

ResearchOnline@JCU

This file is part of the following reference:

Mohamed, Amin Roushdy (2016) *Transcriptomics of coral-algal interactions: novel insights into the establishment of symbiosis*. PhD thesis, James Cook University.

Access to this file is available from:

<http://researchonline.jcu.edu.au/49900/>

The author has certified to JCU that they have made a reasonable effort to gain permission and acknowledge the owner of any third party copyright material included in this document. If you believe that this is not the case, please contact

*ResearchOnline@jcu.edu.au and quote
<http://researchonline.jcu.edu.au/49900/>*

Transcriptomics of coral-algal interactions: novel insights into the establishment of symbiosis

A Doctoral Thesis

Submitted by

Amin Roushdy Mohamed

BSc, Zoology and MSc, Marine Biology

July 2016



For the degree of Doctor of Philosophy (PhD)
Biochemistry and Marine Biology

ARC Centre of Excellence for Coral Reef Studies,
Department of Molecular and Cell Biology,
James Cook University
and

Australian Institute of Marine Science
Townsville, Queensland, Australia



I dedicate this work to the soul of my mother, Nahed O. El-Sawwaf,

You were and will always be in my heart !

ACKNOWLEDGEMENTS

Well, it has been a very long journey to get to this point. It gave me the opportunity to meet many amazing people and work/visit many amazing places around the globe. Despite being very heart broken to lose the most beloved person (my mother) during my studies, there were many ups; I really enjoyed the Aussie life style and the huge experience I gained is priceless.

First and for the most, I feel deeply indebted to Allah, the Lord of the world, the most merciful and the most gracious for assisting and providing me with power and patience to complete this work. A huge thank you goes to my beloved wife Marwa and my little angels Hanin and Habiba. Marwa, thank you so much for being with me. You really sacrificed for me, thank you for love, support and every thing. I also really want to express my gratitude to my dear father (Roushdy Esmail) for being very supportive in all aspects of my life especially financially, every time I had any difficulty you were there and very generous.

I would like to give a big thank you to my supervisory team; Professor David Miller, Professor Bette Willis, Dr David Bourne and Dr Vivian Cumbo. Actually, I was very lucky to work with experts in the fields of coral reef genomics and/or ecology during my studies at James Cook University. David M, I would like to express my deep gratitude to you for accepting me in your lab in the first place. It would take many pages to acknowledge you enough, as you always do when you write recommendation letters for me. I really appreciate that you took the risk of having a PhD student with no prior experience in molecular genetics. It was my dream to do my PhD research in ecological genomics in Australia and your work really inspired me to become a molecular ecologist. Thank you so much for being supportive in many aspects of my life, not only in research. Finally a big thank you for your advice regarding the future and supporting me during my applications for postdoc fellowships/positions. Bette, it was a pleasure to be one of your students. I really enjoyed and learnt a lot during our discussions. Thank you so much for your encouragement, support and help. Thank you so much for inviting my family to your house, we were very delighted to visit you and meet with your husband. David B, I really enjoyed working with you. Thank you so much for being my supervisor. I would never forget the opportunities that I had through your involvement in my work, access to research facilities at AIMS and scholarships/awards from AIMS@JCU. I really appreciate

your help and support during conducting experiments at the CMMG lab in AIMS during GBR spawning in 2014. Vivian, it was a pleasure to work with you during my PhD. I obtained great benefit from your great experience about *Symbiodinium*, *Chromera* and coral-algal symbioses. I really enjoyed our field trip to Okinawa and learnt a lot from you about coral spawning and infection experiments. Thank you so much for being there when I needed to discuss with you my results.

I was very lucky again to have training on next-gen sequencing analysis with Professor Mark Ragan group and Dr Cheong Xin Chan (CX) at the Institute of Molecular Bioscience (IMB), the University of Queensland. Mark, it was a pleasure to be with your group for few weeks to learn basics of bioinformatics. Thank you so much for allowing me to come and visit your group and for being very welcoming and generous. I really appreciate your support in my postdoc applications, despite the short time I spent with your group. A big thank you CX, I was very lucky to meet and work closely with you. My bioinformatics skills were suddenly very good after few days working with you. Thank you so much for your help and patience during my visit at IMB. You really showed me how to think as a bioinformatician, not as a biologist. Those few weeks were a real turning point in my PhD research.

During my PhD I was lucky enough to be funded from different bodies. I'd like to deeply thank the mission department of the Egyptian Ministry of Higher Education, James Cook University and AIMS@JCU for their financial support during my studies in Townsville. Also it doesn't go without saying thank you to many people for their help during field work among them: Chuya Shinzato, Saki Harri and Joana Figueiredo during field work in Sesokok station; David Abrego, Allison Paley, Margaux Hein during coral sampling at Heron Island Research Station (HIRS); Mei fang Lin and Chao-yang Kuo during coral sampling at Orpheus Island Research Station (ORIS); SeaSim staff at AIMS during GBR spawning in 2014. Finally, I would like to thank all the members of the David Miller's lab during 2012/16, especially Bruno Lapeyre and Sussanne Sprungala for their help during my first days in the DM lab and teaching me basic molecular biology techniques.

STATEMENT OF THE CONTRIBUTION OF OTHERS

A supervisory committee consisted of Professor David Miller and Professor Bette Willis James Cook University and Dr David Bourne from the Australian Institute of Marine Science provided intellectual and editorial support.

The research work in the thesis was conducted with collaboration with other research scientists in different institutions. I was responsible for designing and conducting experiments, data generation/analyses and writing the thesis. Dr Vivian Cumbo provided guidance in designing and conducting experiments in Chapters 2 and 3. She also allowed access to *Chromera* cultures used to construct the *Chromera de novo* transcriptome in Chapter 4. Prof Dee Carter provided culture of *Chromera* used in Chapter 3. Dr Saki Harri provided the *Symbiodinium* culture used for infection experiment in Chapter 2 and provided laboratory facilities at the Sesoko Island marine station during *Acropora digitifera* spawning event in June/July 2013. Prof Nori Satoh and Dr Chuya Shinzato at the Okinawa Institute of Marine Science (OIST) in Okinawa, Japan provided reagents for RNA isolation and illumina RNA-Seq libraries preparations and illumina HiSeq sequencing in Chapters 2 and 3. Prof Mark Ragan and Dr Chenog X Chan at the Institute of Molecular Bioscience (IMB) of the University of Queensland (UQ) provided guidance for the bioinformatics analysis required in the thesis specially Chapter 4.

Primary research funding was provided through an Australian Research Council Grant to Prof David Miller. I was supported with a 4-year PhD scholarship from the Egyptian Ministry of Higher Education and Research. I was also supported by a James Cook University Postgraduate Research Scholarship (JCUPRS) and under the joint venture between the Australian Institute of Marine Science and James Cook University AIMS@JCU.

Table of contents

TABLE OF CONTENTS	i
LIST OF FIGURES	vi
LIST OF TABLES	xiii
THESIS ABSTRACT	1
CHAPTER 1 BACKGROUND AND GENERAL INTRODUCTION	3
1.1 Background on coral biology and the coral holobiont	4
1.2 The coral- <i>Symbiodinium</i> symbiosis and its significance	5
1.3 Coral-associated apicomplexan-related lineages	7
1.4 Coral-Chromerida associations	9
1.5 Coral- <i>Chromera</i> symbiosis	10
1.6 <i>Chromera</i> as a proxy for <i>Plasmodium</i>	14
1.7 Coral reef ecological genomics	14
1.8 RNA-Seq approach to study coral reef biology	18
1.9 Aims, objectives and thesis structure	19
1.10 References	21
CHAPTER 2 UNRAVELLING THE MOLECULAR MECHANISMS UNDERLYING ESTABLISHMENT OF CORAL- SYMBIODINIUM SYMBIOSIS USING RNA-SEQ	27
2.1 ABSTRACT	28
2.2 INTRODUCTION	28
2.3 MATERIAL AND METHODS	32
2.3.1 Coral larvae	32
2.3.2 <i>Symbiodinium</i> culture	33
2.3.3 <i>Symbiodinium</i> acquisition experimental design	33
2.3.4 Larval sampling and RNA isolation	34
2.3.5 High-throughput sequencing Illumina RNA-Seq	34
2.3.6 RNA-Seq data analysis	34
2.3.6.1 Reads quality check and reference mapping	34
2.3.6.2 Differential gene expression analysis	35
2.3.6.3 Hierarchical clustering analysis	35
2.3.6.4 Functional annotations and gene ontology (GO) enrichment analyses	36
2.4 RESULTS	37
2.4.1 General results and overall transcriptome changes	37
2.4.2 Differential gene expression analysis	42
2.4.3 Gene Ontology (GO) enrichment analysis	45
2.4.4 GO enrichment analysis reveals suppression of ribosome, Endoplasmic Reticulum and mitochondria functions	45
2.4.5 Focus on symbiosis-related coral genes	51
2.4.5.1 The coral host might recognize symbionts via MAMP-PRR interactions	51
2.4.5.2 Role of cell adhesion proteins during recognition	52

2.4.5.3	Genes involved in vesicular trafficking likely used in symbiosome formation	54
2.4.5.4	Transcription factors and epigenetic tags for silencing were affected	56
2.4.5.5	Genes involved in regulating the host cell cycle during symbiosis establishment	57
2.4.5.6	Immune-related genes are suppressed during onset of coral- <i>Symbiodinium</i> symbiosis	59
2.4.5.7	Establishment of symbiosis alters expression of genes involved in host apoptosis	60
2.4.5.8	Genes involved in host response to reactive oxygen species (ROS), inflammation and stress	61
2.5	DISCUSSION	65
2.5.1	The coral host transcriptome showed rapid and transient changes upon symbiosis onset	65
2.5.2	Metabolic suppression during coral and <i>Symbiodinium</i> initial interactions	66
2.5.3	PRR-MAMP signaling is used to recognize competent <i>Symbiodinium</i> during infection	67
2.5.4	Role of cell adhesion genes during symbiosis establishment	68
2.5.5	The symbiosome is formed as a result of arresting phagosomal maturation	69
2.5.6	Regulation of the host cell cycle during symbiont tolerance	71
2.5.7	Regulation of the host apoptotic repertoire during symbiont tolerance	72
2.5.8	Suppression of host immunity during onset of symbiosis	73
2.5.9	Responses to stress and ROS during symbiont tolerance	74
2.5.10	Conclusion	75
2.6	REFERENCES	75
CHAPTER 3 DECIPHERING THE NATURE OF THE CORAL- CHROMERA ASSOCIATION VIA NEXT GENERATION SEQUENCING		81
3.1	ABSTRACT	82
3.2	INTRODUCTION	82
3.3	MATERIALS AND METHODS	84
3.3.1	<i>Chromera velia</i> culture	84
3.3.2	Coral larvae and <i>Chromera</i> infection experiment	85
3.3.3	Larval sampling and RNA isolation	85
3.3.4	High-throughput next generation sequencing	85
3.3.5	RNA-Seq data analysis	85
3.3.5.1	Reads quality check and reference mapping	85
3.3.5.2	Differential gene expression analysis	86
3.3.5.3	Hierarchical clustering analysis	86
3.3.5.4	Functional annotations, gene ontology (GO) and KEGG pathway enrichment analyses	86
3.4	RESULTS	87
3.4.1	General results and overall transcriptome changes	87
3.4.2	Differential gene expression analysis	92

3.4.3 Functional profiles	96
3.4.4 GO enrichment in the late response to <i>Chromera</i>	99
3.4.5 Focus on genes involved in host-microbe interactions	105
3.4.5.1 Coral immune response against <i>Chromera</i>	107
3.4.5.2 Genes involved in phagocytosis and the host endocytic pathway, an in particular phagosome maturation, were affected	112
3.4.5.3 Genes involved in host apoptosis were affected	117
3.5 DISCUSSION	121
3.5.1 Coral responses common to <i>Chromera</i> and <i>Symbiodinium</i> infection	122
3.5.2 The late response of coral to <i>Chromera</i> infection	122
3.5.3 Suppression of the host immune response during <i>Chromera</i> infection	123
3.5.4 Modulation of the endocytic pathway and phagosome maturation in the coral host response to <i>Chromera</i> infection	124
3.5.5 Apoptosis as a double-edged sword for both coral host and <i>Chromera</i> survival	126
3.5.6 Conclusion	127
3.6 REFERENCES	128
CHAPTER 4 CHROMERA TRANSCRIPTOMICS: A FUNCTIONAL GENOMIC RESOURCE FOR A CHROMERID ALGA ISOLATED FROM THE GREAT BARRIER REEF AND COMPARATIVE TRANSCRIPTOMIC ANALYSES WITH PARASITIC AND PHOTOSYNTHETIC RELATIVES	132
4.1 ABSTRACT	133
4.2 INTRODUCTION	134
4.3 MATERIAL AND METHODS	135
4.3.1 <i>Chromera velia</i> culture	135
4.3.2 <i>Chromera</i> culturing conditions for transcriptome construction	135
4.3.3 RNA isolation and high-throughput next generation sequencing	136
4.3.4 RNA-Seq data quality control, processing and quality filtering	137
4.3.5 Transcriptome <i>de novo</i> assembly using the TRINITY software	137
4.3.6 Assessing the quality of the <i>de novo</i> assembly	138
4.3.7 Transcriptome functional annotations	138
4.3.7.1 Gene Ontology (GO) terms	138
4.3.7.2 KEGG pathways	139
4.3.8 <i>Chromera</i> strains comparative transcriptome analysis	140
4.3.8.1 Identification of the orthologous genes between two <i>Chromera</i> strains	140
4.3.8.2 <i>Chromera</i> orthologs functional profile	141
4.3.8.3 Estimation of substitution rates between <i>Chromera</i> strains	141
4.3.9 <i>Chromera</i> , <i>Symbiodinium</i> and <i>Plasmodium</i> comparative transcriptome analyses	142
4.3.9.1 <i>Symbiodinium</i> and <i>Plasmodium</i> sequences and functional annotation	142
4.3.9.2 GO enrichment and KEGG pathways analyses based on shared genes	142
4.4 RESULTS	143
4.4.1 Illumina sequencing and <i>de novo</i> assembly of <i>Chromera</i> transcriptome	143

4.4.2 Transcriptome annotations	145
4.4.2.1 GO terms distribution	146
4.4.2.2 KEGG pathway analysis	146
4.4.3 Assessing the quality of the <i>de novo</i> transcriptome assembly	151
4.4.4 Comparative transcriptomics of <i>Chromera</i> stains	151
4.4.4.1 Identification of orthologous genes between Sydney and GBR <i>Chromera</i> strains	151
4.4.4.2 <i>Chromera</i> orthologs functional profile	152
4.4.4.3 Analysis of Ka/Ks, a test for selection	153
4.4.5 Comparative analyses of <i>Chromera</i> , <i>Symbiodinium</i> and <i>Plasmodium</i> transcriptome data	155
4.4.5.1 Functional profile for the set of genes shared between Sydney and GBR <i>Chromera</i>	156
4.4.5.2 Functional profile for the set of genes shared between <i>Chromera</i> and <i>Symbiodinium</i>	162
4.4.5.3 Functional profile for the set of genes shared between <i>Chromera</i> and <i>Plasmodium</i>	165
4.4.6 KEGG pathways comparative analysis	168
4.4.6.1 Glycan biosynthesis in <i>Chromera</i> , <i>Symbiodinium</i> and <i>Plasmodium</i>	171
4.4.6.2 Transcription machinery in <i>Chromera</i> , <i>Symbiodinium</i> and <i>Plasmodium</i>	174
4.5 DISCUSSION	180
4.5.1 Transcriptome assembly and completeness	180
4.5.2 Identification of <i>Chromera</i> orthologs and genes under selection	181
4.5.3 Functional profile of shared genes among <i>Chromera</i> strains	182
4.5.4 Functional profiles of genes shared between <i>Chromera</i> and <i>Symbiodinium</i> or <i>Chromera</i> and <i>Plasmodium</i>	183
4.5.5 KEGG pathways comparison reveals similarity between <i>Chromera</i> and <i>Symbiodinium</i> suggesting the potential of <i>Chromera</i> being a coral symbiont	184
4.5.6 Variation in N-Glycan biosynthesis enzymes might explain the coral host recognition specificity	184
4.5.7 Reduction of the transcription machinery in <i>Symbiodinium</i> ; a unique feature of dinoflagellates	185
4.5.8 Conclusion	186
4.6 REFERENCES	186
CHAPTER 5 GENERAL DISCUSSION, MAJOR THESIS FINDINGS AND FUTURE RESEARCH	189
5.1 General discussion	189
5.2 Major thesis findings	194
5.3 Future research	195
5.4 Conclusion and impact	197
5.5 REFERENCES	198
APPENDICES	200
Appendix I	200
The transcriptomic response of the coral <i>Acropora digitifera</i> to a competent	

<i>Symbiodinium</i> strain: the symbiosome as an arrested early phagosome	
Appendix II	216
Detecting the presence of chromerids/ apicomplexans in corals of the Great Barrier Reef, Australia using PCR/ sequencing assays	
Appendix III Chapter 3.0 Supplementary Information	224
Appendix IV Chapter 4.0 Supplementary Information	242

List of Figures

- Figure 1.1** Coral as a meta-organism (holobiont) consisting of the coral animal and associated microorganisms (bacteria, viruses and microbial eukaryotes including algal symbionts), modified from Bosch and McFall-Ngai (2011). 5
- Figure 1.2** Phylogeny of novel apicomplexan-related lineages (ARLs) and their distribution. ARLs were previously misidentified as of bacterial origin in 16S rDNA. Some of these ARLs appear to be tightly associated with coral tissue such as ARL-1 *Vitrella* clade, ARL-III *Chromera* clade, ARL-V, and ARL-VIII unidentified apicomplexan sequences (Janouskovec *et al.* 2012). 8
- Figure 1.3** Phylogenetic relationships among chromerids (*Chromera* and *Vitrella*), parasitic apicomplexans and photosynthetic dinoflagellates (Obornik & Lukes 2013). The three families (Dinophyta, Apicomplexa and Chromerida) are highlighted with three different background colours. White rectangles indicate the loss of photosynthesis; black rectangle indicates the loss of plastid. 10
- Figure 1.4** Location map, showing sites and hard coral species that host *Chromera* in Australia including corals on the great barrier reef (GBR). Coral photos from websites; <http://bie.ala.org.au/> and <http://coral.aims.gov.au/info/factsheets.jsp>. 11
- Figure 1.5** *Chromera* life cycle: a) predominant immotile coccoid cells; b, c) cells in division state; d, e) autospores released to start a new cycle; f) bi-flagellated motile *Chromera* cells, which occur during favorable conditions (Obornik & Lukes 2013). 12
- Figure 1.6** Light and electron micrographs of *Chromera* (Obornik *et al.* 2011). Part A shows coccoid *Chromera* cells during cell division (scale bar = 5 µm). Part B shows a scanning electron micrograph the bi-flagellated “motile” form of *Chromera*, with two heterodynamic flagella. The shorter flagellum contains a typical finger-like projection (scale bar = 1 µm). Part C shows a transmission electron micrograph of a section of *Chromera*, showing the unique organelle chromerosome, which is a projection-like organelle with a central bundle of fibers whose function is unknown (scale bar = 1 µm). 13
- Figure 1.7** Phylogenetic tree based on nuclear and plastid gene sequences, showing relationships within the alveolates. Note that *Chromera* and *Vitrella* (CCMP3155) are two independent photosynthetic lineages that are more closely related to apicomplexan parasites than to the dinoflagellates (Janouskovec *et al.* 2010). 13
- Figure 2.1** Colony of the coral *A. digitifera* collected in front of the Sesoko Station, the Tropical Biosphere Research Center of the University of the Ryukyus in Sesoko Island, Okinawa. 32
- Figure 2.2** Bioinformatic analysis flowchart. Illumina Hiseq reads were mapped onto the *A. digitifera* transcriptome assembly using the mapping software BOWTIE, then RSEM was used to estimate transcript abundance in each sample. Differential gene expression was inferred based on the mapping counts using the edgeR package in the R environment. Using the Uniprot accession IDs of the up- and down-regulated DEGs, Gene Ontology (GO) enrichment analysis was performed using DAVID database. Annotated DEGs were classified based on literature search of genes involved in cnidarian-algal symbioses. 37
- Figure 2.3** Images of *A. digitifera* larvae and *Symbiodinium* uptake in symbiotic larva under fluorescence microscopy. Left image shows a control “uninfected” larva fluorescing green, while right image shows an infected *Symbiodinium* sp. cladeB1-infected larva, the red fluorescence confirming successful symbiont uptake (infection). 38
- Figure 2.4** A screenshot of *A. digitifera* transcripts and reads alignments of both control and *Symbiodinium* infected reads at 4 h post-infection. Using the Integrated Genomics Viewer (IGV), the sorted BAM files containing the aligned reads were uploaded as well as the reference transcriptome sequence data. 40
- Figure 2.5** Level of agreement among the biological replicates. The heat map shows the hierarchically clustered Spearman correlation matrix resulting from comparing the 41

transcript expression values (TMM-normalized FPKM) for all samples against one another. Sample clustering indicates the consistency between the biological replicates of the *Symbiodinium* sp. clade B1 infection (samples S1, S2, S3) and negative control condition (samples C1, C2, C3) at 4 h post infection. The level of correlation represented by a colored field that ranges from green (correlation coefficient 0.85) to red (correlation coefficient 1.0).

Figure 2.6 A multidimensional scaling (MDS) plot produced by edgeR showing the relationship between all replicates of *Symbiodinium* clade B infection (S1, S2, S3) and negative control condition (C1, C2, C3). The distances shown on the plot are the biological coefficient of variation (BCV) between samples for the 500 genes that best distinguish the samples. **41**

Figure 2.7 *A. digitifera* transcriptome changes during onset of symbiosis with *Symbiodinium* sp. clade B1. Percentages of transcripts up- (blue) and down-regulated (red) in response to *Symbiodinium* infection. Only 2.91% of the coral transcriptome showed differential expression at the 4 h time point (adjusted $P \leq 0.05$), whereas no differences were observed in the coral transcriptome at 12 and 48 h post-infection. **42**

Figure 2.8 Volcano and MA plots comparing gene expression profiles between the *Symbiodinium*-infected and control larvae at 4, 12 and 48 h (Parts A, B and C). The red dots in part A represent the significant differentially expressed genes (DEGs) detected only at the 4-h time point (adjusted $P \leq 0.05$). Parts B and C show no significant DEGs using the same P -value. **43**

Figure 2.9 Differential gene expression profiles at 4 h post *Symbiodinium* infection. The heat map shows the expression profiles of *Symbiodinium*-infected and control samples. The hierarchical clustering was obtained by comparing the expression values (Fragment Per Kilobase of transcript per Million; FPKM) for *Symbiodinium*-infected samples compared against the control at 4 h post-infection. Expression values are \log_2 -transformed and then median-centered by transcript. Relative expression levels are shown in red (up) and green (down). Samples S1, S2 and S3 are the biological replicates of *Symbiodinium* sp. clade B1 infection while samples C1, C2, C3 are the biological replicates of the control condition. **44**

Figure 2.10 Heat map of annotated DEGs ($FDR \leq 0.05$) likely involved in pattern recognition, cell adhesion and vesicle trafficking in *Symbiodinium*-infected (S1, S2, S3) and control larvae (C1, C2, C3) at 4h post infection. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Symbiodinium*-infected samples against the control at 4 h post-infection. Expression values are \log_2 -transformed and then median-centered by transcript. The blue scale represents the relative expression values as \log_2 (fold-change). **55**

Figure 2.11 Heat map of annotated DEGs ($FDR \leq 0.05$) likely involved in transcriptional control and the cell cycle in *Symbiodinium*-infected (S1, S2, S3) and control larvae (C1, C2, C3) at 4h post infection. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Symbiodinium*-infected samples against the control at 4 h post-infection. Expression values are \log_2 -transformed and then median-centered by transcript. The blue scale represents the relative expression values as \log_2 (fold-change). **59**

Figure 2.12 Heat map of annotated DEGs ($FDR \leq 0.05$) likely involved in immunity, apoptosis and stress response in *Symbiodinium*-infected (S1, S2, S3) and control larvae (C1, C2, C3) at 4h post infection. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Symbiodinium*-infected samples against the control at 4 h post-infection. Expression values are \log_2 -transformed and then median-centered by transcript. The blue scale represents the relative expression values as \log_2 (fold-change). **63**

Figure 2.13 An integrative model of the molecular mechanisms during the onset and establishment of symbiosis in larvae of *Acropora digitifera*. Genes of interest and their **64**

roles are discussed; up-regulated genes are in blue text while genes that were down-regulated are in red. The initial uptake process (left panel) involves differential expression of a number of cell adhesion genes, including the up-regulation of a C-type lectin and down-regulation of the GP2 glycoprotein, as well as suppression of immunity, translation and mitochondrial processes. The establishment phase (central panel) involves blocking the maturation of the early phagosome containing the symbiont. Rabenosyn 5 and the Rab5-specific GEF ALS2 are likely to prevent the displacement of RAB5 by RAB7 and hence interfere with phagosome maturation. Differential expression of vitamin K epoxide reductase (VKOR) and calumenin is significant because both are thought to modify Sym32, a symbiosis-specific fasciclin (Ganot et al., 2011). The symbiont tolerance phase (right panel) involves complex changes required to handle the production of reactive oxygen species (ROS) by symbionts, as well as transient modifications in cell cycle regulation and the apoptotic network.

Figure 3.1 Images of *A. digitifera* larva and *Chromera* infection under fluorescent microscopy (20X objective). Part A shows control “uninfected” larva fluorescing green while part B shows *Chromera*-infected larva with red fluorescing observed during the infection process. **88**

Figure 3.2 A screenshot of *Acropora digitifera* transcripts and read alignments of both control and *Chromera* infection samples at the 48 h time point. Using the Integrated Genomics Viewer (IGV), the sorted BAM files containing the aligned reads were uploaded as well as the reference transcriptome sequence data. **89**

Figure 3.3 Level of agreement among the biological replicates at 4 h post *Chromera* infection. The heat map shows the hierarchically clustered Spearman correlation matrix resulting from comparing the transcript expression values (TMM-normalized FPKM) for all samples against one another. Samples clustering indicate the consistency between the biological replicates of the *Chromera* infection (samples I1, I2, I3) and negative control conditions (samples C1, C2, C3) at the 4 h time point. A color field that ranges from green to red presents the level of correlation. **90**

Figure 3.4 Level of agreement among the biological replicates at 48 h post *Chromera* infection. The heat map shows the hierarchically clustered Spearman correlation matrix resulting from comparing the transcript expression values (TMM-normalized FPKM) for all samples against one another. Samples clustering indicate the consistency between the biological replicates of the *Chromera* infection (samples I1, I2, I3) and negative control conditions (samples C1, C2, C3) at the 4h time point. A color field that ranges from green to red presents the level of correlation. **91**

Figure 3.5 A multidimensional scaling (MDS) plot produced by edgeR showing the relationship between all the replicates of *Chromera* infection (samples I1, I2, I3) and control conditions (samples C1, C2, C3) at the 48-h time point. The distances shown on the plot are the coefficient of variation of expression between samples for the 500 genes that best distinguish the samples. **92**

Figure 3.6 Venn diagram showing the number of differentially expressed genes (adjusted $P \leq 0.05$) in response to *Chromera* infection at the 48-h time point based on edgeR (yellow) and DESeq (blue) analyses. **93**

Figure 3.7 Changes in the *Acropora digitifera* transcriptome during *Chromera* infection. The pie charts show percentages of transcripts up- (blue) and down-regulated (red). 16% of the coral transcripts were differentially expressed at 48 h (Part C) (adjusted $P \leq 0.05$), while very small differences were observed at 4 h post-infection (Part A) and no changes were detected at the 12- h time point (Part B). **93**

Figure 3.8 MA and Volcano plots for the gene expression profiles between *Chromera*-infected and control samples at 4 h (Part A) and 48 h (Part B). The red dots represent clusters differentially expressed at adjusted $P \leq 0.05$. **94**

Figure 3.9 Differential gene expression profiles at 4 h post-*Chromera* infection. The heat map shows the expression profiles of 48 DEGs in *Chromera*-infected versus control samples. The hierarchical clustering obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against controls at 4 h post-infection. Expression values are \log_2 -transformed and then median-centered by transcript. Relative expression levels are shown in red (up) and green (down). Samples I1, I2, I3 are the biological replicates of *Chromera* infection while samples C1, C2, C3 are the biological replicates of the control condition. 95

Figure 3.10 Differential gene expression profiles at 48 h post-*Chromera* infection. The heat map shows the expression profiles of the 1086 highly differentially expressed genes (FDR < 0.001, 4-fold) in *Chromera*-infected versus control samples. The hierarchical clustering obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against the control at 4 h post-infection. Expression values are \log_2 -transformed and then median-centered by transcript. Relative expression levels are shown in red (up) and green (down). Samples I1, I2, I3 are the biological replicates of *Chromera* infection while samples C1, C2, C3 are the biological replicates of the control condition. 96

Figure 3.11 Venn diagram showing the overlap among DEGs with Swiss-Prot database annotation in *Chromera* and *Symbiodinium* infection compared to control at the 4 h time point. 98

Figure 3.12 An interactive model of genes and pathways discussed in the text during interaction between larvae of *Acropora digitifera* and *Chromera*. Up- and down-regulated genes/functions are in blue and red text respectively. The initial contact phase (left panel) involves up-regulation of ribosome and mitochondria functions (based on GO enrichment), suppression of host immune response including down regulation of the pattern recognition receptors (PRRs) “that can recognize signature microbial compounds microbe-associated molecular patterns (MAMPs)”, the detected PRRs included TLRs, toll-like receptors; Nod, nucleotide-binding oligomerization domain protein; SRs, scavenger receptors; lectins; and the complement protein (C3). Moreover, the pancreatic secretory granule membrane major glycoprotein GP2, which serves as an uptake receptor for pathogenic bacteria in humans was highly down regulated. Suppression of PRRs- MAMPs interaction leads to inactivation of nuclear factor-B (NF-kappa-B), which is master regulator of immunity. The phagosome maturation phase (central panel) involves down-regulation of genes involved in phagocytosis and actin remodelling. As well as differential expression of genes implicated in endosomal trafficking that enhances the maturation of the phagosome and lysosome fusion (see Figure 4 for more details about those genes). The *Chromera* tolerance phase (right panel) involves complex changes in the apoptotic network. Genes implicated both anti-and pro-apoptotic functions were differentially expressed. 106

Figure 3.13 Heat map of differentially expressed genes likely involved in immunity, inflammatory and stress responses in *Chromera*-infected (I1, I2, I3) and control larvae (C1, C2, C3) at the 48 h time point. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against the control at 48h post-infection. Expression values were \log_2 -transformed and median-centered by transcript. 111

The blue scale represents the relative expression values.

Figure 3.14 The coral host modulates the endocytic pathway and enhances the phagosome maturation and lysosome fusion in response to *Chromera*. Upon phagocytosis, the phagosome acquires the Rab5 GTPase via fusion with early endosomes. During *Chromera* infection, the Rab5 effector ALS2 was up-regulated. Rab5 acts to recruit phosphatidylinositol-3-kinase (PI3K) to generate phosphatidylinositol-3-phosphate (PI3P) and recruit the early endosomal antigen (EEA1) from endosomes. EEA1 is a Rab5 effector protein and was up-regulated to trigger fusion of phagosome with late endosome. During the phagosomal maturation process, Rab7 replaces Rab5 and the intermediate phagosome fuses with late endosomal vesicles and acquires proteins such as the proton-ATPase pump (V-ATPase), lysosome-associated membrane glycoprotein 1 (LAMP1), CD63, and lysosomal hydrolases. Vacuoles containing *Chromera* fused with late endosomes/lysosomes as indicated by the up-regulation of genes encoding Rab7, LAMP1, and CD63 (plus late endosomal/ lysosomal adaptor, SNAP-associated protein, vesicle-associated membrane protein 7 (VAMP7); not shown). Moreover, the lysosome V-ATPase was up-regulated, indicating that lysosome acidification was activated in order to kill *Chromera* (the phagolysosome is rich in hydrolytic enzymes and has a very low pH). Genes highlighted in red color are down-regulated, while those in blue are up-regulated. **113**

Figure 3.15 Heat map of differentially expressed genes involved in phagosome maturation in *Chromera*-infected (I1, I2, I3) and control larvae (C1, C2, C3) at the 48-h time point. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against the control at 48 h post-infection. Expression values were \log_2 -transformed and median-centered by transcript. The blue scale represents the relative expression values. **114**

Figure 3.16 Heat map of differentially expressed genes involved in pro apoptosis in *Chromera*-infected (I1, I2, I3) and control larvae (C1, C2, C3) at the 48-h time point. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against the control at 48 h post-infection. Expression values were \log_2 -transformed and median-centered by transcript. The blue scale represents the relative expression values. **118**

Figure 3.17 Heat map of differentially expressed genes involved in anti apoptosis in *Chromera*-infected (I1, I2, I3) and control larvae (C1, C2, C3) at the 48-h time point. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against the control at 48h post infection. Expression values were \log_2 -transformed and median-centered by transcript. The blue scale represents the relative expression values. **120**

Figure 4.1 *Chromera* transcriptome construction workflow. The schematic flowchart shows the three-steps process of raw data processing, *de novo* transcriptome assembly and functional annotation. **140**

Figure 4.2 Distribution of contig length (≥ 500 nt) of GBR *Chromera de novo* transcriptome assembly. **144**

Figure 4.3 BLASTX top-hit species distribution of gene annotations with high homologies to species with known protein sequences in the Swiss-Prot database with E-value cut off of $1.0E^{-3}$. **145**

Figure 4.4 Gene Ontology (GO) assignment (2nd level GO terms) of transcripts from the GBR *Chromera* strain. Biological processes (A) constituted that majority of GO assignment of contigs (22,205 counts, 58.02%), followed by cellular components (C) (11,149 counts, 29.1%) and molecular function (B) (4,917 counts, 12.8%). **147**

Figure 4.5 Main KEGG pathway category representation and percentages in the case **148**

of the GBR <i>Chromera</i> strain. Numbers above the bars give percent of annotated sequences in each category.	
Figure 4.6 Distribution of KEGG pathways in transcriptome of the GBR <i>Chromera</i> strain. The charts show the percentages of the sequences assigned with each category.	149
Figure 4.7 Strategy for identification of orthologous gene pairs between Sydney and GBR <i>Chromera</i> strains. The best reciprocal blast hit method was used to identify putative orthologs. BLASTX with a threshold of 10^{-15} against the Swiss-Prot database was used to filter paralogs.	152
Figure 4.8 Distribution of nonsynonymous (Ka) and synonymous (Ks) substitution rates in 321 Sydney and GBR <i>Chromera</i> orthologous pairs. The threshold of Ka/Ks < 1 indicates negative selection at P-value ≤ 0.05 . The analysis was performed using the KaKs calculator software (Zhang <i>et al.</i> 2006).	154
Figure 4.9 Ka/Ks distribution in ortholog pairs between the Sydney and GBR <i>Chromera</i> strains. The frequency distribution of Ka/Ks rates is shown on the x-axis while the KaKs values are shown on the y-axis.	154
Figure 4.10 Venn diagram showing overlap between Sydney/ GBR <i>Chromera</i> , <i>Symbiodinium</i> and <i>Plasmodium</i> Swiss-Prot annotated genes at BLASTX E-value $\leq 10^{-15}$. Note the number of non-redundant hits for each species is shown in parentheses.	156
Figure 4.11 Genes-to-terms heat map of the significant FAC “plastid” showing the plastid-related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.	158
Figure 4.12 Genes-to-terms heat map of the significant FAC “transport” showing the membrane genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively (previous page).	159
Figure 4.13 Genes-to-terms heat map of the significant FAC “cellular response to stress” showing the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.	160
Figure 4.14 Genes-to-terms heat map of the significant FAC “sugar transport” showing the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.	161
Figure 4.15 Genes-to-terms heat map of the significant FAC “plastid” showing the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.	163
Figure 4.16 Genes-to-terms heat map of the significant FAC “organelle envelope” showing the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.	164
Figure 4.17 Genes-to-terms heat map of the significant FAC “phosphatidylinositol-4-phosphate 5-kinase” showing the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.	164
Figure 4.18 KEGG pathway Proteasome “ath03050” enriched in the list of 240 genes shared between <i>Chromera</i> and <i>Plasmodium</i> . The red stars represent similar components found in both datasets.	165
Figure 4.19 Genes-to-terms heat map of the significant FAC “proteasome” showing the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.	166
Figure 4.20 Genes-to-terms heat map of the significant FAC “proteolysis the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.	167
Figure 4.21 Overall distribution of the main KEGG categories in Sydney/ GBR	169

Chromera, *Symbiodinium* and *Plasmodium*. The pie charts show the percentages of the sequences assigned with the six categories; metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems and human diseases.

Figure 4.22 Distribution of the KEGG pathways in Sydney/ GBR *Chromera*, *Symbiodinium* and *Plasmodium*. The charts show the percentages of the sequences assigned with each category. **170**

Figure 4.23 Genes mapped to the KEGG pathway N-Glycan biosynthesis “ko00510” in Sydney, GBR *Chromera*, *Symbiodinium* and *Plasmodium*. **173**

Figure 5.1 Diagram showing phagosome maturation and its arrest in coral larvae during interaction with a parasite (*Chromera*; right side) and symbiont (competent *Symbiodinium*; left side). Initially a stage of recognition occurs between the host and invading microbes and immediately after phagocytosis the host endocytic pathway is responsible for either killing invading microbes or protecting beneficial ones from host degradation enzymes. In case of *Chromera*, coral larvae showed increase in expression of genes implicated in phagosome maturation; late phagosome formation (specially RAB7) and phago-lysosome fusion in order to kill *Chromera*. On the other hand, in case of competent *Symbiodinium* the phagosome maturation is arrested primarily via increasing of expression of RAB5-activating genes, and consequently preventing the presence of RAB7. **192**

List of Tables

Table 1.1 Available genome and transcriptome data for reef-building corals in chronological order	16
Table 1.2 Available genome and transcriptome data for <i>Symbiodinium</i> and chromerids (<i>Chromera/Vitrella</i>)	17
Table 2.1 Summary of published data on gene expression analysis of coral larvae during onset and establishment of coral- <i>Symbiodinium</i> symbiosis.	31
Table 2.2 Raw Illumina Hi-Seq sequencing reads. 16 cDNA libraries were sequenced and produced a total of 333 millions of reads (<i>Symbiodinium</i> infected and control) in 3 time points; 4, 12, and 48 h post infection. NA = no data and the absence of data at 12h are attributed to RNA quality issues. The low number of reads in negative control 1 at 48 h is due to a low RNA yield.	39
Table 2.3 Percent of reads successfully mapped onto the <i>A. digitifera</i> transcriptome assembly.	39
Table 2.4 Summary of the differential gene expression profiling at 4 h post- <i>Symbiodinium</i> infection. An adjusted $P \leq 0.05$ and E-value of $\leq 10^{-15}$ were used to filter differentially expressed genes (DEGs) in BLASTX searches against the Swiss-Prot database.	45
Table 2.5 Enriched Gene Ontology (GO) categories (GO-MF, CC) at adjusted $P \leq 0.05$ in the set of down regulated genes.	47
Table 2.6 Differential expression of <i>A. digitifera</i> DEGs (n=19) (FDR ≤ 0.05) likely involved in voltage gated (ion) channel activity and trans-membrane transporters.	48
Table 2.7 Down regulation of <i>A. digitifera</i> DEGs (n=25) (FDR ≤ 0.05) likely involved in protein synthesis and translation.	49
Table 2.8 Down regulation of <i>A. digitifera</i> DEGs (n=19) (FDR ≤ 0.05) likely involved in mitochondria functions. IV and I refer to complexes of the electron transport chain.	50
Table 2.9 Down regulation of <i>A. digitifera</i> DEGs (n=8) (FDR ≤ 0.05) likely involved in Endoplasmic Reticulum functions.	51
Table 2.10 Differential expression of <i>A. digitifera</i> DEGs (n=17) (FDR ≤ 0.05) likely involved in pattern recognition and cell adhesion. Up-regulated (n=10) and down-regulated (n=7) genes are shaded blue and red respectively. Data for these genes are shown in Figure 2.10.	53
Table 2.11 Up-regulation of <i>A. digitifera</i> DEGs (n=6) (FDR ≤ 0.05) likely involved in vesicular trafficking and symbiosome formation. Data for these genes are also shown in Figure 2.10 and in the central panel of Figure 2.13 (symbiosis establishment).	55
Table 2.12 Differential expression of <i>A. digitifera</i> DEGs (n=21) (FDR ≤ 0.05) likely involved in transcription regulatory machinery. Up- and down-regulated genes are shaded blue and red respectively.	57
Table 2.13 Up regulation of <i>A. digitifera</i> DEGs (n=10) (FDR ≤ 0.05) likely involved in cell cycle. Data for these genes are also shown in Figure 2.11 and in the right panel of Figure 2.13.	58
Table 2.14 Differential expression of <i>A. digitifera</i> DEGs (n=5) (FDR ≤ 0.05) likely involved in immune response. Data for these genes are also shown in Figure 2.12 and in the left panel of Figure 2.13.	60
Table 2.15 Differential expression of <i>A. digitifera</i> DEGs (n=11) (FDR ≤ 0.05) likely involved in regulation of Apoptosis. Up- and down-regulated genes are shaded blue and red respectively. Data for these genes are also shown in Figure 2.12 and in the right panel of Figure 2.13.	61
Table 2.16 Differential expression of <i>A. digitifera</i> DEGs (n=13) (FDR ≤ 0.05) likely involved in responses to stress. Up- and down-regulated genes are shaded blue and red respectively. Data for these genes are also shown in Figure 2.12 and in the right panel	62

of Figure 2.13.

