceans (Adachi et al., 2003). Haemocyanin might be related to hemocyte aggregation, which is a
defense mechanism in *P. canaliculata* (Cueto *et al*., 2013). In addition, proteins involved in “responding to stimulus” were greatly up-regulated in both species under both acute and chronic thermal stress, indicating the important role of molecular chaperones during chronic heat exposure.

### 3.3.3 Identification of orthologous proteins, and estimating Ka/Ks of differentially expressed orthologs

Reciprocal best hits from the protein database of the two species revealed 13,308 putative orthologs. Among them, 310 were differentially expressed between the same treatment in different species and treatments of the same species (Fig. 3.5). Analysis of genes encoding these differentially expressed proteins revealed nine genes with Ka/Ks values > 0.5 (two Ka/Ks > 1; seven Ka/Ks between 0.5 and 1; Fig. 3.5). Ka/Ks > 1 is usually considered as clear evidence of rapid positive Darwinian selection (Yang & Bielawski, 2000). Whereas genes with Ka/Ks > 0.5 have been considered as good candidates for determining whether certain section of the sequences has undergone positive selection (Swanson *et al*., 2004). Among the nine genes, three were classified into ‘others’ due to unknown GO functional annotations. Three of these genes are related to ‘energy metabolic process’, including nucleoside diphosphate-linked moiety X motif 19 (mitochondrial), placental protein 11 and solute carrier family 2 (facilitated glucose transporter member 1). Two of these genes are involved in “cellular redox homeostasis”, annotated as glutathione S-transferase theta-1 and NADH dehydrogenase
[ubiquinone] iron-sulfur protein 4 (mitochondrial). One protein-coding gene, annotated as midline-2, is involved in “signaling and binding”.

3.3.4 Pathway enrichment analysis

Pathway enrichment analysis, conducted with the differentially expressed proteins between the two species and different thermal treatments of the same species, resulted in the enrichment of 80 GO terms, including 42 biological process GO terms (Table 3.1). The enriched pathways include energy metabolism (e.g. ATP synthesis and hydrolysis coupled proton transport, fatty acid beta-oxidation using acyl-CoA dehydrogenase), protein folding and degradation (e.g. protein polyubiquitination, protein folding), cellular redox homeostasis (e.g. cell redox homeostasis, oxidation-reduction process), transcription and translation process (e.g. translation, cytoplasmic mRNA processing body assembly), transport and localization (e.g. intracellular protein transmembrane transport, establishment of protein localization), amino acid metabolism (e.g. regulation of cellular amino acid metabolic process) and immune response (antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent).

3.4 Discussion

Responses to stressors including heat at the cellular level involve several groups of evolutionarily conserved genes, particularly those regulating energy utilization,
DNA and protein structural stabilization and repair, cell cycle, scavenging reactive oxygen species and removal of damaged proteins (Storey, 2004; Kültz, 2005; Somero, 2011). Comparing the protein expression profiles between the two apple snails and between different thermal treatments of the same snail has allowed us to gain insights into the complex protein networks that are coordinated to cope with heat stress in these congeners. While the expression of stress responsive genes and proteins has been relatively well characterized in a few congeneric marine animals (Johns & Somero, 2004; Dong & Somero, 2009; Tomanek & Zuzow, 2010; Lockwood & Somero, 2011; Fields et al., 2012), little is known about the genetic basis (e.g. sequence divergence between orthologs) of the differential expression.

Since our study is the first to compare congeneric aquatic animals using iTRAQ coupled LC-MS/MS as well as estimate substitution rate of orthologous sequences, we will first discuss the differentially expressed functional protein groups within species. Within each protein functional group, we will point out the differential expression pattern between species. Then, we will analyze the base substitution rate of orthologous protein-coding genes and pathway enrichment of the differentially expressed proteins. In addition, we will point out the technical challenges for determining orthologs in non-model organisms, and insights from such analysis in the apple snail congeners.

### 3.4.1 Molecular chaperones
Exposure to high temperature or other stresses can trigger the synthesis of heat shock proteins (HSPs). HSPs act as chaperones of other proteins, playing an important role in protein folding and preventing unwanted protein aggregation. Protein disulfide isomerase (PDI), a thioredoxin superfamily oxidoreductase, also contributes to the refolding of denatured proteins in addition to its role in maintaining cellular redox homeostasis (Laurindo et al., 2012).

Consistent with their role in stabilizing protein structures, all these molecular chaperones in both species were up-regulated in response to the acute/chronic heat stress. However, there were notable differences in the expression of chaperones between the two species and between the acute and chronic exposure (Figs. 3.3 & 3.4). Acute and chronic heat exposure appeared to induce the expression of different sets of chaperones, with several high molecular weight HSPs (HSP90 and PDIs) being up-regulated only in the chronic exposure treatment in both species. Under both acute and chronic heat exposure, more HSPs were up-regulated in *P. diffusa* than in *P. canaliculata*. In addition, low molecular weight HSPs including HSP20.7 and putative alpha-B-crystallin behaved differently in the two species. In *P. canaliculata*, they were only up-regulated in the acute exposure treatment, while in *P. diffusa* they were up-regulated in both acute and chronic exposure treatments. Up-regulation of HSPs due to the acute exposure might indicate that thermal stress had reached a point that constitutively expressed HSPs can no longer repair unfolded proteins, therefore requiring the synthesis of more HSPs. In other words, in response to the same acute or chronic
heat exposure, the invasive species *P. canaliculata* might suffer a lower level of protein structural changes than its non-invasive congener *P. diffusa*. Although it is beneficial for an organism to induce HSPs to cope with stress as an acute response, maintaining high HSP expression level for a long time can be deleterious as HSP synthesis is a highly energy-consuming process (Feder & Hofmann, 1999). For *P. diffusa*, investing a large amount of resources into HSP synthesis would likely result in the down-regulation of many other proteins, therefore interfering with normal cellular functions (Feder & Hofmann, 1999).

### 3.4.2 Oxidative stress related proteins

High temperature accelerates mitochondrial respiration therefore increases the formation of reactive oxygen species (ROS). In response to ROS production, animals can activate the scavenging system which involves the actions of several enzymes such as superoxide dismutase (SOD). In the present study, we detected the up-regulation of manganese superoxide dismutase (MnSOD) in *P. diffusa* under chronic heat exposure, and this might contribute to reducing oxidative damage in this species. The accumulation of ROS can cause oxidative damage to DNA, proteins and lipids. Since many HSPs were up-regulated, we expected to see no damage in the cellular environment of *P. diffusa*. However, cathepsin L and cathepsin L-like cysteine proteinase which help to degrade proteins were down-regulated in *P. diffusa*. Accumulation of un-degraded proteins would cause endoplasmic reticulum (ER) stress (Walter & Ron, 2011). This process was
implicated by the up-regulation of ER protein ERp29 and ARMET protein precursor in *P. diffusa* under chronic exposure. One approach to deal with ER stress is to control transcription and translation which can decrease the influx of nascent proteins into ER (Walter & Ron, 2011). In this regard, in our study, eukaryotic translation initiation factor 4 gamma 2, tyrosyl-tRNA synthetase (cytoplasmic) and putative pre-mRNA-splicing factor ATP-dependent RNA helicase mog-5 were all down-regulated. Interestingly, this regulation pattern was only detected in *P. diffusa* under chronic thermal stress. In addition, omega class glutathione S-transferase (GST) was down-regulated under both acute and chronic heat stress in *P. canaliculata*, but GST omega-1 was up-regulated under chronic heat stress in *P. diffusa*. GST omega class belongs to the GST protein family and is critical in cellular redox homeostasis and detoxification.

As a whole, the differences in expression pattern of these oxidative stress related proteins indicated that *P. diffusa* suffered more from oxidative stress and cellular damage than *P. canaliculata* under the same heat exposure.

### 3.4.3 Energy metabolism related proteins

Consistent with physiological parameters at the whole organism level, the relationship between temperature and metabolic rate and energy usage at the cellular level should also be biphasic (Pörtner, 2002). Before reaching the tolerance limit, there is usually a positive correlation between temperature and metabolic rate (Hochachka & Somero, 2002; Cherkasov *et al.*, 2006). Beyond the
tolerance limit, however, maintenance processes to ensure cell survival become a priority. This phase is usually characterized by metabolic rate depression (MRD), an adaptation that can extend the survival time of an organism. Marshall & McQuaid (1992) reported that some of the effects of heat-induced MRD are similar to those of estivation. Consistent with this prediction, in *P. diffusa*, chitotriosidase-1-like protein, a carbohydrate metabolic enzyme, was down-regulated in the acute exposure treatment. However, in *P. canaliculata*, heparanase isoform 1 (an enzyme that cleaves heparan sulfate proteoglycans into core proteoglycans and heparan sulfate side chains) was up-regulated, and there was a lack of significant changes in lipid metabolism related enzymes under acute heat stress, indicating that the major energy source was carbohydrate. In addition, the remarkable up-regulation of the respiratory pigment haemocyanin indicated a high demand for aerobic metabolism during the acute heat exposure. We therefore believed that exposure to 37 °C for 3 h has only induced MRD only in *P. diffusa*. Under chronic heat stress, in *P. canaliculata*, carbohydrate was still the main energy source. Beta-D-xylosidase 2 and 5 increased their expression while fatty acid-binding-like protein 5 decreased its expression. In addition, cytochrome P450 3A24, an enzyme in the electron transport chain (ETC), was down-regulated. As ETC is the main source of ROS, down-regulation of the enzyme likely indicated a limited amount of oxidative damage. In *P. diffusa*, carbohydrate metabolic enzymes such as cellulase were down-regulated, but several fatty acid transport and metabolic enzymes including fatty acid-binding protein (adipocyte),
acetyl-CoA acetyltransferase B (mitochondria) and acyl-Coenzyme A dehydrogenase (C-4 to C-12 straight chain) were all up-regulated. In particular, in *P. diffusa*, the *de novo* purine synthesis pathway was greatly activated under chronic heat exposure, as indicated by the up-regulation of six out of the ten enzymes in this pathway. As the final product of purine metabolism is uric acid, the up-regulation of purine synthesis might contribute to antioxidation because uric acid is a powerful antioxidant. Since purine synthesis is energy demanding, this process may compromise other critical biological functions (Giraud-Billoud *et al.*, 2011; Sun *et al.*, 2013).

### 3.4.4 Immune responsive proteins

Environmental stressors such as high temperature can affect immune functions (Rolff & Siva-Jothy, 2003). It was reported that a 15-min mechanical disturbance caused significantly down-regulation of immune genes in *Crassostrea gigas* during the stress period, but in the following 0.5-1.5 h, immune responses increased greatly (Lacoste *et al.*, 2002). Currently it is unknown why, under chronic thermal stress, these immune related proteins were up-regulated in *P. canaliculata* but not in *P. diffusa*. But it is possible that the unchanged expression of immune proteins in *P. diffusa* was related to the lack of energy resource allocated to immunity under chronic environmental stress, while there was no obvious energy-shortage in *P. canaliculata* for maintaining immune surveillance and immune protein expression to cope with potential pathogen attack.
3.4.5 Comparison of the orthologous proteins between congeneric: challenges and insights

It is challenging to compare orthologs in non-model species (Chen et al., 2010): 1, the proteome database is usually incomplete, and many of the proteins are unannotated; 2, the copy number of a gene may not be identical between proteome databases, making unambiguous match difficult; 3, when only partial gene sequences are available, the overlapping sequences may be too short to allow for the establishment of orthologous relationship; 4, when a gene is spilt into different ESTs during assembly, it is difficult to find a match in the other species’ proteome database.

Despite these limitations, we were able to identify a large set of orthologs between the two congeneric apple snails, which resulted in the identification of two differentially expressed proteins (DEPs) with clear signal of rapid divergent selection, and seven proteins with Ka/Ks > 0.5 that deserve further analysis of divergent selection. Three of these DEPs are unannotated, which are candidates for functional studies. Of the six annotated DEPs, it is remarkable that they are implicated in energy metabolism, cell oxidative homeostasis, signaling and binding, which have all been identified previously as critical stress responses (Wei et al., 2008; Elmer et al., 2010; Zhao et al., 2014). Since an increase in heat stress will likely trigger stress signaling, disrupt metabolism, cause accumulation of ROS and denatured proteins (Zhang et al., 2012; Meng et al., 2013), it is not
surprising that some of the genes in these functional groups are targets of positive selection. Pathway enrichment analysis has provided additional evidence that these biological processes may play an important role in stress resistance (Table 3.1). For instance, the enriched pathways include ATP synthesis coupled proton transport, cell redox homeostasis, DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest, all have been known to be involved in stress responses (Elmer et al., 2010; Zhao et al., 2014). It should be noted, however, that some of the pathways involved in the heat responses in apple snails may not be stress specific. For example, Zhao et al. (2014), who analyzed the differentially expressed genes (DEGs) after exposing two congeneric oysters to hypoxia, found that several genes involved in metabolic process, oxidase activity and ion binding (i.e. 5'-AMP-activated protein kinase subunit beta-2, cytochrome c oxidase subunit 6A1 (mitochondrial), and tripartite motif-containing protein 2) might have undergone positive selection. Our study therefore has provided evidence for the genetic basis of differential resistance to heat stress in the two congeneric apple snails, which involves not only the regulation of protein expression in conserved metabolic pathways, but also structural changes in specific proteins functioning in energy metabolism, cell oxidative homeostasis, and signaling and binding processes.
Table 3.1. The significantly enriched biological process gene ontology (GO) terms for differentially expressed orthologs between species and treatments.

<table>
<thead>
<tr>
<th>GO ID</th>
<th>Term</th>
<th>Level</th>
<th>No. genes in gene group</th>
<th>No. genes in transcriptome</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0006977</td>
<td>DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest</td>
<td>12</td>
<td>5</td>
<td>25</td>
<td>9.67E-05</td>
</tr>
<tr>
<td>0006098</td>
<td>Pentose-phosphate shunt</td>
<td>12</td>
<td>3</td>
<td>13</td>
<td>1.73E-03</td>
</tr>
<tr>
<td>0015986</td>
<td>ATP synthesis coupled proton transport</td>
<td>12</td>
<td>2</td>
<td>6</td>
<td>5.23E-03</td>
</tr>
<tr>
<td>0031145</td>
<td>Anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process</td>
<td>11</td>
<td>5</td>
<td>31</td>
<td>7.46E-04</td>
</tr>
<tr>
<td>0051436</td>
<td>Negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle</td>
<td>11</td>
<td>5</td>
<td>27</td>
<td>1.42E-04</td>
</tr>
<tr>
<td>0033539</td>
<td>Fatty acid beta-oxidation using acyl-CoA dehydrogenase</td>
<td>11</td>
<td>2</td>
<td>5</td>
<td>3.53E-03</td>
</tr>
<tr>
<td>0042062</td>
<td>Long-term strengthening of neuromuscular junction</td>
<td>11</td>
<td>2</td>
<td>5</td>
<td>3.53E-03</td>
</tr>
<tr>
<td>0000209</td>
<td>Protein polyubiquitination</td>
<td>10</td>
<td>7</td>
<td>101</td>
<td>3.64E-03</td>
</tr>
<tr>
<td>0051443</td>
<td>Positive regulation of ubiquitin-protein ligase activity</td>
<td>10</td>
<td>6</td>
<td>34</td>
<td>4.01E-05</td>
</tr>
<tr>
<td>0015991</td>
<td>ATP hydrolysis coupled proton transport</td>
<td>10</td>
<td>5</td>
<td>12</td>
<td>1.77E-06</td>
</tr>
<tr>
<td>0046033</td>
<td>AMP metabolic process</td>
<td>10</td>
<td>2</td>
<td>8</td>
<td>9.51E-03</td>
</tr>
<tr>
<td>0006749</td>
<td>Glutathione metabolic process</td>
<td>9</td>
<td>4</td>
<td>27</td>
<td>1.64E-03</td>
</tr>
<tr>
<td>0007308</td>
<td>Oocyte construction</td>
<td>9</td>
<td>3</td>
<td>12</td>
<td>3.51E-03</td>
</tr>
<tr>
<td>0030727</td>
<td>Germarium-derived female germ-line cyst formation</td>
<td>9</td>
<td>3</td>
<td>13</td>
<td>1.73E-03</td>
</tr>
<tr>
<td>0006573</td>
<td>Valine metabolic process</td>
<td>9</td>
<td>2</td>
<td>5</td>
<td>3.53E-03</td>
</tr>
<tr>
<td>0000084</td>
<td>S phase of mitotic cell cycle</td>
<td>8</td>
<td>5</td>
<td>67</td>
<td>9.17E-03</td>
</tr>
<tr>
<td>0019388</td>
<td>Galactose catabolic process</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>6.78E-05</td>
</tr>
<tr>
<td>0006105</td>
<td>Succinate metabolic process</td>
<td>8</td>
<td>2</td>
<td>8</td>
<td>9.51E-03</td>
</tr>
<tr>
<td>0019859</td>
<td>Thymine metabolic process</td>
<td>8</td>
<td>2</td>
<td>5</td>
<td>3.53E-03</td>
</tr>
<tr>
<td>0046135</td>
<td>Pyrimidine nucleoside catabolic process</td>
<td>8</td>
<td>2</td>
<td>8</td>
<td>9.51E-03</td>
</tr>
<tr>
<td>0051260</td>
<td>Protein homooligomerization</td>
<td>7</td>
<td>7</td>
<td>115</td>
<td>6.96E-03</td>
</tr>
<tr>
<td>0051262</td>
<td>Protein tetramerization</td>
<td>7</td>
<td>6</td>
<td>49</td>
<td>3.28E-04</td>
</tr>
<tr>
<td>0065002</td>
<td>Intracellular protein transmembrane transport</td>
<td>7</td>
<td>4</td>
<td>26</td>
<td>1.42E-03</td>
</tr>
<tr>
<td>0007097</td>
<td>Nuclear migration</td>
<td>7</td>
<td>3</td>
<td>11</td>
<td>1.03E-03</td>
</tr>
<tr>
<td>0005997</td>
<td>Xylulose metabolic process</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>3.53E-03</td>
</tr>
<tr>
<td>0033962</td>
<td>Cytoplasmic mRNA processing body</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>7.23E-03</td>
</tr>
<tr>
<td>GO:ID</td>
<td>Description</td>
<td>GO Count</td>
<td>-log10(p)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0006412</td>
<td>Translation</td>
<td>6</td>
<td>311</td>
<td>2.29E-10</td>
<td></td>
</tr>
<tr>
<td>0006457</td>
<td>Protein folding</td>
<td>6</td>
<td>96</td>
<td>1.38E-09</td>
<td></td>
</tr>
<tr>
<td>0002479</td>
<td>Antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent</td>
<td>6</td>
<td>17</td>
<td>4.85E-07</td>
<td></td>
</tr>
<tr>
<td>0006521</td>
<td>Regulation of cellular amino acid metabolic process</td>
<td>6</td>
<td>14</td>
<td>4.33E-06</td>
<td></td>
</tr>
<tr>
<td>0033574</td>
<td>Response to testosterone stimulus</td>
<td>6</td>
<td>11</td>
<td>1.35E-03</td>
<td></td>
</tr>
<tr>
<td>0034418</td>
<td>Urate biosynthetic process</td>
<td>6</td>
<td>5</td>
<td>3.53E-03</td>
<td></td>
</tr>
<tr>
<td>0060612</td>
<td>Adipose tissue development</td>
<td>6</td>
<td>6</td>
<td>5.23E-03</td>
<td></td>
</tr>
<tr>
<td>0045454</td>
<td>Cell redox homeostasis</td>
<td>5</td>
<td>22</td>
<td>8.20E-03</td>
<td></td>
</tr>
<tr>
<td>0006662</td>
<td>Glycerol ether metabolic process</td>
<td>5</td>
<td>8</td>
<td>9.44E-03</td>
<td></td>
</tr>
<tr>
<td>006775</td>
<td>Fat-soluble vitamin metabolic process</td>
<td>5</td>
<td>6</td>
<td>5.23E-03</td>
<td></td>
</tr>
<tr>
<td>0045184</td>
<td>Establishment of protein localization</td>
<td>4</td>
<td>520</td>
<td>6.00E-04</td>
<td></td>
</tr>
<tr>
<td>0006730</td>
<td>One-carbon metabolic process</td>
<td>4</td>
<td>16</td>
<td>2.01E-04</td>
<td></td>
</tr>
<tr>
<td>0030721</td>
<td>Spectosome organization</td>
<td>4</td>
<td>6</td>
<td>5.23E-03</td>
<td></td>
</tr>
<tr>
<td>0045478</td>
<td>Fusome organization</td>
<td>4</td>
<td>7</td>
<td>7.23E-03</td>
<td></td>
</tr>
<tr>
<td>0055114</td>
<td>Oxidation-reduction process</td>
<td>3</td>
<td>548</td>
<td>6.91E-07</td>
<td></td>
</tr>
<tr>
<td>0016032</td>
<td>Viral reproduction</td>
<td>3</td>
<td>127</td>
<td>1.18E-03</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.1. A photograph of a harvested vegetable garden in Long Yuen, Hong Kong, showing *P. canaliculata* on the sediment surface and their pink egg clutches on a side wall, with an inset showing an enlarged individual (A); and a photograph of an individual of *P. diffusa* purchased from a local aquarium shop (B). Photograph courtesy of Jian-Wen Qiu, Jin Sun and Huawei Mu. Copyright 2015.
Figure 3.2 Percent mortality of *P. canaliculata* (*n* = 30) and *P. diffusa* (*n* = 35) after exposure to increasing temperatures. The mortality-temperature relationship was modelled using a logistic function.
Figure 3.3 Expression patterns of DEPs under acute and chronic heat exposure (37 °C/25 °C and 35 °C/25 °C) in *P. canaliculata* (Pc). The color scale bar on the left represents the fold change (red = up-regulation, green = down-regulation). On the left are level 2 biological processes GO terms.
Figure 3.4 Expression patterns of DEPs under acute and chronic heat exposure (37 °C/25 °C and 35 °C/25 °C) in P. diffusa (Pd). The color scale bar on the left represents the fold change (red = up-regulation, green = down-regulation. On the left are level 2 biological processes GO terms.
Figure 3.5. Ka/Ks values in different functional protein groups. Orthologs with Ka/Ks > 1 fall above the dotted line while orthologs with Ka/Ks between 0.5 and 1 fall between the dotted and solid lines.
Chapter 4 Molecular mechanisms of tolerance to hypoxia stress in congeneric snails