Table 3.1 Raw Illumina Hi-Seq sequencing reads. 17 cDNA libraries were sequenced and produced a total about 346.2 millions reads (*Chromera*-infected and uninfected larvae) in 3 time points; 4, 12 and 48 h post infection. NA= no data. The absence of data the 12 h for negative control1 is due to RNA quality issues and the low number of reads in negative control 1, *Chromera*-infected 2 and 3 are consequences of low RNA yield. **88**

Table 3.2 Percent of illumina reads successfully mapped onto the *A. digitifera* assembly. **89**

Table 3.3 Summary of differential gene expression profiles in *Chromera*-infected compared to control larvae at 4, 12, and 48 h. An adjusted $P \leq 0.05$ and E-value cut off $\leq 10^{-10}$ were used to filter differentially expressed genes and for BLASTX searches against the Swiss-Prot database. **94**

Table 3.4 Down regulated *A. digitifera* transcripts differentially expressed in *Chromera* infected-larvae at the 4 h time point with corrected $P \leq 0.05$. Columns correspond to coral cluster name, best BLASTX result, E-value and the \log_2 fold-change values. **97**

Table 3.5 Down regulation of *A. digitifera* DEGs at adjusted $P \leq 0.05$ common to *Chromera* and *Symbiodinium* infections at 4 h time point. **98**

Table 3.6 Significant KEGG pathway enrichments among the set of down regulated genes in *Chromera*-infected larvae at 48 h post infection with corrected $P \leq 0.05$ after multiple testing correction by the Benjamini procedure. **99**

Table 3.7 Biological process (BP) categories significantly enriched among the set of down regulated genes in *Chromera* infection compared to control at 48 h post infection with corrected $P \leq 0.05$. **100**

Table 3.8 Cellular component (CC) categories significantly enriched among the set of down regulated genes in *Chromera* infection compared to control at 48 h post infection with $P \leq 0.05$. **101**

Table 3.9 Molecular function (MF) categories significantly enriched among the set of down regulated genes in *Chromera* infection compared to control at 48 h post infection with corrected $P \leq 0.05$ **102**

Table 3.10 Biological processes (BP) categories significantly enriched among the set of up regulated genes in *Chromera* infection compared to control at 48 h post infection with corrected $P \leq 0.05$. **103**

Table 3.11 Cellular component (CC) categories significantly enriched among the set of up regulated genes in *Chromera* infection compared to control at 48 h post infection with corrected $P \leq 0.05$. **104**

Table 3.12 Molecular function (MF) categories significantly enriched among the set of up regulated genes in *Chromera* infection compared to control at 48 h post infection with corrected $P \leq 0.05$. **105**

Table 3.13 Down regulated *A. digitifera* transcripts likely involved in suppression of the host immune response in *Chromera*-infected larvae at 48 h post infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the \log_2 fold change values. **109**

Table 3.14 Differential expression of *A. digitifera* clusters likely involved in phagocytosis in *Chromera*-infected larvae at 48 h post infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the \log_2 fold change values. **112**

Table 3.15 Differential expression of *A. digitifera* transcripts likely involved in early and/or late endosome formation and phagosomal maturation in *Chromera*-infected larvae at 48 h post infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the \log_2 fold change values. **115**

Table 3.16 Differential expression of *A. digitifera* transcripts likely involved in **116**

autophagy and lysosomal functions in <i>Chromera</i> -infected larvae at 48 h post infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the \log_2 fold change values.	
Table 3.17 Differential expression of <i>A. digitifera</i> clusters likely have pro-apoptotic functions in <i>Chromera</i> -infected larvae at 48 h post infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the \log_2 fold change values.	119
Table 3.18 Differential expression of <i>A. digitifera</i> transcripts likely have anti-apoptotic functions in <i>Chromera</i> -infected larvae at 48 h post infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the \log_2 fold change values.	121
Table 4.1 Raw and processed Illumina reads after quality filtering and trimming.	140
Table 4.2 Summary statistics of <i>Chromera de novo</i> transcriptome assembly using Trinity software.	144
Table 4.3 Summary of KEGG orthology data for the GBR <i>Chromera</i> strain.	150
Table 4.4 Assessing the read content of the transcriptome using the percentages of the mapped reads.	151
Table 4.5 Selected KEGG pathways/protein complexes identified in the GBR <i>Chromera</i> transcriptome.	151
Table 4.6 GO categories enriched among the <i>Chromera</i> ortholog pairs (Benjamini-corrected P -value ≤ 0.05).	153
Table 4.7 Candidate gene under positive selection in <i>Chromera</i> .	155
Table 4.8 Functional annotation of <i>Chromera</i> ortholog pair under positive selection.	155
Table 4.9 Genes mapped to the KEGG pathway N-Glycan biosynthesis “ko00510” in Sydney, GBR <i>Chromera</i> , <i>Symbiodinium</i> and <i>Plasmodium</i> .	172
Table 4.10 Genes mapped to the KEGG pathway RNA polymerase “ko03020” in Sydney, GBR <i>Chromera</i> , <i>Symbiodinium</i> and <i>Plasmodium</i> .	175
Table 4.11 Genes mapped to the KEGG pathway basal transcription factors “ko03022” in Sydney, GBR <i>Chromera</i> , <i>Symbiodinium</i> and <i>Plasmodium</i> .	176
Table 4.12 Genes mapped to the KEGG pathway spliceosome “ko03040” in Sydney, GBR <i>Chromera</i> , <i>Symbiodinium</i> and <i>Plasmodium</i> .	177

Thesis Abstract

Reef-building corals are considered as meta-organisms where the coral animal lives in symbiosis with a wide array of microorganisms. While mutualistic association between corals and *Symbiodinium* is crucial for the functioning and success of the coral reef ecosystems, surprisingly little is currently known about its molecular basis and this is especially true of the events leading to establishment of the relationship. A morphologically similar alga to *Symbiodinium* was discovered in Australian corals and has been identified as *Chromera*. The discovery of *Chromera* is very significant as it holds a unique position in evolution, between the photosynthetic dinoflagellates and the parasitic apicomplexans. The nature of the association between *Chromera* and corals is currently unclear. In this thesis, I used high throughput next generation sequencing technology (Illumina RNA-Seq) to explore the molecular mechanisms underlying establishment of coral-algal symbiosis between coral larvae and a competent strain of *Symbiodinium*. I examined also the nature of the poorly understood relationship between corals and the newly described photosynthetic apicomplexan alga *Chromera* using RNA-Seq. Finally, I present a functional genomic resource (transcriptome) for a *Chromera* strain isolated from a Great Barrier Reef coral, and use a comparative transcriptomic approach to examine sharing of functions and pathways among *Chromera*, *Symbiodinium kawagutii* and *Plasmodium falciparum*.

To better understand the molecular mechanisms underlying the initial coral-*Symbiodinium* interactions, *Acropora digitifera* larvae were inoculated with a competent *Symbiodinium* strain and the responses of the coral whole transcriptome were investigated 4, 12 and 48 h post-*Symbiodinium* infection using RNA-Seq. Although previous studies (based on use of cDNA microarrays) did not detect host signals during establishment of coral-*Symbiodinium* symbiosis, using the RNA-Seq approach, transient changes in gene expression, involving 1073 differentially expressed genes (DEGs), were observed early in the *Symbiodinium* uptake (infection) process. This is the first report of differential expression of a significant number of genes during *Symbiodinium* uptake by corals. The list of DEGs allowed the construction of a model for the molecular mechanisms that operate during onset and establishment of coral-*Symbiodinium* symbiosis, including suppression of host immunity, protein synthesis and

oxidative metabolism. More importantly, the data provided support for the formation of the symbiosome as an arrested early phagosome, a mechanism thought also to apply to the process by which *Symbiodinium* colonises some sea anemones.

To determine the nature of the relationship between corals and *Chromera* “the closest relative to apicomplexan parasites”, *A. digitifera* larvae were inoculated with *Chromera* CCMP2878 strain and the coral whole transcriptome responses were investigated at 4, 12 and 48 h post-*Chromera* infection using RNA-Seq. Stress, disease and immune challenge (in corals) have distinct transcriptomic signatures as does the process of infection by a competent *Symbiodinium* strain. Analysis of the transcriptomic impact of *Chromera* infection shed some light on the nature of the coral-*Chromera* association and provided novel insights into host-parasite/pathogen interactions. Based on the transcriptomic data, I suggest that the coral-*Chromera* relationship may be parasitic, thus the assumption that *Chromera* is a coral symbiont requires re-evaluation.

In order to provide a functional genomic resource for a chromerid alga and explore its gene catalogue, a *de novo* transcriptome assembly was generated for a *Chromera* strain isolated from *Montipora digitata* on the GBR and the obtained contigs were annotated. This novel dataset was compared with coding sequence data for another *Chromera* strain (isolated from different host and geographic location) and 664 orthologous gene pairs were identified. The overwhelming majority of these orthologs were under purifying selection, only one pair being under positive selection; this gene encoded a homolog of the human tetratricopeptide TTC21B. Overall KEGG pathway distributions were very similar between *Chromera* and *Symbiodinium* the largest proportion of genes in both cases being assigned to metabolism. Comparing KEGG pathways involved in glycan biosynthesis and transcription machinery, revealed the genetic uniqueness of the symbiotic dinoflagellate *Symbiodinium*.

In conclusion, coral-algal symbioses are the basis for coral reef ecosystems thus understanding these relationships at a molecular level is very important especially for reef management and fighting against coral bleaching. The work presented in this thesis provides novel insights into the molecular events occurring during onset of coral-*Symbiodinium* symbiosis that enabled better mechanistic understanding of algal symbioses in corals. Knowledge derived from the thesis contributes to better

understanding of the symbiont infection process and that will help in coral reef management especially when engineering coral symbioses towards increased coral thermotolerance/resilience and better understanding how symbiosis breakdown (coral bleaching) occurs, thus understanding the mechanisms of coral symbiosis is a step forward in order to combat coral bleaching. In addition, the thesis showed that coral responses to *Chromera* have similarities to the responses of vertebrates to parasites and provided insights into host-pathogen/parasite interactions that will enhance our understanding how host cells defend themselves against infectious organisms. Moreover, the thesis provided a genomic resource for a *Chromera* strain that can be used as a reference for large-scale gene expression and comparative analyses to better understand the biology of these newly discovered algae and suggested the potential use of *Chromera* as a model organism in developing anti-malarial drugs.

Chapter 1.0 Background and general introduction

Coral reefs are often referred to as the rain forests of the ocean since they possess highly productive marine communities with rich biodiversity (Richmond 1993). They are estimated to harbor approximately one-third of all described marine organisms, and their productivity supports approximately one quarter of marine fisheries. Thus, in addition to their significant ecological value, coral reefs have great economic value (Moberg & Folke 1999). In the past decades, coral reefs have been deteriorating because of both natural and anthropogenic factors (Harvell *et al.* 2004). These factors include: mass coral bleaching events (Hoegh-Guldberg 1999), increased prevalence of coral diseases (Willis *et al.* 2004) increased levels of pollution (McCulloch *et al.* 2003), and over-exploitation of commercial marine species (Pandolfi *et al.* 2003). The past ten years have been the warmest on record globally since the mass coral bleaching events in 1998 and 2002, and consequently corals have been under continuous stress (Hoegh-Guldberg 1999). More recently the worst bleaching event on record hit corals in 2016 where (for example) in the Great Barrier Reef (GBR) 93% of the reefs have been affected by bleaching as a result of increasing the SST by 1.5° C to 2° C above the long-term average of the period 1971-2000 (NOAA 2016). Whether corals will be able to adapt fast enough to that rapid rate of temperature increase currently being experienced is an important and open question.

It is crucial for effective coral reef conservation and management that we understand the molecular mechanisms underlying the biology of reef organisms including reef-building corals. For example, many possible causes of coral bleaching have been proposed, including exposure to high temperature, heavy metals, pollution, pathogenic bacteria and darkness (Douglas 2003). In that case with the absence of a clear cause, marine conservation plans cannot be initiated. Molecular techniques can provide diagnostic tests to identify candidate molecules in response to stress such as alterations in gene expression in response to bleaching, diseases or other environmental/anthropogenic stressors (Forêt *et al.* 2007). This information could really help reef managers to identify corals under stressors well in advance before any morphological signs of stress and/or disease. Thus, genomic approaches hold great promise to improve our understanding of basic biological processes in corals (including health, symbioses and relationships with pathogenic and/or symbiotic microorganisms) as well as understanding responses to environmental stress, bleaching and disease.

1.1 Background on coral biology and the coral holobiont

Hermatypic or reef-building corals (Phylum Cnidaria) have essential roles in shallow tropical seas particularly due to their ability to deposit calcium carbonate as an external skeleton, thus providing the foundation of the extensive shallow-water coral reefs found in the tropics. Reef-building corals are colonial animals consisting of what are assumed to be genetically identical polyps. The coral polyps depend largely on energy produced by microalgae of the genus *Symbiodinium* (comprised of many diverse clades/lineages) that inhabit coral tissues in a mutualistic relationship. In addition to *Symbiodinium*, diverse array of microbes are associated with corals (Mouchka *et al.* 2010; Thurber *et al.* 2009), some of which are thought to be host specific (Rohwer *et al.* 2002). These microbial communities inhabit various microhabitats within the coral (Ainsworth *et al.* 2010). Recent advances in molecular microbiology have enabled researchers to begin to characterise the coral microbiota, which encompass viruses (Claverie *et al.* 2009), fungi (Kirkwood *et al.* 2010), bacteria (Kimes *et al.* 2010; Rosenberg *et al.* 2007), Archaea (Rohwer & Kelley 2004) and protozoans (Toller *et al.* 2002). These associations are thought to play important, but as yet poorly understood, roles in the overall fitness of the coral colony. Corals harbor diverse and abundant types of bacteria and Archaea in specific compartments such as tissue, surface mucus and

skeleton (Bourne *et al.* 2009; Rohwer *et al.* 2002; Rosenberg *et al.* 2007). Those bacterial/archeal assemblages may benefit corals by producing antimicrobial compounds (Nissimov *et al.* 2009; Ritchie 2006) and cycling nutrients (Beman *et al.* 2007; Lesser *et al.* 2007; Wegley *et al.* 2007). However, some of these coral-associated bacteria can be harmful to the coral such as pathogenic and/or opportunistic bacteria that cause disease (Ainsworth *et al.* 2010; Harvell *et al.* 2007; Rosenberg *et al.* 2007). A collective term for the coral animal host, its symbiotic dinoflagellates and these diverse microbial associates is ‘the coral holobiont’ (Knowlton & Rohwer 2003). Members of the coral holobiont interact with one another and with the environment to maintain the complex biological system (Figure 1.1). The holobiont perspective had led to a holistic view on studying coral biology instead of focusing on each component (coral, algal symbionts, and bacteria) separately.

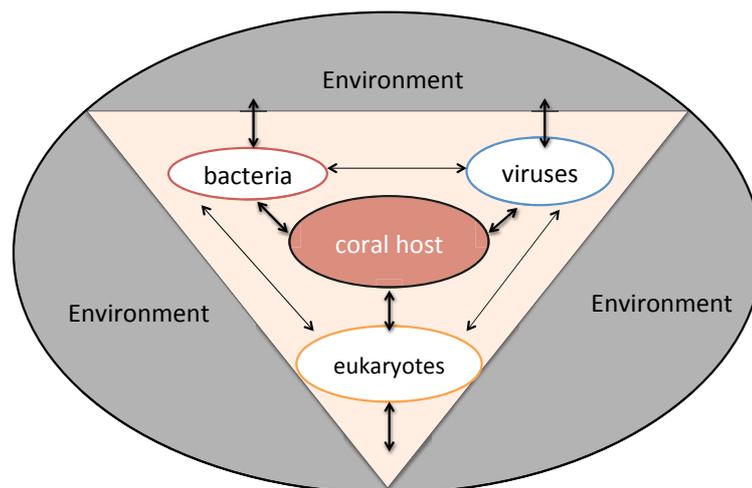


Figure 1.1 Coral as a meta-organism (holobiont) consisting of the coral animal and associated microorganisms (bacteria, viruses and microbial eukaryotes including algal symbionts), modified from Boschand McFall-Ngai (2011).

1.2 *The coral-Symbiodinium symbiosis and its significance*

The high productivity and extensive accretion of skeletal CaCO_3 by shallow-water tropical reef ecosystems is attributed to the symbiotic relationship of hermatypic corals with an alveolate dinoflagellates of the genus *Symbiodinium* (Grigg 1995). Using molecular genetics techniques, it became obvious that the genus *Symbiodinium* is much

more diverse than initially thought; see review by Baker (2003). Multiple molecular markers from the *Symbiodinium* nucleus and chloroplast, including the small/large subunits of the rDNA (18 and 28S regions), internal transcribed spacer regions 1 and 2 (ITS1/2), *psbA* gene and the mitochondrial *cox1* gene revealed that the genus *Symbiodinium* consists of nine distinct and divergent lineages (clades A-I), with multiple types within each clade (Lesser *et al.* 2013; Pochon & Gates 2010; Pochon *et al.* 2006). This diversity can be explained by the large number of taxa that live in symbiosis with *Symbiodinium* including members of four different invertebrate phyla (Cnidaria, Mollusca and Porifera) and the single-celled Foraminifera (Stat *et al.* 2006). Despite the huge genetic diversity, little is known about ecological and/or functional diversity. The functional diversity of different coral-*Symbiodinium* associations is a barely studied area that needs more consideration, and this is especially true when understanding coral reef resilience and predicting the fate of corals under the threat of global warming. Some of the genetic diversity appeared to reflect in functional diversity such as (for example) in coral thermal tolerance as clade D *Symbiodinium* was shown to be dominant in the coral host during bleaching (Jones & Berkelmans 2011). In addition, clade A *Symbiodinium* was shown experimentally to be functionally less beneficial to corals than clade C *Symbiodinium* (Stat *et al.* 2008), this led the authors to conclude that clade A *Symbiodinium* may be parasitic symbionts rather than mutualistic.

This relationship with *Symbiodinium* is very common among reef invertebrates, particularly cnidarians. The majority of symbiotic cnidarians acquire *Symbiodinium* from the surrounding environment via horizontal transmission (Richmond & Hunter 1990). Host endodermal cells take up the symbionts by phagocytosis. Whereas this process normally leads to digestion of these invaders (parasites), in the case of compatible symbionts, these are then harbored in intracellular host-derived vacuoles known as “symbiosomes” (Colley & Trench 1983). Symbiosome membranes are thought to originate from the plasma membrane of host cells during acquisition of symbionts via phagocytosis (Hohman *et al.* 1982); at some point the phagosome membrane (containing the symbiont) transforms into the symbiosome membrane, which protects the symbiont from host lysosomal degradation and provides an interface through which nutrient transport can occur (Davy *et al.* 2012). Studies on the *Aiptasia pulchella*-*Symbiodinium* symbiosis indicated that sorting of membrane proteins might be involved in the process of symbiosome membrane formation as a late endosome

marker (membrane protein) was sorted away from the developed symbiosome (Chen *et al.* 2003). More recently, Peng *et al.* (2010) investigated the protein components of pure and intact symbiosomes of *A. pulchella* and identified 17 proteins implicated in cellular functions such as cell recognition, cytoskeletal remodeling, transport, stress responses and programmed cell death (apoptosis).

Once the symbiosis is established, the photosynthetic algae contribute significantly to host nutrition via organic carbon translocation, which can account for 90% of the host's daily energy requirements. In return, the algae are thought to benefit by having access to the relatively high nutrient levels of host tissue in addition to a high light environment optimal for photosynthesis (Muscatine 1990). However, recent work has demonstrated that carbon fixation capacity and the amount of carbon translocated are highly variable among *Symbiodinium* clades (Leal *et al.* 2015).

1.3 Coral-associated apicomplexan-related lineages (ARLs)

A recent meta-analysis by Janouskovec *et al.* (2012) has implicated several lineages of apicomplexans as specifically associated with corals (Figure 1.2). A variety of sequences previously misidentified as of bacterial origin through 16S rDNA profiling have recently been shown to originate from eight distinct novel apicomplexan-related lineages (ARLs), some of which appear to be tightly associated with coral tissues. Interestingly, two ARLs include the only known photosynthetic chromerids, *Chromera velia* (Moore *et al.* 2008) and the more recently described *Vitrella brassicaformis* (Obornik *et al.* 2012) which fall into the ARL-III and ARL-I clades respectively. The ARL-I and -III clades are conspicuous because of their intermediate position between parasitic apicomplexans and the dinoflagellates (lineages known to include free-living photosynthetic species), and their apparent association with corals. Consequently, whether ARL-I and ARL-III clades are intracellular endosymbionts, parasites or surface contaminants (that associate with the surface during sampling) of corals are an important and open questions that merit further investigation. It was determined that 90% of the examined *Orbicella annularis* colonies harbored apicomplexans, based on DNA-based molecular Markers, and these symbionts/associates were named "genotype N" (Toller *et al.* 2002). The unidentified coral associate falls into the ARL-VIII clade. This group of organisms is particularly interesting because it belongs to the

1.4 *Coral-Chromerida associations*

Although the association between the photosynthetic dinoflagellate *Symbiodinium* and corals has been known for many years, recent work has established that a number of other related alveolates are also intimately associated with corals (Janouskovec *et al.* 2012). The alveolates are a major lineage of protists that are defined by the possession of subsurface flattened vesicles supported by microtubules called “alveoli”, as well as the presence of micropores and mitochondria with tubular cristae (Adl *et al.* 2005). Recent taxonomic schemes continue to retain the alveolate group as either phylum Alveolata or infrakingdom Alveolata (Adl *et al.* 2005). The recently identified structural protein alveolin is a molecular synapomorphy of Alveolata and is associated with the alveoli (Gould *et al.* 2008).

Chromerida is a newly defined phylum of photoautotrophic alveolate protists that has been isolated from corals in Australia (Cumbo *et al.* 2013; Janouskovec *et al.* 2012; Moore *et al.* 2008) and the Caribbean (Visser *et al.* 2012). This phylum currently includes only two species *Chromera* (Moore *et al.* 2008) and *Vitrella brassicaformis* (Obornik *et al.* 2012). Hence, the alveolates comprise four diverse phyla of primarily single-celled eukaryotes: ciliates, dinoflagellates, apicomplexans (Leander & Keeling 2003) and chromerids (Moore *et al.* 2008). Phylum Chromerida is closely related to the apicomplexan parasites (Figure 1.3), but also related to the photosynthetic dinoflagellates including the coral symbiont *Symbiodinium*. This new phylum of algae is very important as it holds a unique position in evolution, between the photosynthetic dinoflagellates and the parasitic apicomplexans, but very little information regarding their biology, ecology and genetics is currently available. Moreover, the nature of their associations with corals and their impact on fitness are important open questions.

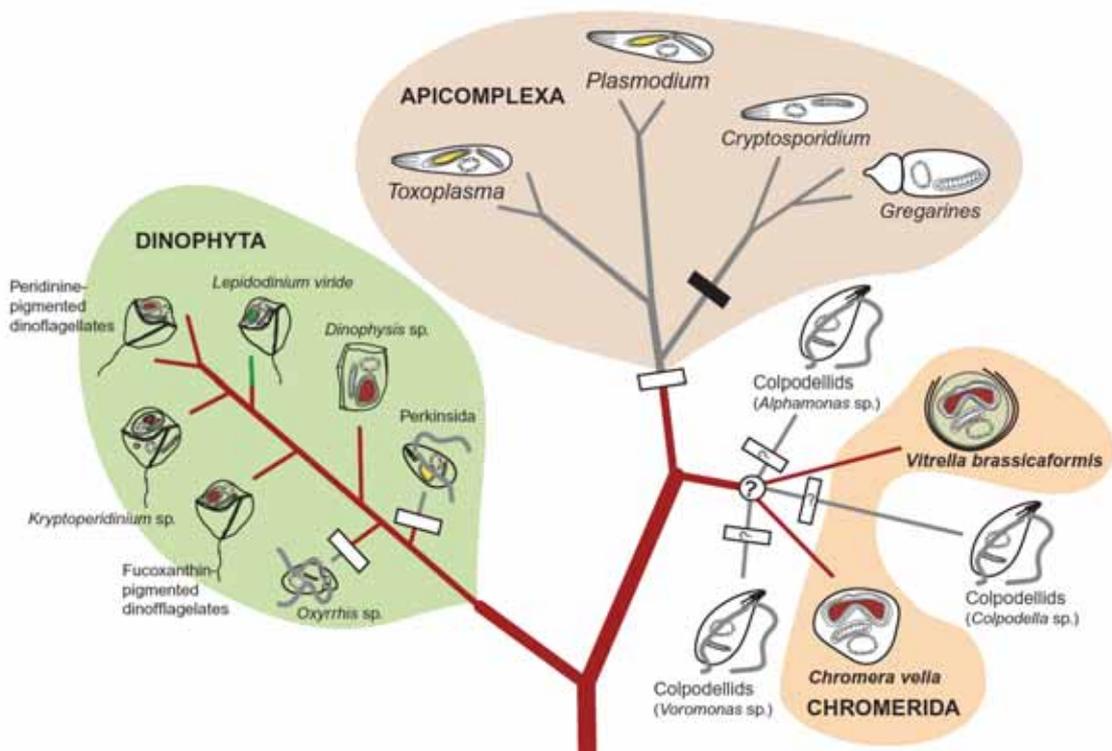


Figure 1.3 Phylogenetic relationships among chromerids (*Chromera* and *Vitrella*), parasitic apicomplexans and photosynthetic dinoflagellates (Obornik & Lukes 2013). The three families (Dinophyta, Apicomplexa and Chromerida) are highlighted with three different background colours. White rectangles indicate the loss of photosynthesis; black rectangle indicates the loss of plastid.

1.5 Coral-Chromera symbiosis

The coral holobiont is a nutrient-rich system involving complex interactions between the predominant coral host and symbiotic zooxanthellae partners together with a poorly defined milieu of associated microorganisms such as bacteria, viruses and eukaryotes, including *Chromera* (Bourne *et al.* 2009; Janouskovec *et al.* 2012). *Chromera* occupies a significant phylogenetic position as an evolutionary link between dinoflagellates and apicomplexans. It is a sister species to the Apicomplexa and is the closest photosynthetic relative of these obligate parasites (Moore *et al.*, 2008). *Chromera* was isolated for the first time from the scleractinian corals *Plesiastrea versipora* and *Leptastrea purpurea* (Faviidae) from Sydney Harbor and One Tree Island in Australia (Moore *et al.* 2008). More recently, Cumbo *et al.* (2013) isolated *Chromera* from the scleractinian coral *Montipora digitata* (Acroporidae) from Nelly Bay, Magnetic Island in the central inner region of the Great Barrier Reef (GBR). Hence *Chromera* was isolated from Australian corals over a wide geographic range (Figure

1.4). Moreover, *Chromera* was also isolated from the Caribbean coral *Agaricia agaricites* and its presence was confirmed by sequencing of the small subunit rRNA (Visser *et al.* 2012), thus the coral-*Chromera* association might be widespread.

Using 454-amplicon pyrosequencing of eukaryotic small-subunit rDNA, Slapeta and Linares (2013) detected the presence of apicomplexan type-N and *Chromera* sequences in corals of the southern GBR. The apicomplexan type-N sequences were detected in *Acropora palifera*, *Montipora digitata*, *Porites cylindrica* (Heron Island) and *Seriatopora hystrix* (One Tree Island). On the other hand, *Chromera* was detected only in *M. digitata*. In this context, in an attempt to study the biogeographic distribution of *Chromera* in corals of the GBR using PCR/sequencing assays, *Chromera*-specific PCR primers were designed and tested in multiple corals from Heron, Orpheus and Magnetic Islands. *Chromera* was successfully detected using PCR in *M. digitata* (Magnetic Island) and the soft coral *Lobophytum pauciflorum* (Orpheus Island) (more information are provided in Appendix II).

Visser *et al.* (2012) found that *Chromera* cultures had higher growth rates compared to *Symbiodinium* cultures when exposed to increased temperatures (up to 35 °C). These results led the authors to conclude that *Chromera* might have an advantage over *Symbiodinium* during periods of increased seawater temperatures that might result in a shift in the relative abundances of *Chromera* and *Symbiodinium* in the coral holobiont if *Chromera* is an endosymbiont in corals. More recently, infection and stable association of *Chromera* with larvae of *A. digitifera* and *A. tenuis* has been experimentally demonstrated (Cumbo *et al.* 2013) where it was located in both ectoderm and endoderm.

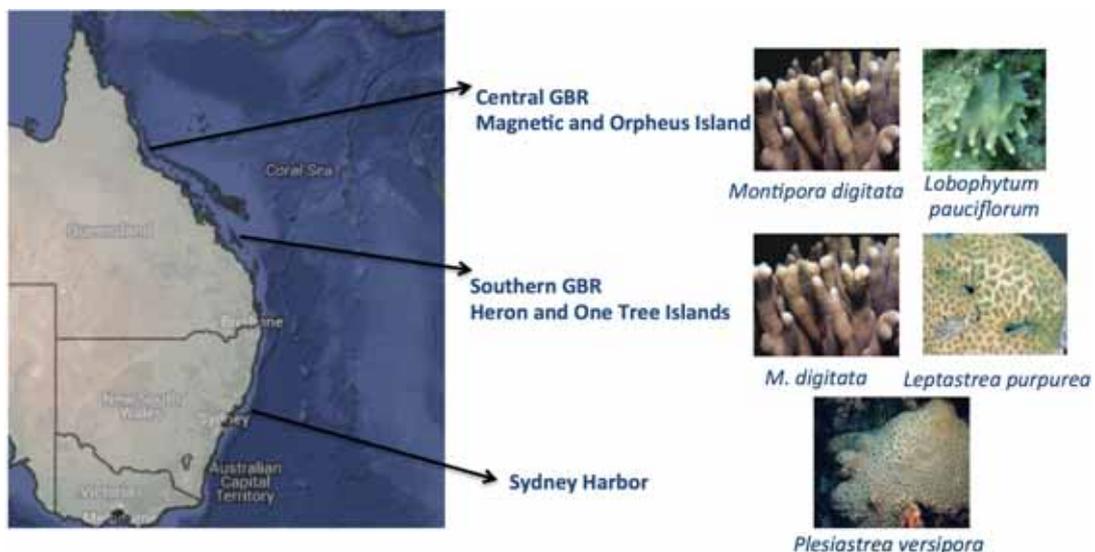


Figure 1.4 Location map, showing sites and hard coral species that host *Chromera* in Australia including corals on the great barrier reef (GBR). Coral photos from websites; <http://bie.ala.org.au/> and <http://coral.aims.gov.au/info/factsheets.jsp>.

Chromera undergoes a diurnal transformation between motile and immotile forms (Figures 1.5 and 1.6), similar to that seen in the abundant coral symbiont, *Symbiodinium* (Guo *et al.* 2010; Obornik *et al.* 2011). When *Chromera* was initially isolated from corals it was in the immotile stage, oval in shape, and brownish-green in color, and superficially resembled the coral symbiont *Symbiodinium* (Moore *et al.* 2008). However, despite the morphological similarity between *Chromera* and symbiotic dinoflagellates, phylogenetic analyses based on several nuclear and plastid genes provide strong support for a close relationship with the Apicomplexa (Figure 1.7) (Janouskovec *et al.* 2010; Moore *et al.* 2008). Compared to the photosynthetic dinoflagellates, *Chromera* has a relatively simple pigment composition, lacking chlorophyll c and possessing only chlorophyll a and three carotenoids (Moore *et al.* 2008). Sutak *et al.* (2010) found that one strain/isolate of *Chromera* in culture has a high iron requirement, although its natural environment is poor in iron. The photosynthetic system of *Chromera* is highly efficient and adaptable to a wide range of light conditions owing to its specialized photo-acclimation strategies (Quigg *et al.* 2012). *Chromera* possesses a prominent subcellular structure known as the chromerosome, which was originally misidentified as a mitochondrion (Figure 1.6; part C) (Obornik *et al.* 2011). The function of this organelle is unknown, however, based on the similarity of the chromerosome with the trichocysts of dinoflagellates, Obornik *et al.* (2011) speculated that it could be used for hunting algae or other prey.

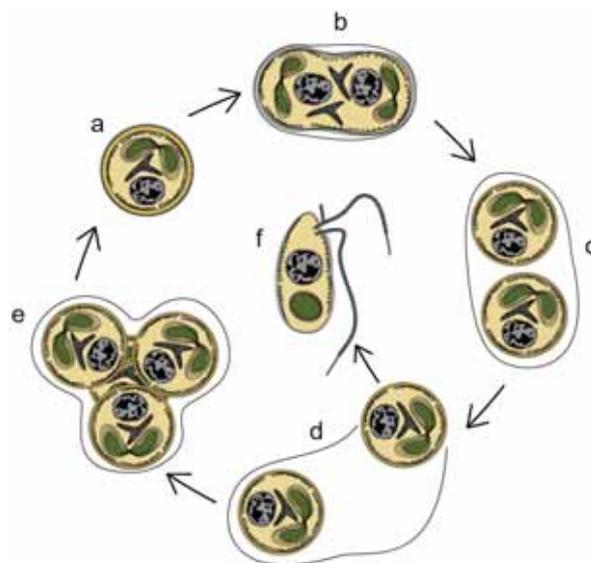


Figure 1.5 *Chromera* life cycle: a) predominant **immotile** coccoid cells; b, c) cells in division state; d, e) autospores released to start a new cycle; f) bi-flagellated **motile** *Chromera* cells, which occur during favorable conditions (Obornik & Lukes 2013).

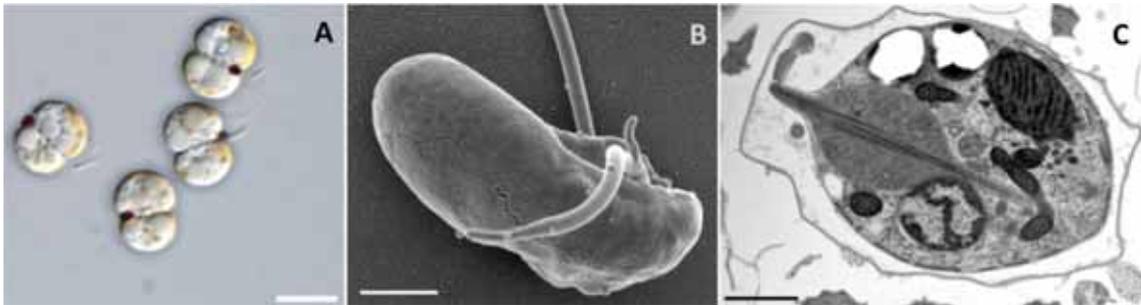


Figure 1.6 Light and electron micrographs of *Chromera* (Obornik *et al.* 2011). Part A shows coccoid *Chromera* cells during cell division (scale bar = 5 μm). Part B shows a scanning electron micrograph the bi-flagellated “motile” form of *Chromera*, with two heterodynamic flagella. The shorter flagellum contains a typical finger-like projection (scale bar = 1 μm). Part C shows a transmission electron micrograph of a section of *Chromera*, showing the unique organelle chromerosome, which is a projection-like organelle with a central bundle of fibers whose function is unknown (scale bar = 1 μm).

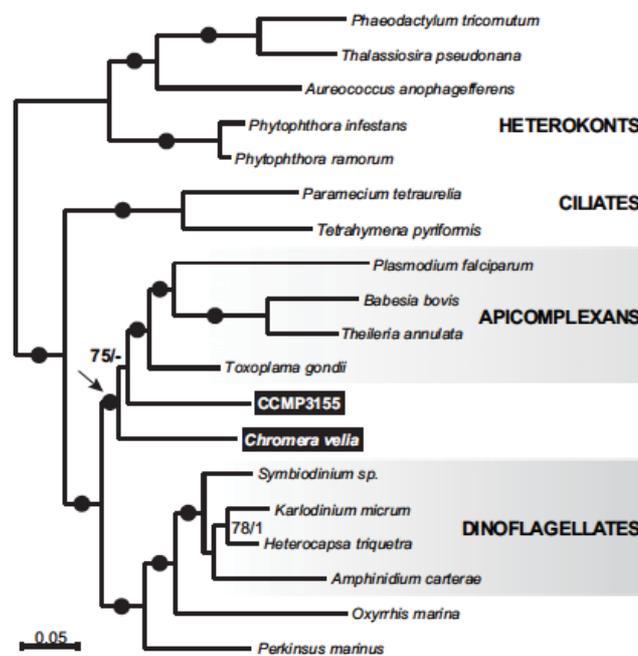


Figure 1.7 Phylogenetic tree based on nuclear and plastid gene sequences, showing relationships within the alveolates. Note that *Chromera* and *Vitrella* (CCMP3155) are two independent photosynthetic lineages that are more closely related to apicomplexan parasites than to the dinoflagellates (Janouskovec *et al.* 2010).

In summary, *Chromera* associates with different hard and soft corals (and potentially also with other marine invertebrates) at widely spaced locations on the east coast of Australia (see, Figure 1.4) as well as in the Caribbean. Hence, the association

between corals and *Chromera* appears to be quite common and widespread, and it may be a cosmopolitan member of coral holobionts. However, the nature of the relationship of *Chromera* with corals remains to be determined; *Chromera* can establish stable relationships with coral larvae and is widely associated with adult corals, but its close relationship to apicomplexans suggests that it might be a facultative coral parasite rather than a true symbiont.

1.6 *Chromera as a proxy for Plasmodium*

Chromera is essentially a missing link in the evolution of parasitism within the Alveolata. The new alga that represents a common origin for apicoplasts (a relict, non-photosynthetic plastid found in parasitic Apicomplexa) and plastids (in dinoflagellates) and provides an intermediate state that is promising to decipher the difference between parasitic apicomplexa and symbiotic dinoflagellates (Moore *et al.* 2008). Apicomplexan parasites are a massive global health problem – for example, about 500 million people suffer from malaria annually, and infection with mosquito-borne *Plasmodium* is particularly common in Africa (Sachs & Malaney 2002). Working with apicomplexan parasites is particularly challenging due to their obligatory parasitism, which complicates culture under laboratory conditions. By contrast, *Chromera* can be cultured autotrophically in simple inorganic growth media (Guo *et al.* 2010; Moore *et al.* 2008; Obornik *et al.* 2011). Taking into account its close relationship with the Apicomplexa, *Chromera* could therefore be a useful laboratory model organism for anti-malarial drug development. Consequently, studying genes involved in signaling and metabolic pathways in *Chromera* might pave the way towards screening methods for novel therapeutics intended for human healthcare.