4.1 Introduction

The rapid globalization in recent decades has greatly facilitated species introduction around the world (Strayer, 2010; Moran et al., 2016). Successfully established introduced species, so-called invasive species, have threatened the invaded ecosystems by reducing biodiversity, altering community structure and compromising ecosystem services (Pimentel et al., 2005; Simberloff et al., 2013). While it is difficult to predict whether a species will become invasive following its introduction, several niche- and trait-based characteristics are known to be associated with establishment success (Hayes & Barry, 2008). Climate suitability, resource availability, presence of potential competitors, community diversity and frequency and scale of disturbances are some of the niche-related characteristics (Davis et al., 2000; Blumenthal, 2006; Bellard et al., 2013); whereas growth rate, life span, fecundity, dispersal ability, dietary spectrum, and tolerance of environmental stressors are some of the trait-based characteristics (Williamson & Fitter, 1996). Indeed, a review by Kolar & Lodge (2001) showed that 86% of the invasive species could be predicted by a subset of life-history traits including fast growth, high fecundity, high dispersal ability, wide food spectrum, phenotypic plasticity and tolerance of a wide range of extreme abiotic conditions.
There has been substantial interest in studying how invasive species respond to extreme environmental conditions because the results can help not only predict their range expansion and manage their impact, but also reveal the adaptive mechanisms underlying their invasiveness (Moran et al., 2016). Comparative studies using closely related species, when combined with transcriptomic and proteomic techniques, have provided scientists with unprecedented opportunities to study the molecular mechanisms regulating animal responses to environmental stressors including hypoxia (Somero, 2011). Such studies have advantages over single-species studies in that they can distinguish effects of adaptive variation from those of phylogeny (Dong & Somero, 2009; Tomanek & Zuzow, 2010; Lockwood & Somero, 2011). High-throughput transcriptomic and proteomic methods are especially suitable for such studies because it is expected that environmental stressors will alter the expression of many genes and proteins involved various critical biological processes such as stabilization of protein structure, repairing damaged DNA, and regulating protein turnover, energy production and redox homeostasis (Kültz, 2005).

In the present study, we aimed to understand the genetic basis of hypoxia tolerance in the congeneric snails *Pomacea canaliculata* and *Pomacea diffusa* (Gastropoda: Ampullariidae). *Pomacea canaliculata*, a native of freshwater wetlands of South America, has invaded East and Southeast Asia, the southern United States, and Hawaii in the 1980s (Hayes et al., 2015). In its invaded
regions, this species has become a pest, causing tremendous loss to rice and taro farming (Cowie, 2002), and affecting wetland biodiversity and function by herbivory (Carlsson et al., 2004; Fang et al., 2010). Tolerance of environmental stressors, especially desiccation, heat, cold, and hypoxia, has been considered as an important trait determining the success of invasion in P. canaliculata (Yusa et al., 2006; Matsukura et al., 2009; Yamanishi et al., 2012; Giraud-Billoud et al., 2011, 2013; Sun et al., 2013; Yoshida et al., 2014; Mu et al., 2015). In southern China where the climate is divided into a warm and wet summer and a cold and dry winter (Kwong et al., 2008), this species is often found in shallow-water bodies that experience high temperature in summer and hypoxia in winter. Our field surveys found that P. canaliculata is often present in stagnant shallow water bodies, especially ponds and abandoned farmlands with dissolved oxygen $< 1 \text{ mg L}^{-1}$ (Kwong et al., 2008). It is apparently even more tolerant to hypoxia than the Nile tilapia Oreochromis niloticus, a hypoxia tolerant fish that has invaded southern China as well. However, Pomacea diffusa, another species of native snail in the freshwater wetlands of South America, is not considered an invasive species. This species is a common ornamental and tank cleaning species in the aquarium trade around the world, and the only species in the family Ampullariidae allowed to be transported across states without a permit issued by the United States Department of Agriculture (Rawlings et al., 2007). There have been scattered reports of P. diffusa in Asia (Hayes & Barry, 2008), New Zealand (Collier et al., 2011), and southern United States (Rawlings et al.,
2007), but this species has not been found to build up large populations or become an agricultural pest. Although the two species of *Pomacea* are both of South American origin, the native range of *P. canaliculata* is a small area of southern South America from northeastern Argentina to southern Uruguay (Hayes *et al.*, 2015), whereas that of *P. diffusa* is central Bolivia to the western Amazon basin of Brazil (Kenneth Hayes, personal communication). Therefore, *P. canaliculata* evolved in an area with greater climate extremes than *P. diffusa* and may have evolved to have higher hypoxia tolerance.

We conducted two experiments using these emerging models in ecological genomics. The first experiment aimed to determine the mortality of the two *Pomacea* species exposed to different levels of hypoxia. The second experiment aimed to identify protein biomarkers that are responsive to hypoxia stress in both species and those that are differentially expressed between the two species. To understand whether adaptive evolution has contributed to the differential protein expression (Gallego-Romero *et al.*, 2012), we analyzed the sequences of genes encoding the differentially expressed proteins. Since the statistical power of such base substitution analysis is weak when only two species are compared (Clark & Swanson, 2005), we adopted a phylogenetic comparative approach (Whitehead & Crawford, 2006) to determine genes that have contributed to the adaptive divergence in hypoxia tolerance in apple snails by including the transcriptome data from two additional ampullariid species in the analysis.
4.2 Material and methods

4.2.1 Snail maintenance and hypoxia exposure system

Adults of *Pomacea canaliculata* were collected from a drainage channel in Yuen Long (22°15’N, 114°10’E), a vegetable farming area, and adults of *Pomacea diffusa* were purchased from an aquarium shop in Hong Kong. The two species were separately reared in 250-L aquaria with a submerged glass heater to maintain the water temperature at 25 ± 1 °C. The aquaria were each supplied with a submerged pump and canister filter to circulate water, and remove and food waste and snail feces. Oxygen in the water was supplemented by continuous aeration with an air pump. Snails were fed with carrot, lettuce and fish feed and leftovers were removed daily. The culture room was illuminated by fluorescent light with a photoperiod of 14 h light: 10 h dark to facilitate the development of biofilm. Field-collected snails were used in experiments after acclimation in the laboratory for at least one month.

During the experiments, both species of snails were kept in aquaria filled with 3.5 L water. The normoxia control was established by pumping compressed air into the water continuously to maintain the ambient dissolved oxygen (DO) level of 6.7 (range 6.6 to 6.8) mg O₂ L⁻¹. Different hypoxia treatments were created by mixing compressed nitrogen gas and air before introducing the mixed gas into the exposure water (Cheung *et al.*, 2008). The discharge rates of the two gases were adjusted to reach the following dissolved oxygen (DO) levels: 4.1 mg O₂ L⁻¹ (range 3.9 to 4.3 mg O₂ L⁻¹), 2.0 mg O₂ L⁻¹ (range 1.8 to 2.2 mg O₂
L\(^{-1}\)) and 1.0 mg O\(_2\) L\(^{-1}\) (range 0.8 to 1.2 mg O\(_2\) L\(^{-1}\)). The DO levels in the experimental chambers were monitored continuously using a Stable Optical Oxygen System (TauTheta Instruments LLC, Colorado, USA) coupled with an optical sensor inserted into the exposure water. Since apple snails can ventilate their lung by crawling on aquarium wall near water surface and extending their respiratory siphon into air (Seuffert & Martín, 2010), a plastic grid (pore size 1.5 cm) was fixed horizontally 10 cm below the water surface, a distance longer than their respiratory siphon, to make sure the snails were truly under the different oxygen treatments. The experiments were conducted at 25 ± 1°C.

4.2.2 Experiment 1. Differential mortality under hypoxia

Cumulative mortality rate was determined for both species following Cheung et al. (2008). Twelve snails of both species (2.3 to 2.5 cm shell length) were starved for one day before use in experiment. They were then transferred into experimental aquaria and exposed to the different normoxia and hypoxia treatments for 72 hours, during which the snails were checked for their survival every a few hours. Dead snails, defined as those that did not move and had no response to physical stimuli, were removed promptly to avoid fouling due to tissue decomposition. These presumably dead snails were confirmed by transferring them to normoxia condition; none of them recovered. The cumulative mortality in each treatment was plotted against time to show the treatment and time effects.
4.2.3 Experiment 2. Responses of the proteome to acute hypoxia exposure

4.2.3.1 Exposure to hypoxia, protein extraction and labelling

Snails of *P. canaliculata* (shell length 25.9 to 32.8 mm, *n* = 9, three replicates) and *P. diffusa* (shell length 28.6–37.6 mm, *n* = 9, three replicates) were exposed to the normoxia and hypoxia treatments for 8 hours, after which they were dissected, and their gills were placed in 8 M urea and stored in -80°C until use. During protein extraction, gill samples were placed on ice and homogenized using a plastic pestle. A mild sonication was applied to break cell membranes. Sample solutions were centrifuged at 15,000 g for 15 min under 4°C. Supernatant containing proteins was collected and purified with a 2D-cleanup kit (Bio-Rad, Hercules, California, USA). The purified protein pellets in each sample were reconstituted with 8 M urea and quantified with a RC-DC kit (Bio-Rad). Afterwards, an aliquot containing 200 μg protein was used for Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) labelling. Proteins were firstly reduced using 5 mM triscarboxyethyl phosphine hydrochloride at 60°C for 1 hour and then alkylated using 10 mM methylethanethiosulfonate at room temperature for 20 min. Protein solutions were diluted 8-fold using 50 mM triethylammonium bicarbonate, then digested with sequencing-grade trypsin (Promega, Madison, WI) with an enzyme-to-protein ratio of 1:50 (w/w) for 16 h at 37°C. After digestion, samples were desalted with Sep-Pak C18 cartridges (Waters, Milford, MA) and dried in an Eppendorf vacuum concentrator. Peptides derived from the 6.7, 4.1, 2.0 and 1.0 mg O₂ L⁻¹ treatments were labeled with iTRAQ reagents 114, 115, 116.
and 117 (AB Sciex, Framingham, MA), respectively. Labeled peptides which were from one biological replicate were pooled together and dried in an Eppendorf concentrator.

4.2.3.2 Fractionation and LC-MS/MS analysis

Buffer A (10 mM KH$_2$PO$_4$ and 20% acetonitrile (ACN), pH 3.0) was firstly applied to reconstitute dried samples. Sample solutions were then fractionated with a PolySULFOETHYL strong cation-exchange column (200 x 4.6 mm, 5 μm particle size, 200-Å pore size, PolyLC, Columbia, Maryland) on a Waters 2695 high performance liquid chromatography. A 50-min fractionation gradient was used to fractionate samples with a speed of 1 ml/min: 100% Buffer A for 5 min, 0-30% Buffer B (10 mM KH$_2$PO$_4$, 0.5 M KCl and 20% ACN, pH 3.0) with a linear changing gradient for the following 28 min, 30-100% (linear changing gradient) Buffer B for 5 min, 100% Buffer B for 5 min, and 100% Buffer A for the final 7 min. Fractions were collected, dried and desalted.

Each fraction was reconstituted with 0.1% formic acid and analyzed twice with LTQ-Orbitrap Elite coupled with an Easy-nLC 1000 (Thermo Fisher, Bremen, Germany). Peptides were then separated with a C18 capillary column (50 μm × 15 cm, packed with Acclaim PepMap RSLC C18, 2 μm, 100 Å, nanoViper, Thermo scientific) by using a 90 min gradient: 100% Solution A (0.1% formic acid in Milli-Q water) for the initial 5 min, followed by a 55 min linear changing gradient of 0-30% Solution B (0.1% formic acid in ACN), 30-98%
Solution B (linear changing gradient) for 10 min, and maintained at 98% Solution B for 10 min, then re-equilibrated at 100% Solution A for the final 10 min. Mass spectrometry scans (350-1,600 \( m/z \) ranges) were acquired in the positive mode with a resolution of 60,000. The top five most abundant multiple-charged ions with a combined abundance threshold > 500.0 were selected for fragmentation using both high-energy collision-induced dissociation (HCD) and collision-induced dissociation (CID) methods. HCD spectra were used for iTRAQ quantification in C-trap while CID spectra were used for identifying peptides in LTQ. In order to remove duplicate precursors, the isolation width was set to 2.0 (\( m/z \)) for both HCD and CID scannings which had activated dynamic exclusion. The parameters of HCD fragmentation were 10 ms of activation time, 45% of normalized collision energy, and a full scan with FTMS at a resolution of 15,000 in a centroid mode. The parameters of CID fragmentation were 10 ms of activation time, activation Q of 0.25, and 35% of normalized collision energy.

4.2.3.3 Protein identification

Raw mass spectrometry data were converted into .mgf files using Thermo Proteome Discoverer 1.3.0.339 (Thermo Finnigan, California). Each file was filtered by a python script to remove unpaired scans and separated into the HCD and CID data (Mu et al., 2015). For HCD and CID data which belong to the same precursor, reporter groups (i.e. 114-117.5 Da) in CID scans were replaced
by those of normalized HCD scans. Both HCD data file and modified CID files were separately searched against a *P. canaliculata* (Sun *et al.*, 2012) and a *P. diffusa* database using Mascot version 2.3.2 (Matrix Sciences, London, UK). The searching parameters were set as no more than two missed cleavages, ± 5 ppm for precursor ions, ± 0.6 Da for fragments, fixed modification for methylthio of cysteine, and variable modification for oxidation (methionine) and deamidation (asparagine and glutamine). Searching criteria for HCD files were similar to those of CID files with an exception of 20 mmu fragment ions tolerance. Data with ion scores less than 28 were removed. The remaining data were exported as a .csv file. Peptides which contained < 7 amino acids were further removed because short peptides might easily have match in decoy database. In the database search, false discovery rate was set as 1%. Peptides which matched more than one protein were deleted because these peptides might have matched to conserved regions in protein groups and caused bias in quantitation. In addition, peptides which were unlabeled, matched to decoy database and had erratic ratio between iTRAQ reporters (e.g., 115/114 < 0 or > 100) were removed. Median normalization was performed in each replicate. Proteins detected in all three biological replicates and contained at least 4 summed peptides were used for quantification based on the summed intensity of matched peptides.

**4.2.3.4 Determination of differentially expressed proteins**
Orthologs in the two species were identified by searching the translated transcriptome of the two species using the reciprocal best hit method (RBH) with BLASTp (NCBI-blast 2.2.26) using an E-value threshold of $1.0 \times 10^{-5}$ (Sun et al., 2012; Mu et al., 2015). For each pair of ortholog, a two-way analysis of variance (ANOVA) was performed to determine the hypoxia and species treatment effects. Because some of the proteins without an ortholog in the other species were also likely responsive to hypoxia, a separate one-way ANOVA was conducted for each species, followed by the Tukey tests to determine the hypoxia treatment effects. For both two-way and one-way ANOVA, we chose a $P$ value of 0.02 to control false positives (Tomanek et al., 2010; Fields et al., 2012; Mu et al., 2015). Analyses were performed using an R script (R Development Core Team, 2009; Lockwood et al., 2010 & 2011).

4.2.3.5 Functional classification

The differentially expressed proteins (DEPs) were annotated based on their Gene Ontology (GO). With further manual correction, the DEPs were assigned into several Biological Process GO terms. Proteins without GO annotation or known functions were grouped as ‘Others’. The expressional patterns of DEPs were compared using Spearman Rank correlation similarity matrix and the centroid linkage method, and finally visualized using Java Treeview (Sun et al., 2013).
4.2.3.6 Evolutionary analysis of orthologous protein-coding genes

We adopted a phylogenetic comparative approach (Whitehead & Crawford, 2006) to determine the genes that have contributed to the adaptive divergence in hypoxia tolerance in the apple snails by including the transcriptome data from two additional species in the same family (P. maculata and Marisa cornuarietis) in the analysis. Using the RBH method, the DEOs between P. canaliculata and P. diffusa were searched against the translated transcriptomes of P. maculata and M. cornuarietis, which were sequenced by Illumina Hiseq 2000 and assembled by Trinity. The orthologs between P. canaliculata and the other species were used for the following multiple-species ω calculation.

Orthologous sequences among the four species were aligned using ParaAT 1.0 (Zhang et al., 2012). Gaps in the aligned codons were removed. A likelihood ratio test (LRT) was then applied to compare different ω ratios under different pairs of models using codeml tool implemented in PAMLx (Xu & Yang, 2013). Briefly, one neutral model hypothesized no codons with an ω value >1 (ℓ0), and the other alternative model contained a subset of sites which could have an ω value >1 (ℓ1). Then, the number of twice the difference of log-likelihood values between alternative and null models (2Δℓ) was calculated and compared with the chi-square distribution with a degree of freedom of 2 (number estimated from the difference of parameters between compared models). Two pairs of site models were used. The first pair included a null model M1a (two classes of sites with rates: 0 < ω0 < 1 and ω1 = 1) and a positive selection model M2a (three
classes of sites with rates: $0 < \omega_0 < 1$, $\omega_1 = 1$, and $\omega_2 > 1$). The second pair compared the M7 model which expects a beta distribution for $\omega$ ($0 < \omega < 1$) and the M8 model which has one more class of sites with $\omega > 1$. In both the M2a and M8 models, the Bays Empirical Bayes (BEB) approach which accounted for sampling error was applied to calculate the posterior probability of each amino acid site belonging to the site class of diversifying selection.

4.3 Results

4.3.1 Differential tolerance to hypoxia

The 72 h experiment showed that *P. canaliculata* is much more tolerant to hypoxia than *P. diffusa* (Fig. 4.1). For *P. canaliculata*, all individuals in the 6.7, 4.1 and 2.0 mg O$_2$ L$^{-1}$ treatments survived through the experiment period. In the 1.0 mg O$_2$ L$^{-1}$ treatment, dead snails (41.6%) were first observed at 46 h, and all snails were dead by 54 h. For *P. diffusa*, all individuals exposed to 6.7 mg O$_2$ L$^{-1}$ survived through the exposure period. At 4.1 mg O$_2$ L$^{-1}$, dead snails (8.3%) were first observed at 18 h; by the end of the exposure period the mortality only reached 41.7%. At 2.0 and 1.0 mg O$_2$ L$^{-1}$, dead snails were first found at 18 h; by 65 h all individuals in both treatments were found dead.

4.3.2 Proteome analysis

In *P. canaliculata*, a total of 2759 proteins were identified, including 2105, 2049, and 2060 in three biological replicates, respectively. There were 1430 shared
proteins among the three replicates. In *P. diffusa*, a total of 2550 proteins were identified, including 1951, 1989, and 1973 from the three biological replicates, respectively. There were 1378 shared proteins among the three replicates.