1.7 *Coral reef ecological genomics*

Advances in the fields of genomics and bioinformatics have generated new tools for studying the biology of reef-corals at the cellular and molecular levels. For example, high-throughput methods for examining gene expression such as DNA microarray and next generation sequencing (NGS) (Miller *et al.* 2011) allow biologists to experimentally test genome-wide gene expression levels. Adopting such an approach allows the identification of differentially expressed genes and enriched gene categories

and/ or pathways that operate in particular conditions compared to their corresponding controls. While these methods were initially developed for biomedical research, they can now be applied to ecological and environmental questions. The field of ecological genomics is rapidly expanding, from characterization of single genes (Miller *et al.* 2000) using expression sequence tags (ESTs) as developed for the Indo-Pacific coral, *Acropora millepora* with more than 3,000 ESTs (Kortschak *et al.* 2003). Later on, several transcriptomic studies were conducted and EST data sets became available for both corals and *Symbiodinium* spp. (Tables 1.1, 1.2). These approaches enabled complex questions to be addressed, such as (for example) how corals respond at the molecular level to different stressors (Dupont *et al.* 2007) and how *Symbiodinium* is able to establish symbioses with corals (Schnitzler & Weis 2010; Schwarz *et al.* 2008a; Shinzato *et al.* 2014b; Voolstra *et al.* 2009a).

The first draft genome for a coral was decoded in 2011 for *Acropora digitifera* (Shinzato *et al.* 2011), and in 2013 a draft genome of *Symbiodinium minutum* was decoded (Shoguchi *et al.* 2013). In 2015, draft genome sequences of *Symbiodinium kawaguti* (Lin *et al.* 2015) and the chromerids *Chromera* and *Vitrella* became available (Woo *et al.* 2015). In addition, transcriptomes for several different corals, *Symbiodinium* and *Chromera* strains are available (Tables 1.1, 1.2). Moreover, the Reef Future Genomics 2020 (ReFuGe 2020) Consortium plan to sequence hologenomes (genome, transcriptome, meta-genome/transcriptome) for ten coral species having different physiological capabilities that will provide enormous genomic recourses for coral reef studies utilizing the “omics” approaches to explore the resilience of coral holobionts to environmental stressors and understand their adaptability (Voolstra *et al.* 2015).

Table 1.1 Available genome and transcriptome data for reef-building corals in chronological order

Assembly	Species	Sequencing technology	Reference
Genome	<i>Acropora digitifera</i>	454, Illumina	(Shinzato <i>et al.</i> 2011)
Transcriptome	<i>Orbicella faveolata</i>	Sanger	(Schwarz <i>et al.</i> 2008b)
	<i>Acropora palmata</i>	Sanger, 454	(Polato <i>et al.</i> 2011)
	<i>Pocillopora damicornis</i>	454	(Traylor-Knowles <i>et al.</i> 2011)
	<i>Acropora millepora</i>	454, Illumina	(Moya <i>et al.</i> 2012)
	<i>Acropora cervicornis</i>	Illumina	(Libro <i>et al.</i> 2013)
	<i>Acropora hyacinthus</i>	Illumina	(Barshis <i>et al.</i> 2014)
	<i>Porites australiensis</i>	Illumina	(Shinzato <i>et al.</i> 2014b)
	<i>Stylophora pistillata</i>	454	(Karako-Lampert <i>et al.</i> 2014)
	<i>Orbicella faveolata</i>	Illumina	(Pinzon <i>et al.</i> 2015)
	<i>Orbicella faveolata</i>	Illumina	(Anderson <i>et al.</i> 2016)

Table 1.2 Available genome and transcriptome data for *Symbiodinium* and chromerids (*Chromera/Vitrella*)

Assembly	Species (strain)	Clade	Host	Sequencing technology	Reference
Genome	<i>Symbiodinium kawagutii</i>	F	<i>Montipora verrucosa</i>	Illumina	(Lin <i>et al.</i> 2015)
	<i>Symbiodinium minutum</i> (Mf 1.05b.01)	B1	<i>Orbicella faveolata</i>	454, Illumina	(Shoguchi <i>et al.</i> 2013)
	<i>Chromera velia</i> (CCMP2878)	na	<i>Plesiastrea versipora</i>	Illumina	(Woo <i>et al.</i> 2015)
	<i>Vitrella brassicaformis</i> (CCMP3155)	na	<i>Leptastrea purpurea</i>	Illumina	(Woo <i>et al.</i> 2015)
Transcriptome	<i>Symbiodinium</i> sp.	C3	<i>Acropora aspera</i>	Sanger	(Leggat <i>et al.</i> 2007)
	<i>Symbiodinium</i> sp.	A	<i>Aiptasia pallida</i>	Sanger	(Sunagawa <i>et al.</i> 2009)
	<i>Symbiodinium</i> sp. (CassKB8)	A	<i>Cassiopea</i> sp.	Sanger	(Voolstra <i>et al.</i> 2009b)
	<i>Symbiodinium</i> spp.	C	<i>Acropora hyacinthus</i>	Illumina	(Ladner <i>et al.</i> 2012)
	<i>Symbiodinium</i> spp.	D	<i>Acropora hyacinthus</i>	Illumina	(Ladner <i>et al.</i> 2012)
	<i>Symbiodinium</i> sp. (Mf1.05b)	B1	<i>Montastraea faveolata</i>	454	(Bayer <i>et al.</i> 2012)
	<i>Symbiodinium</i> sp. (CassKB8)	A	<i>Cassiopea</i> sp.	454	(Bayer <i>et al.</i> 2012)
	<i>Symbiodinium minutum</i> (Mf 1.05b.01)	B1	<i>Orbicella faveolata</i>	Illumina	(Shoguchi <i>et al.</i> 2013)
	<i>Symbiodinium microadriaticum</i> (CCMP2467)	A1	<i>Stylophora pistillata</i>	Illumina	(Baumgarten <i>et al.</i> 2013)
	<i>Symbiodinium kawagutii</i> (CCMP2468)	F1	<i>Montipora verrucosa</i>	Illumina	(Xiang <i>et al.</i> 2015)
	<i>Symbiodinium minutum</i> (Mf 1.05b.01)	B1	<i>Orbicella faveolata</i>	Illumina	(Shoguchi <i>et al.</i> 2013)
	<i>Symbiodinium</i> sp.	C3K	<i>Acropora hyacinthus</i>	Illumina	(Barshis <i>et al.</i> 2014)
	<i>Symbiodinium</i> sp.	D2	<i>Acropora hyacinthus</i>	Illumina	(Barshis <i>et al.</i> 2014)
	<i>Symbiodinium</i> sp.	C15	<i>Porites australiensis</i>	Illumina	(Shinzato <i>et al.</i> 2014a)
	<i>Chromera velia</i> (CCMP2878)	na	<i>Plesiastrea versipora</i>	Illumina	(Woo <i>et al.</i> 2015)
	<i>Vitrella brassicaformis</i> (CCMP3155)		<i>Leptastrea purpurea</i>	Illumina	(Woo <i>et al.</i> 2015)
	<i>Symbiodinium</i> (aims-aten-C1-WSY)	C1	<i>Acropora tenuis</i>	Illumina	(Levin <i>et al.</i> 2016)
	<i>Symbiodinium</i> (aims-aten-C1-MI)	C1	<i>Acropora tenuis</i>	Illumina	(Levin <i>et al.</i> 2016)
	<i>Chromera velia</i> (Mdig03)	na	<i>Montipora digitata</i>	Illumina	Mohamed <i>et al.</i> (Unpublished)

1.8 RNA-Seq approach to study coral reef biology

Transcriptomic studies of corals in the last decade have been largely based on cDNA microarray technology. However, microarrays have many drawbacks as they are limited to detect transcripts that correspond to existing sequences spotted onto the chip. In many cases, those transcripts might be dominated by housekeeping genes, which are expressed in all cells under normal and abnormal cellular conditions. Microarray-based analyses have been utilized to study different aspects of coral reef biology at the molecular level including (for example) development, symbiosis, thermal stress and bleaching (Bellantuono *et al.* 2012; Grasso *et al.* 2008; Schwarz *et al.* 2008a; Schwarz *et al.* 2008c). With the advent of next generation sequencing technologies, unbiased and quantitative means of exploring the transcriptomes for non-model organisms have become available. Specifically, the use of Illumina RNA-Seq technology to profile changes in gene expression has enabled unbiased large-scale analyses in many non-model organisms including corals. RNA-Seq approach is ideal for discovery-based experiments as novel sequences are obtained. In addition to the ability to quantify large ranges of expression levels across the whole genome with absolute values without the need for a reference genome, a *de novo* transcriptome can be assembled for the differential gene expression analyses.

RNA-Seq based approach has been successfully to study corals at the transcriptional level. In *Acropora millepora*, RNA-seq was used to explore the coral responses to various stimuli or treatments, including elevated CO₂ (Moya *et al.* 2012; Moya *et al.* 2015) and immune stimuli (Weiss *et al.* 2013). The same technology has recently been applied to investigate the responses of corals to, for example, various diseases (Daniels *et al.* 2015; Libro *et al.* 2013; Wright *et al.* 2015), thermal stress and bleaching (Pinzon *et al.* 2015; Seneca & Palumbi 2015). The RNA-Seq approach is really powerful by comparison with array technology; mRNA from coral samples is converted to complementary DNA (cDNA) and then is subjected to illumina sequencing. The inclusion of internal standards makes possible quantification of gene expression by simply counting the sequence reads corresponding to each transcript in the reference genome and/or transcriptome. Consequently, RNA-Seq provides an efficient way for unbiased large-scale gene expression analyses in non-model organisms including corals. This technology can be applied to gain insight into the coral

transcriptomic response to *Symbiodinium* and *Chromera* infections.

1.9 Aims, objectives and thesis structure

The primary aim of this study was to investigate the poorly understood relationship between corals and *Chromera*, a photosynthetic relative of apicomplexans, using transcriptomic approaches. In order to set a base line to what are the coral responses during symbiosis, infection experiment was conducted using competent strain of *Symbiodinium*. This knowledge was then used to compare with the coral responses to *Chromera*; in order to check whether *Chromera* fits into the same model or it is more parasitic and that would help to determine the nature of the coral interaction with *Chromera*.

To accomplish this global aim, the work was broken down into a number of specific project objectives as follows:

- 1) **To better understand the molecular events underlying the establishment of coral-*Symbiodinium* symbiosis.** A time-course infection experiment was carried out, whereby larvae of the coral *A. digitifera* were infected with a competent strain of *Symbiodinium* (clade B1). Illumina RNA-Seq was then used to follow transcriptional changes and profile gene expression landscapes in the coral at 4, 12, and 48 h post-infection. Whereas previous studies, based on the use of microarrays, implied that the host transcriptome is largely unchanged during establishment of symbiosis with a competent strain of *Symbiodinium*, a transient and early response was detected in the present study. The results of this experiment, described in Chapter 2, provide novel insights into the molecular events underlying the establishment of the coral-algal symbiosis, and an integrative model of the process is presented.

- 2) **To understand the impact of *Chromera* infection on coral larvae.** As in the case of the *Symbiodinium* infection experiment (aim #1, above), larvae of the coral *A. digitifera* were infected with *Chromera* (CCMP2878 strain). Illumina RNA-Seq was then used to follow the transcriptomic

changes in the coral host during infection of *A. digitifera* with *Chromera* at 4, 12 and 48 h post-infection. In order to allow valid comparisons with the *Symbiodinium* infection experiment described in Chapter 2, the *Chromera* infection experiment was conducted under the same experimental conditions. Transcriptome profiling revealed a massive transcriptional response of the coral host to *Chromera* infection, which was very different to that of the same coral host to *Symbiodinium*. The results presented in this Chapter provide novel insights into the relationship between *Chromera* and corals, and suggest that this organism may be parasitic on, rather than symbiotic with, corals.

- 3) **To clarify the nature of *Chromera* being a symbiont or a parasite,** a *de novo* transcriptome assembly was generated for a *Chromera* strain (Mdig3 strain) isolated from *Montipora digitata* on the GBR. RNA-Seq data derived from *Chromera* grown under a variety of conditions, was assembled using trinity and the *de novo* transcriptome compared with the coding sequences (CDS) for the *Chromera* strain isolated from *Plesiastrea versipora* from Sydney harbor, the coral symbiont *Symbiodinium kawaguti* and the malaria parasite *Plasmodium falciparum*. This approach led to the identification of genes and pathways shared between *Chromera* and its symbiotic and parasitic relatives.

The above aims are addressed in data Chapters 2, 3 and 4.

1.10 References

- Adl SM, Simpson AG, Farmer MA, *et al.* (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology* **52**, 399-451.
- Ainsworth TD, Thurber RV, Gates RD (2010) The future of coral reefs: a microbial perspective. *Trends in Ecology & Evolution* **25**, 233-240.
- Anderson DA, Walz ME, Weil E, Tonellato P, Smith MC (2016) RNA-Seq of the Caribbean reef-building coral *Orbicella faveolata* (Scleractinia-Merulinidae) under bleaching and disease stress expands models of coral innate immunity. *PeerJ* **4**, e1616.
- Baker AC (2003) Flexibility and specificity in coral-algal symbiosis: Diversity, ecology, and biogeography of Symbiodinium. *Annual Review of Ecology Evolution and Systematics* **34**, 661-689.
- Barshis DJ, Ladner JT, Oliver TA, Palumbi SR (2014) Lineage-Specific Transcriptional Profiles of Symbiodinium spp. Unaltered by Heat Stress in a Coral Host. *Molecular Biology and Evolution* **31**, 1343-1352.
- Baumgarten S, Bayer T, Aranda M, *et al.* (2013) Integrating microRNA and mRNA expression profiling in Symbiodinium microadriaticum, a dinoflagellate symbiont of reef-building corals. *Bmc Genomics* **14**.
- Bayer T, Aranda M, Sunagawa S, *et al.* (2012) Symbiodinium Transcriptomes: Genome Insights into the Dinoflagellate Symbionts of Reef-Building Corals. *Plos One* **7**.
- Bellantuono AJ, Granados-Cifuentes C, Miller DJ, Hoegh-Guldberg O, Rodriguez-Lanetty M (2012) Coral Thermal Tolerance: Tuning Gene Expression to Resist Thermal Stress. *Plos One* **7**, e50685.
- Beman JM, Roberts KJ, Wegley L, Rohwer F, Francis CA (2007) Distribution and diversity of archaeal ammonia monooxygenase genes associated with corals. *Applied and Environmental Microbiology* **73**, 5642-5647.
- Bosch TCG, McFall-Ngai MJ (2011) Metaorganisms as the new frontier. *Zoology* **114**, 185-190.
- Bourne DG, Garren M, Work TM, *et al.* (2009) Microbial disease and the coral holobiont. *Trends in Microbiology* **17**, 554-562.
- Chen CA, Lam KK, Nakano Y, Tsai WS (2003) A stable association of the stress-tolerant zooxanthellae, Symbiodinium clade D, with the low-temperature-tolerant coral, *Oulastrea crispata* (Scleractinia : Faviidae) in subtropical non-reefal coral communities. *Zoological Studies* **42**, 540-550.
- Claverie JM, Grzela R, Lartigue A, *et al.* (2009) Mimivirus and Mimiviridae: Giant viruses with an increasing number of potential hosts, including corals and sponges. *Journal of Invertebrate Pathology* **101**, 172-180.
- Colley NJ, Trench RK (1983) Selectivity in phagocytosis and persistence of symbiotic algae in the scyphistoma stage of the jellyfish *Cassiopeia xamachana*. *Proc R Soc Lond B Biol Sci* **219**, 61-82.
- Cumbo VR, Baird AH, Moore RB, *et al.* (2013) *Chromera velia* is Endosymbiotic in Larvae of the Reef Corals *Acropora digitifera* and *A. tenuis*. *Protist* **164**, 237-244.
- Daniels C, Baumgarten S, Yum LK, *et al.* (2015) Metatranscriptome analysis of the reef-building coral *Orbicella faveolata* indicates holobiont response to coral disease. *Frontiers in Marine Science* **2**.
- Davy SK, Allemand D, Weis VM (2012) Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiol Mol Biol Rev* **76**, 229-261.

- Douglas AE (2003) Coral bleaching—how and why? *Marine Pollution Bulletin* **46**, 385-392.
- Dupont S, Wilson K, Obst M, *et al.* (2007) Marine ecological genomics: when genomics meets marine ecology. *Marine Ecology Progress Series* **332**, 257-273.
- Forêt S, Kassahn KS, Grasso LC, *et al.* (2007) Genomic and microarray approaches to coral reef conservation biology. *Coral Reefs* **26**, 475.
- Gould SB, Tham WH, Cowman AF, McFadden GI, Waller RF (2008) Alveolins, a new family of cortical proteins that define the protist infrakingdom Alveolata. *Molecular Biology and Evolution* **25**, 1219-1230.
- Grasso LC, Maindonald J, Rudd S, *et al.* (2008) Microarray analysis identifies candidate genes for key roles in coral development. *Bmc Genomics* **9**, 540.
- Grigg RW (1995) Corals in-Space and Time - the Biogeography and Evolution of the Scleractinia - Veron, Jen. *Science* **269**, 1893-1894.
- Guo JT, Weatherby K, Carter D, Slapeta J (2010) Effect of Nutrient Concentration and Salinity on Immotile-Motile Transformation of *Chromera velia*. *Journal of Eukaryotic Microbiology* **57**, 444-446.
- Harvell D, Aronson R, Baron N, *et al.* (2004) The rising tide of ocean diseases: unsolved problems and research priorities. *Frontiers in Ecology and the Environment* **2**, 375-382.
- Harvell D, Jordan-Dahlgren E, Merkel S, *et al.* (2007) Coral Disease, Environmental Drivers, and the Balance between Coral and Microbial Associates. *Oceanography* **20**, 172-195.
- Hoegh-Guldberg O (1999) Climate change, coral bleaching and the future of the world's coral reefs. *Marine and Freshwater Research* **50**, 839-866.
- Hohman TC, McNeil PL, Muscatine L (1982) Phagosome-lysosome fusion inhibited by algal symbionts of *Hydra viridis*. *J Cell Biol* **94**, 56-63.
- Janouskovec J, Horak A, Barott KL, Rohwer FL, Keeling PJ (2012) Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. *Current Biology* **22**, R518-R519.
- Janouskovec J, Horak A, Obornik M, Lukes J, Keeling PJ (2010) A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci U S A* **107**, 10949-10954.
- Jones AM, Berkelmans R (2011) Tradeoffs to Thermal Acclimation: Energetics and Reproduction of a Reef Coral with Heat Tolerant Symbiodinium Type-D.
- Karako-Lampert S, Zoccola D, Salmon-Divon M, *et al.* (2014) Transcriptome analysis of the scleractinian coral *Stylophora pistillata*. *Plos One* **9**, e88615.
- Kimes NE, Van Nostrand JD, Weil E, Zhou JZ, Morris PJ (2010) Microbial functional structure of *Montastraea faveolata*, an important Caribbean reef-building coral, differs between healthy and yellow-band diseased colonies. *Environmental Microbiology* **12**, 541-556.
- Kirkwood M, Todd JD, Rypien KL, Johnston AWB (2010) The opportunistic coral pathogen *Aspergillus sydowii* contains dddP and makes dimethyl sulfide from dimethylsulfoniopropionate. *Isme Journal* **4**, 147-150.
- Knowlton N, Rohwer F (2003) Multispecies microbial mutualisms on coral reefs: The host as a habitat. *American Naturalist* **162**, S51-S62.
- Kortschak RD, Samuel G, Saint R, Miller DJ (2003) EST analysis of the cnidarian *Acropora millepora* reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Current Biology* **13**, 2190-2195.

- Ladner JT, Barshis DJ, Palumbi SR (2012) Protein evolution in two co-occurring types of Symbiodinium: an exploration into the genetic basis of thermal tolerance in Symbiodinium clade D. *Bmc Evolutionary Biology* **12**.
- Leal MC, Hoadley K, Pettay DT, *et al.* (2015) Symbiont type influences trophic plasticity of a model cnidarian-dinoflagellate symbiosis. *Journal of Experimental Biology* **218**, 858-863.
- Leander BS, Keeling PJ (2003) Morphostasis in alveolate evolution. *Trends in Ecology & Evolution* **18**, 395-402.
- Leggat W, Hoegh-Guldberg O, Dove S, Yellowlees D (2007) Analysis of an EST library from the dinoflagellate (Symbiodinium sp.) symbiont of reef-building corals. *Journal of Phycology* **43**, 1010-1021.
- Lesser MP, Bythell JC, Gates RD, Johnstone RW, Hoegh-Guldberg O (2007) Are infectious diseases really killing corals? Alternative interpretations of the experimental and ecological data. *Journal of Experimental Marine Biology and Ecology* **346**, 36-44.
- Lesser MP, Stat M, Gates RD (2013) The endosymbiotic dinoflagellates (Symbiodinium sp.) of corals are parasites and mutualists. *Coral Reefs* **32**, 603-611.
- Levin RA, Beltran VH, Hill R, *et al.* (2016) Sex, Scavengers, and Chaperones: Transcriptome Secrets of Divergent Symbiodinium Thermal Tolerances. *Molecular Biology and Evolution* **33**, 3032.
- Libro S, Kaluziak ST, Vollmer SV (2013) RNA-seq Profiles of Immune Related Genes in the Staghorn Coral *Acropora cervicornis* Infected with White Band Disease. *Plos One* **8**.
- Lin S, Cheng S, Song B, *et al.* (2015) The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* **350**, 691-694.
- McCulloch M, Fallon S, Wyndham T, *et al.* (2003) Coral record of increased sediment flux to the inner Great Barrier Reef since European settlement. *Nature* **421**, 727-730.
- Miller DJ, Ball EE, Forêt S, Satoh N (2011) Coral genomics and transcriptomics — Ushering in a new era in coral biology. *Journal of Experimental Marine Biology and Ecology* **408**, 114-119.
- Miller DJ, Hayward DC, Reece-Hoyes JS, *et al.* (2000) Pax gene diversity in the basal cnidarian *Acropora millepora* (Cnidaria, Anthozoa): implications for the evolution of the Pax gene family. *Proc Natl Acad Sci U S A* **97**, 4475-4480.
- Moberg F, Folke C (1999) Ecological goods and services of coral reef ecosystems. *Ecological Economics* **29**, 215-233.
- Moore RB, Obornik M, Janouskovec J, *et al.* (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**, 959-963.
- Mouchka ME, Hewson I, Harvell CD (2010) Coral-Associated Bacterial Assemblages: Current Knowledge and the Potential for Climate-Driven Impacts. *Integrative and Comparative Biology* **50**, 662-674.
- Moya A, Huisman L, Ball EE, *et al.* (2012) Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO₂-driven acidification during the initiation of calcification. *Molecular Ecology* **21**, 2440-2454.
- Moya A, Huisman L, Foret S, *et al.* (2015) Rapid acclimation of juvenile corals to CO₂-mediated acidification by upregulation of heat shock protein and Bcl-2 genes. *Molecular Ecology* **24**, 438-452.

- Muscatine L (1990) The role of symbiotic algae in carbon and energy flux in reef corals. *Dubinsky Z (ed) Ecosystems of the world, vol. 25. Coral reefs. Elsevier, Amsterdam* **25**, 75–87.
- Nissimov J, Rosenberg E, Munn CB (2009) Antimicrobial properties of resident coral mucus bacteria of *Oculina patagonica*. *Fems Microbiology Letters* **292**, 210-215.
- Obornik M, Lukes J (2013) Cell biology of chromerids: autotrophic relatives to apicomplexan parasites. *Int Rev Cell Mol Biol* **306**, 333-369.
- Obornik M, Modry D, Lukes M, *et al.* (2012) Morphology, ultrastructure and life cycle of *Vitrella brassicaformis* n. sp., n. gen., a novel chromerid from the Great Barrier Reef. *Protist* **163**, 306-323.
- Obornik M, Vancova M, Lai DH, *et al.* (2011) Morphology and Ultrastructure of Multiple Life Cycle Stages of the Photosynthetic Relative of Apicomplexa, *Chromera velia*. *Protist* **162**, 115-130.
- Pandolfi JM, Bradbury RH, Sala E, *et al.* (2003) Global trajectories of the long-term decline of coral reef ecosystems. *Science* **301**, 955-958.
- Peng SE, Wang YB, Wang LH, *et al.* (2010) Proteomic analysis of symbiosome membranes in Cnidaria-dinoflagellate endosymbiosis. *Proteomics* **10**, 1002-1016.
- Pinzon JH, Kamel B, Burge CA, *et al.* (2015) Whole transcriptome analysis reveals changes in expression of immune-related genes during and after bleaching in a reef-building coral. *R Soc Open Sci* **2**, 140214.
- Pochon X, Gates RD (2010) A new Symbiodinium clade (Dinophyceae) from soritid foraminifera in Hawai'i. *Molecular Phylogenetics and Evolution* **56**, 492-497.
- Pochon X, Montoya-Burgos JI, Stadelmann B, Pawlowski J (2006) Molecular phylogeny, evolutionary rates, and divergence timing of the symbiotic dinoflagellate genus Symbiodinium. *Molecular Phylogenetics and Evolution* **38**, 20-30.
- Polato NR, Vera JC, Baums IB (2011) Gene discovery in the threatened elkhorn coral: 454 sequencing of the *Acropora palmata* transcriptome. *Plos One* **6**, e28634.
- Quigg A, Kotabova E, Jaresova J, *et al.* (2012) Photosynthesis in *Chromera velia* Represents a Simple System with High Efficiency. *Plos One* **7**.
- Richmond RH (1993) Coral-Reefs - Present Problems and Future Concerns Resulting from Anthropogenic Disturbance. *American Zoologist* **33**, 524-536.
- Richmond RH, Hunter CL (1990) Reproduction and Recruitment of Corals - Comparisons among the Caribbean, the Tropical Pacific, and the Red-Sea. *Marine Ecology Progress Series* **60**, 185-203.
- Ritchie KB (2006) Regulation of microbial populations by coral surface mucus and mucus-associated bacteria. *Marine Ecology Progress Series* **322**, 1-14.
- Rohwer F, Kelley S (2004) Culture-independent analyses of coral-associated microbes. *Coral Health and Disease*, 265-277.
- Rohwer F, Seguritan V, Azam F, Knowlton N (2002) Diversity and distribution of coral-associated bacteria. *Marine Ecology Progress Series* **243**, 1-10.
- Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I (2007) The role of microorganisms in coral health, disease and evolution. *Nature Reviews Microbiology* **5**, 355-362.
- Sachs J, Malaney P (2002) The economic and social burden of malaria. *Nature* **415**, 680-685.
- Schnitzler CE, Weis VM (2010) Coral larvae exhibit few measurable transcriptional changes during the onset of coral-dinoflagellate endosymbiosis. *Mar Genomics* **3**, 107-116.

- Schwarz J, Brokstein P, Manohar C, *et al.* (2008a) Coral Reef Genomics: Developing tools for functional genomics of coral symbiosis.
- Schwarz JA, Brokstein PB, Voolstra C, *et al.* (2008b) Coral life history and symbiosis: functional genomic resources for two reef building Caribbean corals, *Acropora palmata* and *Montastraea faveolata*. *Bmc Genomics* **9**, 97.
- Schwarz JA, Brokstein PB, Voolstra C, *et al.* (2008c) Coral Life History and Symbiosis: functional genomic resources for two reef building Caribbean corals, *Acropora palmata* and *Montastraea faveolata*. *Bmc Genomics* **9**.
- Seneca FO, Palumbi SR (2015) The role of transcriptome resilience in resistance of corals to bleaching. *Molecular Ecology* **24**, 1467-1484.
- Shinzato C, Inoue M, Kusakabe M (2014a) A snapshot of a coral "holobiont": a transcriptome assembly of the scleractinian coral, *Porites*, captures a wide variety of genes from both the host and symbiotic zooxanthellae. *Plos One* **9**, e85182.
- Shinzato C, Mungpakdee S, Satoh N, Shoguchi E (2014b) A genomic approach to coral-dinoflagellate symbiosis: studies of *Acropora digitifera* and *Symbiodinium minutum*. *Frontiers in Microbiology* **5**.
- Shinzato C, Shoguchi E, Kawashima T, *et al.* (2011) Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**, 320-323.
- Shoguchi E, Shinzato C, Kawashima T, *et al.* (2013) Draft Assembly of the *Symbiodinium minutum* Nuclear Genome Reveals Dinoflagellate Gene Structure. *Current Biology* **23**, 1399-1408.
- Stat M, Carter D, Hoegh-Guldberg O (2006) The evolutionary history of *Symbiodinium* and scleractinian hosts - Symbiosis, diversity, and the effect of climate change. *Perspectives in Plant Ecology Evolution and Systematics* **8**, 23-43.
- Stat M, Morris E, Gates RD (2008) Functional diversity in coral-dinoflagellate symbiosis. *Proc Natl Acad Sci U S A* **105**, 9256-9261.
- Sunagawa S, Wilson EC, Thaler M, *et al.* Generation and analysis of transcriptomic resources for a model system on the rise: the sea anemone *Aiptasia pallida* and its dinoflagellate endosymbiont.
- Sunagawa S, Wilson EC, Thaler M, *et al.* (2009) Generation and analysis of transcriptomic resources for a model system on the rise: the sea anemone *Aiptasia pallida* and its dinoflagellate endosymbiont. *Bmc Genomics* **10**, 258.
- Sutak R, Slapeta J, San Roman M, Camadro JM, Lesuisse E (2010) Nonreductive Iron Uptake Mechanism in the Marine Alveolate *Chromera velia*. *Plant Physiology* **154**, 991-1000.
- Thurber RV, Willner-Hall D, Rodriguez-Mueller B, *et al.* (2009) Metagenomic analysis of stressed coral holobionts. *Environmental Microbiology* **11**, 2148-2163.
- Toller WW, Rowan R, Knowlton N (2002) Genetic evidence for a protozoan (phylum Apicomplexa) associated with corals of the *Montastraea annularis* species complex. *Coral Reefs* **21**, 143-146.
- Traylor-Knowles N, Granger BR, Lubinski TJ, *et al.* (2011) Production of a reference transcriptome and transcriptomic database (PocilloporaBase) for the cauliflower coral, *Pocillopora damicornis*. *Bmc Genomics* **12**, 585.
- Upton SJ, Peters EC (1986) A new and unusual species of coccidium (Apicomplexa: Agamococcidiorida) from Caribbean scleractinian corals. *Journal of Invertebrate Pathology* **47**, 184-193.

- Voolstra C, Miller D, Ragan M, *et al.* (2015) The ReFuGe 2020 Consortium—using “omics” approaches to explore the adaptability and resilience of coral holobionts to environmental change. *Frontiers in Marine Science* **2**.
- Voolstra CR, Schwarz JA, Schnetzer J, *et al.* (2009a) The host transcriptome remains unaltered during the establishment of coral-algal symbioses. *Molecular Ecology* **18**, 1823-1833.
- Voolstra CR, Sunagawa S, Schwarz JA, *et al.* (2009b) Evolutionary analysis of orthologous cDNA sequences from cultured and symbiotic dinoflagellate symbionts of reef-building corals (Dinophyceae: Symbiodinium). *Comparative Biochemistry and Physiology D-Genomics & Proteomics* **4**, 67-74.
- Wegley L, Edwards R, Rodriguez-Brito B, Liu H, Rohwer F (2007) Metagenomic analysis of the microbial community associated with the coral *Porites astreoides*. *Environmental Microbiology* **9**, 2707-2719.
- Weiss Y, Foret S, Hayward DC, *et al.* (2013) The acute transcriptional response of the coral *Acropora millepora* to immune challenge: expression of GiMAP/IAN genes links the innate immune responses of corals with those of mammals and plants. *Bmc Genomics* **14**, 400.
- Willis BL, Page CA, Dinsdale EA (2004) Coral disease on the Great Barrier Reef. *Coral Health and Disease*, 69-104.
- Woo YH, Ansari H, Otto TD, *et al.* (2015) Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *Elife* **4**, e06974.
- Wright RM, Aglyamova GV, Meyer E, Matz MV (2015) Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. *Bmc Genomics* **16**, 371.
- Xiang T, Nelson W, Rodriguez J, Tolleter D, Grossman AR (2015) Symbiodinium transcriptome and global responses of cells to immediate changes in light intensity when grown under autotrophic or mixotrophic conditions. *Plant Journal* **82**, 67-80.

Chapter 2.0 Unraveling the molecular mechanisms underlying the onset and establishment of coral-*Symbiodinium* symbiosis using RNA-Seq

The content of this chapter has been published as an original article:

Mohamed, A.R., Cumbo, V., Harii, S., Shinzato, C., Chan, C.X., Ragan, M.A., Bourne, D.G., Willis, B.L., Ball, E.E., Satoh, N. and Miller, D.J. (2016), The transcriptomic response of the coral *Acropora digitifera* to a competent *Symbiodinium* strain: the symbiosome as an arrested early phagosome. *Molecular Ecology* 25, 3127-3141

The published paper is attached as Appendix I

Amin Mohamed wrote the entire chapter, with co-authors providing intellectual guidance in the design, implementation of the research and editorial contributions to the paper. Amin Mohamed conducted the experiments, collected and analysed the data, and produced all of the tables and figures.

2.1 ABSTRACT

The mutualistic relationship between reef-building corals and intracellular photosynthetic dinoflagellates of the genus *Symbiodinium* provides the foundation of the coral reef ecosystem. Disruption of this relationship leads to coral bleaching which is occurring on a large scale and poses a real threat to the survival of coral reefs. In spite of the great ecological significance of this partnership to coral reefs, little is currently known about the molecular mechanisms involved in the establishment and maintenance of the symbiosis. Previous studies have investigated host gene expression during the establishment of coral-*Symbiodinium* symbiosis, but these have failed to detect host symbiosis-related signals. To better understand the early events occurring during the establishment of symbiosis, I conducted infection experiments during spawning in Sesoko Island, Okinawa, Japan in June 2013 using *Acropora digitifera* larvae and a competent strain of symbionts (*Symbiodinium* sp clade B1). Next generation sequencing (Illumina RNA-Seq) was used for the first time to follow coral transcriptome-wide gene expression after exposure to competent *Symbiodinium* at 4, 12, 48 h post-infection. These experiments allowed the detection of transient transcriptomic signals in the coral (< 3% of the coral transcriptome) during the initial infection period. This phenomenon has not previously been detected due to the lower sensitivity of both the microarray technology used and the sampling times employed. The transcriptomic data imply that the establishment of symbiosis involves cross talk between the partners; an active response is required on the part of the host in order to recognize appropriate partners but at the same time the symbionts appear to suppress some host responses including: oxidative metabolism, protein synthesis, and immunity. The results are also consistent with the hypothesis that the symbiosome is a phagosome that has undergone early arrest, raising the possibility of a common mechanism of symbiont infection in corals and symbiotic sea anemones.

2.2 INTRODUCTION

The relationship between corals and photosynthetic dinoflagellates of the genus *Symbiodinium* (referred to as zooxanthellae) is one of the most important symbiotic relationships found in the marine environment. These partnerships provide the structural foundation of the coral reef ecosystem. The endosymbiotic algae inhabit the endodermal

cells of the host in a phagosome-derived vacuole called the “symbiosome” (Davy *et al.* 2012). Symbiosome membranes are thought to originate from the plasma membrane of coral cells during symbiont uptake via phagocytosis (Hohman *et al.* 1982); the phagosome membrane surrounding the symbiont becomes the symbiosome membrane, which confers protection from host lysosomal degradation and controls nutrient transport (Davy *et al.* 2012; Fitt & Trench 1983). The endosymbionts contribute to the nutrition of their hosts by providing up to 90% of host energy requirements in the form of fixed carbon and, in return, the symbionts receive protection and nutrients from the host (Muscatine 1990). The ability of corals to build such massive reefs, despite growing in nutrient-poor water, is attributed to the energy acquired via their symbiotic partners.

Reef-building corals are declining globally due to both natural and anthropogenic impacts including increasing sea surface temperature and pollution (De'ath *et al.* 2012). Such stresses can disrupt that important symbiotic partnership, as algae can be lost from their host through a process known as coral bleaching. There is growing concern about the consequences of the disruption of this important relationship on the future of coral reefs, as in severe cases bleaching can result in death of the coral host and destruction of the reef (Hoegh-Guldberg 1999; Weis 2008). In spite of the great ecological significance of the coral-*Symbiodinium* symbiosis to coral reefs and their survival, little is currently known about the molecular mechanisms involved in the process of onset, establishment and collapse of the symbiosis. Few experiments have investigated host gene expression during the onset of coral-*Symbiodinium* symbiosis, and they involved a limited number of candidate genes expressed during initiation of symbiosis with competent *Symbiodinium* strains. Studies focused on symbiosis establishment in coral larvae are summarized in Table 1, with further discussion below. Approximately 15% of coral species maternally inherit their symbionts, while 85% acquire their symbionts from the environment at each generation, a process known as horizontal transmission (Baird *et al.* 2009). Although adult corals typically associate with specific strains of *Symbiodinium*, the larvae are relatively promiscuous (Cumbo *et al.* 2013). Nevertheless, a degree of specificity exists; for example, Voolstra *et al.* (2009) reported that the *Symbiodinium* strain EL1 is “incompetent” to establish symbioses with *Acropora palmata* or *Orbicella* (formerly *Montastraea*) *faveolata*. The establishment of an association via horizontal transmission therefore might require

communication between the partners during the early stages of the interaction (Wood-Charlson *et al.* 2006), in addition to mechanisms for nutrient exchange in the symbiotic state (Meyer & Weis 2012).

Few studies investigated the host gene expression changes during onset and establishment of coral-*Symbiodinium* symbiosis. Woolstra *et al.* (2009) used cDNA microarray to investigate expression differences between aposymbiotic and newly symbiotic larvae of the Caribbean corals *Acropora palmata* and *Orbicella faveolata*. Larvae were challenged with a competent strain of *Symbiodinium* (a strain able to establish symbiosis) and an incompetent strain of *Symbiodinium* (a strain that had previously failed to establish symbiosis). Samples were taken at 2 different time points, (early) at 30 min and (late) at 6 days post-infection. In the case of the competent strain, very few changes were detected in the coral transcriptome at either time point. On the other hand, with the incompetent strain dramatic differences in the host transcriptome were observed in both species at 6 d post-infection. These findings led the authors to suggest that competent *Symbiodinium* strains might enter the hosts by avoiding host recognition, and possibly symbionts either circumvent or modulate the host immune response. In addition, Schnitzler and Weis (2010) used cDNA microarray technology to study the infection of *Fungia scutaria* larvae with competent *Symbiodinium*. This study detected very few transcripts that were differentially expressed between infected/uninfected larvae. They suggested that these results might be due to the inability to detect signals from small number of infected cells or that a window of symbiosis-related expression could have been missed due to the timing of sampling.