Reciprocal best hit identified 13308 orthologous proteins between the two species. A two-way ANOVA of these orthologs revealed a total of 82 proteins with significant expression changes due to hypoxia (17 proteins), species (69 proteins) or their interaction (6 proteins) (Fig. 4.2). The DEPs were classified into 13 GO terms: binding, amino acid metabolism, cell fate, cytoskeletal organization, energy metabolism, localization, oxidation-reduction, protein folding, proteolysis, transcriptional and translational processes, protein modification, signal transduction, immune response and others. The following discussion of two-way ANOVA results focused on the species effects, because the oxygen effects will be discussed in one-way ANOVA results. The 69 proteins that showed species-specific responses could be roughly classified into two categories and within each category, the two species showed overall opposite expression trends. These proteins belonged to several functional groups: energy metabolism (10 proteins), oxidation-reduction (9 proteins), proteolysis (9 proteins), transcription and translation (8 proteins), protein modification (5 proteins), cytoskeletal organization (5 proteins), signal transduction (4 proteins), cell fate (4 proteins), protein folding (3 proteins), localization (3 proteins), amino acid metabolism (2 proteins) and others (7 proteins). There were six proteins showing interactions between species and hypoxia treatments. They
belonged to signal transduction, energy metabolism, protein folding, transcription and translation, protein modification and an undefined functional group, with each group having one protein.

One-way ANOVA conducted in each species revealed hypoxia level-dependent expression patterns. In *P. canaliculata*, when compared with the 6.7 O₂ L⁻¹ control, there were 15, 32 and 15 DEPs in response to the 4.1, 2 and 1 mg O₂ L⁻¹ treatment, respectively (Fig. 4.3a). These DEPs were classified into 13 functional groups (Fig. 4.4), which showed substantial treatment dependent differences. Specifically, DEPs in the 2 mg O₂ L⁻¹ treatment had representatives in all the functional groups. Although the number of DEPs in the 4.1 and 1 mg O₂ L⁻¹ treatments was the same, the functional compositions were quite different. For instance, oxidation-reduction (e.g., uncharacterized protein CAPTEDRAFT_224067), proteolysis (e.g., cystatin-A1-like protein) and immune responsive proteins (e.g., molluscan defense molecule precursor) were present at 4.1 mg O₂ L⁻¹ but not at 1 mg O₂ L⁻¹; while several signal transduction proteins (e.g., guanine nucleotide-binding protein subunit beta and calmodulin) were differentially expressed at 1 mg O₂ L⁻¹ but not at 4.1 mg O₂ L⁻¹.

In *P. diffusa*, there were 14, 23 and 11 DEPs that were responsive to the 4.1, 2 and 1 mg O₂ L⁻¹ treatment, respectively (Fig. 4.3b). Similar to *P. canaliculata*, DEPs from the 2 O₂ L⁻¹ treatment in *P. diffusa* also included all the functional categories. Although the slightly hypoxic treatment (4.1 mg O₂ L⁻¹) and the most hypoxic treatment (1 mg O₂ L⁻¹) induced similar numbers of DEPs, their
functional composition was quite different. For example, protein folding (e.g., chaperonin containing tcp1) was only differentially expressed in the 1 O₂ L⁻¹ treatment, whereas oxidation-reduction, cytoskeletal organization and protein modification were only differentially expressed in the 4.1 O₂ L⁻¹ treatment.

4.3.3 Estimation of Ka/Ks
Among the 82 DEOs in *P. canaliculata* and *P. diffusa*, only 71 had orthologs in all four species. Comparison of M1a vs. M2a and M7 vs. M8 indicated that there was significant difference between the two pairs of models (ω > 1, P < 0.05) in two genes (Table 4.1). One is an uncharacterized protein (CAPTEDRAFT_173803). The other is NADH dehydrogenase [ubiquinone] flavoprotein 2 (mitochondrial), which is located in the electron transport chain and involved in cellular oxidation-reduction hemostasis. Analysis of the sequences using the BEB method revealed several positively selected amino acid sites in these two genes, with a posterior probability ≥ 95% or ≥ 99% (Table 4.1). In addition, although the ω value in cytochrome c oxidase subunit VIb was less than 1, applying the BEB method revealed a positively selected amino acid site (threonine, the 35ᵗʰ amino acid, posterior probability ≥ 95%) in this gene which plays a key role in the respiratory electron transport chain of mitochondrion.

4.4 Discussion
Since oxygen is critical for energy production in metazoans, understanding how hypoxia affects energy metabolism and cellular homoeostasis has been a focus of research over the last decades (Hochachka et al., 1996; Childress & Seibel, 1998). It is well known that, in response to hypoxia, animals may decrease metabolism in the oxidative phosphorylation pathway, thus reducing their energy production through lowering protein synthesis and ion pump activities, and/or activating the anaerobic pathway (Hochachka & Lutz, 2001; Hochachka & Somero, 2002). Several groups of evolutionarily conserved genes, particularly those regulating DNA and protein structural stability, cell cycle and oxidation-reduction, are also involved in hypoxia responses (Gracey et al., 2001; Wu, 2002; Storey et al., 2004; Kültz, 2005; Tomanek, 2011).

Comparing the responses of the two species of *Pomacea* to hypoxia treatments in the present study has allowed us to detect the differential resistance to hypoxia in the two congers, reveal the complex protein expressional networks that are coordinated to cope with the stress, and gain insight into some of the genetic basis (e.g., DNA sequence divergence) of differential protein expression, providing one of the few comparative studies using congeners to distinguish effects of adaptive variation from those of phylogeny. A striking result of the two-way ANOVA was that there were only a few DEPs showing identical expressional trends, and many of the DEPs exhibited opposite responses to hypoxic exposure in the two species. The differentially expressed
orthologs (DEOs) can be classified into several functional groups, whose expression patterns will be discussed below.

4.4.1 Hypoxia sensing and signal transduction

In order to actively avoid hypoxia, many aquatic species can sense the oxygen level quickly, but their oxygen sensors and downstream signal transduction pathways have not been identified (Wu, 2002). Among the candidate oxygen sensors are respiration chain proteins mitochondrial cytochrome a3 and cytochrome c oxidase, whose upregulation can enhance reactive oxygen species (ROS) production, which stabilizes hypoxia induction factors (Chandel et al., 1997; Castello et al., 2006). Support for this mitochondrial model of hypoxia signaling pathway in the present study comes from the discovery of a cytochrome c oxidase subunit VIb, which was down-regulated in P. canaliculata and up-regulated in P. diffusa, and a variety of signaling and transduction related proteins among the DEPs in the two Pomacea species. Among these are a guanine nucleotide-binding protein, which has been reported to act as a molecular switch regulating a variety of cellular processes such as sensual perception, transporting process, and differentiation (Vetter & Wittinghofer, 2001); a RAS-like protein, which can induce the expression of a hypoxia-inducible factor gene under hypoxic exposure (Sheta et al., 2001); as well as several proteins (i.e., calcium-dependent protein kinase isoform 2, phosphoglycerate kinase 1 and tyrosine-protein kinase HTK1) known to
participate in post-translational modifications such as phosphorylation and
demethylation, implying epigenetic control of gene expression, protein synthesis
and intracellular signaling (Engholm-Keller & Larsen, 2013).

4.4.2 Protein synthesis arrest

Since protein synthesis is one of main cellular energy-consuming processes
under standard metabolic rate (Rolfe & Brown, 1997), reducing the energy
demand of protein synthesis to a minimal metabolic rate is a critical survival
strategy, which has been reported in zebrafish and hypoxia-tolerant goby fish
under hypoxic conditions (Gracey et al., 2000; Ton et al., 2003; van der Meer et
al., 2005). In the present study, a variety of translational and transcriptional
chaperones participating protein biosynthesis (e.g., elongation factor 1 alpha,
40S ribosomal protein S19 and ribosomal protein S5) showed significant
changes in both species. For example, putative 60S ribosomal protein L3 and
histone-binding protein Caf1 were down-regulated in P. canaliculata, but
up-regulated in P. diffusa under three hypoxic treatments. When energy supply
is limited, reallocation of cellular energy between essential and non-essential
ATP-demanding processes becomes important, and energy spared by
transcriptional and translational arrest can be reallocated to other critical
biological processes such as ionic homeostasis (Boutilier & St-Pierre, 2000).
The different expression patterns of translational factors between the two species
in our study indicates that P. canaliculata can quickly respond to acute hypoxia
stress by activating translational arrest, but \( P. \textit{diffusa} \) lacks this ability, which could have implications for energy reallocation and protein synthesis.

### 4.4.3 Energy metabolism

In the present study, hypoxia exposure has led to the differential expression of many enzymes in several energy metabolic and amino acid metabolic processes between the two species (Fig. 4.5). For example, malate dehydrogenase precursor and aconitate hydratase (mitochondrial) which are involved in the tricarboxylic acid (TCA) cycle were down-regulated in \( P. \textit{canaliculata} \) and up-regulated in \( P. \textit{diffusa} \). Fructose-biphosphatase aldolase which is an enzyme catalyzing fructose 1,6-bisphosphate into glyceraldehyde 3-phosphate reversely in glycolysis was down-regulated in \( P. \textit{canaliculata} \) but up-regulated in \( P. \textit{diffusa} \) in the three hypoxic treatments. In the glycolytic pathway, glucose is catalyzed into pyruvate, and pyruvate is catalyzed into lactate by lactate dehydrogenase under anoxic condition, or participate into the TCA cycle under normal condition. In \( P. \textit{canaliculata} \), both fructose-biphosphatase aldolase and lactate dehydrogenase, showed significantly down-regulation which may suggest the suppression of anaerobic metabolism in this hypoxia tolerant species, while there was no obvious evidence for activation or suppression of anaerobic metabolism in \( P. \textit{diffusa} \) under the 8-h exposure. Even though anaerobic metabolism could generate ATP in a quite faster speed than aerobic metabolism, ATP production by anaerobiosis could only sustain a short period of time
because prolonged anaerobiosis may rapidly deplete fermentable glycogen store as it is a low efficient energy-generating pathway and leads to excessive accumulation of deleterious end products such as protons, ethanol and lactate (Boutilier & St-Pierre, 2000; Bickler & Buck, 2007). The suppression of anaerobic respiration in *P. canaliculata* might indicate its ability to protect against altered pH and ionic balances.

### 4.4.4 Effects of different oxygen concentrations

One-way ANOVA analysis revealed only three overlapped differentially expressed proteins (DEPs) between different oxygen levels in each of the two species. In *P. canaliculata*, they were fructose-biphosphat aldolase (partial) which was involved in glycolytic process, lethal (2) giant larvae-like protein 1 (cell fate related protein) and a hypothetical protein with transferase activity. In *P. diffusa*, they were G-type lysozyme functioning in immunity, integrin alpha-6 involved in signal transduction, and a protein similar to ubiquitin-conjugating enzyme E2I which participated in binding activity. The shared proteins among the three hypoxic conditions suggested the involvement of several conserved functional groups (i.e., energy metabolism, immune response) in response to hypoxia in the two species.

In our study, a condition of 4.1 mg O₂ L⁻¹ was designed to mimic a slightly low concentration of oxygen in the natural environment. Nevertheless, proteins involved several cellular processes changed their expression levels in both
species. Specifically, DEPs in both species included oxidation-reduction related proteins, indicating the need to reduce the damage by oxidative stress in both species. Although the two *Pomacea* species showed some common trends in responding to the slightly hypoxic condition, they differed in some aspects. For example, signal transduction (e.g., integrin alpha-6) was induced at 4.1 mg O$_2$ L$^{-1}$ in *P. diffusa*, while in *P. canaliculata*, it (e.g., arrestin) was induced at a more hypoxic condition (2 mg O$_2$ L$^{-1}$), which may suggest a relatively less cellular disturbance in *P. canaliculata* under 4.1 mg O$_2$ L$^{-1}$ since there was no need for it to transfer extracellular stress signals to other cellular parts.

A value of 2.0 mg O$_2$ L$^{-1}$ was used as hypoxic threshold in several studies (Vaquer-Sunyer & Duarte, 2008). In the present study, we found that this hypoxia treatment induced the highest number of DEPs which included all the functional categories. Despite the similar functional grouping, the DEPs of the two species showed some different expressional trends. For example, several proteins functioning in protein modification (e.g., tyrosine-protein kinase HTK16 and CAPTEDRAFT_173271) were differentially expressed in *P. diffusa* but not in *P. canaliculata*. Since the protein modification is involved in signaling and downstream gene expression, the significant changes in *P. diffusa* suggested a disturbed cellular homeostasis which needed further repair and recovery. In addition, metabolic depression was more obvious in *P. canaliculata* under 2 mg O$_2$ L$^{-1}$ because most of the enzymes decreased their expression levels.
A value of 1.0 mg O$_2$ L$^{-1}$ represents extreme hypoxia condition (Diaz, 2001; Altieri & Gedan, 2015; Chu & Tunnicliffe, 2015; Jenny et al., 2016). Under the 1 mg O$_2$ L$^{-1}$ treatment, five to six protein groups were differentially expressed, and the DEPs differed in function or expression trend between the two species. For instance, cytoskeletal organization related proteins were down-regulated in *P. canaliculata*, but not in *P. diffusa*. Since proteins involved in cytoskeletal structure and locomotion are highly abundant, reducing their expression in *P. canaliculata* might be an important energy-saving strategy (Ton et al., 2003). In *P. canaliculata*, there was also translation arrest, indicated by the down-regulation of eukaryotic translation initiation factor 3 subunit K. Protein synthesis is a major energy-consuming process, therefore, translation arrest could reduce protein synthesis and help *P. canaliculata* conserve limited energy under severe hypoxia.

4.4.5 Evolutionary analysis of orthologs

As most amino acids in a functional protein are under functional and structural constraints, the Ka/Ks ratio, calculated for the whole protein coding gene sequence, is rarely larger than 1 (Yang & Nielsen, 2002). By adopting codon-based site models which allow the Ka/Ks ratio to vary along codons in the protein (Yang & Nielsen, 2002; Xu & Yang, 2013; Brisolara-Corrêa et al., 2014), we found that three proteins containing positively selected amino acid sites after analyzing 71 orthologous DEOs from four ampullariid species. One of
them is an uncharacterized protein (CAPTEDRAFT_173803), which should be a good candidate for functional analysis to determine its exact hypoxia response. The other two are both involved in the electron transport chain (ETC): NADH dehydrogenase [ubiquinone] flavoprotein 2 (mitochondrial, partial) is part of the ETC complex I and is involved in transferring electrons from NADH to ubiquinone; and cytochrome c oxidase subunit VIb is a subunit of ETC complex IV, and is involved in the transfer of electrons from reduced cytochrome c to oxygen. These results thus indicate the mitochondrial enzymes involved in energy production are targets of positive selection in apple snails that can have implications in their divergence in stress resistance, which is consistent with previous reports of several genes of cytochrome c oxidase subunits and NADH dehydrogenase having undergone positive selection in other animals adapted to hypoxia environments (Luo et al., 2008; Scott et al., 2011).

The combined physiological, proteomic and transcriptomic analyses employed in this study has provided us with a comprehensive understanding about the genetic basis of differential hypoxia tolerance and protein expression profiles between the two congeneric snails. Especially for the proteomic data, we suggested a thorough pathway map (Fig. 4.5) which included all the differentially expressed proteins and provided us with an insight into the underlying molecular mechanism of hypoxia. For example, TCA cycle and glycolysis were greatly affected. Our streamlined comparative approach can be
adopted to understand mechanisms of species invasion in a changing climate for other non-model organisms.
Figure 4.1. Mortality rate of *P. canaliculata* (a) and *P. diffusa* (b) under control (6.7 mg O$_2$ L$^{-1}$) and three hypoxia conditions (4.1, 2.0 and 1.0 mg O$_2$ L$^{-1}$) in a 72-h experiment ($n = 12$ for both species).
Figure 4.2. Expression patterns of 82 orthologous proteins that had significant oxygen concentration, species or their interaction effects (ANOVA, P < 0.02). Values showing significance are highlighted in red. Pc = P. canaliculata, Pd = Pomacea diffusa. 4.1, 2.0 and 1.0 represents the hypoxia treatment of 4.1, 2.0 and 1.0 mg O$_2$ L$^{-1}$, respectively. Color scale bar on the left represents the fold change (red = up-regulation, green = down-regulation).
<table>
<thead>
<tr>
<th>Annotation</th>
<th>Functional classification</th>
<th>P value of Tukey's test</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRAF</td>
<td>draf_2068088</td>
<td>Energy metabolism</td>
</tr>
<tr>
<td>LeuH2</td>
<td>plastid transit peptide 1</td>
<td>Cell fate</td>
</tr>
<tr>
<td>CAPTEDRAFT</td>
<td>_175623</td>
<td>Guanine nucleotide-binding protein subunit beta</td>
</tr>
<tr>
<td>Hypothetical protein similar to Piddin 2</td>
<td>Signal transduction</td>
<td>0.010 0.000 0.048</td>
</tr>
<tr>
<td>Multifunctional defence molecule precursor</td>
<td>Localization</td>
<td>0.062 0.021 0.837</td>
</tr>
<tr>
<td>Chaperonin-containing T-complex polypeptide subunit 2</td>
<td>Immune response</td>
<td>0.005 0.002 0.507</td>
</tr>
<tr>
<td>Unknown</td>
<td>Other</td>
<td>Protein folding</td>
</tr>
<tr>
<td>Unknown</td>
<td>Other</td>
<td>Others</td>
</tr>
<tr>
<td>Alpha helical protein</td>
<td>Cytoplasmic</td>
<td>Energy metabolism</td>
</tr>
<tr>
<td>Transphosphatase isoform</td>
<td>Relia</td>
<td>Transcriptional and translational process</td>
</tr>
<tr>
<td>VHS-DNA synthesis</td>
<td>Translation and translational process</td>
<td>1.000 0.021 0.006</td>
</tr>
<tr>
<td>CAPTEDRAFT</td>
<td>_154349</td>
<td>Other</td>
</tr>
<tr>
<td>Ehxnybtyc translation initiation factor 3 subunit K</td>
<td>Transcriptional and translational process</td>
<td>0.012 0.062 0.005</td>
</tr>
<tr>
<td>Translation elongation factor 4A, isoform</td>
<td>Cell fate</td>
<td>0.081 0.025 0.009</td>
</tr>
<tr>
<td>Putative core histone-protein kinasa C05D10.2, partial</td>
<td>Cytoskeletal organization</td>
<td>0.077 0.022 0.007</td>
</tr>
<tr>
<td>CAPTEDRAFT</td>
<td>_221705</td>
<td>Other</td>
</tr>
<tr>
<td>Sev-related protein p-galactoside trans</td>
<td>Binding</td>
<td>0.735 0.010 0.047</td>
</tr>
<tr>
<td>Fructose-lactose-phosphate aldolase, partial</td>
<td>Energy metabolism</td>
<td>0.000 0.000 0.004</td>
</tr>
<tr>
<td>Flavin-g-reductase</td>
<td>Cytoskeletal organization</td>
<td>0.016 0.001 0.009</td>
</tr>
<tr>
<td>Diaperin binding inhibitor-like</td>
<td>Energy metabolism</td>
<td>0.020 0.007 0.124</td>
</tr>
<tr>
<td>Cytochrome A1</td>
<td>Protein synthesis</td>
<td>0.006 0.016 0.135</td>
</tr>
<tr>
<td>PTS domain-containing engulfment adapter protein 1</td>
<td>Other</td>
<td>0.041 0.017 0.705</td>
</tr>
<tr>
<td>BRAF</td>
<td>draf_211653</td>
<td>Other</td>
</tr>
<tr>
<td>Cytochrome</td>
<td>Immune response</td>
<td>0.090 0.022 0.523</td>
</tr>
<tr>
<td>Malate dehydrogenase precursor</td>
<td>Energy metabolism</td>
<td>0.180 0.004 0.309</td>
</tr>
<tr>
<td>CAPTEDRAFT</td>
<td>_225833</td>
<td>Other</td>
</tr>
<tr>
<td>Arrin</td>
<td>Signaling transduction</td>
<td>0.040 0.018 0.149</td>
</tr>
<tr>
<td>Flavin-g-reductase</td>
<td>Cytoskeletal organization</td>
<td>0.285 0.013 0.115</td>
</tr>
<tr>
<td>Myosin-2 essential light chain</td>
<td>Cytoskeletal organization</td>
<td>0.000 0.026 0.104</td>
</tr>
<tr>
<td>CAPTEDRAFT</td>
<td>_173933</td>
<td>Other</td>
</tr>
<tr>
<td>CAPTEDRAFT</td>
<td>_224657</td>
<td>Oxidation-reduction process</td>
</tr>
<tr>
<td>Uncharacterized protein K03811.5</td>
<td>Cytoskeletal organization</td>
<td>0.035 0.010 0.158</td>
</tr>
<tr>
<td>Predicted protein</td>
<td>Other</td>
<td>0.059 0.010 0.100</td>
</tr>
<tr>
<td>Elongin C-like</td>
<td>Protolytic</td>
<td>0.160 0.004 0.895</td>
</tr>
<tr>
<td>Dfratac dehydrogenase</td>
<td>Energy metabolism</td>
<td>0.094 0.000 0.143</td>
</tr>
<tr>
<td>Ubiquitin-associated membrane glycoprotein 1-like</td>
<td>Binding</td>
<td>0.005 0.000 0.988</td>
</tr>
<tr>
<td>CAPTEDRAFT</td>
<td>_228707</td>
<td>Signaling transduction</td>
</tr>
<tr>
<td>Lethal A, partial</td>
<td>Protein metabolism</td>
<td>0.040 0.003 0.998</td>
</tr>
<tr>
<td>DC10205529</td>
<td>Signaling transduction</td>
<td>0.019 0.000 0.001</td>
</tr>
<tr>
<td>Calmin</td>
<td>Signaling transduction</td>
<td>0.390 0.441 0.000</td>
</tr>
<tr>
<td>Predicted protein</td>
<td>Other</td>
<td>0.305 0.017 1.000</td>
</tr>
</tbody>
</table>
Figure 4.3. Expression patterns of differentially expressed proteins between control (6.7 mg O₂ L⁻¹) and three other treatment groups (4.1, 2.0 and 1.0 mg O₂ L⁻¹; ANOVA, P < 0.02) in P. canaliculata (a) and P. diffusa (b). Values showing significance are highlighted in red. Color scale bar on the left represents the fold change (red = up-regulation, green = down-regulation). On the right of the heat map are protein annotation, functional classification, and P value of post hoc test result.
Figure 4.4. Percentage composition of different functional groups of DEPs in *P. canaliculata* and *P. diffusa* under 4.1, 2.0 and 1.0 mg O$_2$ L$^{-1}$.
Figure 4.5. Possible pathways involved in differential hypoxia-related protein expression in *P. canaliculata* and *P. diffusa*. Protein names in red color represent differentially expressed proteins (including one-way and two-way ANOVA results) detected in our study. Enzymes in the TCA cycle: (1) Aconitase; (2) Succinyl-CoA synthetase; (3) Dihydrolipoyl dehydrogenase (FAD cofactor); (4) Malate dehydrogenase precursor; (5) Citrate synthase; NDUFV2: NADH dehydrogenase [ubiquinone] flavoprotein 2; NDUFA5: NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5; COX6B: cytochrome c oxidase subunit V1b.
Table 4.1. Log-likelihood values and parameter estimates for 3 orthologous protein-coding genes under two pairs of site models.