Table 2.1 Summary of published data on gene expression analysis of coral larvae during onset and establishment of coral-*Symbiodinium* symbiosis

Authors	Coral host/species	Technique	Genes of Interest	Results
Grasso <i>et al.</i> (2008)	<i>Acropora millepora</i>	cDNA microarray	C-type lectin Carbonic anhydrase	Despite focusing on development, the study described some symbiosis-related transcripts
Schwarz <i>et al.</i> (2008)	<i>Orbicella faveolata</i> and <i>A. palmata</i>	EST analysis	Lectins Tachylectins Thrombospondin type I repeats Ferritin Scavenger receptors	Transcripts associated with innate immunity were identified as candidate genes of interest in symbiosis
Voolstra <i>et al.</i> (2009)	<i>Orbicella faveolata</i> and <i>A. palmata</i>	cDNA microarray	NA	Larvae were inoculated with competent and incompetent symbionts. The study found that there were few changes in larval transcriptome in the case of the competent strain, however significant changes in the case of incompetent strain. Authors suggested that initiation of the symbiosis depends on the capability of the symbionts of entering their host in a stealth manner without mounting an active response from the host
Schnitzler and Weis (2010)	<i>Fungia scutaria</i>	cDNA microarray	NA	Larval transcriptome showed very few changes during onset of symbiosis with competent symbionts. The study suggested suppression of host response during early colonization by symbionts
Mohamed <i>et al.</i> (2016) The present Chapter	<i>A. digitifera</i>	Illumina RNA-Seq	1073 genes	Discussed here

In the present chapter, high-throughput next generation sequencing (NGS) (Illumina RNA-Seq) was used to investigate transcriptome-wide gene expression changes in coral larvae during exposure to a competent *Symbiodinium* strain. The gene expression profiles of larvae of the coral *Acropora digitifera* were investigated after exposure to *Symbiodinium* sp. clade B1 culture at 4, 12, and 48 h post-infection. Significant changes in the coral host transcriptome were detected during an early stage of the infection process (i.e. at the 4-h time point). In contrast, the host transcriptome remained unchanged at the subsequent time points. Our data suggest that establishment of the symbiosis might depend on both partners as there might be an active response

from the host in recognizing the symbiotic partners and at the same time the symbionts might be able to suppress the host immune response. Regulation of host protein synthesis/ transport, oxidative metabolism, transcription, and apoptosis might be key processes during onset and establishment of coral-algal symbiosis. Most significantly, the results are consistent with published hypotheses about the establishment of symbiosis in sea anemones, implying that, despite the deep evolutionary divergence between these two lineages, common molecular mechanisms may apply.

2.3 MATERIALS AND METHODS

2.3.1 *Coral larvae*

Mature *Acropora digitifera* colonies were collected in front of the Sesoko Station, the Tropical Biosphere Research Center at the University of the Ryukyus in Sesoko Island, Okinawa, Japan (Figure 2.1) where they were maintained in flow-through aquaria until spawning occurred (19 June 2013). After fertilization, *A. digitifera* embryos were raised in 0.2 μm filtered seawater (FSW) under ambient conditions to the late actively swimming planula stage. The developed larvae were used for algal inoculation (infection) experiments. Larvae were inspected using light microscopy to verify their aposymbiotic state and *Symbiodinium* infection experiments were set up 6 days post-spawning, by which time the oral pore and coelenteron of the larvae were sufficiently well developed to enable symbiont uptake (Harii *et al.* 2009).



Figure 2.1 Colony of the coral *A. digitifera* collected in front of the Sesoko Station, the Tropical Biosphere Research Center of the University of the Ryukyus in Sesoko Island, Okinawa.

2.3.2 *Symbiodinium* culture

Symbiodinium sp clade B1 (culture ID = CCMP3345), originally isolated from the anemone *Aiptasia pallida*, was used in this experiment. Prior to inoculation, *Symbiodinium* cultures were maintained at 25 °C in Guillard's f/2 Marine Water Enrichment Solution (G0154, Sigma-Aldrich) under a 12/12 h light/dark cycle. The *Symbiodinium* strain used was selected on the basis of it being not only competent with (being able to infect) *A. digitifera*, but also this being the strain for which the most comprehensive sequence data existed. The original intention was to simultaneously study gene expression changes in the coral host and the symbiont. For that reason, I chose that *Symbiodinium* strain, however too few reads mapped to the *Symbiodinium* reference genome (Shoguchi *et al.* 2013) to permit statistically valid differential expression analyses.

2.3.3 *Symbiodinium* acquisition experimental design

Acropora digitifera were divided into 2 treatment groups, with 3 replicates per treatment. In each replicate, 700 larvae were placed into 1L plastic containers containing 700 mL of 0.2 µm FSW. The first group of larvae were inoculated with the *Symbiodinium* sp clade B1 cells and labeled “*Symbiodinium*-infected”. Before inoculation, algae were washed three times in 0.2 µm FSW and added at final concentration of 5×10^3 cells/ml. The algal density used is similar to that in sediment around coral reefs (Littman *et al.* 2008). In the second group, larvae were not inoculated with *Symbiodinium* and labeled “control”. Larvae from each of the replicates were collected and washed by pipetting in 0.2 µm FSW to ensure no algae were attached to the larval surface. Containers were held at 26 °C in a constant temperature room and under fluorescent lamps that provided light (86 ± 2 µmol photon m⁻² s⁻¹ at the surface) on a 12/12 h light/dark cycle. At 4, 12 and 48 h post *Symbiodinium* inoculation, 10 larvae from each replicate were collected and washed by pipetting in 0.2 µm FSW, ensuring that no algae were attached to the larval surface. Larvae were inspected under a fluorescence microscope for the symbiont chlorophyll fluorescence to determine if successful infection with *Symbiodinium* had occurred.

2.3.4 Larval sampling and RNA isolation

At 4, 12 and 48 h post-infection, approximately 150 larvae from each replicate were washed in 0.2 µm FSW and sampled (ensuring as little liquid carry over as possible). Samples were snap frozen immediately in liquid nitrogen and stored at -80 °C until further treatment. Total RNA was isolated using TRIzol® reagent (Life Technologies) according to manufacturer's instructions. Larvae were homogenized for 2 min in 2 ml TRIzol reagent using a Polytron homogenizer (Kinematica AG, Lucerne, Switzerland) followed by phase separation with chloroform that separates RNA from other cellular components. The clear upper aqueous phase containing RNA was carefully transferred into a new RNase-free microtube and RNA was precipitated with equal volume of isopropanol. Then RNA pellet was washed with 75% ethanol. Finally, RNA was dissolved in 40 µL of RNase-free water and stored at -80 °C. RNA quality and quantity were assessed using measurements of NanoDrop ND-1000 spectrometer. RNA integrity was checked using the electrophoretic profiling with an Agilent 2100 Bioanalyzer.

2.3.5 High-throughput sequencing Illumina RNA-Seq

Messenger RNA (mRNA) was isolated from 1 µg high quality total RNA and a total of 16 cDNA libraries were prepared using the TruSeq RNA Sample Preparation Kit (Illumina) according to manufacturer's instructions. RNA-Seq libraries were sequenced using an Illumina HiSeq 2000 platform at the Okinawa Institute of Science and Technology (OIST), Okinawa, Japan. Sequencing produced a total of 333.2 million individual 100 bp paired-end (PE) reads.

2.3.6 RNA-Seq data analysis

2.3.6.1 Reads quality check and reference mapping

Raw Illumina PE reads were checked for quality using the FastQC software (version 0.11.2) (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were mapped onto the *A. digitifera* transcriptome assembly v1.0 (Shinzato *et al.* 2011) (<http://marinegenomics.oist.jp/genomes/>) using the BOWTIE mapping software version

0.12.7 (Langmead *et al.* 2009) (<http://bowtie-bio.sourceforge.net/index.shtml>) with default parameters. Then RSEM software version 1.1.17 (Li & Dewey 2011) (<http://deweylab.biostat.wisc.edu/rsem/>) was used to quantify the abundances of the transcripts for each library. In order to assess the read mapping, read alignments were visualized using the Integrated Genomics Viewer (IGV) software version 2.3.34 (Thorvaldsdottir *et al.* 2013) (<http://www.broadinstitute.org/igv/>). Percentages of mapped reads were obtained by using the flagstat command within the samtools.

2.3.6.2 Differential gene expression analysis

Differential gene expression was inferred based on the mapping counts using the edgeR package (Robinson *et al.* 2010) in the R statistical computing environment (<http://www.r-project.org>). *Symbiodinium*-infected samples were compared the control ones at each of the three time points producing 3 expression profiles at 4, 12 and 48 h post infection (at the 12 h time point only 2 replicates per condition were available). *P*-values for differential gene expression were corrected for multiple testing using the Benjamini and Hochberg's algorithm. Genes with a false discovery rate (FDR) of ≤ 0.05 were considered as significant differentially expressed between treatment and control groups. The plot_MA_and_Volcano function in R was also used to generate MA and volcano plots of the DEGs at the same FDR cut off.

2.3.6.3 Hierarchical clustering analysis

To study expression patterns of genes across samples, raw counts were first normalized using the TMM normalization function (genes were \log_2 -transformed and median-centered by transcript) (Robinson & Oshlack 2010) in edgeR to scale the expression values (Fragments Per Kilobase of transcript per Million mapped reads; FPKM) provided by the RSEM software across all samples. The R package heatmap3 was used to generate sample Spearman correlation and gene clustering heat maps. Clustering analysis was conducted on the 1073 significant DEGs detected at the 4 h time point. A heat map was obtained in order to explore patterns of gene expression across samples.

2.3.6.4 Functional annotations and gene ontology (GO) enrichment analyses

To determine the function of the significant differentially expressed genes (at $FDR \leq 0.05$) for both up- and down-regulated genes with absolute \log_2 fold change > 1 , two stages of analysis were employed. BLASTX ($e \leq 10^{-3}$) was performed against the Swiss-Prot annotated protein database (<http://www.uniprot.org>), then the resulting hits were filtered at $e \leq 10^{-15}$ by searching for close matches in *Nematostella vectensis* and/or *Acropora digitifera*. GO enrichment analyses and the identification of enriched biological themes were performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang *et al.* 2009) (<http://david.abcc.ncifcrf.gov>). The UNIPROT ACCESSION terms of the Swiss-Prot genes were used as identifiers. The same BLASTX was performed to the whole transcriptome dataset and all the Swiss-Prot-annotated genes contributed to the background gene set for the enrichment analysis. DAVID uses the Fisher's exact test to ascertain statistically significant pathway enrichment among differentially expressed genes relative to the background transcriptome. The Benjamini-corrected P -value of ≤ 0.05 to filter the significantly enriched GO categories (Huang *et al.* 2009). Heat maps of specific categories of the DEGs likely involved in coral-algal symbiosis were generated using the R package pheatmap (<https://cran.rproject.org/web/packages/pheatmap/index.html>). The bioinformatics analysis flowchart is summarized in Figure (2.2).

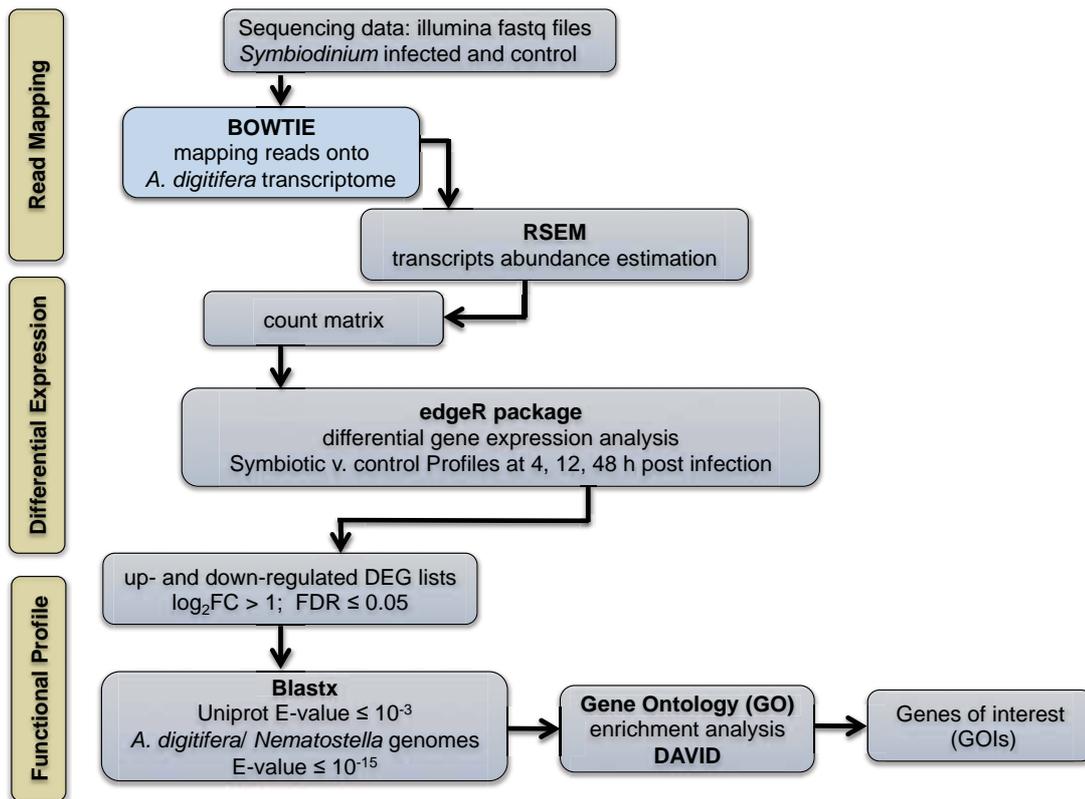


Figure 2.2 Bioinformatic analysis flowchart. Illumina Hiseq reads were mapped onto the *A. digitifera* transcriptome assembly using the mapping software BOWTIE, then RSEM was used to estimate transcript abundance in each sample. Differential gene expression was inferred based on the mapping counts using the edgeR package in the R environment. Using the Uniprot accession IDs of the up- and down-regulated DEGs, Gene Ontology (GO) enrichment analysis was performed using DAVID database. Annotated DEGs were classified based on literature search of genes involved in cnidarian-algal symbioses.

2.4 RESULTS

2.4.1 General results and overall transcriptome changes

Illumina RNA-Seq technology was utilized to investigate the host transcriptome-wide gene expression during onset and establishment of coral-algal symbiosis. *Acropora digitifera* planula larvae were challenged with *Symbiodinium* clade B1 cells as a competent *Symbiodinium* strain. Infection success was followed by fluorescence microscopy over the course of the experiment (Figure 2.3). 16 cDNA libraries were sequenced using the Illumina HiSeq 2000 platform. Sequencing yielded

an average of 20.8 million bp Illumina paired-end reads per library and an average of 36 % of the raw reads were mapped onto the *A. digitifera* transcriptome (see Tables 2.2, 2.3 and Figure 2.4). The RNA-Seq reads and count tables used for the differential expression in the study have been submitted to the NCBI Gene Expression Omnibus (GEO) database under accession number GSE76976.

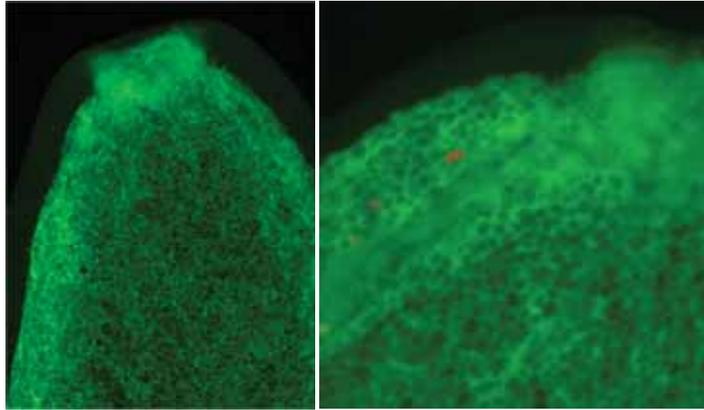


Figure 2.3 Images of *A. digitifera* larvae and *Symbiodinium* uptake in symbiotic larva under fluorescence microscopy. Left image shows a control “uninfected” larva fluorescing green, while right image shows an infected *Symbiodinium* sp. cladeB1-infected larva, the red fluorescence confirming successful symbiont uptake (infection).

Hierarchical clustering analysis of transcript expression values revealed a high level of agreement between the biological replicates, as can be seen in the heat map of the pairwise Spearman correlations between samples in Figure 2.5. Moreover, multidimensional scaling (based on the 500 genes that best differentiate the samples) separates the *Symbiodinium*-infected and control samples along the x coordinate (Figure 2.6). A transient and relatively small transcriptomic response by coral larvae were detected only early in the infection process. Significant changes in the coral host transcriptome were detected only at 4h post infection. At this time, 2.91% of the *A. digitifera* contigs (2.14% down- and 0.77% up-regulated) were differentially expressed, using a cut-off of adjusted P-value ≤ 0.05 . On the other hand, there were no changes in the coral transcriptome at subsequent time points; 12 and 48 h post-infection (Figure 2.7).

Table 2.2 Raw numbers of Illumina Hi-Seq sequencing reads obtained. 16 cDNA libraries were sequenced and produced a total of 333 million reads (*Symbiodinium* infected and control) at 3 time points; 4, 12, and 48 h post-infection. NA = no data and the absence of data at 12h are attributed to RNA quality issues. The low number of reads in negative control 1 at 48h is due to a low RNA yield

Illumina cDNA Libraries	04 h	12 h	48 h	Total # Illumina reads
<i>Symbiodinium</i> -infected 1	26,154,672	NA	18,805,803	44,960,475
<i>Symbiodinium</i> -infected 2	22,040,532	22,705,829	20,055,461	64,801,822
<i>Symbiodinium</i> -infected 3	20,333,953	23,520,684	24,533,716	68,388,353
Negative control 1	22,430,563	NA	2,355,166	24,785,729
Negative control 2	19,542,091	24,035,359	15,729,866	59,307,316
Negative control 3	20,912,751	25,197,830	24,849,108	70,959,689

Table 2.3 Percent of reads successfully mapped onto the *A. digitifera* transcriptome assembly

Illumina cDNA Libraries	04 h	12 h	48 h
<i>Symbiodinium</i> -infected 1	38.38%	NA	34.63%
<i>Symbiodinium</i> -infected 2	37.80%	41.17%	34.05%
<i>Symbiodinium</i> -infected 3	37.02%	33.04%	37.55%
Negative control 1	34.83%	NA	34.83%
Negative control 2	35.53%	34.1%	35.53%
Negative control 3	36.02%	37%	36.02%



Figure 2.4 A screenshot of *A. digitifera* transcripts and reads alignments of both control and *Symbiodinium* infected reads at 4 h post-infection. Using the Integrated Genomics Viewer (IGV), the sorted BAM files containing the aligned reads were uploaded as well as the reference transcriptome sequence data.

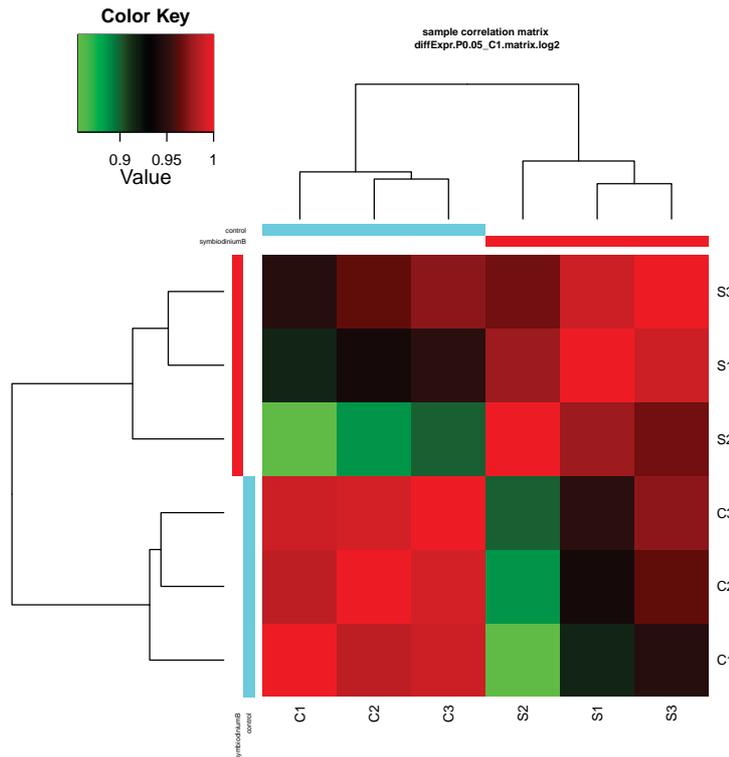


Figure 2.5 Level of agreement among the biological replicates. The heat map shows the hierarchically clustered Spearman correlation matrix resulting from comparing the transcript expression values (TMM-normalized FPKM) for all samples against one another. Sample clustering indicates the consistency between the biological replicates of the *Symbiodinium* sp. clade B1 infection (samples S1, S2, S3) and negative control condition (samples C1, C2, C3) at 4 h post infection. The level of correlation represented by a colored field that ranges from green (correlation coefficient 0.85) to red (correlation coefficient 1.0).

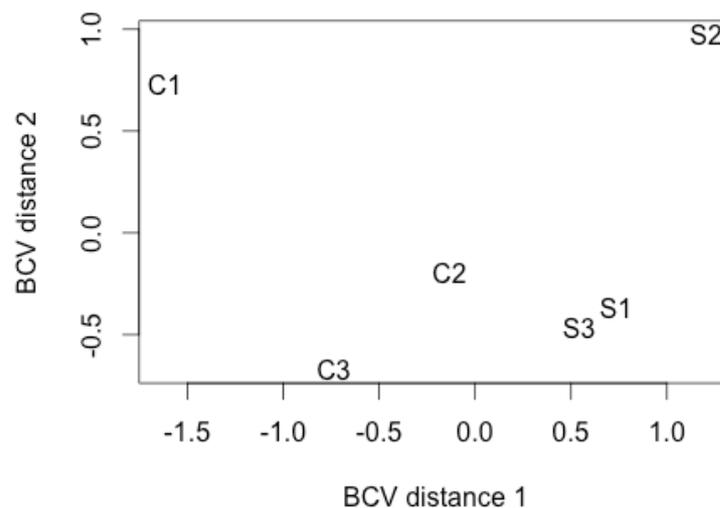


Figure 2.6 A multidimensional scaling (MDS) plot produced by edgeR showing the relationship between all replicates of *Symbiodinium* clade B infection (S1, S2, S3) and negative control condition (C1, C2, C3). The distances shown on the plot are the biological coefficient of variation (BCV) between samples for the 500 genes that best distinguish the samples.

2.4.2 Differential gene expression analysis

Differential gene expression analysis using the edgeR package at 4 h post-infection revealed the presence of 1073 differentially expressed genes (DEGs) (adjusted P -value ≤ 0.05). Amongst these DEGs, 285 (26.6%) and 788 (73.4 %) were up- and down-regulated, respectively ($-3.8 > \log FC < 6.3$, FC medians at -1.2 and 1.3) (Table 2.4, Figures 2.7 and 2.8). Hierarchical clustering of the DEGs at 4h time point reveals 2 distinctive expression profiles for the *Symbiodinium*-infected and control larvae, and the majority of significant DEGs were down regulated in case of *Symbiodinium* infection (Figure 2.9). On the other hand, there were no differences in gene expression between *Symbiodinium*-infected samples and control (adjusted $P \leq 0.05$) for either 12 or 48 h post-infection as shown in the volcano plots (Figures 2.7 and 2.8).

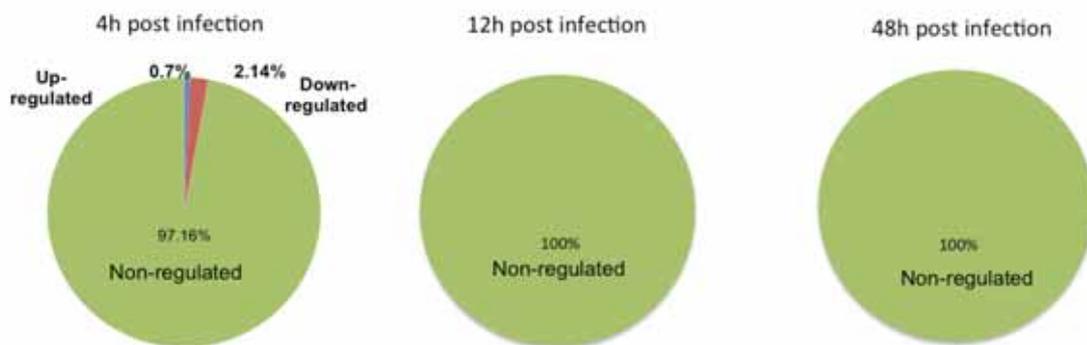


Figure 2.7 *A. digitifera* transcriptome changes during onset of symbiosis with *Symbiodinium* sp. clade B1. Percentages of transcripts up- (blue) and down-regulated (red) in response to *Symbiodinium* infection. Only 2.91% of the coral transcriptome showed differential expression at the 4 h time point (adjusted $P \leq 0.05$), whereas no differences were observed in the coral transcriptome at 12 and 48 h post-infection.

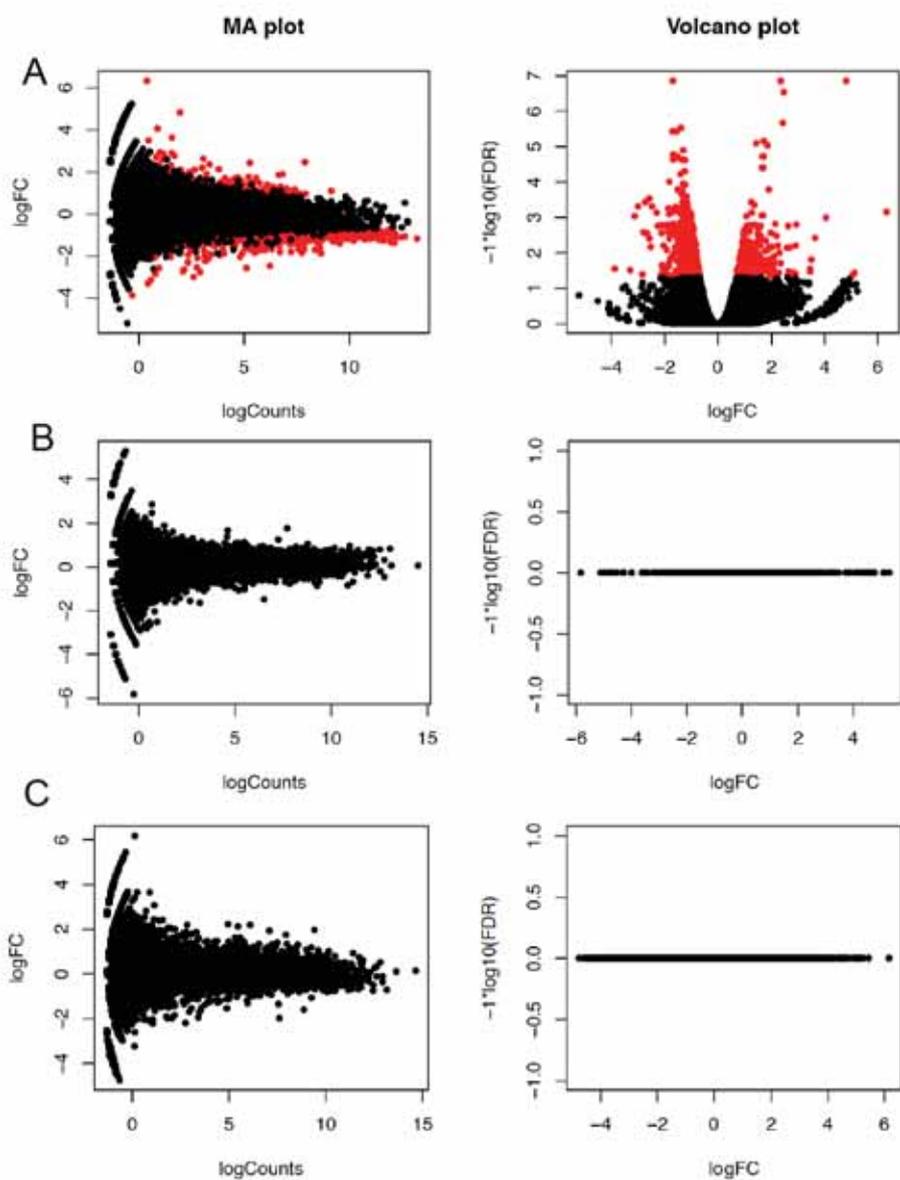


Figure 2.8 Volcano and MA plots comparing gene expression profiles between the *Symbiodinium*-infected and control larvae at 4, 12 and 48 h (Parts A, B and C). The red dots in part A represent the significant differentially expressed genes (DEGs) detected only at the 4-h time point (adjusted $P \leq 0.05$). Parts B and C show no significant DEGs using the same P -value.

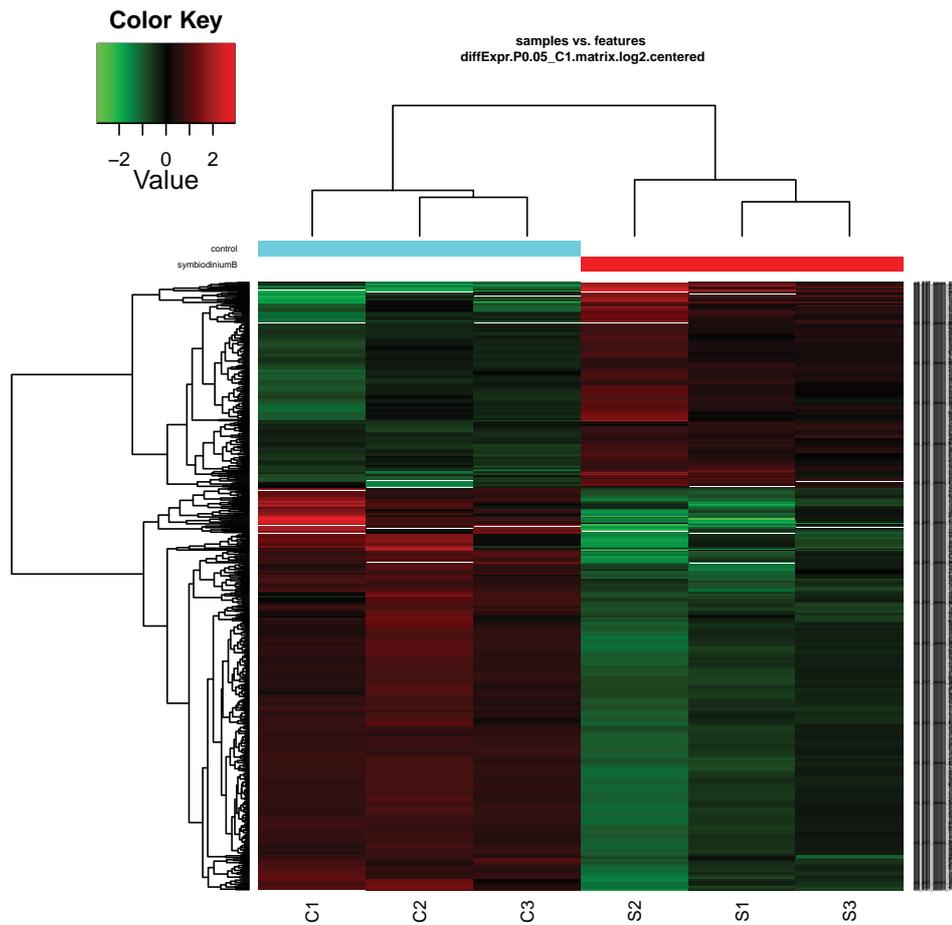


Figure 2.9 Differential gene expression profiles at 4 h post *Symbiodinium* infection. The heat map shows the expression profiles of *Symbiodinium*-infected and control samples. The hierarchical clustering was obtained by comparing the expression values (Fragment Per Kilobase of transcript per Million; FPKM) for *Symbiodinium*-infected samples compared against the control at 4 h post-infection. Expression values are \log_2 -transformed and then median-centered by transcript. Relative expression levels are shown in red (up) and green (down). Samples S1, S2 and S3 are the biological replicates of *Symbiodinium* sp. clade B1 infection while samples C1, C2, C3 are the biological replicates of the control condition.

2.4.3 Gene Ontology (GO) enrichment analysis

As a preliminary approach to data analysis, the DEGs (adjusted P -value ≤ 0.05) were annotated using the GO database. Over-representation of GO terms among differentially expressed genes was evaluated to infer molecular function (GO-MF), cellular component (GO-CC) and biological process (GO-BP) categories that were mostly affected during *Symbiodinium* infection. Of the DEGs, 625 (58.3%) had an absolute $\log_2(\text{fold-changes})$ greater than 1, of which 335 had reliable Swiss-Prot protein annotations (BLASTX e-values $\leq 10^{-15}$) (Table 2.4) that could be used to characterize the transcriptome response of *A. digitifera* during onset and establishment of symbiosis with a competent strain of *Symbiodinium*. Enrichment analyses of the significant DEGs with absolute $\log_2(\text{fold-changes}) > 1$, revealed GO categories that were significantly enriched in the case of up- and down-regulated genes at 4 h post-*Symbiodinium* infection.

Table 2.4 Summary of the differential gene expression profiling at 4 h post-*Symbiodinium* infection. An adjusted $P \leq 0.05$ and E-value of $\leq 10^{-15}$ were used to filter differentially expressed genes (DEGs) in BLASTX searches against the Swiss-Prot database

Coral transcripts	DEGs	Up-regulated	Down-regulated	DEGs >1-fold	Swiss-Prot-annotated DEGs
36,780	1073 2.91%	285 0.77%	788 2.14%	625 58.3% of DEGs	335

2.4.4 GO enrichment analysis reveals suppression of ribosome, Endoplasmic Reticulum and mitochondria functions

Amongst down-regulated genes, there were significant GO terms in both cellular component and molecular function were enriched using a corrected P -value ≤ 0.05 , but not in biological process categories (Table 2.5). However, no significant GO enrichment related to biological process, molecular function or cellular component could be detected amongst up-regulated genes. The GO-MF terms “voltage gated channel activity” (GO:0022832) and “voltage gated ion-channel activity” (GO:0005244) were over-represented amongst up-regulated genes, but with slightly larger corrected P -value (0.058) than the threshold applied (Table 2.6). In the case of the cellular component (GO-CC) genes showed GO enrichment related to 17 categories related to ribosome, mitochondria

and endoplasmic reticulum functions. For the molecular function (GO-MF), there was enrichment related to 7 categories related in some way to membrane transport, although the ranges of transporter activities involved were very broad (Table 3).

The GO-CC categories ribosome, ribo-nucleoprotein complex, cytosolic small ribosomal subunit and the GO-MF categories structural constituent of ribosome categories were enriched in the down-regulated genes. Many genes encoding ribosomal proteins (60S and 40S ribosomal proteins), DNA-directed RNA polymerases (iii subunit RPABC3) and probable translation initiation factor EIF-2b subunit epsilon were down-regulated (Table 2.7). The GO-CC categories mitochondrial membrane, mitochondrial envelope, mitochondrial inner membrane, and mitochondrial membrane part categories were significantly enriched in the down-regulated genes. Components of the electron transport chain including many ATP synthase subunits and proteins from two mitochondrial inner membrane complexes (I and IV) (Table 2.8) were down-regulated. The GO-MF categories hydrogen ion transmembrane transporter activity, transmembrane transporter activity, monovalent inorganic cation transmembrane transporter activity, P-P-bond-hydrolysis-driven protein transmembrane transporter activity, protein transmembrane transporter activity and macromolecule transmembrane transporter activity categories were significantly enriched in the down-regulated genes, for example, 7 trans-membrane proteins were down-regulated including; trans-membrane protein 69, protein asterix, TM2 domain-containing protein, trans-membrane protein 254, oligosaccharyl transferase complex subunit ostc-b, protein rolling stone, and mechano-sensory protein 2.

The GO-CC category rough endoplasmic reticulum membrane was also enriched in the down-regulated genes; despite comprising only three genes it was the most highly over-represented with 44.91 fold enrichment. In addition to the three genes associated with that GO-CC category, five other genes encoding proteins potentially involved in ER functions (regulation of protein translocation in the ER, retention of ER resident proteins and vesicular transport from ER to Golgi apparatus) were down-regulated including translocon-associated protein subunits gamma and delta (part of a complex whose function is binding calcium to the ER membrane and thereby regulating the retention of ER resident proteins), transport protein sec61 subunits gamma and beta (necessary for protein translocation in the ER) and Trafficking protein particle complex subunit 3 (which plays a role in vesicular transport from endoplasmic reticulum to

Golgi) (Table 2.9).

My findings reveal that protein synthesis/transport and oxidative metabolism are likely to be temporarily suppressed, due to host challenge by competent symbionts, during the onset and establishment of the symbiosis. Generally speaking metabolic suppression is an adaptive strategy for survival during short-term energy shortage in many aquatic organisms and might be fulfilled by shutting down processes as protein synthesis (Guppy & Withers 1999).