<table>
<thead>
<tr>
<th>Description</th>
<th>Model type</th>
<th>lnL</th>
<th>2Δℓ</th>
<th>Ka/Ksb</th>
<th>Parameters</th>
<th>BEB residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial, partial</td>
<td>M1a</td>
<td>-1477.213</td>
<td>0.197</td>
<td>P₀ = 0.803, P₁ = 0.197; ω₀ = 0, ω₁ = 1.000</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M2a</td>
<td>-1465.000</td>
<td>24.426</td>
<td>1.044</td>
<td>P₀ = 0.796, P₁ = 0.171, P₂ = 0.033; ω₀ = 0, ω₁ = 1.000, ω₂ = 271V, 275G</td>
<td>214L(<em>), 235Q(<strong>), 271V(</strong>), 275G(</em>)</td>
</tr>
<tr>
<td></td>
<td>M7</td>
<td>-1477.216</td>
<td>0.200</td>
<td>p = 0.005, q = 0.020</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M8</td>
<td>-1465.083</td>
<td>24.266</td>
<td>1.180</td>
<td>P₀ = 0.969 (P₁ = 0.031); p = 0.005, q = 0.021, ω = 31.369</td>
<td>214L(<em>), 235Q(<strong>), 271V(</strong>), 275G(</em>)</td>
</tr>
<tr>
<td>Cytochrome c oxidase subunit VIb</td>
<td>M1a</td>
<td>-1185.528</td>
<td>0.217</td>
<td>P₀ = 0.783, P₁ = 0.217; ω₀ = 0, ω₁ = 1.000</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M2a</td>
<td>-1181.342</td>
<td>8.373</td>
<td>0.365</td>
<td>P₀ = 0.977, P₁ = 0, P₂ = 0.023; ω₀ = 0.144, ω₁ = 1.000, ω₂ = 9.969</td>
<td>28K(**)</td>
</tr>
<tr>
<td></td>
<td>M7</td>
<td>-1185.559</td>
<td>0.206</td>
<td>p = 0.007 q = 0.021</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M8</td>
<td>-1181.347</td>
<td>8.422</td>
<td>0.366</td>
<td>P₀ = 0.978 (P₁ = 0.022); p = 16.709, q = 99.000, ω = 9.989</td>
<td>28K(**), 194S(*)</td>
</tr>
<tr>
<td></td>
<td>M1a</td>
<td>-407.633</td>
<td>0.212</td>
<td>P₀ = 0.788, P₁ = 0.212; ω₀ = 0, ω₁ = 1.000</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M2a</td>
<td>-405.972</td>
<td>3.322</td>
<td>0.490</td>
<td>P₀ = 0.846, P₁ = 0, P₂ = 0.154; ω₀ = 0, ω₁ = 1.000, ω₂ = 3.189</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>M7</td>
<td>-407.647</td>
<td>0.200</td>
<td>p = 0.005, q = 0.020</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M8</td>
<td>-405.971</td>
<td>3.351</td>
<td>0.490</td>
<td>P₀ = 0.846 (P₁ = 0.154); p = 34T(*)</td>
<td>---</td>
</tr>
</tbody>
</table>

---

a Numbers in bold indicate significantly difference between the compared models (M2a vs. M1a and M8 vs. M7).
b This is an average Ka/Ks ratio along all the sites in different aligned genes.
c p and q are shape parameters in the beta distribution. ω₀ is the ratio of Ka/Ks sites with a proportion of P₀, ω₁ is the Ka/Ks of completely neutral sites with a proportion of P₁, ω₂ is the Ka/Ks of positive selection sites with a proportion of P₂.
d Positively selected amino acid sites predicted by Bayes Empirical Bayes (BEB) analysis. * indicates the probability of this amino acid site undergoing positive selection is larger than 95%, Pr (ω > 1) > 95%; ** indicates Pr (ω > 1) > 99%.
Chapter 5 General conclusions and perspectives

My study has shown that functional proteomic and transcriptomic techniques along with comparison between congeneric species is a powerful tool to investigate the molecular basis of stress tolerance and adaptive evolution in *Pomacea* snails. This work could contribute to predicting their biogeographic distribution in changing climate and understanding the phylogenetic evolution and speciation in gastropods.

The first experimental chapter of my study provides us with a comprehensive proteomic profile of the perivitelline fluid (PVF) of *Pomacea maculata*. Transcriptomic expression comparison between albumen gland and other tissues indicates albumen gland is the organ secreting most of these proteins. Functional classification of the 74 PVF proteins reveals that the most abundant proteins are perivitellins followed by immune responsive and energy metabolic proteins which may be involved in protecting the eggs from exogenous pathogens and providing energy source. Evolutionary analysis of the sequence divergence between *P. maculata* and *P. canaliculata* shows that the PVF genes had a weaker evolutionary constrains than the housekeeping genes.

The second experimental chapter of my study provides us insights into the genetic basis of heat resistance and adaptive sequence divergence in *P. canaliculata* and *P. diffusa*. Compared with the non-invasive *P. diffusa*, the invasive *P. canaliculata* had a higher $LT_{50}$, indicating that it can withstand a
higher maximum temperature in the field. The comparative proteomic data showed that chronic heat exposure caused more proteins to differentially express than acute exposure, reflecting significant changes in many biological structures and processes in both species under chronic heat exposure. However, the two species differed greatly in their responses to both acute and chronic heat stress. The number of differentially expressed proteins in *P. diffusa* was twice of that in *P. canaliculata* and there was likely a shift from carbohydrate to lipid catabolism in *P. diffusa* due to the higher heat induced metabolism. Meanwhile, there were also more highly expressed heat shock proteins in *P. diffusa*, which might compromise the energy allocation for other critical biological functions under chronic stress. At the transcriptomic level, base substitution rate analysis of differentially expressed orthologous sequences between these two *Pomacea* species allowed us to identify a set of candidate genes that may have experienced rapid positive selection.

The third experimental chapter provides, using an approach identical to that of the second experimental chapter, provides us with a comprehensive understanding about the genetic basis of differential hypoxia tolerance and protein expression profiles between the two congeneric snails. The proteomic data show a thorough pathway map which include all the differentially expressed proteins and reveal the complex molecular mechanism of hypoxia underlying the differential hypoxia tolerance in the two species. Greatly different between the two species are the TCA cycle and glycolysis, highlighting
the adaptation of the invasive \textit{P. canaliculata} in energy metabolism related pathways.

On a broader context, the combined physiological and high-throughput proteomic techniques, transcriptome-wide orthologous comparison and streamlined comparative approach which were adopted in my study can be applied in studies aiming to understand the genetic basis of stress responses or species invasion in a changing climate for other non-model organisms. For example, the giant African snail \textit{Achatina fulica} is an invasive land snail in Asia and also on the list of ‘100 of the World's Worst Invasive Alien Specie’. Using similar techniques and methods, studies could be performed to examine the differential proteomic and transcriptomic responses of \textit{A. fulica} and native snails to various environmental stressors (e.g., desiccation and cold temperature). With more studies on species invasion, we could better understand the molecular mechanisms of invasiveness.

With respect to apple snails, further studies can be conducted in order to better understand the evolution in molecular mechanisms: (1) examine the proteomic compositions of perivitelline fluid of eggs in other ampullariid species and conduct a phylogenetic analysis. Ampullariidae contain species both depositing egg masses under water and above waterline. Since the aerial oviposition strategy is quite unique, comparison of species which adopted various reproductive strategies from different genera of Ampullariidae would help us understand the evolution of reproductive proteins and their phylogenetic
significance in the species speciation; (2) establish genomic databases. Although we currently identified proteins using translated transcriptomic databases, they showed several drawbacks such as sequence fragmentation and redundancy which are mainly from the alternative splicing of RNA. The genomic resources will facilitate protein identification, functional and evolutionary studies. Those resources and analysis will greatly enhance our understanding about some basic biological process during egg development, evolution of reproductive proteins, phylogeny and speciation of mollusks as well as other animals.
Literature cited


Braby C E, Somero G N. Following the heart: temperature and salinity effects on heart rate in native and invasive species of blue mussels (genus *Mytilus*). *Journal of Experimental Biology*, 2006, 209, 2554–2566.