Table 2.5 Enriched Gene Ontology (GO) categories (GO-MF, CC) at adjusted $P \leq 0.05$ in the set of down-regulated genes

	Annotation Term	Gene Ontology (GO) ID	No. of genes	Fold enrichment
Down-regulated Cellular component (GO-CC) categories	Ribonucleoprotein complex	GO:0030529	28	4
	Ribosome	GO:0005840	24	7.11
	Organelle membrane	GO:0031090	21	2.09
	Organelle envelope	GO:0031967	16	2.40
	Envelope	GO:0031975	16	2.37
	Organelle inner membrane	GO:0019866	13	3.76
	Mitochondrial membrane	GO:0031966	13	3.10
	Mitochondrial envelope	GO:0005740	13	2.86
	Mitochondrial inner membrane	GO:0005743	12	3.55
	Ribosomal subunit	GO:0033279	10	7.04
	Cytosolic part	GO:0044445	9	6.65
	Respiratory chain	GO:0070469	8	11.97
	Cytosolic ribosome	GO:0022626	8	9.03
	Small ribosomal subunit	GO:0015935	7	10.22
	Cytosolic small ribosomal subunit	GO:0022627	6	14.37
	Mitochondrial membrane part	GO:0044455	5	8.31
	Rough endoplasmic reticulum membrane	GO:0030867	3	44.91
Down-regulated Molecular function (GO-MF) categories	Structural constituent of ribosome	GO:0003735	24	9.41
	Structural molecule activity	GO:0005198	24	5.24
	Hydrogen ion transmembrane transporter activity	GO:0015078	6	8.63
	Monovalent inorganic cation transmembrane transporter activity	GO:0015077	6	7.39
	P-P-bond-hydrolysis-driven protein transmembrane transporter activity	GO:0015450	4	30.69
	Protein transmembrane transporter activity	GO:0008320	4	30.69
	Macromolecule transmembrane transporter activity	GO:0022884	4	30.69

Table 2.6 Differential expression of *A. digitifera* DEGs (n=19) (FDR ≤ 0.05) likely involved in voltage gated (ion) channel activity and trans-membrane transporters

Transcript ID	Best BLAST hit	UniProt ID	E-Value	logFC
adi_EST_assem_23905	Sodium-dependent neutral amino acid transporter b at3 (<i>Mus musculus</i>)	O88576	3.44E-23	5.02
adi_EST_assem_25075	Sodium channel protein type 8 subunit alpha (<i>Rattus norvegicus</i>)	O88420	2.12E-75	1.99
adi_EST_assem_141	Sodium channel protein 60e (<i>Drosophila melanogaster</i>)	Q9W0Y8	3.04E-29	1.19
adi_EST_assem_20125	Two pore calcium channel protein 1 (<i>Mus musculus</i>)	Q9EQJ0	3.62E-54	1.74
adi_EST_assem_10127	Voltage-dependent calcium channel type d subunit alpha-1 (<i>Drosophila melanogaster</i>)	Q24270	0	1.34
adi_EST_assem_18020	Voltage-dependent N-type calcium channel subunit alpha-1B (<i>Homo sapiens</i>)	Q00975	0.00E+00	1.29
adi_EST_assem_6942	Chloride channel protein 2 (<i>Homo sapiens</i>)	P51788	0	1.34
adi_EST_assem_3651	Potassium voltage-gated channel protein Shab (<i>Drosophila melanogaster</i>)	P17970	1.51E-118	1.29
adi_EST_assem_4067	Potassium voltage-gated channel subfamily h member 7 (<i>Rattus norvegicus</i>)	O54852	1.75E-85	1.12
adi_EST_assem_8870	Transient receptor potential cation channel subfamily a member 1 (<i>Rattus norvegicus</i>)	Q6RI86	2.17E-113	1.2
adi_EST_assem_15378	Multidrug resistance-associated protein 4 (<i>Homo sapiens</i>) ATP-binding cassette, sub-family C (CFTR/MRP), member 4	O15439	2.90E-151	1.19
adi_EST_assem_10002	H(+) Cl(-) exchange transporter 3 (<i>Rattus norvegicus</i>)	P51792	0	1.06
adi_EST_assem_13911	Transmembrane protein 69 (<i>Mus musculus</i>)	Q3KQJ0	1.19E-18	-1.00
adi_EST_assem_6895	Protein asterix (<i>Sus scrofa</i>)	Q6Q7K0	1.01E-26	-1.10
adi_EST_assem_16094	TM2 domain-containing protein cg10795 (<i>Drosophila melanogaster</i>)	Q9W2H1	1.01E-45	-1.00
adi_EST_assem_1475	Transmembrane protein 254 (<i>Rattus norvegicus</i>)	Q5U220	1.07E-17	-1.10
adi_EST_assem_226	OSTCB_oligosaccharyltransferase complex subunit ostc-b membrane protein (<i>Xenopus laevis</i>)	Q5M9B7	5.83E-76	-1.22
adi_EST_assem_10727	ROST_protein rolling stone (<i>Drosophila melanogaster</i>)	O44252	0	-1.03
adi_EST_assem_1710	MEC2_mechanosensory protein 2 (<i>Caenorhabditis elegans</i>)	Q27433	2.17E-115	-1.13

Table 2.7 Down-regulation of *A. digitifera* DEGs (n=25) (FDR \leq 0.05) likely involved in protein synthesis and translation

Transcript ID	Best BLAST hit	UniProt ID	E-Value	logFC
adi_EST_assem_3874	r115_60s ribosomal protein l15 (<i>Rattus norvegicus</i>)	P61314	1.74E-102	-1.33
adi_EST_assem_3513	rs14_40s ribosomal protein s14 (<i>Podocoryne carnea</i>)	Q08699	3.59E-28	-1.27
adi_EST_assem_2268	rs10b_drome40s ribosomal protein s10b (<i>Drosophila melanogaster</i>)	Q9VWG3	1.21E-58	-1.26
adi_EST_assem_8799	r114_60s ribosomal protein l14 (<i>Lumbricus rubellus</i>)	O46160	1.66E-42	-1.21
adi_EST_assem_3453	rssa_40s ribosomal protein sa (<i>Nematostella vectensis</i>)	A7RKS5	3.68E-138	-1.18
adi_EST_assem_25422	rs271_40s ribosomal protein s27-like (<i>Bos taurus</i>)	Q3T0B7	1.84E-21	-1.16
adi_EST_assem_4095	rs23_40s ribosomal protein s23 (<i>Rattus norvegicus</i>)	P62268	5.75E-71	-1.14
adi_EST_assem_4070	rs193_40s ribosomal protein s19-3 (<i>Arabidopsis thaliana</i>)	Q9FNP8	3.12E-40	-1.14
adi_EST_assem_3374	rs5_40s ribosomal protein s5 (<i>Homo sapiens</i>)	P46782	1.26E-114	-1.14
adi_EST_assem_675	r1391_60s ribosomal protein l39-1 (<i>Arabidopsis thaliana</i>)	P51424	6.06E-21	-1.14
adi_EST_assem_9213	rs27a_ubiquitin-40s ribosomal protein s27a (<i>Gallus gallus</i>)	P79781	1.24E-54	-1.13
adi_EST_assem_6809	rs21_40s ribosomal protein s21 (<i>Sus scrofa</i>)	P63221	3.42E-35	-1.12
adi_EST_assem_8024	rs16_40s ribosomal protein s16 (<i>Heteropneustes fossilis</i>)	Q98TR7	2.66E-54	-1.12
adi_EST_assem_9646	r128_60s ribosomal protein l28 (<i>Rattus norvegicus</i>)	P17702	3.65E-25	-1.06
adi_EST_assem_1997	r127_60s ribosomal protein l27 (<i>Hippocampus comes</i>)	P61359	4.32E-56	-1.05
adi_EST_assem_54	rs15a_40s ribosomal protein s15a (<i>Rattus norvegicus</i>)	P62246	5.41E-53	-1.05
adi_EST_assem_8017	rs12_40s ribosomal protein s12 (<i>Gallus gallus</i>)	P84175	1.77E-57	-1.05
adi_EST_assem_5274	60s ribosomal protein l35a (<i>Caenorhabditis elegans</i>)	P49180	1.25E-24	-1.03
adi_EST_assem_5495	r122_60s ribosomal protein l22 (<i>Sus scrofa</i>)	P67985	2.78E-32	-1.03
adi_EST_assem_3752	r127a_60s ribosomal protein l27a (<i>Xenopus laevis</i>)	P47830	9.15E-62	-1.03
adi_EST_assem_15634	rm15_39s ribosomal protein mitochondrial (<i>Xenopus laevis</i>)	Q6AZN4	3.49E-53	-1.02
adi_EST_assem_7749	r144_60s ribosomal protein l44 (<i>Ochlerotatus triseriatus</i>)	Q9NB33	5.88E-29	-1.01
adi_EST_assem_8917	baf_barrier-to-autointegration factor (<i>Danio rerio</i>)	Q6P026	2.97E-31	-1.22
adi_EST_assem_5753	rpab3_dna-directed rna polymerases and iii subunit rpabc3 (<i>Mus musculus</i>)	Q923G2	8.41E-82	-1.02
adi_EST_assem_6056	ei2be_probable translation initiation factor eif-2b subunit epsilon (<i>Schizosaccharomyces pombe</i>)	P56287	2.51E-39	-1

Table 2.8 Down-regulation of *A. digitifera* DEGs (n=19) (FDR ≤ 0.05) likely involved in mitochondria functions. IV and I refer to complexes of the electron transport chain

Transcript ID	Best BLAST hit	UniProt ID	E-Value	logFC
adi_EST_assem_1086	NADH dehydrogenase (<i>Homo sapiens</i>) (I)	Q86Y39	3.50E-31	-1.06
adi_EST_assem_4309	ndua2_NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2, 8kDa (<i>Homo sapiens</i>) (I)	O43678	6.13E-34	-1.14
adi_EST_assem_20728	ndua7_NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 7, 14.5kDa (<i>Bos taurus</i>) (I)	Q05752	5.85E-20	-1.14
adi_EST_assem_11037	ndub9_NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9, 22kDa (<i>Pan troglodytes</i>) (I)	Q0MQF0	3.24E-23	-1.02
adi_EST_assem_6654	ndus7_NADH dehydrogenase (ubiquinone) Fe-S protein 7 (<i>Mus musculus</i>) (I)	Q9DC70	4.99E-99	-1.13
adi_EST_assem_6313	ATP synthase subunit mitochondrial (<i>Homo sapiens</i>)	O75964	9.97E-19	-1.01
adi_EST_assem_6284	ATP synthase subunit mitochondrial (<i>Bos taurus</i>)	P05630	1.34E-36	-1.22
adi_EST_assem_12987	ATP synthase subunit mitochondrial (<i>Zea mays</i>)	Q41898	7.17E-06	-1.15
adi_EST_assem_17779	Cytochrome b561 (<i>Mus musculus</i>) (IV)	Q60720	5.00E-15	-1.16
adi_EST_assem_17780	Cytochrome b reductase 1 (<i>Xenopus tropicalis</i>) (IV)	Q5CZL8	5.20E-34	-1.06
adi_EST_assem_9537	Dehydrogenase reductase sdr family member 7 (<i>Mus musculus</i>)	Q9CXR1	1.81E-78	-1.06
adi_EST_assem_6764	COX4_cytochrome c oxidase subunit 4 (<i>Acropora digitifera</i>) (IV)	Pfam: aug_v2a.0 5853.t1	9.40E-28	-1.18
adi_EST_assem_238	COX7r_cytochrome c oxidase subunit 7a-related mitochondrial (<i>Homo sapiens</i>) (IV)	O14548	5.55E-17	-1.22
adi_EST_assem_1430	tar1_protein tar1 (<i>Saccharomyces cerevisiae</i>)	Q8TGM6	2.43E-15	-1.4
adi_EST_assem_19733	ptps_6-pyruvoyl tetrahydrobiopterin synthase (<i>Homo sapiens</i>)	Q03393	1.02E-54	-1.16
adi_EST_assem_11343	pdpr_pyruvate dehydrogenase phosphatase regulatory mitochondrial (<i>Bos taurus</i>)	O46504	2.84E-69	-1
adi_EST_assem_5633	emre_essential mcu mitochondrial (<i>Xenopus tropicalis</i>)	Q28ED6	3.44E-15	-1.07
adi_EST_assem_388	tom7_mitochondrial import receptor subunit tom7 homolog (<i>Homo sapiens</i>)	Q9P0U1	3.37E-16	-1.07
adi_EST_assem_11093	tspo_translocator protein (<i>Mus musculus</i>)	P50637	1.78E-41	-1.05

Table 2.9 Down-regulation of *A. digitifera* DEGs (n=8) (FDR \leq 0.05) likely involved in Endoplasmic Reticulum functions

Transcript ID	Best BLAST hit	UniProt ID	E-Value	logFC
adi_EST_assem_3915	sc61b_protein transport protein sec61 subunit beta (<i>Pongo abelii</i>)	Q5RB31	3.07E-16	-1.13
adi_EST_assem_9322	got1b_vesicle transport protein got1b (<i>Homo sapiens</i>)	Q9Y3E0	6.73E-35	-1.06
adi_EST_assem_4958	sc61g_transport protein sec61 subunit gamma (<i>Ciona intestinalis</i>)	Q8I7D9	1.22E-19	-1.02
adi_EST_assem_1236	ssrg_translocon-associated protein subunit gamma (<i>Rattus norvegicus</i>)	Q08013	3.98E-70	-1.02
adi_EST_assem_22402	DPM2_Dolichol phosphate-mannose biosynthesis regulatory protein (<i>Homo sapiens</i>)	O94777	7.12E-16	-1.30
adi_EST_assem_13740	ktap2_keratinocyte-associated protein 2 (<i>Homo sapiens</i>)	Q8N6L1	6.77E-29	-1.14
adi_EST_assem_8065	ssrd_translocon-associated protein subunit delta (<i>Mus musculus</i>)	Q62186	1.43E-24	-1.13
adi_EST_assem_1945	tppc3_trafficking protein particle complex subunit 3 (<i>Gallus gallus</i>)	Q5ZI57	3.30E-96	-1.10

2.4.5 Focus on symbiosis-related coral genes

The GO database is strongly biased towards well-described processes in model organisms. Thus, the GO analysis can provide only a general overview of the host genes involved in establishment of coral-algal symbiosis. Consequently, another phase of data analysis was undertaken. Swiss-Prot annotated DEGs at 4 h post-infection were grouped into categories based on literature searches highlighting functions likely involved in the cnidarian-algal symbiosis. At 4 h post-*Symbiodinium* infection, coral larvae exhibited strong gene expression responses for genes related to: recognition receptors, cell adhesion, symbiosome formation, regulation of transcription, cell cycle, innate immunity, apoptosis and stress response.

2.4.5.1 The coral host might recognize symbionts via MAMP-PRR interactions

Animal innate immunity functions to detect and interact with microbes in the surrounding environment. The initial interaction between the coral host and *Symbiodinium* must involve pattern recognition receptors (Weis *et al.* 2008; Davy *et al.* 2012). Consequently genes involved in microbe recognition are of particular interest in the context of symbiosis onset. Host cells express a wide array of proteins, known as pattern recognition molecules or pattern recognition receptors (PRRs) that recognize

signature microbial compounds, known as microbe-associated molecular patterns (MAMPs) that occur on microbe surfaces, such as peptidoglycans, glycans and lipopolysaccharides (Davy *et al.* 2012). MAMP-PRR interactions start signaling cascades that launch downstream pathways and ultimately either attempt to eliminate pathogenic invaders or allow tolerance of mutualistic organisms. Lectins are diverse group of PRRs and are known to have roles in recognition during cnidarian-*Symbiodinium* interactions (Lin *et al.* 2000; Wood-Charlson *et al.* 2006). However, the identities of the interacting molecules are unclear. In this context, the up-regulation of the mannose receptor 2 (MRC2) gene, a C-type lectin, at 4 h post-infection might be particularly significant (Table 2.10, Figures 2.10 and 2.13)

2.4.5.2 Role of cell adhesion proteins during recognition

Cell adhesion proteins have been shown to be involved in cell signaling processes during cnidarian-dinoflagellate recognition process (Davy *et al.* 2012). Six proteins likely involved in cell adhesion were differentially expressed (Table 2.10 and Figure 2.10) in *Symbiodinium*-infected larvae at the 4-h time point, a number of which (fibrillin-1, nidogen-1, collagen alpha-6 chain, and calumenin) were up-regulated 2.6-, 1.77-, 1.12-, and 1.05-fold, respectively. At the same time, genes encoding vitamin K epoxide reductase complex subunit 1 (VKOR) and calponin-3 were down-regulated 1.2 and 1.1 fold, respectively. Calumenin (Ca² binding EF hand protein) and VKOR participate in regulating the gamma-carboxylation of a range of targets that includes fasciclin I proteins (Coutu *et al.* 2008). The differential expression of calumenin and VKOR is significant in the context of the establishment of symbiosis. It has been suggested that both proteins are involved in modifying the adhesion protein Sym32, a cnidarian fasciclin thought to play a key role in host-*Symbiodinium* interactions (Ganot *et al.* 2011; Reynolds *et al.* 2000).

Table 2.10 Differential expression of *A. digitifera* DEGs (n=17) (FDR ≤ 0.05) likely involved in pattern recognition and cell adhesion. Up-regulated (n=10) and down-regulated (n=7) genes are shaded blue and red respectively. Data for these genes are shown in Figure 2.10.

Transcript ID	Best BLAST hit	UniProt ID	E-Value	logFC
adi_EST_assem_3458	FBN1_fibrillin-1 (Bos taurus)	P98133	7.36E-23	2.60
adi_EST_assem_9121	GDF7 (Homo sapiens)	Q7Z4P5	9.03E-22	1.94
adi_EST_assem_9676	C-type mannose receptor 2_ MRC2 (<i>Mus musculus</i>)	Q64449	5.62E-18	1.17
adi_EST_assem_5646	Von willebrand factor d and EGF domain-containing protein (<i>Homo sapiens</i>)	Q8N2E2	4.40E-41	1.08
adi_EST_assem_3713	Macoilin- transmembrane protein 57 (<i>Mus musculus</i>)	Q7TQE6	9.51E-161	1.08
adi_EST_assem_5440	Reversion-inducing cysteine-rich protein with kazal motifs_ RECK (<i>Homo sapiens</i>)	O95980	0	1.20
adi_EST_assem_9498	Polypeptide n-acetylgalactosaminyl transferase GALT5 (<i>Drosophila melanogaster</i>)	Q6WV17	2.36E-171	1.17
adi_EST_assem_8472	Nidogen-1_ NID1 (<i>Homo sapiens</i>)	P14543	3.77E-88	1.77
adi_EST_assem_5774	Collagen alpha-6 chain_ CO6A6 (<i>Homo sapiens</i>)	A6NMZ7	2.87E-67	1.12
adi_EST_assem_9424	Calumenin_Calu (<i>Homo sapiens</i>)	O43852	2.68E-51	1.05
adi_EST_assem_2644	Uncharacterized protein C9orf135 (<i>Homo sapiens</i>) glycoprotein	Q5VTT2	1.71E-46	-1.20
adi_EST_assem_8996	Upf0565 protein c2orf69 homolog (<i>Danio rerio</i>) glycoprotein	A0JMH2	3.15E-67	-1.20
adi_EST_assem_7512	OSTeonectin-related_ SPARC (<i>Caenorhabditis elegans</i>)	P34714	1.37E-39	-1.20
adi_EST_assem_10377	lamin-b receptor_ LBR (<i>Mus musculus</i>)	Q3U9G9	2.05E-04	-1.10
adi_EST_assem_1377	Gastrointestinal growth factor XP4 (<i>Xenopus laevis</i>)	Q00223	1.06E-15	-1.70
adi_EST_assem_17327	Vitamin k epoxide reductase complex subunit 1 VKOR1 (<i>Bos taurus</i>)	Q6B4J2	1.21E-23	-1.20
adi_EST_assem_5414	Calponin-3_CNN3 (<i>Rattus norvegicus</i>)	P37397	2.18E-19	-1.10

GALT5= polypeptide n-acetyl galactos aminyl transferase, GDF7= growth differentiation factor

7

2.4.5.3 Genes involved in vesicular trafficking likely used in symbiosome formation

Most cnidarian host cells initially acquire *Symbiodinium* cells through phagocytosis. However, instead of being killed by lysosomes fused to phagosomes, symbionts somehow prevent the maturation of the phagosomes and persist in symbiosomes (Davy *et al.* 2012). The normal pathway of phagosomal maturation involves sequential recruitment of specific proteins, including Rab GTPases, at precise times onto the phagosomal membrane (Kinchen & Ravichandran 2008). Many Rab homologs have been characterized in the anemone *Aiptasia pulchella* and found to localize in specific patterns during phagosomal maturation, suggesting that *Symbiodinium* cells modulate phagosome-lysosome fusion to persist in the symbiosome state (Chen *et al.* 2004a; Chen *et al.* 2005; Hong *et al.* 2009a). In this context, the expression of six genes encoding proteins with known or potential roles in vesicle trafficking, specially the processing of early endosome, is significant in *Symbiodinium*-infected larvae at the 4-h time point (Table 2.11, symbiosome formation and Figures 2.10 and 2.13). This category of genes comprises two distinct members of the RAS superfamily of GTPases, with 1.8- and 1.04- fold increase in expression. Other up-regulated genes (amyotrophic lateral sclerosis2 (ALS2), GTP-binding protein RIT1, TBC1 domain family member 9 and rabenosyn-5) have effector activities on Rab proteins. Specifically, rabenosyn-5 is a Rab4/Rab5 effector protein that acts in early phagosome formation and membrane trafficking of recycling endosomes. The small GTPase Rab5 is found on early endosomes and is an early endosomal marker, while Rab7 is mostly found on late endosomes (Gruenberg & van der Goot 2006). Moreover, Hong *et al.* (2009b) provided experimental evidence that the establishment of the symbiosome in host cells of the sea anemone, *Aiptasia pulchella*, involves retention and exclusion of specific Rab proteins.

Table 2.11 Up-regulation of *A. digitifera* DEGs (n=6) (FDR \leq 0.05) likely involved in vesicular trafficking and symbiosome formation. Data for these genes are also shown in Figure 2.10 and in the central panel of Figure 2.13 (symbiosis establishment).

Transcript ID	Best BLAST hit	UniProt ID	E-Value	logFC
adi_EST_assem_26789	Ras-related and estrogen-regulated growth inhibitor_RERG (<i>Bos taurus</i>)	Q0VCJ7	1.41E-15	1.8
adi_EST_assem_22740	RBNS5 (<i>Mus musculus</i>)	Q80Y56	8.21E-85	1.63
adi_EST_assem_27206	GTP-binding protein RIT1 Ras-like without CAAX (<i>Mus musculus</i>)	P70426	8.52E-39	1.37
adi_EST_assem_12817	ALS2 (<i>Rattus norvegicus</i>)	P0C5Y8	0.00E00	1.07
adi_EST_assem_5272	Ras-related and estrogen-regulated growth inhibitor RERG (<i>Mus musculus</i>)	Q8R367	1.99E-33	1.04
adi_EST_assem_559	TBC1 domain family member 9_TBCD9 (<i>Homo sapiens</i>)	Q6ZT07	0.00E00	1

RBNS5=rabenosyn-5, ALS 2= alsin amyotrophic lateral sclerosis 2

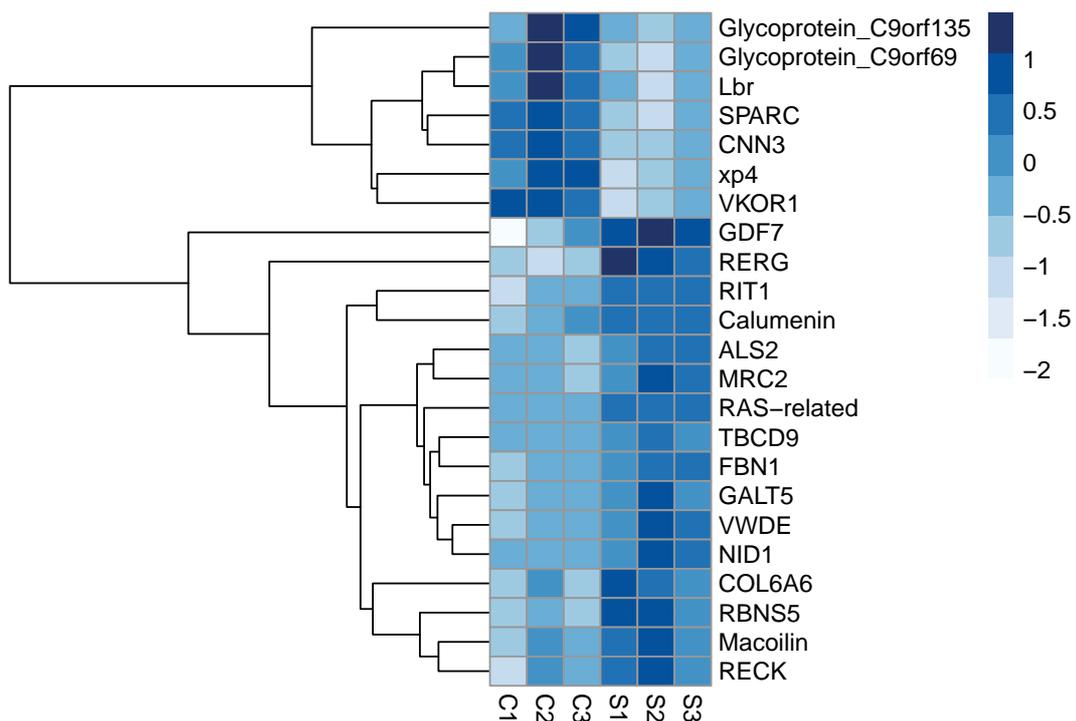


Figure 2.10 Heat map of annotated DEGs (FDR \leq 0.05) likely involved in pattern recognition, cell adhesion and vesicle trafficking in *Symbiodinium*-infected (S1, S2, S3) and control larvae (C1, C2, C3) at 4 h post-infection. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Symbiodinium*-infected samples against the control at 4 h post-infection. Expression values are log₂-transformed and then median-centered by transcript. The blue scale represents the relative expression values as log₂ (fold-change).

2.4.5.4 Transcription factors and epigenetic tags for silencing were affected

Transcription factors are proteins that contain DNA-binding domains and function to regulate levels and timing of gene expression, thus they have important biological roles in many cellular processes. Complex changes occur among the transcription machinery during the initial interaction of larvae with competent symbionts. Eleven genes encoding transcription factors and proteins implicated in silencing were differentially expressed; six genes with transcription activation function were up-regulated while the other five genes were down-regulated (Table 2.12 and Figures 2.11, 2.13),.

Genes encoding transcription repressors and epigenetic silencing factors were also differentially expressed (Tables 2.12). Five genes were up-regulated including: *krueppel-like factor 5 (KLF 5)* (a transcriptional repressor), juxtaposed with another zinc finger protein (*JAZF1*), histone-lysine N-methyl-transferase *PRDM6* (which methylates H4K20 in vitro, H4K20 methylation being a specific epigenetic mark for transcriptional repression) and the heterochromatin protein BAH domain-containing (*BAHD*) protein 1. Other changes imply a general suppression of replication, for example, down-regulation of the core histones H2A and H4 and of *SETD8-A* (a histone-lysine N- methyl transferase). Moreover, the observed down-regulation of the TFIID subunit *TAF10* gene is consistent with suppression of transcription. However, other data are inconsistent with this; for example, the observed down-regulation of *NC2B* (a negative transcription regulator) and up-regulation of *HIRA*, highly conserved chaperone that places H3.3 in nucleosomes; H3.3 is typically found associated with active chromatin.

Table 2.12 Differential expression of *A. digitifera* DEGs (n=21) (FDR ≤ 0.05) likely involved in transcription regulatory machinery. Up- and down-regulated genes are shaded blue and red respectively

Transcript ID	Best BLAST hit	UniProt ID	E-Value	logFC
adi_EST_assem_7419	ETS2_protein c-ets-2 (<i>lytechinus variegatus</i>)	P29773	1.14E-15	2.2
adi_EST_assem_8429	SPDEF_sam pointed domain-containing ets transcription factor (<i>Mus musculus</i>)	Q9WTP3	3.61E-19	1.54
adi_EST_assem_6754	TF3C5_general transcription factor 3c polypeptide 5 (<i>Mus musculus</i>)	Q8R2T8	1.06E-97	1.08
adi_EST_assem_6197	GRHL2_grainyhead-like protein 2 homolog (<i>Mus musculus</i>)	Q8K5C0	1.39E-79	1.02
adi_EST_assem_4854	TFE3_transcription factor e3 (<i>Bos taurus</i>)	Q05B92	4.23E-54	1.01
adi_EST_assem_5829	FOSL1_fos-related antigen 1 (<i>Rattus norvegicus</i>)	P10158	3.19E-09	1.89
adi_EST_assem_21329	KLF5_kruempel-like factor 5 (<i>Homo sapiens</i>)	Q13887	3.51E-26	2.34
adi_EST_assem_14784	JAZF1_juxtaposed with another zinc finger protein 1 (<i>Mus musculus</i>)	Q80ZQ5	1.69E-32	1.67
adi_EST_assem_6801	PRDM6_histone-lysine n-methyltransferase prdm6 (<i>Bos taurus</i>)	A6QPM3	3.41E-95	1.17
adi_EST_assem_945	HIRA_protein hira Tuple1/HirA (<i>Takifugu rubripes</i>)	O42611	0	1.04
adi_EST_assem_3212	BAHD1_BAH domain-containing protein 1 (<i>Mus musculus</i>)	Q497V6	5.78E-47	1
adi_EST_assem_7694	TONSL_tonsoku-like protein (<i>Homo sapiens</i>) NF-kappa-B inhibitor-like protein 2	Q96HA7	3.00E-68	1.3
adi_EST_assem_10387	NC2B_protein dr1 (<i>Rattus norvegicus</i>)	Q5XI68	4.03E-53	-1.05
adi_EST_assem_4175	HES1_transcription factor hes-1 (<i>Homo sapiens</i>)	Q14469	7.84E-23	-1.67
adi_EST_assem_5983	HES1B_transcription factor hes-1-b (<i>Xenopus laevis</i>)	Q8AVU4	5.29E-24	-1.28
adi_EST_assem_15623	SOX14_transcription factor sox-14 (<i>Danio rerio</i>)	Q32PP9	8.06E-21	-1.22
adi_EST_assem_26515	chur_protein churchill (<i>Bos taurus</i>)	Q2HJG7	2.53E-35	-1.30
adi_EST_assem_14524	TAF10_transcription initiation factor TFIID subunit 10 (<i>Mus musculus</i>)	Q8K0H5	1.27E-33	-1.47
adi_EST_assem_8308	H2A_histone h2a (<i>Acropora formosa</i>)	P35061	4.13E-59	-1.17
adi_EST_assem_2270	H4_histone h4 (<i>Dendronephthya klunzingeri</i>)	Q6LAF1	1.05E-49	-1.30
adi_EST_assem_7152	SET8A_Histone-Lysine N-methyltransferase setd8-a (<i>Xenopus laevis</i>)	Q08AY6	6.74E-56	-1.04

2.4.5.5 Genes involved in regulating the host cell cycle during symbiosis establishment

In order to understand how the host regulates cell division and proliferation, the cell cycle and its specific checkpoints that control its progression might be of interest (Davy *et al.* 2012). Ten genes implicated in control of cell proliferation and cell cycle

checkpoints were up-regulated (Table 2.13 and Figure 2.11). The fibroblast growth factor (FGF) 18 gene (which plays an important role in the regulation of cell proliferation) was up-regulated 2.9-fold. Moreover, 8 genes encoding proteins involved in checkpoint-mediated cell cycle arrest and cell cycle progression were up-regulated including TTC28 (a tetratricopeptide repeat protein), cdk5 and ab11 enzyme substrate 1, the MDM4 protein, cell division cycle protein 27, the serine threonine-protein kinase CHK1, parafibromin, MAU2 chromatid cohesion factor homolog and replication factor C subunit 5 (see Discussion).

Table 2.13 Up-regulation of *A. digitifera* DEGs (n=10) (FDR \leq 0.05) likely involved in cell cycle. Data for these genes are also shown in Figure 2.11 and in the right panel of Figure 2.13.

Transcript ID	Best BLAST hit	UniProt ID	E-Value	logFC
adi_EST_assem_18677	FGF18_Fibroblast growth factor 18 (<i>Homo sapiens</i>)	O76093	7.18E-22	2.9
adi_EST_assem_17377	TTC28_tetratricopeptide repeat protein 28 (<i>Homo sapiens</i>)	Q96AY4	1.90E-24	2.7
adi_EST_assem_31292	CABL1_cdk5 and ab11 enzyme substrate 1 (<i>Mus musculus</i>)	Q9ESJ1	5.08E-24	2.64
adi_EST_assem_2540	MDM4 protein (p53-binding protein Mdm4) <i>Homo sapiens</i>	O15151	1.33E-23	1.49
adi_EST_assem_2524	CHK1_serine threonine-protein kinase chk1 (<i>Gallus gallus</i>)	Q8AYC9	0	1.42
adi_EST_assem_6616	CDC27_cell division cycle protein 27 homolog (<i>Homo sapiens</i>)	P30260	0	1.28
adi_EST_assem_6606	CDC73_parafibromin (<i>Gallus gallus</i>)	Q5ZLM0	0	1.2
adi_EST_assem_8125	SCC4_MAU2 chromatid cohesion factor homolog (<i>Nematostella vectensis</i>)	A7SUU7	0	1.19
adi_EST_assem_375	RFC5_replication factor c subunit 5 (<i>Mus musculus</i>)	Q9D0F6	0	1.08
adi_EST_assem_4021	thoc5_THO complex subunit 5 homolog (<i>Danio rerio</i>)	Q6NY52	2.60E-83	1.01

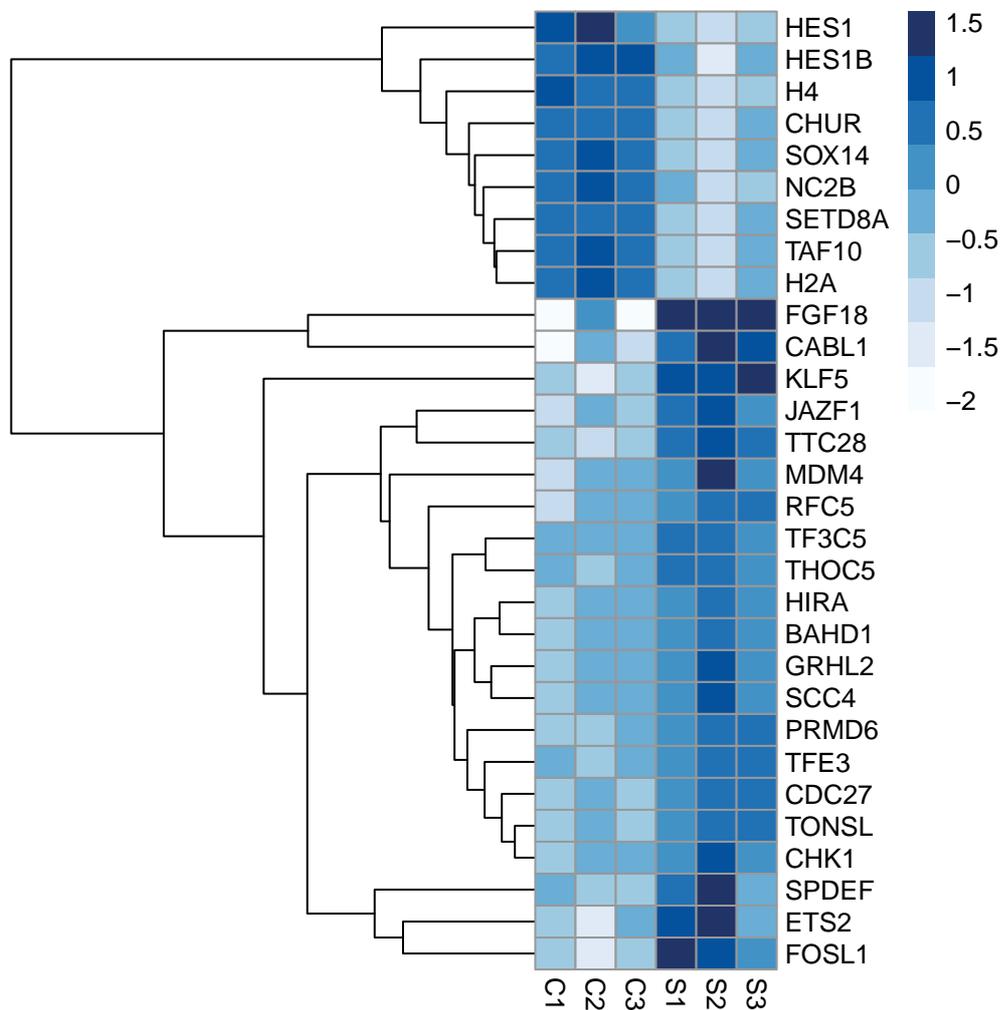


Figure 2.11 Heat map of annotated DEGs ($FDR \leq 0.05$) likely involved in transcriptional control and the cell cycle in *Symbiodinium*-infected (S1, S2, S3) and control larvae (C1, C2, C3) at 4h post infection. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Symbiodinium*-infected samples against the control at 4 h post-infection. Expression values are \log_2 -transformed and then median-centered by transcript. The blue scale represents the relative expression values as $\log_2(\text{fold-change})$.

2.4.5.6 Immune-related genes are suppressed during onset of coral-*Symbiodinium* symbiosis

Corals defend themselves against pathogenic microbes using the innate immune system, a conserved defense system in both invertebrates as well as vertebrates. The coral innate immune repertoire is highly complex and is thought to be more complicated

than that of *Hydra* and *Nematostella* (Hamada *et al.* 2013; Shinzato *et al.* 2011). Coral innate immunity is of special interest not only for self-defense and non-self recognition, but also in relation to the establishment and collapse of the coral-algal symbiosis (Davy *et al.* 2012; Weis 2008). Consistent with roles during establishment of coral-algal symbiosis, five immune-related genes were significantly differentially expressed (Table 2.14, immune response and Figures 2.12, 2.13) in *Symbiodinium*-infected larvae at the 4-h time point including: glycoprotein2 (GP2), cytokine-inducible SH2-containing protein, low affinity immunoglobulin epsilon fc receptor (FCER2), tumor necrosis factor ligand superfamily member 5 (CD40 ligand), and lipopolysaccharide (LPS)-induced tumor necrosis factor (TNF)-alpha factor homolog. Amongst these, GP2 was down regulated by 2.1-fold. The other four immune-related genes were also down regulated with 1- to 1.84-fold decrease. These findings imply that the host immune response is likely to be suppressed during establishment of coral- *Symbiodinium* symbiosis.

Table 2.14 Differential expression of *A. digitifera* DEGs (n=5) (FDR \leq 0.05) likely involved in immune response. Data for these genes are also shown in Figure 2.12 and in the left panel of Figure 2.13.

Transcript ID	Best BLAST hit	UniProt ID	E-Value	logFC
adi_EST_assem_1403	Glycoprotein 2 (zymogen granule membrane) 2 GP2 (<i>Homo sapiens</i>)	P55259	2.00E-22	-2.1
adi_EST_assem_22228	Cytokine-inducible SH2-containing protein_CISH (<i>Homo sapiens</i>)	Q9NSE2	1.85E-15	-1.83
adi_EST_assem_5336	Low affinity immunoglobulin epsilon fc receptor_FCER2 (<i>Mus musculus</i>)	P20693	1.09E-15	-1.4
adi_EST_assem_16400	CD40 ligand TNF05_Tumor necrosis factor ligand superfamily member 5 (<i>Gallus gallus</i>)	Q9I8D8	3.00E-95	-1
adi_EST_assem_4310	Lipopolysaccharide (LPS)-induced tumor necrosis factor (TNF)-alpha factor homolog_LITAF (<i>Gallus gallus</i>)	Q8QGW7	1.44E-23	-1.04

2.4.5.7 Establishment of symbiosis alters expression of genes involved in host apoptosis

Apoptosis of host cells is considered to be one of the mechanisms that plays an important role in the specificity of cnidarian-dinoflagellate associations (Lehnert *et al.* 2014; Meyer & Weis 2012; Voolstra *et al.* 2009). Eleven genes with apoptosis-related functions were significant differentially expressed between *Symbiodinium*-infected and

uninfected larvae (Table 2.15). Five genes encoding proteins implicated in apoptosis regulation were up-regulated (with 1.2- to 6.3-fold increase) including TNF receptor-associated factor 4 (TRAF4), adenosine receptor a2b, fork head box protein O3 (FOXO3), TGF-beta receptor type-1 and protein kinase c delta type. On the other hand, six genes were down-regulated with 1- to 1.2-fold decrease including: baculoviral IAP repeat-containing protein 5 (BIRC5), peroxiredoxin, defender against cell death-1 (DAD1), gamma-secretase subunit PEN2, Histone-Lysine N- methyltransferase SET8A, and testis expressed TEX261. It was very interesting to note not only that not only were large fold-changes in expression observed among up-regulated genes, but also for most of the down-regulated genes that have apoptosis inhibition functions. Based on these results, it seems that host cells are more sensitive to apoptotic stimuli during the infection (see Discussion).

Table 2.15 Differential expression of *A. digitifera* DEGs (n=11) (FDR \leq 0.05) likely involved in regulation of Apoptosis. Up- and down-regulated genes are shaded blue and red respectively. Data for these genes are also shown in Figure 2.12 and in the right panel of Figure 2.13

Transcript ID	Best BLAST hit	UniProt ID	E-Value	logFC
adi_EST_assem_30449	TRAF4 (<i>Homo sapiens</i>)	Q9BUZ4	9.67E-47	6.3
adi_EST_assem_26435	Adenosine receptor a2b_AA2BR (<i>Gallus gallus</i>)	O13076	1.10E-17	1.2
adi_EST_assem_8165	FOXO3 (<i>Homo sapiens</i>)	O43524	2.80E-30	1.58
adi_EST_assem_19964	TGFR1 (<i>Sus scrofa</i>)	Q5CD18	1.94E-173	1.9
adi_EST_assem_4877	protein kinase c delta type_KPCD (<i>Rattus norvegicus</i>)	P09215	00E00	1.3
adi_EST_assem_15760	BIRC5 (<i>Felis catus</i>)- Apoptosis Inhibitor	Q6I6F4	5.05E-36	-1.03
adi_EST_assem_7793	Peroxiredoxin- mitochondrial_PRDX5 (<i>Homo sapiens</i>)	P30044	6.09E-59	-1.2
adi_EST_assem_9908	DAD1 (<i>Sus scrofa</i>)	Q29036	2.39E-51	-1.03
adi_EST_assem_17635	Gamma-secretase subunit PEN2 (<i>Danio rerio</i>)	Q8JHF0	6.96E-45	-1.12
adi_EST_assem_7152	Histone-Lysine N- methyltransferase SET8A(<i>Xenopus laevis</i>)	Q08AY6	6.74E-56	-1.04
adi_EST_assem_18666	protein testis expressed tex261_TX261 (<i>Rattus norvegicus</i>)	Q5BJW3	3.17E-53	-1

TRAF4= TNF receptor-associated factor 4, FOXO3= Fork head box protein O3, TGFR1= TGF-beta receptor type-1, BIRC5_baculoviral IAP repeat-containing protein 5, DAD1= defender against cell death1

2.4.5.8 Genes involved in host responses to reactive oxygen species (ROS), inflammation and stress

Photosynthesis by *Symbiodinium* imposes oxidative stress on host cells and, in the symbiotic state, cnidarians typically show elevated expression of anti-oxidant genes (Richier *et al.* 2005). In this context, six genes implicated in oxidative stress responses were differentially expressed at the 4-h time point (Table 2.16). Glutathione metabolism might be of particular importance during infection, as a glutathione synthetase gene was up-regulated, whereas a glutathione S-transferase was down-regulated. Genes down-regulated included hematopoietic prostaglandin d synthase, peroxiredoxin, sulfiredoxin and soma ferritin. Moreover, three genes implicated in responses to stress were up-regulated, including map kinase-activated protein kinase 2 and two members of the serine/threonine protein kinase family: serine threonine-protein kinase1 and mitogen-activated protein kinase kinase kinase 12. In addition, two genes encoding proteins involved in inflammatory and UV-protective responses, arachidonate 5-lipoxygenase and ultraviolet-b receptor (uvr8), were up-regulated 2.05- and 1.79-fold respectively (see Table 2.16).

Table 2.16 Differential expression of *A. digitifera* DEGs (n=13) (FDR \leq 0.05) likely involved in responses to stress. Up- and down-regulated genes are shaded blue and red respectively. Data for these genes are also shown in Figure 2.12 and in the right panel of Figure 2.13.

Transcript ID	Best BLAST hit	UniProt ID	E-Value	logFC
adi_EST_assem_15984	lox5_arachidonate 5-lipoxygenase (<i>Homo sapiens</i>)	P09917	9.68E-45	2.05
adi_EST_assem_13458	gshb_glutathione synthetase (<i>Bos taurus</i>)	Q5EAC2	8.97E-134	1.05
adi_EST_assem_24281	uvr8_ultraviolet-b receptor uvr8 (<i>Arabidopsis thaliana</i>) AT5G63860	Q9FN03	2.62E-38	1.79
adi_EST_assem_10531	npff2_neuropeptide ff receptor 2 (<i>Homo sapiens</i>)	Q9Y5X5	2.98E-32	1.45
adi_EST_assem_8870	trpa1_transient receptor potential cation channel subfamily a member 1 (<i>Rattus norvegicus</i>)	Q6RI86	2.17E-113	1.2
adi_EST_assem_11763	dnaj1_drome protein homolog 1 (<i>Drosophila melanogaster</i>)	Q24133	2.54E-101	1.02
adi_EST_assem_711	oxsr1_serine threonine-protein kinase osr1(<i>Mus musculus</i>)	Q6P9R2	0	1.15
adi_EST_assem_1285	MAPK2_map kinase-activated protein kinase 2 (<i>Mus musculus</i>)	P49138	1.11E-156	1.11
adi_EST_assem_621	m3k12_mitogen-activated protein kinase kinase kinase 12 (<i>Homo sapiens</i>)	Q12852	7.24E-163	1.3
adi_EST_assem_11737	srx_sulfiredoxin (<i>Drosophila melanogaster</i>)	Q9VX10	4.63E-43	-1.02
adi_EST_assem_5095	gstol_glutathione s-transferase omega-1 (<i>Homo sapiens</i>)	P78417	1.37E-52	-1.18
adi_EST_assem_6095	HPGDS_hematopoietic prostaglandin d synthase Glutathione S-transferase (<i>Gallus gallus</i>)	O73888	3.32E-35	-1.31
adi_EST_assem_31632	fris_soma ferritin (<i>Lymnaea stagnalis</i>)	P42577	1.09E-50	-1.11

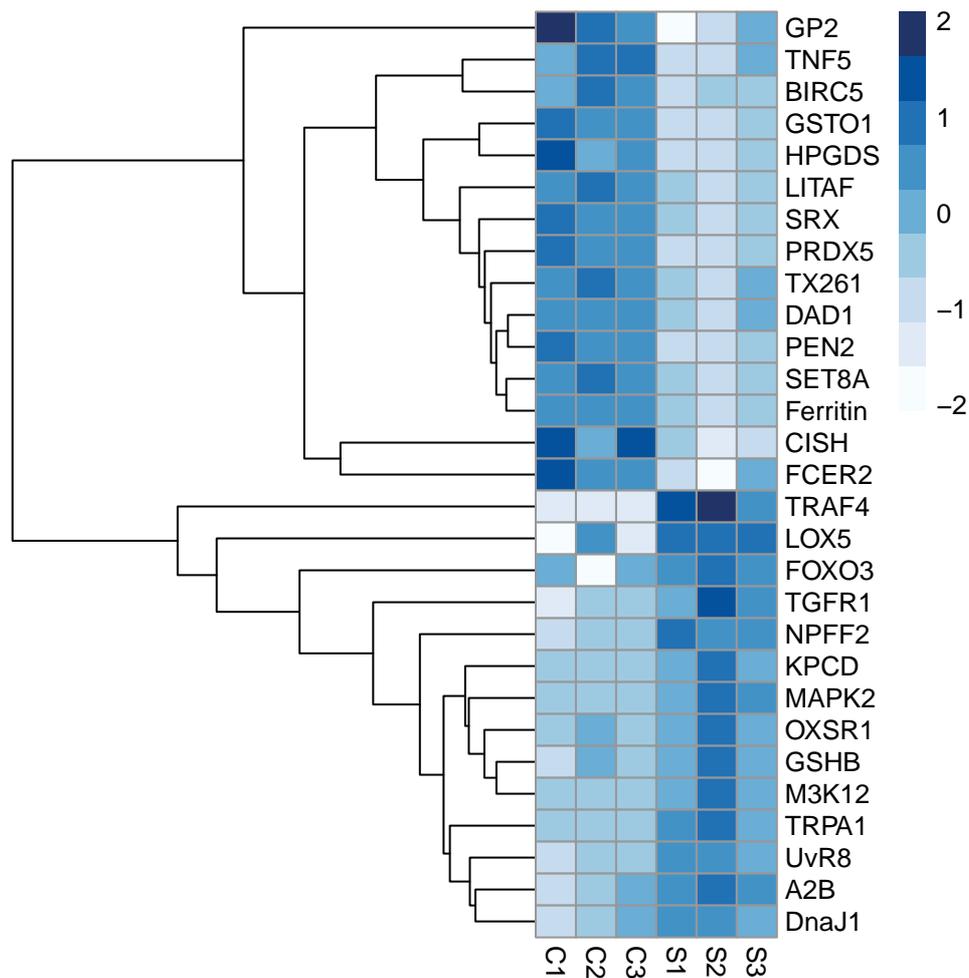


Figure 2.12 Heat map of annotated DEGs ($FDR \leq 0.05$) likely involved in immunity, apoptosis and stress response in *Symbiodinium*-infected (S1, S2, S3) and control larvae (C1, C2, C3) at 4h post infection. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Symbiodinium*-infected samples against the control at 4 h post-infection. Expression values are \log_2 -transformed and then median-centered by transcript. The blue scale represents the relative expression values as \log_2 (fold-change).

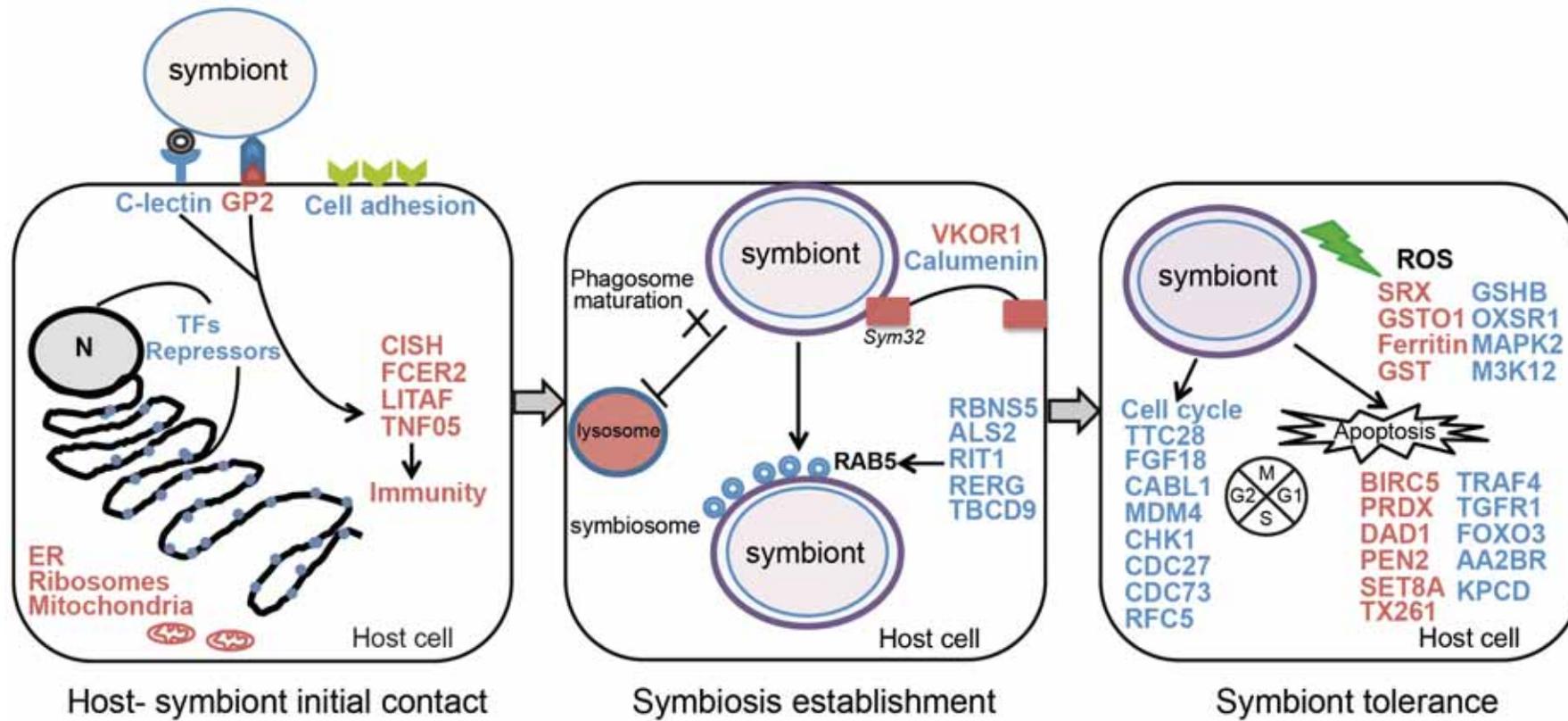


Figure 2.13 An integrative model of the molecular mechanisms during the onset and establishment of symbiosis in larvae of *Acropora digitifera*. Genes of interest and their roles are discussed; up-regulated genes are in blue text while genes that were down-regulated are in red. The initial uptake process (left panel) involves differential expression of a number of cell adhesion genes, including the up-regulation of a C-type lectin and down-regulation of the GP2 glycoprotein, as well as suppression of immunity, translation and mitochondrial processes. The establishment phase (central panel) involves blocking the maturation of the early phagosome containing the symbiont. Rabenosyn 5 and the Rab5-specific GEF ALS2 are likely to prevent the displacement of RAB5 by RAB7 and hence interfere with phagosome maturation. Differential expression of vitamin K epoxide reductase (VKOR) and calumenin is significant because both are thought to modify Sym32, a symbiosis-specific fasciclin (Ganot et al., 2011). The symbiont tolerance phase (right panel) involves complex changes required to handle the production of reactive oxygen species (ROS) by symbionts, as well as transient modifications in cell cycle regulation and the apoptotic network. The figure is modified from Meyer and Weis (2012).

2.5 DISCUSSION

In this Chapter I used Illumina RNA-Seq, a next generation sequencing technology, to address some critical questions regarding the onset and establishment of coral-*Symbiodinium* symbiosis. For example, are there any host signals involved in host-symbiont recognition? Are any of these recognition signals detectable at the transcriptional level? Does the coral host mount an immune response upon onset of symbiosis with competent *Symbiodinium* strains? How do symbionts persist in the host-derived symbiosome without being attacked by the host? Whereas previous studies used cDNA microarray technology to investigate changes in the host transcriptome during establishment of coral-algal symbiosis and detected very few changes (Schnitzler & Weis 2010; Voolstra *et al.* 2009), here I used RNAseq to profile global gene expression landscapes in *A. digitifera* larvae after inoculation with competent *Symbiodinium* (clade B1 strain). I was able to identify significant differentially expressed genes in response to symbiosis that allowed characterizing the host transcriptome response during establishment of coral-*Symbiodinium* symbiosis. A major factor in this difference is the use of high throughput next-gen sequencing, rather than cDNA microarray technology, but timing is also clearly important.

2.5.1 The coral host transcriptome showed rapid and transient changes upon symbiosis onset

The transcriptomic data identified significant changes in the coral host transcriptome at 4 h post-*Symbiodinium* infection where 1073 significant DEGs were detected in symbiotic larvae compared to uninfected control larvae whereas the coral transcriptome remained unchanged in the 12- and 48-h time points. This finding is consistent with what has been found in other cnidarian-algal symbioses (Ganot *et al.* 2011; Lehnert *et al.* 2014; Rodriguez-Lanetty *et al.* 2006) as well as other mutualistic relationships (Chun *et al.* 2008; Dale & Moran 2006; Heller *et al.* 2008). For example, Chun *et al.* (2008) identified a similar number of DEGs during initiation of the symbiosis between the squid *Euprymna scolopes* and the luminous bacterium *Vibrio fischeri*, where 781 genes were detected as differentially expressed genes.

On the other hand, few studies have investigated changes in the coral transcriptome during *Symbiodinium* infection (Schnitzler & Weis 2010; Voolstra *et al.* 2009). In the Schnitzler and Weis study, very few signals were detected in response to symbiosis between the coral *Fungia scutaria* and competent strain of *Symbiodinium*. In the previous study, the authors inoculated larvae with very high concentrations of symbionts mixed with *Artemia* to induce a feeding behavior and collected larval samples for RNA sequencing 48 h post infection. The inability to find coral signals to *Symbiodinium* infection in the later studies might be due to missing the window of symbiosis-related expression as I showed in this chapter; the coral response is subtle and transient. Another possibility is the use of cDNA microarray that might be dominated with housekeeping genes, in this case the transcriptomic results will be biased to only genes spotted onto that array.

2.5.2 Metabolic suppression during coral and Symbiodinium initial interactions

Most of the DEGs detected (73.4%) were down-regulated in infected larvae relative to controls and GO analysis revealed that many of these genes are associated with ribosomes, the rough endoplasmic reticulum and respiratory chain/ mitochondria. GO terms related to ribosomes and rough endoplasmic reticulum were significantly enriched in the down regulated genes. For example, 22 ribosomal proteins and a probable translation initiation factor were down regulated. While these results imply that protein synthesis (translation) is suppressed, the down-regulation of RNA polymerase subunits and the up-regulation of several transcription repressors, including

those imposing epigenetic marks for silencing, imply an overall down-regulation of transcription during the initial interaction between corals and *Symbiodinium*. Meyer *et al.* (2011) reported that the long-term exposure of *Acropora millepora* larvae to elevated temperatures induced the down regulation of ribosomal proteins. Also, protein synthesis was down regulated in the Caribbean coral *Orbicella faveolata* during thermal stress and bleaching (DeSalvo *et al.* 2008).

Mitochondrial metabolism appears to be down regulated during the interaction, as GO terms related to mitochondria and respiratory chain were significantly enriched among down regulated genes. Consistent with the idea that mitochondrial metabolism was suppressed, down-regulation of Tom7, a key component of the complex that transports proteins into mitochondria, was observed. It is unclear why such suppression of metabolism should occur during establishment of coral-*Symbiodinium* symbiosis. Although there are no direct precedents for these findings, Moya *et al.* (2012) reported the suppression of metabolism in *A. millepora* juveniles exposed to elevated levels of CO₂ and suggested that metabolic suppression might indirectly suppress the calcification process. Also, Nguyen and Pieters (2005) suggested that the symbionts might modulate host metabolism by modifying or suppression of their response. In the case of the Moya *et al.* study, it was suggested that acute oxidative metabolic suppression might enable the redirection of energy towards stress and immune responses, but in the present study suppression occurs in the absence of an immune response.

2.5.3 PRR-MAMP signaling was used to recognize competent Symbiodinium during infection

Many host PRRs are involved in recognition of microbial MAMPs, such as glycans and lipopolysaccharides. It is likely that some host PRRs are capable of recognition of both pathogens and potential symbionts (Kvennefors *et al.* 2008). MAMP-PRR signaling and innate immune processes in cnidarian-dinoflagellate associations are extensively reviewed by Davy *et al.* (2012). Initiation of coral-algal symbiosis and the acquisition of the symbionts are mediated by PRRs (Weis *et al.* 2008) that are utilized in the recognition of symbionts and pathogens, as in many cases of host-microbe interactions. For example, the squid-bacterial symbiosis and plant-nitrogen fixing bacteria associations are initiated through MAMP-PRR interactions (Cullimore & Denarie 2003; Nyholm & McFall-Ngai 2004). In the present case the

candidate PRR mannose receptor 2 (MRC2; a C-type lectin) was highly up-regulated during infection. This protein family includes many PRRs that recognize mannose and fucose residues on glycoproteins of bacteria and eukaryotic pathogens (Stahl & Ezekowitz 1998). The first coral PRR to be identified was a distinct C-type lectin known as millectin (Kvennefors *et al.* 2008). This protein has been shown to bind to both the pathogenic bacterium *Vibrio coralliilyticus* and the dinoflagellate *Symbiodinium*, suggesting a dual function in recognizing both pathogenic and mutualistic microorganisms (Kvennefors *et al.* 2010; Kvennefors *et al.* 2008). Consistent with a role in recognition, MRC2 and three other homologs of mammalian macrophage receptors were up-regulated in *Acropora cervicornis* after infection with white band disease (WBD), presumably indicating activation of phagocytic activity in WBD-infected corals (Libro *et al.* 2013). There are a number of reports of CTL genes being differentially expressed during the onset of cnidarian-dinoflagellate symbioses, and this gene family are candidates for key roles in the establishment of a diverse range of symbioses (Grasso *et al.* 2008; Schwarz *et al.* 2008; Sunagawa *et al.* 2009).

Previous work directly investigated the role of glycan-lectin interactions during establishment of the cnidarian-dinoflagellate symbiosis. Lin *et al.* (2000) and Wood-Charlson *et al.* (2006) altered *Symbiodinium* cell surfaces by enzymatic removal of glycans and by glycan masking before algae were inoculated into aposymbiotic sea anemones *Aiptasia pulchella* and mushroom corals *Fungia scutaria*. In both studies, glycan removal significantly decreased infection success, measured in *A. pulchella* by quantifying algal cells per tentacle and in *F. scutaria* larvae by quantifying both the percentage of larvae infected and the density of algae in larvae.

2.5.4 Role of cell adhesion genes during symbiosis establishment

Cell adhesion proteins are thought to be involved in cnidarian-dinoflagellate symbiosis during recognition (Ganot *et al.* 2011; Meyer & Weis 2012). A fasciclin I homolog known as Sym32 was found to be highly expressed in symbiotic individuals of the anemone *Anthopleura elegantissima* compared to aposymbiotic individuals (Weis & Levine 1996). Reynolds *et al.* (2000) and Ganot *et al.* (2011) later confirmed this finding. Moreover, Schwarz and Weis (2003) were able to localize Sym32 to the symbiosome surrounding resident *Symbiodinium* in the endoderm of symbiotic anemones and hypothesized it to be a component of an inter-partner signaling system.

Another gene called Calumenin was found to be the most up regulated gene in the tissue layer harboring dinoflagellate symbionts in the symbiotic anemone *Anemonia viridis* (Ganot *et al.* 2011). Genes coding for proteins likely involved in cell adhesion were differentially expressed in our data. Amongst them, calumenin and VKOR are of particular interest in symbiosis establishment, as the corresponding proteins are thought to interact with Sym32 protein, a fasciclin I protein. Ganot *et al.* (2011) have developed a model to investigate the roles of both sym32 and calumenin proteins in recognition and tolerance of symbionts. They suggested that calumenin is involved in host and symbiont recognition via regulation of Sym32 proteins, where the presence of symbionts might signal calumenin to promote gamma-carboxylation maturation of Sym32 on symbiosomes. However, how those proteins participate in signaling interactions between host and symbiont still a big question that needs further investigations. It is worth noting that Sym32 was not differentially expressed during *A. digitifera* infection, although it has been reported in many corals including *A. digitifera* (Meyer & Weis 2012).

2.5.5 The symbiosome is formed as a result of arresting phagosomal maturation

During initial contact between algae and coral host, symbionts enter the host cell by phagocytosis and are contained in host intracellular membrane-enclosed compartments known as the “symbiosomes”, which are thought to originate from host cell plasma membranes during phagocytosis. After that time, the phagosome membrane (containing symbionts) is transformed into the symbiosome membrane, rather than the contents ultimately being digested by host lysosomal enzymes during phagosome maturation (Davy *et al.* 2012). Newly formed phagosomes normally undergo “maturation” - a process that involves a series of interactions with the endocytic machinery before fusing with lysosomes (Desjardins *et al.* 1994). In mammals, phagosomal maturation involves the sequential acidification and acquisition of specific proteins, including Rab GTPases, at precise times onto the phagosomal membrane (Kinchen & Ravichandran 2008). The normal process of endocytosis involves internalization and fusion with host endosomes, resulting in the formation of an early phagosome that is characterized by the presence of Rab5, a protein critical to endocytic trafficking (Gruenberg & van der Goot 2006). Maturation to a late phagosome involves loss of Rab5 and recruitment of Rab7, a process brought about by Mon1/SAND-1 (Poteryaev *et*

al. 2010). Rab7 alone is unable to execute the early/late phagosome transition; other components are required including the LAMP-1 and -2 proteins. The content of the late phagosome is in the range pH 4.5 - 5, and its fusion with lysosomes brings about the digestion of the contents.

In mutualistic relationships, symbionts alter host cellular behavior in order to persist in the symbiosome by interrupting host membrane trafficking, so that host lysosomes fail to fuse with the vacuoles housing the symbionts (Chen *et al.* 2005). The same strategy is employed by various intracellular protozoan parasites, including apicomplexans, to persist within macrophages (Sibley 2011). Studies on the symbiotic sea anemone *Aiptasia pulchella* have provided evidence that the presence of *Symbiodinium* cells in phagosomes alters phagosomal maturation and endosomal trafficking. Cnidarian orthologs of the vertebrate Rab proteins have been shown to be present at endosomal locations corresponding to those of their vertebrate orthologs, i.e. Rab5 was located to early endosomes and Rab7 to late endosomes. Chen *et al.* (2003) and Chen *et al.* (2004b) detected human Rab5 homolog (ApRab5) on phagosomes harboring live newly ingested *Symbiodinium*, while the Rab7 homolog (ApRab7) was detected on phagosomes when heat-killed *Symbiodinium* were ingested, but not on phagosomes harboring live *Symbiodinium*. These results imply that, in the case of the symbiotic sea anemone, live *Symbiodinium* cells somehow prevent maturation beyond the early phagosome stage by stabilizing the presence of Rab5 and excluding Rab7 (on late endosome) thus preventing late phagosome formation. On the other hand, it has been shown that the mechanism that arrests phagosome maturation, permitting the establishment of symbiosis, is deactivated during thermally induced bleaching of *Pocillopora damicornis* (Downs *et al.* 2009). In this case, the symbiosome is transformed into a digestive organelle via fusion with host lysosomes, a process that involves recruitment of the late phagosome marker Rab7 and has been referred to as “symbiophagy” (Downs *et al.* 2009). In this context, the up-regulation in the present case of 6 genes, including ALS2 and rabenosyn-5, involved in the processing of early endosomes during the infection process is of particular interest. ALS2 acts as a specific GEF (guanosine nucleotide exchange factor) for Rab5 (Hadano *et al.* 2007), maintaining it in the active state and thus contributing to the stabilization of early phagosomes. Rabenosyn-5 acts early in mammalian endocytosis and plays a critical role in trafficking of endosomes for recycling (Navaroli *et al.* 2012), modulating the Rab5 GTPase activity by binding

specifically to the active form of Rab5, and thus regulating the processes of docking and fusion of endosomal membranes, motility of endosomes and intracellular signal transduction. Four other genes that may be involved in endosome trafficking were significantly up-regulated in the larvae exposed to *Symbiodinium*, however the literature on the mammalian orthologs of these is limited, thus their connection with symbiosome formation is more tenuous. The TBC1D9 / MDR1 protein is thought to act as a Rab GTPase activator, RIT1 is a representative of a highly divergent Ras subfamily, and RERG (two different transcripts up-regulated) proteins are small Ras GTPases predicted to have higher affinity for GDP than GTP.

Based on my results, it appears that similar mechanisms operate during the infection of *Acropora* and *Aiptasia* by competent *Symbiodinium* strains, but whether these developed before the deep evolutionary divergence between corals and sea anemones (Shinzato *et al.* 2011) or by convergent evolution is unclear. Many parasites (including prokaryotes) survive in host cells by subverting phagocytosis, so some similarities in how this is achieved are inevitable.

2.5.6 Regulation of the host cell cycle during symbiont tolerance

The regulation of proliferation of host cells is pivotal in maintaining a successful mutualistic relationship; synchronized division of the symbionts should be accompanied by an appropriate division of host cells, so it is not surprising that genes involved in regulating the host cell cycle should be differentially expressed in symbiosis (Davy *et al.* 2012). Wilkerson *et al.* (1988) provided evidence for the positive correlation between host and symbiont growth. Cell cycle progression is regulated by a control system of cyclin-dependent protein kinases and features a series of checkpoints where specific proteins assess the status of the cell prior to allowing progress. In this context, up-regulation of genes implicated in progression of different stages of the cell cycle (cell division cycle protein 27, parafibromin, MAU2 chromatid cohesion factor homolog, Mdm4 and CHK) is significant. Cell division cycle protein 27 is a component of the anaphase promoting complex/cyclosome that controls progression through mitosis and the G1 phase of the cell cycle. Parafibromin and MAU2 chromatid cohesion factor homolog might be involved in cell cycle progression at the stage of normal progression through prometaphase. Moreover, the Mdm4 protein functions to inhibit

cell cycle arrest by binding its transcriptional activation domain. CHK1 is a serine/threonine-protein kinase that is required for checkpoint-mediated cell cycle arrest and activation of DNA repair in response to the presence of DNA damage or unreplicated DNA. In the sea anemone *Anthopleura elegantissima*, Rodriguez-Lanetty *et al.* (2006) found sphingosine 1-phosphate phosphatase (SPPase) to be differentially expressed in the symbiotic state. The sphingosine rheostat is a key regulator in cell fate in animals (Spiegel & Milstien 2003).

2.5.7 Regulation of the host apoptotic repertoire during symbiont tolerance

The potential involvement of apoptosis and cell death has been proposed in coral bleaching and corals under stress including diseases (Ainsworth *et al.* 2007; Barshis *et al.* 2013; Dunn *et al.* 2007; Tchernov *et al.* 2011) as well as in cnidarian-dinoflagellate mutualism (Lehnert *et al.* 2014; Rodriguez-Lanetty *et al.* 2006; Voolstra *et al.* 2009). Some genes implicated in apoptosis showed differential expression in our data, including those encoding apoptosis inhibitors such as BIRC5, PRDX5, and SET8A. BIRC5 is a member of the inhibitor of apoptosis (IAP) gene family that encodes negative regulatory proteins that prevent apoptotic cell death. PRDX5, defender against cell death 1 and SET8A might act as apoptosis inhibitor as well. TX261 was the only exception among apoptosis-related down regulated genes, as it might be involved in the positive regulation of apoptosis.

Some genes that code for proteins that might act to trigger apoptosis (TRAF4, FOXO3, AA2BR, TGFR1 and KPCD) were up-regulated during *Symbiodinium* infection. Amongst these, FOXO3 is a transcriptional activator that triggers apoptosis in the absence of survival factors, resulting, for example, in neuronal cell death upon oxidative stress. Consistent with our data, Lehnert *et al.* (2014) reported the up-regulation of eight apoptosis/cell death related- genes in symbiotic relative to aposymbiotic anemone. It has previously been suggested that the increased apoptotic activity might be required so that the animal host can cope with the presence of the symbionts. Other studies have also suggested that apoptotic pathways might be important in maintaining the dynamic equilibrium between host and symbiont cell proliferation (Davy *et al.* 2012; Fitt 2000). Dunn and Weis (2009) have also suggested that apoptosis might be the basis of the winnowing mechanism thought to operate

during the post-phagocytic selection of preferred symbionts. Nevertheless, the roles of apoptosis in establishment and maintenance of a stable symbiotic relationship have not been experimentally tested.

Tumor-necrosis factor (TNF) family members and their associated proteins are key regulators of proliferation and cell survival (Aggarwal *et al.* 2012). TRAF4 was highly up-regulated in *Symbiodinium*-infected larvae. As an adapter protein and signal transducer that links members of the tumor necrosis factor receptor (TNFR) family to different signaling pathways, TRAF4 might play a role in the activation of NF-kappa-B and JNK, and in the regulation of cell survival and apoptosis. Lehnert *et al.* (2014) found a TNF-family ligand, a TNF receptor, and a TNF receptor-associated factor to be up regulated in symbiotic anemones leading to caspase-dependent apoptosis. During *Symbiodinium* infection, a type-1 TGF-beta receptor was also up-regulated. This gene family contains number of signaling proteins with key roles in cell differentiation, apoptosis and immune system in model organisms (Heldin *et al.* 2009). Detournay *et al.* (2012) provided evidence that the TGF-beta pathway was up-regulated during the onset of cnidarian-dinoflagellate mutualisms.

2.5.8 Suppression of host immunity during the onset of symbiosis

Progress has been made towards understanding the coral immune system and how corals use it to identify and respond to microbes (Miller *et al.* 2007; Mydlarz *et al.* 2010) and during establishment and breakdown of coral-algal symbiosis (Weis & Allemand 2009; Weis *et al.* 2008). In this context, five genes with immune-related functions (GP2, FCER2, CD40 ligand and LITAF) were down-regulated at 4 h post-*Symbiodinium* infection. The potential roles of these immune-related genes are discussed below. GP2 encodes a homolog of the mammalian pancreatic secretory granule membrane major glycoprotein GP2 that binds pathogens such as enterobacteria, thereby playing an important role in the (mammalian) innate immune response. FCER2 is a low-affinity receptor for immunoglobulin E (IgE) and has essential roles in regulation of IgE production and in B-cell differentiation. The CD40 ligand is a cytokine member of the tumor necrosis factor ligand superfamily that binds to TNFRSF5. It also mediates B-cell proliferation as well as IgE production. LITAF encodes lipopolysaccharide-induced TNF-alpha factor that can mediate the TNF-alpha

expression by direct binding to the promoter region of the TNF-alpha gene. It also may regulate through NFKB1 the expression of the CCL2/MCP-1 chemokine. Lipopolysaccharide is a potent stimulator of monocytes and macrophages, causing secretion of tumor necrosis factor-alpha (TNF-alpha) and other mediators. The experimental results presented here are consistent with the idea of host immune suppression during establishment of coral-algal symbiosis, as suggested in a number of previous studies (Davy *et al.* 2012; Schwarz *et al.* 2008; Voolstra *et al.* 2009; Weis *et al.* 2008). These studies suggested that the ability of the algal symbionts to suppress some components of the host immune system is a key process in maintenance of the partnership.

2.5.9 Responses to stress and ROS during symbiont tolerance

The presence of photosynthetic endosymbionts inside the endodermal cells of the coral host results in increased production of reactive oxygen species (ROS) that can cause oxidative damage to the host. Consequently, it is not surprising that the cnidarian hosts have a large repertoire of genes encoding antioxidants in order to survive the oxidative stress (Richier *et al.* 2005). Some genes coding for proteins involved in stress responses, particularly to oxidative stress, were differentially expressed in our data including proteins involved in glutathione metabolism, members of the serine/threonine protein kinase family, ferritin, arachidonate 5-lipoxygenase and the ultraviolet-b receptor *uvr8*. Glutathione has many roles in oxidative stress response, where it detoxifies H₂O₂ and acts as electron donor in a reaction catalyzed by glutathione peroxidase (Dringen 2000). MAPKs are involved in signal transduction pathways that regulate cellular responses to wide array of stimuli including pathogen infection, exocytosis, and redox signaling (Ramachandran *et al.* 2002). Arachidonate 5-lipoxygenase and ultraviolet-b receptor *uvr8* genes play roles in inflammatory and UV-protective responses. Ferritins are iron-regulated proteins that might act as an antioxidant by binding free iron (Andrews & Schmidt 2007), a gene coding for soma ferritin was down regulated in our data. Ferritin homologs were frequently detected in EST datasets of symbiotic anthozoans (Levy *et al.* 2011; Schwarz *et al.* 2008). Previous studies have found host genes involved in ROS response, such as glutathione S-transferase, to be down regulated in symbiotic to aposymbiotic individuals of sea anemone (Ganot *et al.* 2011; Rodriguez-Lanetty *et al.* 2006).

2.5.10 Conclusion

The data presented here provided the first insights into coral host genes differentially expressed during onset and establishment of symbiosis with a competent strain of *Symbiodinium*. Importantly, significant transcriptome changes were only detected early in the infection process. The transcriptome data imply that translation and oxidative metabolism are suppressed in the coral host during the infection process and altered expression of some apoptosis-related genes was also observed. The results also suggest the involvement of both partners in the establishment of the symbiosis; there is an active response on the part of the host in recognizing the symbiotic partner, but an apparent suppression of host immune responses may be initiated by the symbiont. The results are also consistent with the hypothesis that the symbiosome is a phagosome that has undergone early arrest, raising the possibility of common mechanisms in corals and symbiotic sea anemones.

2.6 REFERENCES

- Ainsworth TD, Kvennefors EC, Blackall LL, Fine M, Hoegh-Guldberg O (2007) Disease and cell death in white syndrome of Acroporid corals on the Great Barrier Reef. *Marine Biology* **151**, 19-29.
- Andrews NC, Schmidt PJ (2007) Iron homeostasis. *Annual Review of Physiology* **69**, 69-85.
- Baird AH, Guest JR, Willis BL (2009) Systematic and Biogeographical Patterns in the Reproductive Biology of Scleractinian Corals. *Annual Review of Ecology Evolution and Systematics* **40**, 551-571.
- Barshis DJ, Ladner JT, Oliver TA, *et al.* (2013) Genomic basis for coral resilience to climate change. *Proc Natl Acad Sci U S A* **110**, 1387-1392.
- Chen MC, Cheng YM, Hong MC, Fang LS (2004a) Molecular cloning of Rab5 (ApRab5) in *Aiptasia pulchella* and its retention in phagosomes harboring live zooxanthellae. *Biochemical and Biophysical Research Communications* **324**, 1024-1033.
- Chen MC, Cheng YM, Hong MC, Fang LS (2004b) Molecular cloning of Rab5 (ApRab5) in *Aiptasia pulchella* and its retention in phagosomes harboring live zooxanthellae. *Biochem Biophys Res Commun* **324**, 1024-1033.
- Chen MC, Cheng YM, Sung PJ, Kuo CE, Fang LS (2003) Molecular identification of Rab7 (ApRab7) in *Aiptasia pulchella* and its exclusion from phagosomes harboring zooxanthellae. *Biochem Biophys Res Commun* **308**, 586-595.
- Chen MC, Hong MC, Huang YS, *et al.* (2005) ApRab11, a cnidarian homologue of the recycling regulatory protein Rab11, is involved in the establishment and maintenance of the *Aiptasia-Symbiodinium* endosymbiosis. *Biochemical and Biophysical Research Communications* **338**, 1607-1616.

- Chun CK, Troll JV, Koroleva I, *et al.* (2008) Effects of colonization, luminescence, and autoinducer on host transcription during development of the squid-vibrio association. *Proc Natl Acad Sci U S A* **105**, 11323-11328.
- Coutu DL, Wu JH, Monette A, *et al.* (2008) Periostin, a member of a novel family of vitamin K-dependent proteins, is expressed by mesenchymal stromal cells. *Journal of Biological Chemistry* **283**, 17991-18001.
- Cullimore J, Denarie J (2003) Plant sciences. How legumes select their sweet talking symbionts. *Science* **302**, 575-578.
- Cumbo VR, Baird AH, van Oppen MJH (2013) The promiscuous larvae: flexibility in the establishment of symbiosis in corals. *Coral Reefs* **32**, 111-120.
- Dale C, Moran NA (2006) Molecular interactions between bacterial symbionts and their hosts. *Cell* **126**, 453-465.
- Davy SK, Allemand D, Weis VM (2012) Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiol Mol Biol Rev* **76**, 229-261.
- De'ath G, Fabricius KE, Sweatman H, Puotinen M (2012) The 27-year decline of coral cover on the Great Barrier Reef and its causes. *Proc Natl Acad Sci U S A* **109**, 17995-17999.
- DeSalvo MK, Voolstra CR, Sunagawa S, *et al.* (2008) Differential gene expression during thermal stress and bleaching in the Caribbean coral *Montastraea faveolata*. *Molecular Ecology* **17**, 3952-3971.
- Desjardins M, Huber LA, Parton RG, Griffiths G (1994) Biogenesis of Phagolysosomes Proceeds through a Sequential Series of Interactions with the Endocytic Apparatus. *Journal of Cell Biology* **124**, 677-688.
- Detournay O, Schnitzler CE, Poole A, Weis VM (2012) Regulation of cnidarian-dinoflagellate mutualisms: Evidence that activation of a host TGFbeta innate immune pathway promotes tolerance of the symbiont. *Dev Comp Immunol* **38**, 525-537.
- Downs CA, Kramarsky-Winter E, Martinez J, *et al.* (2009) Symbiophagy as a cellular mechanism for coral bleaching. *Autophagy* **5**, 211-216.
- Dringen R (2000) Metabolism and functions of glutathione in brain. *Prog Neurobiol* **62**, 649-671.
- Dunn SR, Schnitzler CE, Weis VM (2007) Apoptosis and autophagy as mechanisms of dinoflagellate symbiont release during cnidarian bleaching: every which way you lose. *Proc Biol Sci* **274**, 3079-3085.
- Dunn SR, Weis VM (2009) Apoptosis as a post-phagocytic winnowing mechanism in a coral-dinoflagellate mutualism. *Environmental Microbiology* **11**, 268-276.
- Fitt WK (2000) Cellular growth of host and symbiont in a cnidarian-zooxanthellar symbiosis. *Biological Bulletin* **198**, 110-120.
- Fitt WK, Trench RK (1983) Endocytosis of the Symbiotic Dinoflagellate *Symbiodinium-Microadriaticum* Freudenthal by Endodermal Cells of the Scyphistomae of *Cassiopeia-Xamachana* and Resistance of the Algae to Host Digestion. *Journal of Cell Science* **64**, 195-212.
- Ganot P, Moya A, Magnone V, *et al.* (2011) Adaptations to endosymbiosis in a cnidarian-dinoflagellate association: differential gene expression and specific gene duplications. *PLoS Genet* **7**, e1002187.
- Grasso LC, Maindonald J, Rudd S, *et al.* (2008) Microarray analysis identifies candidate genes for key roles in coral development. *Bmc Genomics* **9**, 540.
- Gruenberg J, van der Goot FG (2006) Mechanisms of pathogen entry through the endosomal compartments. *Nat Rev Mol Cell Biol* **7**, 495-504.