Cheesman D F. Ovorubin, a chromoprotein from the eggs of the gastropod mollusc *Pomacea canaliculata*. *Proceedings of the Royal Society of London B: Biological Sciences*, 1958, 149, 571–587.


Das T, Stickle W B. Sensitivity of crabs *Callinectes sapidus* and *C. similis* and the gastropod *Stramonita haemastoma* to hypoxia and anoxia. *Marine Ecology Progress Series*, 1993, 98, 263–263.


Dong Y, Somero G N. Temperature adaptation of cytosolic malate dehydrogenases of limpets (genus *Lottia*): differences in stability and


EFSA PLH Panel (EFSA Panel on Plant Health). Scientific opinion on the assessment of the potential establishment of the apple snail in the EU. *EFSA Journal*, 2013, 11, 3487, 49.


mitten crab *Eriocheir sinensis*. Developmental & Comparative Immunology, 2014, 46, 255–266.


Lach L, Cowie R H. The spread of the introduced freshwater apple snail
*Pomacea canaliculata* (Lamarck) (Gastropoda: Ampullariidae) on Oahu,

Lacoste A, Malham S K, Gélébart F, Cueff A, Poulet S A. Stress-induced
immune changes in the oyster *Crassostrea gigas*. *Developmental &
Comparative Immunology*, 2002, 26, 1–9.

Laurindo F R M, Pescatore L A, de Castro Fernandes D. Protein disulfide
isomerase in redox cell signaling and homeostasis. *Free Radical Biology

Lejeusne C, Latchere O, Petit N, Rico C, Green A J. Do invaders always
perform better? Comparing the response of native and invasive shrimps to
temperature and salinity gradients in south-west Spain. *Estuarine, Coastal

Li B, Dewey C N. RSEM: accurate transcript quantification from RNA-Seq data

two sublittoral nassariid gastropods to hypoxia. *Marine Ecology Progress

Lockwood B L, Sanders J G, Somero G N. Transcriptomic responses to heat
stress in invasive and native blue mussels (genus *Mytilus*): molecular
correlates of invasive success. *Journal of Experimental Biology*, 2010, 213,
3548–3558.

Lockwood B L, Somero G N. Transcriptomic responses to salinity stress in
invasive and native blue mussels (genus *Mytilus*). *Molecular Ecology*, 2011,
20, 517–529.

Mitochondrial genome analysis of *Ochotona curzoniae* and implication of
cytochrome *c* oxidase in hypoxic adaptation. *Mitochondrion*, 2008, 8,
352–357.


Matsukura K, Tsumuki H, Izumi Y, Wada T. Physiological response to low


Rossi V. Scientific Opinion on the assessment of the potential establishment of
the apple snail in the EU. *The EFSA Journal*, 2013, 11, 1–49.


Silvestro D, Michalak I. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution*, 2012, 12, 335–337.


Strayer, D L. Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. *Freshwater Biology*, 2010, 55,
152–174.


Tomanek L, Somero G N. Interspecific-and acclimation-induced variation in levels of heat-shock proteins 70 (hsp70) and 90 (hsp90) and heat-shock


Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology*


Appendix 1. A python script used for filtering unpaired scans and separating data generated by the CID and HCD scans in mass spectrometry.

Note: Please kindly note that the starting target file is .mgf file which is converted from raw mass spectrometry data using Proteome Discoverer 1.3.0.339 (Thermo Finnigan, CA).

```python
# Delete the first line 'MASS=Monoisotopic' manually.
import gc
from time import clock

start = clock()  # Record the start time point.

# Part I: filter failed ions, then divide the target file into CID and HCD files.

file = open('targetfile.mgf', 'r')  # Open the target file.
hcdfilteredfile = open('targetfile_HCDfiltered.txt', 'w')  # Open the HCD filtered file.
cidfilteredfile = open('targetfile_cidfiltered.txt', 'w')  # Open the CID filtered file.

index = -1
list = []
hcdlist = []
cidlist = []

for line in file:
    if 'BEGIN' in line:
        index = index + 1
        list.insert(index, '')  # Transfer the target file into a list.
        list[index] = line
    else:
        list[index] = list[index] + line

    if not line:
        break

listlength = len(list)

print 'List length is', listlength

for randomnum in range(listlength - 1):
    findmassspace = list[randomnum].find('PEPMASS')
    findmass2space = list[randomnum].find('CHARGE')
```

143
x = list[randomnum][findmassspace:(findmass2space-1)]
find2massspace = list[randomnum+1].find('PEPMASS')
find2mass2space = list[randomnum+1].find('CHARGE')
y = list[randomnum+1][find2massspace:(find2mass2space-1)]
a = list[randomnum].find('Spectrum')
b = list[randomnum].find('scans')
num = list[randomnum][(a+8):(b-1)] # Pick up the spectrum number.
if x == y:                 # Determine whether the adjacent pepmass numbers are
    # same.
    if int(num) % 2 == 0:   # Determine spectrum number-'num' is an odd or even
        hcdlist.append(list[randomnum])
        cidlist.append(list[randomnum+1])
    if int(num) % 2 == 1:
        cidlist.append(list[randomnum])
        hcdlist.append(list[randomnum+1])
hcdfilteredline = '\n'.join(hcdlist)
cidfilteredline = '\n'.join(cidlist)
hcdfilteredfile.write(hcdfilteredline)
cidfilteredfile.write(cidfilteredline)
print 'Filtered and sort done'
file.close()
cidfilteredfile.close()
hcdfilteredfile.close()

del list, hcdlist, cidlist
gc.collect()  # Release RAM.

# Part II: Delete 114 to 117.5 in CIDfile; Select only 114 to 117.5 into HCDfile.

# Part II: Delete 114 to 117.5 in CIDfile; Select only 114 to 117.5 into HCDfile.
cidfile = open('targetfile_cidfiltered.txt','r')
hcdfile = open('targetfile_HCDfiltered.txt','r')
newcidfile = open('targetfile_cidfiltered_no114to117.txt','w')
newhcdfile = open('targetfile_hcdfiltered_only114to117_division2.txt','w')
i = 0
for line in cidfile:
    i = i + 1
    if line[0] == '1':
        c = line.find(' )
        d = float(str(line[0:c]))
        if d > 114 and d < 117.5:
            newcidfile.write('')
        else:
newcidfile.write(line)
else:
    newcidfile.write(line)
print 'Delete 114 to 117.5 in CIDfile.'
for line in hcdfile:
    i = i + 1
    if line[0] == '1':
        x = line.find(' ')
        y = float(str(line[0:x]))
        if y > 114 and y < 117.5:
            z = float(str(line[(x+1):-1])/2)  # For 14 to 117.5, division 2 of intensity,
            newline = str(line[0:x])+' '+str(z)+'
            newhcdfile.write(newline)
else:
    newhcdfile.write('')
if line[0] in ['2','3','4','5','6','7','8','9']:
    newhcdfile.write('')
if line[0] in ['B','T','P','C','R','S','E','\n']:
    newhcdfile.write(line)
print 'Select only114 to 117.5 into HCDfile.'
cidfile.close()
hcdfile.close()
newcidfile.close()
newhcdfile.close()

# Part III: Combine CIDfile with 114 to 117.5 of HCDfile.
oddfile = open ('targetfile_cidfiltered_no114to117.txt','r')
evenfile = open ('targetfile_hcdfiltered_only114to117_division2.txt','r')
combinefile = open ('targetfile_CIDwith114to117Division2.fasta','w')
evenlist = []
oddlist = []
combinelist = []
evenindex = -1
oddindex = -1
for line in evenfile:
    if 'BEGIN' in line:
        evenindex = evenindex + 1
        evenlist.insert(evenindex,'')
        evenlist[evenindex] = line
    else:
        evenlist[evenindex] = evenlist[evenindex] + line
a = len(evenlist)
print 'selected hcd list length is ', a

for line in oddfile:
    if 'BEGIN' in line:
        oddindex = oddindex + 1
        oddlist.insert(oddindex,"")
        oddlist[oddindex] = line
    else:
        oddlist[oddindex] = oddlist[oddindex] + line
b = len(oddlist)
print 'deleted cid list length is ',b

if a == b:
    for i in range(a):
        even_num1 = evenlist[i].find('Spectrum')
        even_num2 = evenlist[i].find('scan')
        even_num3 = evenlist[i].find('SCANS')
        even_num4 = evenlist[i].find('\n',even_num3)
        even_num5 = evenlist[i].find('END')
        even_num = float(evenlist[i][(even_num1 + 8):(even_num2 - 1)])
        odd_num1 = oddlist[i].find('Spectrum')
        odd_num2 = oddlist[i].find('scan')
        odd_num3 = oddlist[i].find('SCANS')
        odd_num4 = oddlist[i].find('\n',odd_num3)
        odd_num5 = oddlist[i].find('IONS')
        odd_num = float(oddlist[i][(odd_num1 + 8):(odd_num2 - 1)])
        if even_num -1 == odd_num:
            combinelist = oddlist[i][0:(odd_num4 + 1)] + evenlist[i][(odd_num4 + 1):even_num5] + oddlist[i][(odd_num4 + 1):-1]
            combineline = str(combinelist)
            combinefile.write(combineline)
    print 'Final done.'
else:
    print 'Wrong: CID and HCD files are not equal length!'

#Part IV: remove redundant '\n' in HCDfile.
file = open('targetfile_HCDfiltered.txt','r')
wfile = open('targetfile_HCDfiltered_RemoveEnter.fasta', 'w')
index = 0
for line in file:
    index = index + 1
    if line[0] == 'n':
        wfile.write('')
    if line[0] == 'E':
        line = line[0:-1]
        wfile.write(line)
    else:
        wfile.write(line)
file.close()
wfile.close()

finish = clock()
print 'Total running time is ', finish - start, ' seconds.'   # Print total running seconds.

# Add 'MASS=Monoisotopic' at the first line in each file manually.

CURRICULUM VITAE

Academic qualifications of the thesis author, Ms. MU Huawei:

- Received the degree of Bachelor of Science in Biological Science, College of Marine Life Science from Ocean University of China, June 2012.

September 2016