- Guppy M, Withers P (1999) Metabolic depression in animals: physiological perspectives and biochemical generalizations. *Biol Rev Camb Philos Soc* **74**, 1-40.
- Hadano S, Kunita R, Otomo A, Suzuki-Utsunomiya K, Ikeda JE (2007) Molecular and cellular function of ALS2/alsin: implication of membrane dynamics in neuronal development and degeneration. *Neurochem Int* **51**, 74-84.
- Hamada M, Shoguchi E, Shinzato C, *et al.* (2013) The complex NOD-like receptor repertoire of the coral *Acropora digitifera* includes novel domain combinations. *Molecular Biology and Evolution* **30**, 167-176.
- Harii S, Yasuda N, Rodriguez-Lanetty M, Irie T, Hidaka M (2009) Onset of symbiosis and distribution patterns of symbiotic dinoflagellates in the larvae of scleractinian corals. *Marine Biology* **156**, 1203-1212.
- Heldin CH, Landstrom M, Moustakas A (2009) Mechanism of TGF-beta signaling to growth arrest, apoptosis, and epithelial-mesenchymal transition. *Curr Opin Cell Biol* **21**, 166-176.
- Heller G, Adomas A, Li G, *et al.* (2008) Transcriptional analysis of *Pinus sylvestris* roots challenged with the ectomycorrhizal fungus *Laccaria bicolor*. *BMC Plant Biol* **8**, 19.
- Hoegh-Guldberg O (1999) Climate change, coral bleaching and the future of the world's coral reefs. *Marine and Freshwater Research* **50**, 839-866.
- Hohman TC, McNeil PL, Muscatine L (1982) Phagosome-lysosome fusion inhibited by algal symbionts of *Hydra viridis*. *J Cell Biol* **94**, 56-63.
- Hong MC, Huang YS, Lin WW, Fang LS, Chen MC (2009a) ApRab3, a biosynthetic Rab protein, accumulates on the maturing phagosomes and symbiosomes in the tropical sea anemone, *Aiptasia pulchella*. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **152**, 249-259.
- Hong MC, Huang YS, Song PC, *et al.* (2009b) Cloning and Characterization of ApRab4, a Recycling Rab Protein of *Aiptasia pulchella*, and Its Implication in the Symbiosome Biogenesis. *Marine Biotechnology* **11**, 771-785.
- Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57.
- Kinchen JM, Ravichandran KS (2008) Phagosome maturation: going through the acid test. *Nature Reviews Molecular Cell Biology* **9**, 781-795.
- Kvennefors EC, Leggat W, Kerr CC, *et al.* (2010) Analysis of evolutionarily conserved innate immune components in coral links immunity and symbiosis. *Dev Comp Immunol* **34**, 1219-1229.
- Kvennefors ECE, Leggat W, Hoegh-Guldberg O, Degnan BM, Barnes AC (2008) An ancient and variable mannose-binding lectin from the coral *Acropora millepora* binds both pathogens and symbionts. *Developmental and Comparative Immunology* **32**, 1582-1592.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25.
- Lehnert EM, Mouchka ME, Burriesci MS, *et al.* (2014) Extensive differences in gene expression between symbiotic and aposymbiotic cnidarians. *G3 (Bethesda)* **4**, 277-295.
- Levy O, Kaniewska P, Alon S, *et al.* (2011) Complex diel cycles of gene expression in coral-algal symbiosis. *Science* **331**, 175.
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.

- Libro S, Kaluziak ST, Vollmer SV (2013) RNA-seq Profiles of Immune Related Genes in the Staghorn Coral *Acropora cervicornis* Infected with White Band Disease. *Plos One* **8**.
- Lin KL, Wang JT, Fang LS (2000) Participation of glycoproteins on zooxanthellal cell walls in the establishment of a symbiotic relationship with the sea anemone, *Aiptasia pulchella*. *Zoological Studies* **39**, 172-178.
- Littman RA, van Oppen MJH, Willis BL (2008) Methods for sampling free-living Symbiodinium (zooxanthellae) and their distribution and abundance at Lizard Island (Great Barrier Reef). *Journal of Experimental Marine Biology and Ecology* **364**, 48-53.
- Meyer E, Aglyamova GV, Matz MV (2011) Profiling gene expression responses of coral larvae (*Acropora millepora*) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Molecular Ecology* **20**, 3599-3616.
- Meyer E, Weis VM (2012) Study of cnidarian-algal symbiosis in the "omics" age. *Biological Bulletin* **223**, 44-65.
- Miller DJ, Hemmrich G, Ball EE, *et al.* (2007) The innate immune repertoire in cnidaria--ancestral complexity and stochastic gene loss. *Genome Biol* **8**, R59.
- Moya A, Huisman L, Ball EE, *et al.* (2012) Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO₂-driven acidification during the initiation of calcification. *Molecular Ecology* **21**, 2440-2454.
- Muscatine L (1990) The role of symbiotic algae in carbon and energy flux in reef corals. *Dubinsky Z (ed) Ecosystems of the world, vol. 25. Coral reefs. Elsevier, Amsterdam* **25**, 75-87.
- Mydlarz LD, McGinty ES, Harvell CD (2010) What are the physiological and immunological responses of coral to climate warming and disease? *Journal of Experimental Biology* **213**, 934-945.
- Navaroli DM, Bellve KD, Standley C, *et al.* (2012) Rabenosyn-5 defines the fate of the transferrin receptor following clathrin-mediated endocytosis. *Proc Natl Acad Sci U S A* **109**, E471-480.
- Nguyen L, Pieters J (2005) The Trojan horse: survival tactics of pathogenic mycobacteria in macrophages. *Trends Cell Biol* **15**, 269-276.
- Nyholm SV, McFall-Ngai MJ (2004) The winnowing: establishing the squid-vibrio symbiosis. *Nature Reviews Microbiology* **2**, 632-642.
- Poteryaev D, Datta S, Ackema K, Zerial M, Spang A (2010) Identification of the switch in early-to-late endosome transition. *Cell* **141**, 497-508.
- Ramachandran A, Moellering D, Go YM, *et al.* (2002) Activation of c-Jun N-terminal kinase and apoptosis in endothelial cells mediated by endogenous generation of hydrogen peroxide. *Biol Chem* **383**, 693-701.
- Reynolds WS, Schwarz JA, Weis VM (2000) Symbiosis-enhanced gene expression in cnidarian-algal associations: cloning and characterization of a cDNA, sym32, encoding a possible cell adhesion protein. *Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology* **126**, 33-44.
- Richier S, Furla P, Plantivaux A, Merle PL, Allemand D (2005) Symbiosis-induced adaptation to oxidative stress. *Journal of Experimental Biology* **208**, 277-285.
- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.
- Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25.

- Rodriguez-Lanetty M, Phillips WS, Weis VM (2006) Transcriptome analysis of a cnidarian-dinoflagellate mutualism reveals complex modulation of host gene expression. *Bmc Genomics* **7**, 23.
- Schnitzler CE, Weis VM (2010) Coral larvae exhibit few measurable transcriptional changes during the onset of coral-dinoflagellate endosymbiosis. *Mar Genomics* **3**, 107-116.
- Schwarz JA, Brokstein PB, Voolstra C, *et al.* (2008) Coral life history and symbiosis: functional genomic resources for two reef building Caribbean corals, *Acropora palmata* and *Montastraea faveolata*. *Bmc Genomics* **9**, 97.
- Schwarz JA, Weis VM (2003) Localization of a symbiosis-related protein, sym32, in the *Anthopleura elegantissima*-*Symbiodinium muscatinei* association. *Biological Bulletin* **205**, 339-350.
- Shinzato C, Shoguchi E, Kawashima T, *et al.* (2011) Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**, 320-323.
- Shoguchi E, Shinzato C, Kawashima T, *et al.* (2013) Draft Assembly of the *Symbiodinium minutum* Nuclear Genome Reveals Dinoflagellate Gene Structure. *Current Biology* **23**, 1399-1408.
- Sibley LD (2011) Invasion and Intracellular Survival by Protozoan Parasites. *Immunological Reviews* **240**, 72-91.
- Spiegel S, Milstien S (2003) Sphingosine-1-phosphate: an enigmatic signalling lipid. *Nat Rev Mol Cell Biol* **4**, 397-407.
- Stahl PD, Ezekowitz RA (1998) The mannose receptor is a pattern recognition receptor involved in host defense. *Curr Opin Immunol* **10**, 50-55.
- Sunagawa S, Wilson EC, Thaler M, *et al.* (2009) Generation and analysis of transcriptomic resources for a model system on the rise: the sea anemone *Aiptasia pallida* and its dinoflagellate endosymbiont. *Bmc Genomics* **10**, 258.
- Tchernov D, Kvitt H, Haramaty L, *et al.* (2011) Apoptosis and the selective survival of host animals following thermal bleaching in zooxanthellate corals. *Proc Natl Acad Sci U S A* **108**, 9905-9909.
- Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192.
- Voolstra CR, Schwarz JA, Schnetzer J, *et al.* (2009) The host transcriptome remains unaltered during the establishment of coral-algal symbioses. *Molecular Ecology* **18**, 1823-1833.
- Weis VM (2008) Cellular mechanisms of Cnidarian bleaching: stress causes the collapse of symbiosis. *Journal of Experimental Biology* **211**, 3059-3066.
- Weis VM, Allemand D (2009) Physiology. What determines coral health? *Science* **324**, 1153-1155.
- Weis VM, Davy SK, Hoegh-Guldberg O, Rodriguez-Lanetty M, Pringle JR (2008) Cell biology in model systems as the key to understanding corals. *Trends in Ecology & Evolution* **23**, 369-376.
- Weis VM, Levine RP (1996) Differential protein profiles reflect the different lifestyles of symbiotic and aposymbiotic *Anthopleura elegantissima*, a sea anemone from temperate waters. *Journal of Experimental Biology* **199**, 883-892.
- Wilkerson FP, Kobayashi D, Muscatine L (1988) Mitotic Index and Size of Symbiotic Algae in Caribbean Reef Corals. *Coral Reefs* **7**, 29-36.

Wood-Charlson EM, Hollingsworth LL, Krupp DA, Weis VM (2006) Lectin/glycan interactions play a role in recognition in a coral/dinoflagellate symbiosis. *Cell Microbiol* **8**, 1985-1993.

Chapter 3.0 Deciphering the nature of the coral-*Chromera* association via next generation sequencing

The content of the chapter will be submitted as manuscript entitled:

Coral transcriptome profiling during *Chromera* infection provides insights into host-pathogen interactions, by Mohamed AR, Cumbo VR, Shinzato C, Chan CX, Bourne DG, Ragan MA, Satoh N, Miller DJ. 2016

Target journal: Nature Communications

Amin Mohamed wrote the entire chapter, with co-authors providing intellectual guidance in the design and implementation of the research and editorial contributions to the paper. Amin Mohamed conducted the experiments, collected and analysed the data, and produced all of the tables and figures.

3.1 ABSTRACT

Since the discovery of *Chromera velia* as a novel coral-associated microalga, this organism has attracted a great deal of interest because of its unique evolutionary position between the photosynthetic dinoflagellates and the parasitic apicomplexans. Although the literature implies that *Chromera* is a coral symbiont, the nature of this relationship remains to be established. To better understand the interaction, cultured *Chromera* (CCMP2878 strain) was used to infect larvae of the common Indo-Pacific reef-building coral, *Acropora digitifera*, in the period following the June, 2013 coral spawning event in Okinawa, Japan. The impact of *Chromera* infection on the transcriptome of coral larvae at 4, 12, and 48 h post-infection was investigated using Illumina RNA-Seq technology. The impact of *Chromera* was particularly extensive at the last time point (48 h) where 5,748 transcripts (about 16% of the transcriptome) were differentially expressed. The transcriptomic response of the coral to *Chromera* was complex and implies that host immunity is strongly suppressed, and both phagosome maturation and the apoptotic machinery modified. These responses resemble those of both vertebrate and invertebrate hosts to parasites and/or pathogens including *Plasmodium* sp. and *Mycobacterium tuberculosis*. Based on the interactions between *A. digitifera* larvae and *Chromera* at the transcriptional level, the coral-*Chromera* relationship is likely to be parasitic rather than symbiotic.

3.2 INTRODUCTION

Although the association between photosynthetic algae of the genus *Symbiodinium* and corals has been known for many years, a recent meta-analysis has implicated several lineages of apicomplexans specifically associated with corals. A variety of 16S ribosomal RNA (rRNA) sequences previously misidentified as of bacterial origin have recently been shown to originate from eight distinct novel apicomplexan-related lineages (ARLs), several of which appear to be tightly associated with coral tissues (Janouskovec *et al.* 2012). Two of these ARLs, *Chromera velia* and *Vitrella brassicaformis*, are members of the family Chromerida (Janouskovec *et al.* 2012). *Chromera* is the closest known photosynthetic relative of the apicomplexan parasites and, at the same time, related to the photosynthetic dinoflagellates including the coral symbiont *Symbiodinium* (Moore *et al.* 2008). *Chromera* was isolated for the

first time from the scleractinian coral *Plesiastrea versipora* from Sydney harbour (Moore *et al.* 2008). More recently, Cumbo *et al.* (2013) isolated *Chromera* from another scleractinian coral, *Montipora digitata* (Acroporidae), from Nelly Bay, Magnetic Island in the inner central region of the Great Barrier Reef (GBR). The previous study has established that *Chromera* can infect larvae of both *Acropora digitifera* and *A. tenuis*. Given that *Chromera* has been isolated from taxonomically distinct scleractinian corals at widely spaced locations on the east coast of Australia (Sydney harbour and the GBR), the association appears to be widespread. Whilst the photosynthetic capacity of *Chromera* is consistent with the assumption of it having a beneficial relationship with corals (symbiosis), some form of parasitism remains a possibility. Although little is known about the coral response to pathogens, in general, host transcriptional responses to potential symbionts or parasites differ markedly (Jenner & Young 2005). Whilst the innate immune repertoire of corals is surprisingly complex and vertebrate-like (Miller *et al.* 2007), few studies have addressed immune responses (Ocampo *et al.* 2015; Vidal-Dupirol *et al.* 2014; Weiss *et al.* 2013). However, the response of *Acropora* to a competent *Symbiodinium* strain is both subtle (Voolstra *et al.* 2009) and transient (Mohamed *et al.* 2016) whereas non-competent strains triggered immune responses (Voolstra *et al.*, 2009). The host transcriptional response may therefore provide clues as to the nature of the interaction between coral and *Chromera*.

To investigate the coral response to *Chromera* infection, transcriptome-wide gene expression levels in *Chromera*-infected larvae were compared to the control, uninfected larvae by mapping Illumina RNA-Seq reads onto the *A. digitifera* transcriptome assembly (Shinzato *et al.* 2011). To our knowledge this is the first investigation of the coral response to this novel apicomplexan-related microalga, *Chromera*. In this chapter, larvae of the coral *A. digitifera* were challenged with *Chromera* and the coral gene expression landscapes were profiled at 4, 12, and 48 h post-infection. In Chapter 2 I showed that the transcriptional response to a competent *Symbiodinium* strain was rapid, transient and subtle (Mohamed *et al.* 2016). On the other hand, the response to *Chromera* took place on a longer time scale and involved varied expression of larger numbers of genes. The coral response to *Chromera* was complex and involved immune suppression, phagocytosis and modulation of the endocytic pathway. The coral response to *Chromera* has some similarities with that of vertebrates to the parasite *Plasmodium* and the pathogen *Mycobacterium tuberculosis*.

The assumption that *Chromera* is a coral symbiont may therefore require re-evaluation in the light of the data presented here.

3.3 MATERIALS AND METHODS

3.3.1 *Chromera velia* culture

Chromera velia CCMP2878 (subsequently referred to as *Chromera*) was used in this experiment and was originally isolated from *Plesiastrea versipora* (Faviidae) from Sydney harbour (Moore *et al.* 2008). Cultures were grown in axenic f/2 medium (Guillard & Ryther 1962) and maintained at 25 °C under a 12/12 h light/dark cycle before they were used for inoculation of coral larvae.

3.3.2 *Coral larvae and Chromera infection experiment*

Mature *Acropora digitifera* colonies were collected in front of the Sesoko marine station, the University of the Ryukyus in Okinawa, Japan before spawning, where they were maintained in flow-through aquaria. Coral colonies spawned on the 19th of June 2013. After fertilization, embryos were raised in 0.2 µm filtered seawater (FSW) under ambient conditions to the late actively swimming planula stage. *Chromera* infection experiments were set up at 6 days post spawning. Approximately 700 larvae were distributed into 1L plastic containers each containing 700 mL of 0.2 µM FSW for each replicate; three replicates of untreated control and *Chromera*-infected larvae were used. *Chromera* cultures were washed three times in 0.2 µm FSW and added at final concentration of 5×10^3 cells/ml. Containers were held at 26 °C in a constant temperature room and under fluorescent lamps that provided light (86 ± 2 µmol photon m⁻² s⁻¹ at the surface) on a 12/12 h light/dark cycle. At 4, 12 and 48 h post-*Chromera* infection, 10 larvae from each replicate were collected and washed by pipetting in 0.2 µm FSW to ensure that no algae were attached to the larval surface. Larvae were checked under fluorescence microscopy over the course of the experiment to determine the success of the infection process.

3.3.3 Larval sampling and RNA isolation

~150 larvae were washed in 0.2 µm FSW from each replicate at 3 time points; 4, 12, and 48 h post infection and sampled with as little water carryover as possible. Samples were snap frozen immediately in liquid nitrogen and stored at -80 °C until further treatment. Total RNA was isolated from fixed larvae using TRIzol® reagent (Ambion Life Technologies, Austin, TX, USA) (according to manufacturer's instructions and as described in Chapter 2). Finally, RNA was dissolved in 40 µL of RNase-free water and stored at -80 °C. RNA quality and quantity were assessed using measurements of NanoDrop ND-1000 spectrometer (Wilmington, DE, USA). RNA integrity was checked using the electrophoretic profiling with an Agilent 2100 Bioanalyzer (Santa Clara, CA, USA).

3.3.4 High-throughput next generation sequencing

Messenger RNA (mRNA) was isolated from 1 µg high quality total RNA. A total of 17 cDNA libraries were prepared using Illumina TruSeq RNA Sample Preparation Kit (San Diego, CA, USA). cDNA libraries were sequenced using the Illumina HiSeq 2000 platform at the Okinawa Institute of Science and Technology (OIST), Japan. Sequencing produced a total of 346 million 100 bp paired-end reads from the 17 RNA-Seq libraries.

3.3.5 RNA-Seq data analysis

3.3.5.1 Read quality check and reference mapping

Reads were mapped onto the *A. digitifera* transcriptome (Shinzato *et al.* 2011) (<http://marinegenomics.oist.jp/genomes/>) using the BOWTIE mapping software version 0.12.7 (Langmead *et al.* 2009) (<http://bowtie-bio.sourceforge.net/index.shtml>) with default parameters. The alignment (bam) files were fed to RSEM software version 1.1.17 (<http://deweylab.biostat.wisc.edu/rsem/>) to generate abundance estimation data for each sample (Li & Dewey 2011). The read alignments and the reference sequences were visualized using the Integrated Genomics Viewer (IGV) software version 2.3.34

(Thorvaldsdottir *et al.* 2013) (<http://www.broadinstitute.org/igv/>). The percentages of mapped reads were obtained by using the samtools flagstat command.

3.3.5.2 Differential gene expression analysis

The R packages edgeR (Robinson *et al.* 2010) and DESeq (Anders & Huber 2010) were firstly evaluated for the differential gene expression analysis and edgeR results were considered. EdgeR measures gene expression (transcript counts) modeled with a Negative Binomial (NB) distribution and determines differential expression using empirical Bayesian estimation and exact tests based on the NB model. *Chromera*-infected samples were compared to control samples at each of the three time points; 4, 12 and 48 h post-infection. *P*-values for differential gene expression were corrected for multiple testing using the Benjamini and Hochberg's algorithm, false discovery rate (FDR) ≤ 0.05 . The plot_MA_and_Volcano function in R was also used to generate MA and volcano plots of the DEGs at the same FDR cut off.

3.3.5.3 Hierarchical clustering analysis

To study expression patterns of genes across samples, raw counts were first normalized using the TMM normalization function (genes were \log_2 -transformed and median-centered by transcript) (Robinson & Oshlack 2010) in edgeR to scale the expression values (Fragments Per Kilobase of transcript per Million mapped reads; FPKM) provided by the RSEM software across all samples. The R package heatmap3 was used to generate sample Spearman correlation and gene clustering heat maps. Clustering analysis was conducted on the 48 DEGs detected at the 4-h time point and the 1086 most highly expressed genes detected at the 48-h time point. A heat map was obtained in order to explore patterns of gene expression across samples.

3.3.5.4 Functional annotations, gene ontology (GO) and KEGG pathway enrichment analyses

To determine the functions of the up- and down-regulated genes with absolute \log_2 fold change > 1 , two stages of analysis were conducted. Initially BLASTX ($E \leq 10^{-3}$) was performed against the Swiss-Prot database, then the resulting hits were then filtered at $E \leq 10^{-10}$ by searching for close matches in *Nematostella vectensis* and/or

Acropora digitifera. GO enrichment analysis was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang da *et al.* 2009). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (Kanehisa & Goto 2000) was used to visually map the up- and down-regulated DEGs involved in common processes. The UNIPROT ACCESSIONS of the Swiss-Prot genes were used as identifiers (Huang da *et al.* 2009) for DAVID database. The same BLASTX was performed to the whole transcriptome and all the Swiss-Prot-annotated genes contributed to the background gene set for the enrichment analysis. DAVID uses the Fisher's exact test to ascertain statistically significant pathway enrichment among differentially expressed clusters relative to the background. The Benjamini-corrected *P*-value was applied with a cutoff ≤ 0.05 to filter the significantly enriched KEGG pathways and GO categories. Heat maps of specific categories of the DEGs likely involved in host-microbe interactions were generated using the R package pheatmap (<https://cran.rproject.org/web/packages/pheatmap/index.html>).

3.4 Results

3.4.1 General results and overall transcriptome changes

Illumina RNA-Seq technology was utilized to investigate coral transcriptomic changes during *Chromera* infection. *A. digitifera* larvae were inoculated with *Chromera* (Figure 3.1). Total RNA was isolated and a total of 17 cDNA libraries were sequenced using the Illumina HiSeq 2000 platform. Sequencing yielded an average of 20 million Illumina paired end (PE) reads per library (Table 3.1). An average of 34 % of reads were successfully mapped onto the *A. digitifera* transcriptome assembly (Table 3.2 and Figure 3.2). I investigated the gene expression profiles of *Chromera*-infected larvae at 4, 12, and 48 h post infection compared to controls. Analyses of the data revealed a close correlation among the expression levels of transcripts in the experimental conditions and the three biological replicates (Figures 3.3, 3.4). Moreover, multidimensional scaling (based on the top 500 DEGs that best differentiate the samples) separates the *Chromera*-infected and control samples along the x coordinate (see MSD plot Figure 3.5). These data were compared with results of a previous infection experiment with competent *Symbiodinium* strain conducted on the same species and in the same experimental format (Chapter 2). Both experiments allowed

comparison of gene expression levels in coral larvae infected with *Symbiodinium* and *Chromera* at the same time points relative to their controls.

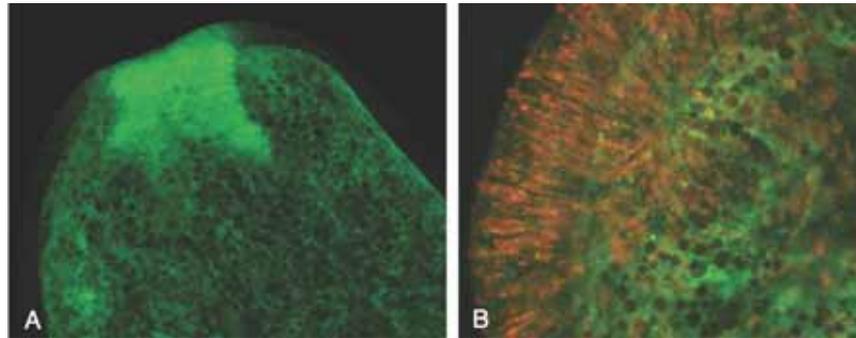


Figure 3.1 Images of *A. digitifera* larva and *Chromera* infection under fluorescent microscopy (20X objective). Part A shows control “uninfected” larva fluorescing green while part B shows *Chromera*-infected larva with red fluorescing observed during the infection process.

Table 3.1 Raw Illumina Hi-Seq sequencing reads. 17 cDNA libraries were sequenced and produced a total about 346.2 millions reads (*Chromera*-infected and uninfected larvae) in 3 time points; 4, 12 and 48 h post-infection. NA= no data. The absence of data the 12-h time point for negative control1 is due to RNA quality issues and the low number of reads in negative control 1, *Chromera*-infected 2 and 3 are consequences of low RNA yield

RNAseq library	04 h	12 h	48 h	Total # Illumina reads
<i>Chromera</i> -infected 1	26,204,349	25,557,352	26,530,067	78,291,768
<i>Chromera</i> -infected 2	25,721,449	26,639,362	2,336,789	54,697,600
<i>Chromera</i> -infected 3	31,698,400	23,800,967	2,663,580	58,162,947
Negative control 1	22,430,563	NA	2,355,166	24,785,729
Negative control 2	19,542,091	24,035,359	15,729,866	59,307,316
Negative control 3	20,912,751	25,197,830	24,849,108	70,959,689
Total # Illumina reads	146,509,603	125,230,870	74,464,576	346,205,049

Table 3.2 Percentages of Illumina reads successfully mapped onto the *A. digitifera* assembly.

Illumina RNA-Seq libraries	04 h	12 h	48 h
<i>Chromera</i> -infected 1	33.18%	36.24%	34.71%
<i>Chromera</i> - infected 2	32.98%	32.40%	30.08%
<i>Chromera</i> - infected 3	34.83%	30.08%	32.40%
Negative control-1	34.83%	NA	35.48%
Negative control-2	35.53%	39.09%	34.50%
Negative control-3	36.02%	35.87%	38.93%



Figure 3.2 A screenshot of *Acropora digitifera* transcripts and read alignments of both control and *Chromera* infection samples at the 48 h time point. Using the Integrated Genomics Viewer (IGV), the sorted BAM files containing the aligned reads were uploaded as well as the reference transcriptome sequence data.

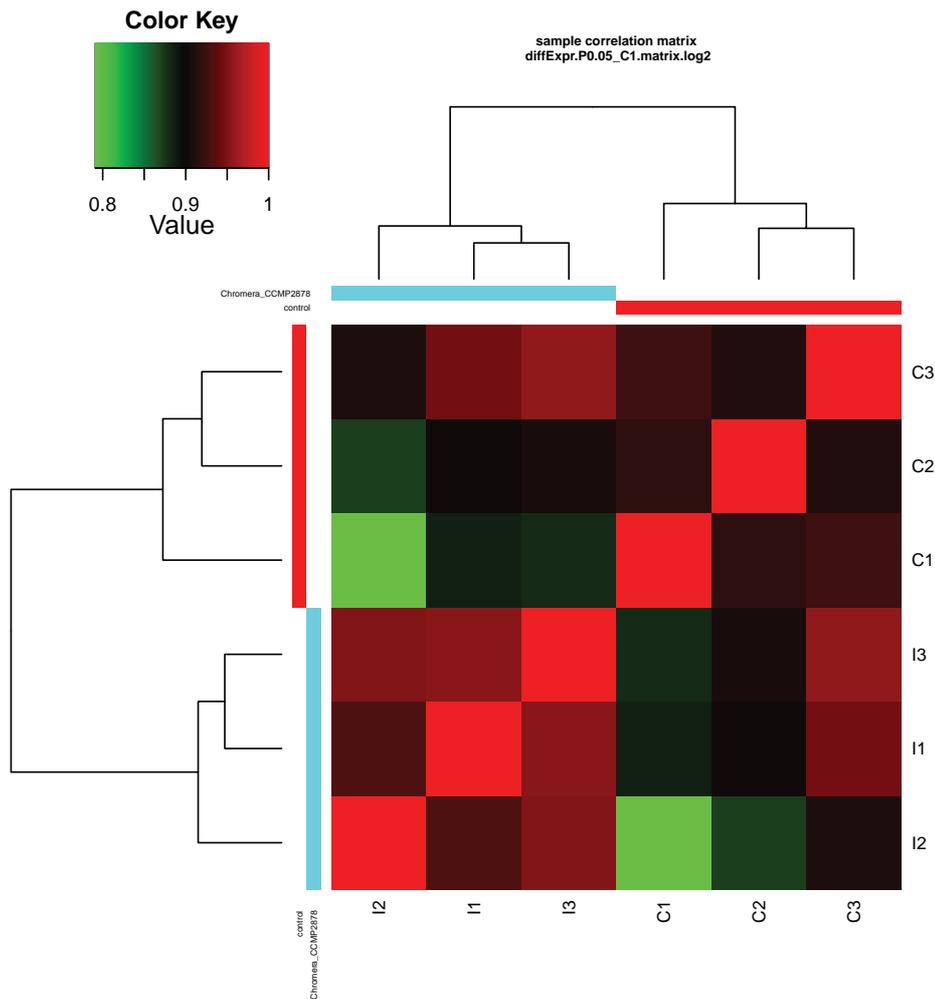


Figure 3.3 Level of agreement among the biological replicates at 4 h post *Chromera* infection. The heat map shows the hierarchically clustered Spearman correlation matrix resulting from comparing the transcript expression values (TMM-normalized FPKM) for all samples against one another. Samples clustering indicate the consistency between the biological replicates of the *Chromera* infection (samples I1, I2, I3) and negative control conditions (samples C1, C2, C3) at the 4 h time point. A color field that ranges from green to red presents the level of correlation.

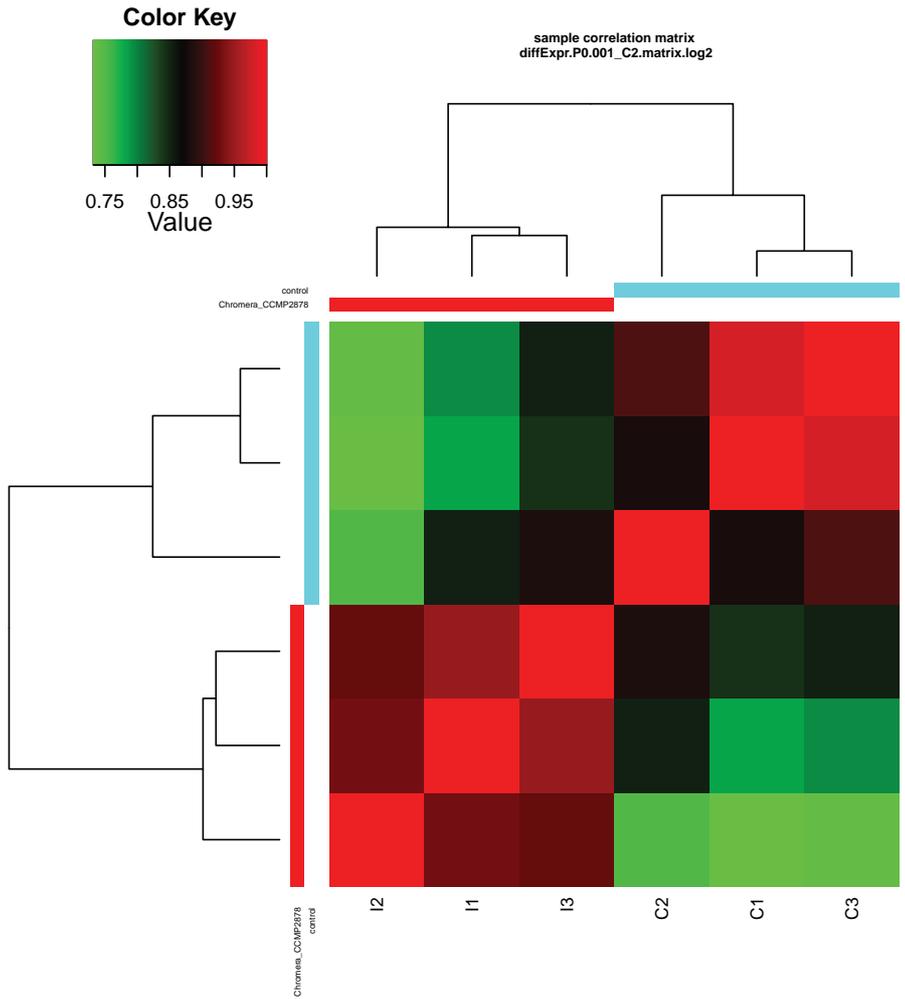


Figure 3.4 Level of agreement among the biological replicates at 48 h post *Chromera* infection. The heat map shows the hierarchically clustered Spearman correlation matrix resulting from comparing the transcript expression values (TMM-normalized FPKM) for all samples against one another. Samples clustering indicate the consistency between the biological replicates of the *Chromera* infection (samples I1, I2, I3) and negative control conditions (samples C1, C2, C3) at the 4h time point. A color field that ranges from green to red presents the level of correlation.

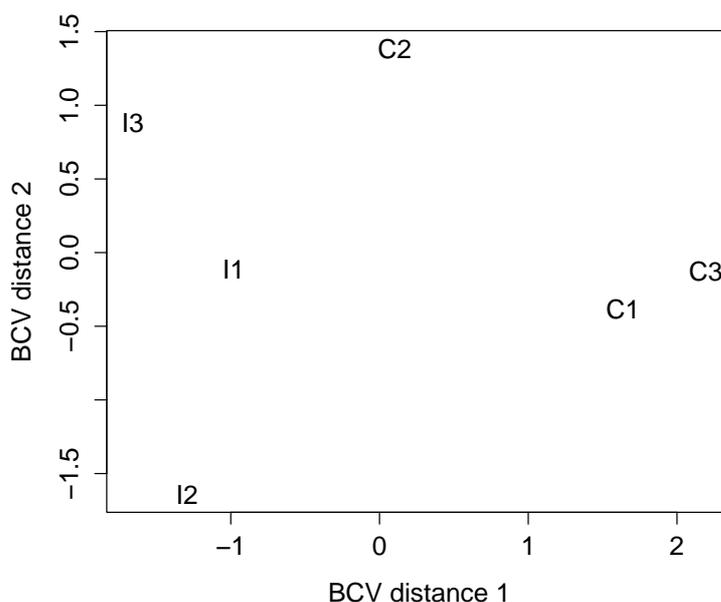


Figure 3.5 A multidimensional scaling (MDS) plot produced by edgeR showing the relationship between all the replicates of *Chromera* infection (samples I1, I2, I3) and control conditions (samples C1, C2, C3) at the 48-h time point. The distances shown on the plot are the coefficient of variation of expression between samples for the 500 genes that best distinguish the samples.

3.4.2 Differential gene expression analysis

The R packages edgeR and DESeq were first evaluated to test for differential expression and edgeR detected a higher number of DEGs at adjusted $P \leq 0.05$ than did DESeq (Figure 3.6), hence edgeR results were considered for downstream analysis. At 4 h time point only 48 DEGs (0.13% of the coral transcriptome) were down regulated at adjusted $P \leq 0.05$, logFC ranged from -1.2 to -5.8 (Figures 3.7, 3.8 and Table 3.3). However, at 48 h there were dramatic changes in the coral transcriptome, where 5,748 clusters (about 16% of the transcriptome) were differentially expressed at adjusted $P \leq 0.05$, logFC ranged from -6.9 to 5.8. The 48 h response consisted of 3,594 down- and 2,154 up-regulated genes (Figure 3.7, 3.8 and Table 3.3). On the other hand, no significant changes in the coral transcriptome were detected at 12 h at adjusted $P \leq 0.05$. Hierarchical clustering of the 48 DEGs at the 4-h time point and the 1086 most highly differentially expressed genes at the 48-h time point revealed distinctive expression profiles for *Chromera*-infected and uninfected larvae (Figures 3.9, 3.10).

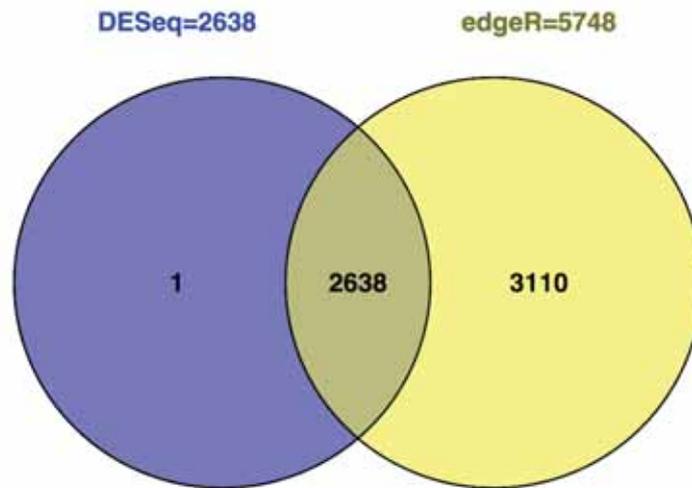


Figure 3.6 Venn diagram showing the number of differentially expressed genes (adjusted $P \leq 0.05$) in response to *Chromera* infection at the 48-h time point based on edgeR (yellow) and DESeq (blue) analyses.

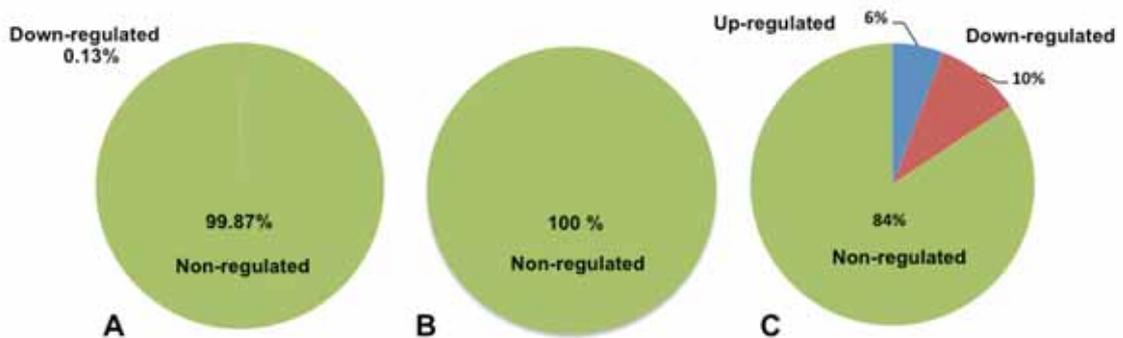


Figure 3.7 Changes in the *Acropora digitifera* transcriptome during *Chromera* infection. The pie charts show percentages of transcripts up- (blue) and down-regulated (red). 16% of the coral transcripts were differentially expressed at 48 h (Part C) (adjusted $P \leq 0.05$), while very small differences were observed at 4 h post-infection (Part A) and no changes were detected at the 12-h time point (Part B).

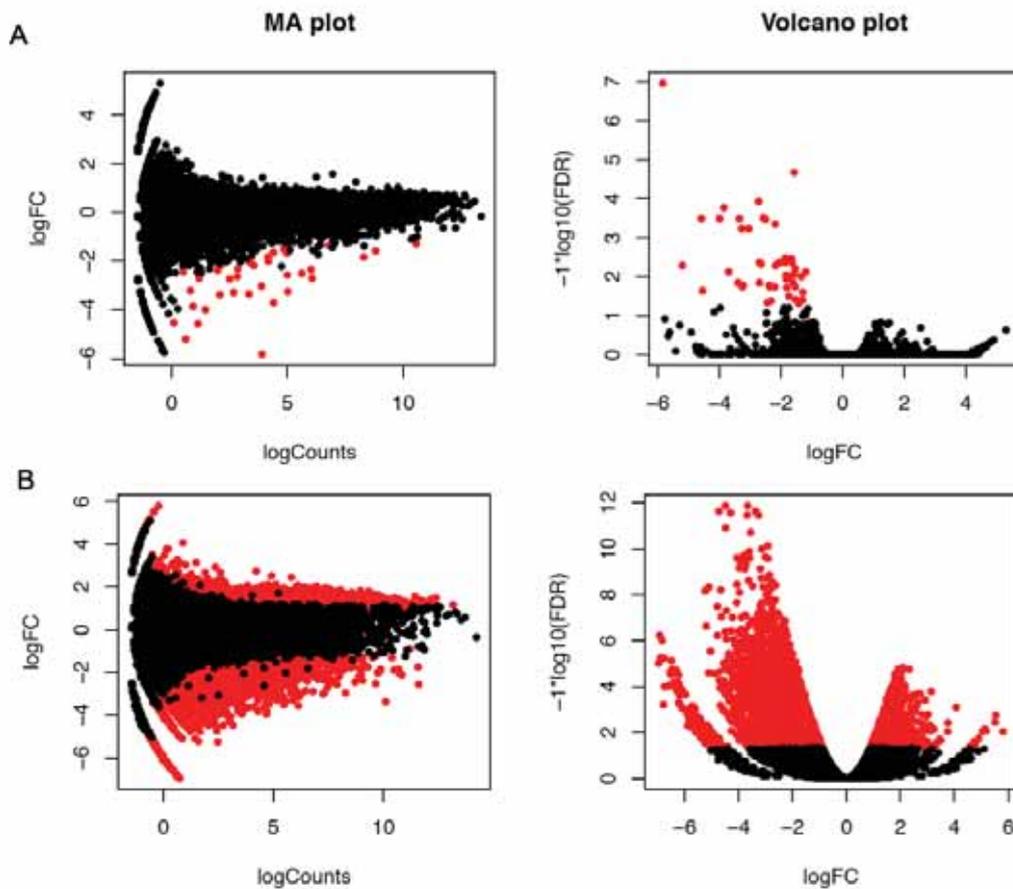


Figure 3.8 MA and Volcano plots for the gene expression profiles between *Chromera*-infected and control samples at 4 h (Part A) and 48 h (Part B). The red dots represent clusters differentially expressed at adjusted $P \leq 0.05$.

Table 3.3 Summary of differential gene expression profiles in *Chromera*-infected compared to control larvae at 4, 12, and 48 h. An adjusted $P \leq 0.05$ and E-value cut off $\leq 10^{-10}$ were used to filter differentially expressed genes and for BLASTX searches against the Swiss-Prot database

<i>Chromera</i> -infected vs control larvae			
Time point	04 h	12 h	48 h
DEGs	48 (0.13%)	0	5,748 (15.6%)
Up- regulated	0	NA	2,154 (5.8%)
Down- regulated	48	NA	3,594 (9.8%)
DEGs $>1 \log_2\text{FC}$	48	NA	5,611
Swiss-Prot-annotated DEGs	20	NA	4,004
E-value $\leq 10^{-10}$			

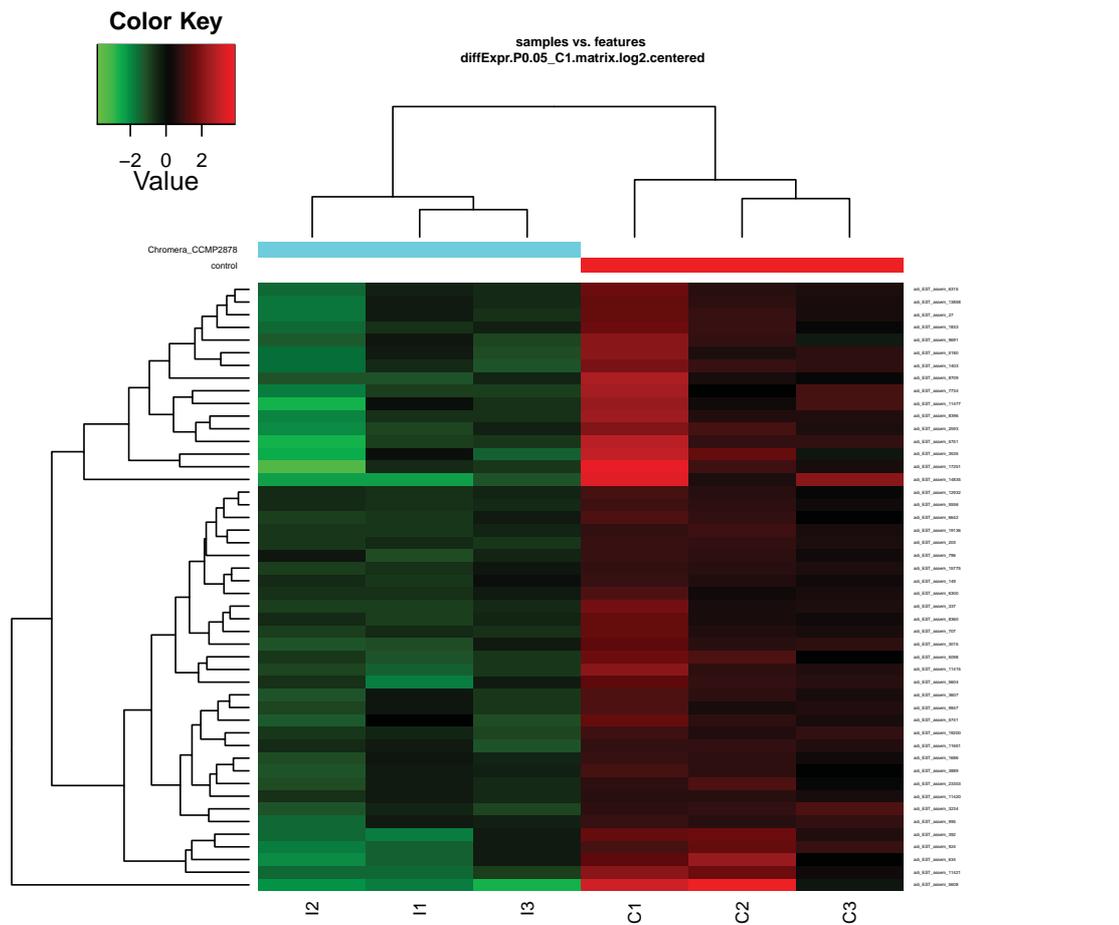


Figure 3.9 Differential gene expression profiles at 4 h post-*Chromera* infection. The heat map shows the expression profiles of 48 DEGs in *Chromera*-infected versus control samples. The hierarchical clustering obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against controls at 4 h post-infection. Expression values are \log_2 -transformed and then median-centered by transcript. Relative expression levels are shown in red (up) and green (down). Samples I1, I2, I3 are the biological replicates of *Chromera* infection while samples C1, C2, C3 are the biological replicates of the control condition.

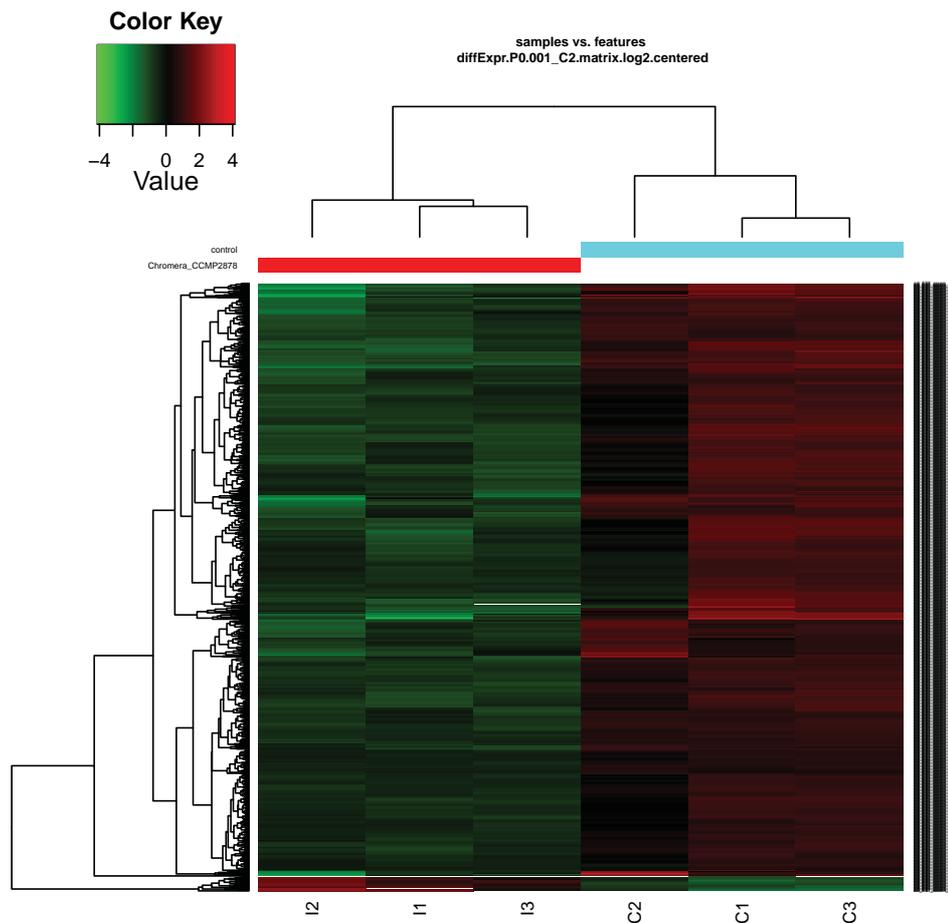


Figure 3.10 Differential gene expression profiles at 48 h post-*Chromera* infection. The heat map shows the expression profiles of the 1086 highly differentially expressed genes (FDR<0.001, 4-fold) in *Chromera*-infected versus control samples. The hierarchical clustering obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against the control at 4 h post-infection. Expression values are \log_2 -transformed and then median-centered by transcript. Relative expression levels are shown in red (up) and green (down). Samples I1, I2, I3 are the biological replicates of *Chromera* infection while samples C1, C2, C3 are the biological replicates of the control condition.

3.4.3 Functional profiles

The coral transcriptomic response to *Chromera*-infection was complex and consists of 04 h (early) and 48 h (late) responses where significant DEGs were detected. Twenty of the DEGs detected at 4h had reliable Swiss-Prot annotations (BLASTX e-values $\leq 10^{-10}$) (Table 3.3). These genes encode a number of membrane proteins, members of the immunoglobulin superfamily, and glycoproteins (Table 3.4). Amongst these the transcript, *adi_EST_assem_1403*, which encodes a homolog of the pancreatic

secretory granule membrane major glycoprotein GP2, was down regulated with 2.72-fold. Three transcripts encoding homologs of hemicentin-1, a member of the immunoglobulin superfamily, were down regulated with 4.57-, 2.7-, 1.68-fold decreases, respectively. One gene encoding fibronectin type III domain-containing protein was also down regulated with 1.53-fold.

Table 3.4 Down-regulated *A. digitifera* transcripts differentially expressed in *Chromera* infected-larvae at the 4-h time point with corrected $P \leq 0.05$. Columns correspond to coral cluster name, best BLASTX result, E-value and the \log_2 fold-change values

Cluster ID	Best BLAST hit	E-Value	logFC
adi_EST_assem_337	Hmcn1_hemicentin-1 (<i>Homo sapiens</i> ; Q96RW7)	3.92E-72	-4.57
adi_EST_assem_5098	Mfrp_membrane frizzled-related protein (<i>Acropora digitifera</i> ; aug_v2a.12459)	1.40E-13	-3.98
adi_EST_assem_3535	Cytochrome P450, Family 17, Subfamily A (<i>Gallus gallus</i> ; P12394)	3.32E-130	-3.26
adi_EST_assem_6300	Iod3_type iii iodothyronine deiodinase (<i>Gallus gallus</i> ; O42412)	3.12E-44	-3.23
adi_EST_assem_1403	Gp2_pancreatic secretory granule membrane major glycoprotein gp2 (<i>Homo sapiens</i> ; P55259)	5.49E-25	-2.72
adi_EST_assem_634	Hmcn1_humanhemicentin-1 (<i>Homo sapiens</i> ; Q96RW7)	5.12E-67	-2.7
adi_EST_assem_2593	Bhmt1_danrebetaine--homocysteine s-methyltransferase 1 (<i>Danio rerio</i> ; Q32LQ4)	1.72E-153	-2.5
adi_EST_assem_707	Apcd1_pelsiprotein apcdd1 (<i>Pelodiscus sinensis</i> ; Q5R2J4)	3.70E-73	-2.18
adi_EST_assem_5604	CUB domain-containing protein (<i>Acropora digitifera</i> ; aug_v2a.10941)	6.00E-55	-2.05
adi_EST_assem_12932	Ap2a_rattranscription factor ap-2-alpha (<i>Rattus norvegicus</i> ; P58197)	5.25E-78	-1.86
adi_EST_assem_6642	Rho1_ashgogtp-binding protein rho1 (<i>Ashbya gossypii</i> ; Q9HF54)	1.87E-38	-1.84
adi_EST_assem_8360	Hmcn1_humanhemicentin-1 (<i>Homo sapiens</i> ; Q96RW7)	6.97E-72	-1.68
adi_EST_assem_3607	Mot12_xentrmonocarboxylate transporter 12 (<i>Xenopus tropicalis</i> ; Q6P2X9)	1.82E-13	-1.65
adi_EST_assem_995	Asomp_acrmiacidic skeletal organic matrix protein (<i>Acropora millepora</i> ; B3EWY7)	2.64E-177	-1.63
adi_EST_assem_23303	Mdga1_mousemam domain-containing glycosylphosphatidylinositol anchor protein 1 (<i>Mus musculus</i> ; Q0PMG2)	1.35E-18	-1.54
adi_EST_assem_149	Fp_acrmifibronectin type iii domain-containing protein (<i>Acropora millepora</i> ; B8VIW9)	0	-1.53
adi_EST_assem_3889	Cd151_humancd151 antigen (<i>Homo sapiens</i> ; P48509)	2.02E-35	-1.42
adi_EST_assem_796	Cah2_humancarbonic anhydrase 2 (<i>Homo sapiens</i> ; P00918)	4.90E-72	-1.34
adi_EST_assem_1686	Hem1_rat5-aminolevulinat mitochondrial (<i>Rattus norvegicus</i> ; P13195)	4.84E-172	-1.28
adi_EST_assem_11420	Yrbe_bacsuuncharacterized oxidoreductase (<i>Bacillus subtilis</i> ; O05389)	4.11E-54	-1.19

The transcriptomic response of *A. digitifera* larvae to *Chromera* differed markedly from that to competent *Symbiodinium*, as described in Chapter 2. The coral transcriptome showed significant changes upon competent *Symbiodinium* infection only at the 4 h time point, where 1073 genes were differentially expressed, after which it remained unchanged until the end of the experiment (48 h). By contrast, in the case of *Chromera* infection, significant changes in gene expression were detected at both the 4- and 48-h time points (Supplementary Figure S3.1). Detecting significant gene expression changes at the 4-h time point in response to *Chromera* infection as well as to competent *Symbiodinium* infection (chapter 2) allowed direct comparison at the same time point. Only four transcripts were common to both *Chromera* and competent *Symbiodinium* infections (Figure 3.11), amongst which was *adi_EST_assem_1403*, encoding a homolog of the human pancreatic secretory granule membrane major glycoprotein GP2, which was down-regulated with 2.72- and 2.1-fold, respectively (Table 3.5) suggesting an important role for this gene during initial interactions of coral larvae with algal symbionts.

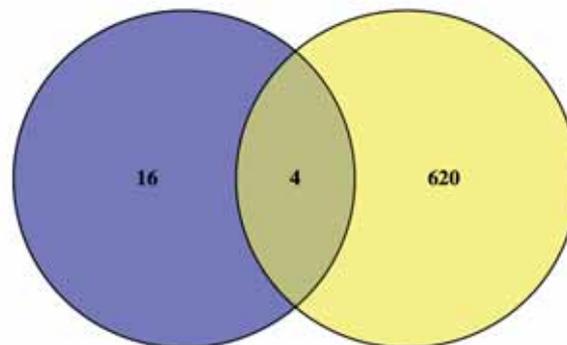


Figure 3.11 Venn diagram showing the overlap among DEGs with Swiss-Prot database annotation in *Chromera* (blue) and *Symbiodinium* (yellow) infection compared to control at the 4-h time point.

Table 3.5 Down-regulation of *A. digitifera* DEGs at adjusted $P \leq 0.05$ common to *Chromera* and *Symbiodinium* infections at the 4-h time point

Transcript ID	Best BLAST hit	logFC <i>Chromera</i> infection vs control	logFC <i>Symbiodinium</i> infection vs control
<i>adi_EST_assem_1403</i>	GP2_pancreatic secretory granule membrane major glycoprotein GP2 (<i>Homo sapiens</i> ; P55259)	-2.72	-2.1
<i>adi_EST_assem_2593</i>	BHMT1_betaine--homocysteine s-methyltransferase 1 (<i>Danio rerio</i> ; Q32LQ4)	-2.5	-1.7
<i>adi_EST_assem_707</i>	APCD1_protein apcdd1 (<i>Pelodiscus sinensis</i> ; Q5R2J4)	-2.18	-1.3
<i>adi_EST_assem_3607</i>	MO12_monocarboxylate transporter 12 (<i>Xenopus tropicalis</i> ; Q6P2X9)	-1.65	-1.3

Of the genes differentially expressed at 48 h, 4004 had reliable Swiss-Prot annotations (BLASTX E-values $\leq 10^{-10}$) (Table 3.3). Three pathways were enriched among the set of down regulated genes (using the Benjamini-corrected $P \leq 0.05$): regulation of actin cytoskeleton, ECM-receptor interaction, and focal adhesion (Table 3.6 and Supplementary Tables S3.1, S3.2, S3.3). Mapping against the KEGG database provides immediate visualization of the genes potentially involved in the enriched pathways using a background of all genes in the transcriptome (Supplementary Figures S3.2, S3.3, S3.4).

Table 3.6 Significant KEGG pathway enrichment among the set of down-regulated genes in *Chromera*-infected larvae at 48 h post-infection with corrected $P \leq 0.05$ after multiple testing correction by the Benjamini procedure

Significant KEGG pathway	KEGG pathway ID	No. of genes	Fold Enrichment
Regulation of actin cytoskeleton	hsa04810	19	2.93
Focal adhesion	hsa04510	18	2.87
ECM-receptor interaction	hsa04512	12	3.19

3.4.4 GO enrichment in the late response to *Chromera*

Down regulated genes at 48 h showed significant GO enrichment to 27 categories related to biological process including regulation of small GTPase mediated signal transduction, transcription and RNA metabolic process (Table 3.7), 43 categories related to cellular component including cytoskeleton (Table 3.8) and 45 categories related to molecular function including GTPase regulator activity (Table 3.9). The molecular function ‘GTPase regulator activity’ was the most highly over-represented category among down regulated genes. A total of 122 *A. digitifera* genes fall into this category, including homologs of RAB GTPase activating and binding proteins and members of TBC1 domain family, which are known to play important roles during early endosome formation in mammals (Supplementary Table S3.4).

Amongst up-regulated genes at 48 h, 14 GO categories related to biological process showed significant enrichment, including translation and electron transport chain (Table 3.10). Thirty six GO categories related to cellular component, including mitochondrion and ribosome (Table 3.11), and 7 categories related to molecular function including structural constituent of ribosome (Table 3.12) were significantly enriched at the 48 h time point. The cellular component ‘mitochondrion’ and the

molecular function ‘structural constituent of ribosome categories’ were the most highly over-represented categories among up-regulated genes. Several genes encoding proteins involved in ribosome functions and translation including ribosomal proteins were also up regulated (Ribosome, Supplementary Table S3.5). Many genes encoding mitochondrial ribosomal proteins and components of the electron transport chain including many ATP synthase subunits and proteins of the mitochondrial inner membrane complexes were up regulated (Mitochondrion, Supplementary Table S3.6).

Table 3.7 Biological process (BP) categories significantly enriched among the set of down-regulated genes in *Chromera* infection compared to controls at 48 h post-infection with corrected $P \leq 0.05$

Annotation term	Gene Ontology (GO) IDs	No. of genes	Fold Enrichment
Regulation of small GTPase mediated signal transduction	GO:0051056	72	2.06
Regulation of transcription and RNA metabolic process	GO:0045449, GO:0006350, GO:0006355, GO:0051252	301	1.34
Regulation of Rho protein signal transduction	GO:0035023	33	2.65
Regulation of Ras protein signal transduction and Ras GTPase activity	GO:0046578, GO:0032318	57	2.03
Phosphate, phosphorus metabolic processes, and phosphorylation	GO:0006796, GO:0006793, GO:0006468, GO:0016310	187	1.41
Regulation of GTPase activity	GO:0043087	37	2.17
Microtubule-based movement	GO:0007018	48	1.92
Chromatin modification	GO:0016568	67	1.71
Biological, cell adhesion	GO:0022610, GO:0007155	104	1.51
Transmembrane receptor protein tyrosine kinase signaling pathway	GO:0007169	44	1.95
Cell projection organization	GO:0030030	80	1.58
Microtubule-based process	GO:0007017	72	1.51
Intracellular signaling cascade	GO:0007242	163	1.28
Enzyme linked receptor protein signaling pathway	GO:0007167	53	1.61
Neuron development, differentiation	GO:0030182, GO:0048666, GO:0031175	79	1.45
Regulation of hydrolase activity	GO:0051336	47	1.62
Detection of stimulus involved in sensory perception	GO:0050906	12	2.95

Table 3.8 Cellular component (CC) categories significantly enriched among the set of down-regulated genes in *Chromera* infection compared to control at 48 h post-infection with $P \leq 0.05$

Annotation Term	Gene Ontology (GO) ID	No. of genes	Fold Enrichment
Cytoskeleton, cytoskeletal part, microtubule, microtubule cytoskeleton, microtubule organizing center	GO:0005856, GO:0044430, GO:0005874, GO:0015630, GO:0005815, GO:0005874, GO:0005875	251	1.64
Extrinsic to membrane	GO:0019898	116	1.70
Cell projection	GO:0042995	127	1.61
Intracellular non-membrane-bounded organelle	GO:0043232, GO:0043228	394	1.25
Centrosome	GO:0005813	58	1.79
Plasma membrane	GO:0005886, GO:0044459, GO:0005624	371	1.19
Cell fraction	GO:0000267	127	1.40
Cell junction	GO:0030054	74	1.54
Cell projection part	GO:0044463	54	1.68
Neuron projection	GO:0043005	61	1.60
Adherens junction	GO:0005912	27	2.04
Insoluble fraction	GO:0005626	105	1.38
Anchoring junction	GO:0070161	27	2.01
Cell leading edge	GO:0031252	29	1.93
Nuclear envelope	GO:0005635	45	1.66
Nuclear lumen	GO:0031981	199	1.23
Perinuclear region of cytoplasm	GO:0048471	44	1.66
Extracellular matrix part	GO:0044420	33	1.81
Cell cortex	GO:0005938	29	1.87
Ruffle	GO:0001726	15	2.53
Cilium	GO:0005929	37	1.71
Axon	GO:0030424	30	1.83
Lamellipodium	GO:0030027	18	2.26
Actin cytoskeleton	GO:0015629	42	1.63
Axoneme	GO:0005930	20	2.14
Vesicle	GO:0031982	101	1.34
DNA-directed RNA polymerase II, holoenzyme	GO:0016591	25	1.92
Cell-cell junction	GO:0005911	23	1.98
Nucleoplasm part	GO:0044451	91	1.35
Dendrite	GO:0030425	31	1.72
Nuclear pore	GO:0005643	24	1.85
Cytoplasmic vesicle	GO:0031410	95	1.30
Pore complex	GO:0046930	25	1.77
Basolateral plasma membrane	GO:0016323	29	1.67

Table 3.9 Molecular function (MF) categories significantly enriched among the set of down-regulated genes in *Chromera* infection compared to control at 48 h post-infection with corrected $P \leq 0.05$

Annotation Term	Gene Ontology (GO) IDs	No. of genes	Fold Enrichment
GTPase regulator activity	GO:0030695, GO:0060589, GO:0005083, GO:0005096	122	2.33
Guanyl-nucleotide exchange factor activity	GO:0005085	54	2.51
Protein kinase activity	GO:0004672	150	1.58
Enzyme activator activity	GO:0008047	70	2.01
Rho guanyl-nucleotide exchange factor activity	GO:0005089	32	2.81
Calcium ion binding	GO:0005509	198	1.45
Ras guanyl-nucleotide exchange factor activity	GO:0005088	34	2.7
Nucleotide binding, purine nucleotide binding, ribonucleotide binding, adenylyl nucleotide binding	GO:0000166, GO:0017076, GO:0032555, GO:0001883, GO:0030554, GO:0001882, GO:0005524, GO:0032559	498	1.21
Motor activity	GO:0003774, GO:0003777	59	1.95
Gtpase binding, small gtpase binding	GO:0051020, GO:0031267	29	2.5
Enzyme binding	GO:0019899	78	1.65
Ras gtpase activator activity, Ras gtpase binding,	GO:0005099, GO:0017016	29	2.69
DNA binding	GO:0003677	261	1.25
Cytoskeletal protein binding	GO:0008092	87	1.51
Transcription regulator activity	GO:0030528	165	1.33
Atpase activity	GO:0016887	104	1.44
Ion binding, metal ion binding	GO:0043167, GO:0043169, GO:0046872	673	1.11
Protein domain specific binding	GO:0019904	55	1.65
Calmodulin binding	GO:0005516	40	1.78
Protein serine/threonine kinase activity	GO:0004674, GO:0004713, GO:0004714	100	1.4
Diacylglycerol binding	GO:0019992	22	2.22
Actin binding, actin filament binding	GO:0003779, GO:0051015	62	1.54
Transcription factor binding	GO:0008134	60	1.46
Rab gtpase activator activity	GO:0005097, GO:0017137	18	2.12
Zinc ion binding	GO:0008270	322	1.14
SH3 domain binding	GO:0017124	21	1.93

Table 3.10 Biological processes (BP) categories significantly enriched among the set of up-regulated genes in *Chromera* infection compared to control at 48 h post-infection with corrected $P \leq 0.05$

Annotation term	Gene Ontology (GO) ID	No. of genes	Fold Enrichment
Translation	GO:0006412	63	2.19
Electron transport chain	GO:0022900	28	2.97
Protein targeting to membrane	GO:0006612	11	6.95
Protein transport, localization	GO:0008104, GO:0045184, GO:0015031	80	1.57
Protein import into mitochondrial inner membrane	GO:0045039, GO:0007007	6	10.74
Oxidative phosphorylation	GO:0006119	14	3.41
Intracellular transport	GO:0046907	54	1.68
Mitochondrial membrane organization	GO:0007006	9	5.08
Mitochondrion organization	GO:0007005	17	2.8
Intracellular protein transport	GO:0006886	35	1.89
Protein targeting	GO:0006605	22	2.3

Table 3.11 Cellular component (CC) categories significantly enriched among the set of up regulated genes in *Chromera* infection compared to control at 48 h post-infection with corrected $P \leq 0.05$

Annotation Term	Gene Ontology (GO) IDs	No. of genes	Fold Enrichment
Mitochondrion	GO:0005739, GO:0044429	165	2.19
Ribosome	GO:0005840	54	2.77
Mitochondrial envelope/ membrane	GO:0005740, GO:0031966	64	2.44
Ribonucleoprotein complex	GO:0030529	81	2
Mitochondrial inner membrane	GO:0005743, GO:0044455	49	2.51
Organelle membrane	GO:0031090, GO:0019866, GO:0031967	100	1.73
Respiratory chain, Respiratory chain complex I, NADH dehydrogenase complex	GO:0070469, GO:0045271, GO:0030964	18	4.67
Mitochondrial ribosome	GO:0005761	13	5.87
Organellar ribosome	GO:0000313	13	5.87
Envelope	GO:0031975	70	1.79
Ribosomal subunit	GO:0033279	24	2.93
Mitochondrial inter-membrane space	GO:0005758, GO:0042719	12	4.61
Mitochondrial small ribosomal subunit	GO:0005763, GO:0000314, GO:0016272	6	7.78
Srb-mediator complex	GO:0016592	9	4.45
Organelle envelope lumen	GO:0031970	12	3.36
Mitochondrial matrix and lumen	GO:0005759, GO:0031980	27	1.94
Small ribosomal subunit	GO:0015935	12	3.03
Proteasome core complex	GO:0005839	6	5.66
Endoplasmic reticulum	GO:0005783	72	1.38
Large ribosomal subunit	GO:0015934	12	2.7
Mitochondrial respiratory chain	GO:0005746	6	5.19
Mitochondrial large ribosomal subunit	GO:0005762, GO:0000315	6	5.19
Organellar large ribosomal subunit			
Mitochondrial respiratory chain complex I	GO:0005747	5	6.49
DNA-directed RNA polymerase II, core complex	GO:0005665	5	6.49

Table 3.12 Molecular function (MF) categories significantly enriched among the set of up-regulated genes in *Chromera* infection compared to control at 48 h post-infection with corrected $P \leq 0.05$

Annotation term	Gene Ontology (GO) ID	No. of genes	Fold Enrichment
Structural constituent of ribosome	GO:0003735	50	3.32
Structural molecule activity	GO:0005198	54	2
Peptidyl-prolyl cis-trans isomerase activity	GO:0003755	10	4.5
RNA polymerase II transcription mediator activity	GO:0016455	9	5
Cis-trans isomerase activity	GO:0016859	10	4.34
General RNA polymerase II transcription factor activity	GO:0016251	10	4.18
Hydrogen ion transmembrane transporter activity	GO:0015078	13	3.17

3.4.5 Focus on genes involved in host-microbe interactions

The GO and KEGG analyses provided a general overview of the host pathways enriched in the coral data, after which another phase of data analysis was undertaken by grouping the annotated DEGs into categories based on literature searches highlighting coral related functions likely involved in host-microbe interactions. During the late response to *Chromera* infection, strong responses were detected in categories of genes involved in immunity (PRRs, complement system, antimicrobial activities, toll-like receptors (TLRs), NOD-like receptors (NLRs) signaling pathways), the endocytic pathway (phagocytosis, early/late endosome formation, phagosome maturation, endosomal trafficking), and apoptosis (anti- and pro-apoptotic functions) (Figure 3.12). The response of each of these categories of genes is explored in more detail below.

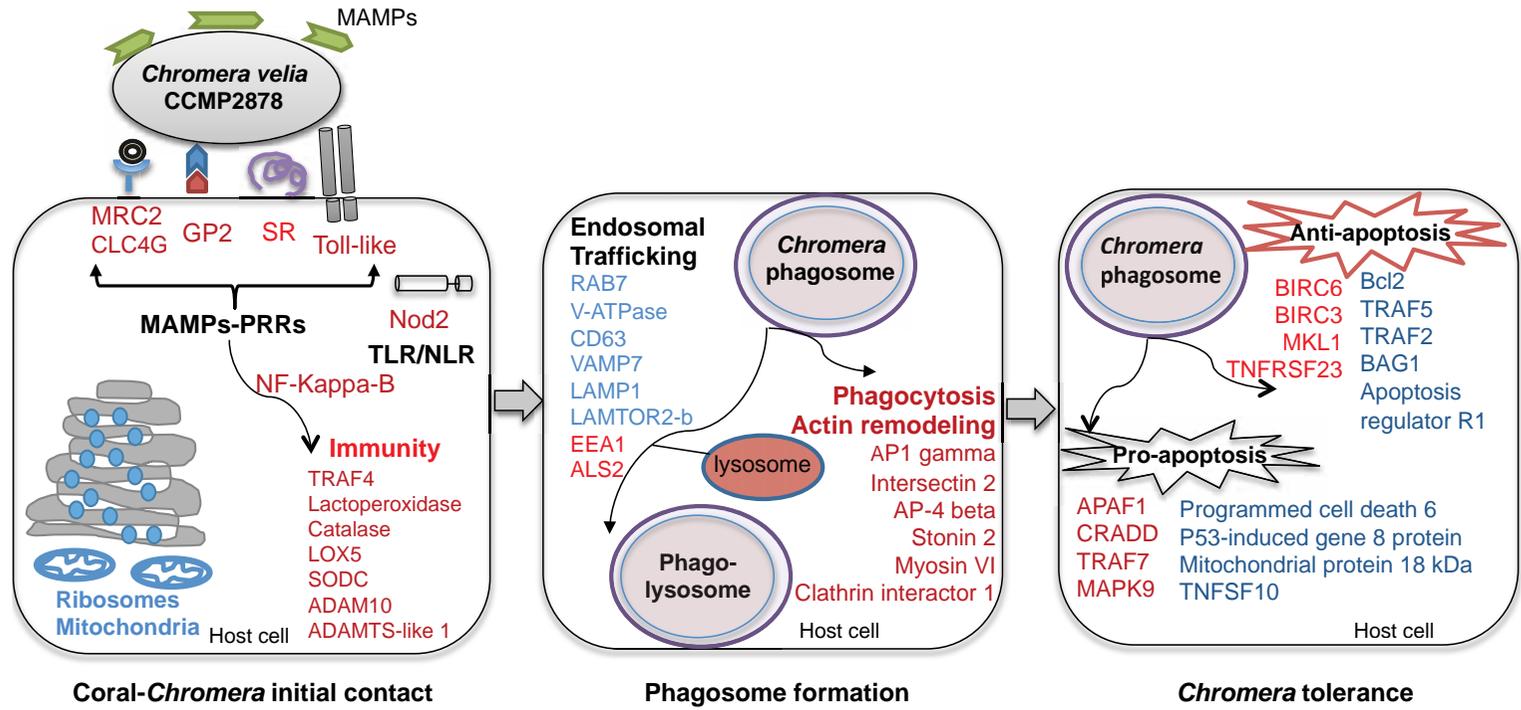


Figure 3.12 An interactive model of genes and pathways discussed in the text during interaction between larvae of *Acropora digitifera* and *Chromera*. Up- and down-regulated genes/functions are in blue and red text respectively. The initial contact phase (left panel) involves up-regulation of ribosome and mitochondria functions (based on GO enrichment), suppression of host immune response including down regulation of the pattern recognition receptors (PRRs) “that can recognize signature microbial compounds microbe-associated molecular patterns (MAMPs)”, the detected PRRs included TLRs, toll-like receptors; Nod, nucleotide-binding oligomerization domain protein; SRs, scavenger receptors; lectins; and the complement protein (C3). Moreover, the pancreatic secretory granule membrane major glycoprotein GP2, which serves as an uptake receptor for pathogenic bacteria in humans was highly down regulated. Suppression of PRRs- MAMPs interaction leads to inactivation of nuclear factor-B (NF-kappa-B), which is master regulator of immunity. The phagosome maturation phase (central panel) involves down-regulation of genes involved in phagocytosis and actin remodelling. As well as differential expression of genes implicated in endosomal trafficking that enhances the maturation of the phagosome and lysosome fusion (see Figure 4 for more details about those genes). The *Chromera* tolerance phase (right panel) involves complex changes in the apoptotic network. Genes implicated both anti- and pro-apoptotic functions were differentially expressed.

3.4.5.1 Coral immune response against *Chromera*

In the late response to *Chromera*, sixty *A. digitifera* genes encoding proteins implicated in immunity were down regulated, with fold-changes ranging from -5.19 to -1.2. Genes in this category include pattern recognition receptors (PRRs), components of the complement system, genes involved in the TLR/NLR signaling pathways, antimicrobial peptides, and genes involved in ROS and inflammatory responses (Table 3.13 and Figure 3.13).

PRRs are secreted or cell surface proteins expressed by the host in order to recognize compounds occurring on microbe surfaces known as microbe associated molecular patterns (MAMPs). In this context, down-regulation of a c-type lectin, mannose receptor 2 (MRC2) and lectin domain family 4-member G (CLC4G) is significant. Scavenger receptors are common PRRs in phagocytes that are able to bind a variety of MAMPs. Two genes encoding scavenger receptor cysteine-rich type 1 protein M160 and CD163 molecule-like 1 (that have scavenger receptor activity) were down-regulated. Three components of the complement system (a biochemical pathway involved in innate and acquired immune responses) were highly down-regulated,

including complement C3 homologs and component C6 precursor, which is a constituent of the membrane attack complex that plays important roles in both innate and adaptive immune responses via forming pores in target cell's plasma membrane. Genes encoding glycoproteins including pancreatic secretory granule membrane major glycoprotein GP2 (that might have roles in recognizing pathogenic bacteria) and dystroglycan 1 (glycoprotein that might act as a receptor for certain adenoviruses) were also down-regulated. Genes implicated in TLR and NLR signaling pathways including nucleotide-binding oligomerization domain containing-protein 2 (NOD2), toll-like receptors, and their downstream signaling molecules such as TNF alpha-induced protein 3 (TNFAIP3), TNF receptor-associated factor 4 (TRAF4) and nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NF-kappa-B) were all down-regulated. Specifically TNFR4 is signal transducer that links members of the TNFR family to different signaling pathways. It regulates activation of NF-kappa-B in response Toll-like receptors signaling by modulates TNF receptor associated factor 6 (TRAF6) that also mediate signals from members of the TNF receptor superfamily.

Genes encoding antimicrobial proteins that have antimicrobial activities including RNA polymerase III polypeptides A and B (which are key players in sensing and limiting bacterial and DNA viral infection), guanylate binding protein 7 (which promotes oxidative killing and deliver antimicrobial peptides to phagolysosomes, thus providing broad host protection against many pathogens), and lactoperoxidase (which uses hydrogen peroxide and thiocyanate to generate the antimicrobial substance hypothiocyanous acid) were all down-regulated (Table 3.13 and Figure 3.13). Genes involved in ROS and inflammatory responses including catalase (which acts to protect cells from the toxic effects of hydrogen peroxide), superoxide dismutase (which destroys toxic radicals within the cells), arachidonate 5-lipoxygenase (which play roles in inflammatory processes by catalyzing leukotriene biosynthesis) and genes encoding immunoglobulin and fibrinogen domains containing proteins were also down regulated (Table 3.13).

Table 3.13 Down regulated *A. digitifera* transcripts likely involved in suppression of the host immune response in *Chromera*- infected larvae at 48 h post infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the log₂fold change values

Cluster ID	UniProt ID/ gene model ID	Protein Name	Species	E-value	logFC
adi_EST_assem_5384	P01027	Complement C3	<i>Mus musculus</i>	1.66E-19	-5.12
adi_EST_assem_15402	P13671	Complement component C6 precursor	<i>Homo sapiens</i>	8.95E-50	-4.84
adi_EST_assem_31512	Q91132	Cobra venom factor Complement C3 homolog	<i>Naja kaouthia</i>	5.20E-28	-4.10
adi_EST_assem_14742	Q9NR16	Scavenger receptor cysteine-rich type 1 protein M160	<i>Homo sapiens</i>	6.00E-43	-4.47
adi_EST_assem_22790	P30205	CD163 molecule-like 1	<i>Bos taurus</i>	4.98E-12	-1.28
adi_EST_assem_8256	P55259	Speract/scavenger receptor glycoprotein 2 (zymogen granule membrane)	<i>Homo sapiens</i>	2.48E-15	-4.18
adi_EST_assem_4678	Q14118	dystroglycan 1 (dystrophin-associated glycoprotein 1)	<i>Homo sapiens</i>	2.89E-27	-2.37
adi_EST_assem_10380	Q9TSZ6	dystroglycan 1 (dystrophin-associated glycoprotein 1)	<i>Canis lupus</i>	7.30E-18	-2.81
adi_EST_assem_14075	aug_v2a.04319	toll-like receptor	<i>Acropora digitifera</i>	3.10E-26	-1.80
adi_EST_assem_5555	aug_v2a.02686	toll-like receptor	<i>Acropora digitifera</i>	3.20E-24	-1.60
adi_EST_assem_23391	aug_v2a.03526	C-type lectin domain family 4 member g (clc4g)	<i>Acropora digitifera</i>	3.20E-17	-3.70
adi_EST_assem_9676	Q64449	C-type mannose receptor 2 (MRC2)	<i>Mus musculus</i>	5.62E-18	-1.84
adi_EST_assem_13491	P19838	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	<i>Homo sapiens</i>	2E-167	-1.28
adi_EST_assem_8683	Q8K3Z0	nucleotide-binding oligomerization domain containing 2	<i>Mus musculus</i>	2.32E-16	-2.00
adi_EST_assem_461	Q9BUZ4	TNF receptor-associated factor 4	<i>Homo sapiens</i>	2.44E-84	-1.63
adi_EST_assem_3655	P21580	tumor necrosis factor, alpha-induced protein 3	<i>Homo sapiens</i>	1.00E-64	-1.20
adi_EST_assem_9748	O95163	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein	<i>Homo sapiens</i>	9.26E-94	-3.50
adi_EST_assem_6149	Q9ESE1	LPS-responsive beige-like anchor	<i>Mus musculus</i>	0	-2.38
adi_EST_assem_18500	Q91880	suppressor of hairless protein 1	<i>Xenopus laevis</i>	3.01E-40	-3.19
adi_EST_assem_10795	Q8N8V2	guanylate binding protein 7	<i>Homo sapiens</i>	3.46E-42	-2.48
adi_EST_assem_4061	P36888	fms-related tyrosine kinase 3 CD_antigen: CD135	<i>Homo sapiens</i>	4.7E-42	-2.37
adi_EST_assem_10633	Q9NW08	polymerase (RNA) III (DNA directed) polypeptide B	<i>Homo sapiens</i>	1.13E-156	-2.67
adi_EST_assem_8717	O14802	polymerase (RNA) III (DNA directed) polypeptide A, 155kDa	<i>Homo sapiens</i>	0	-2.50
adi_EST_assem_3715	P80025	perl_lactoperoxidase	<i>Bos taurus</i>	1.22E-12	-2.54
adi_EST_assem_5779	Q9JJ22	endoplasmic reticulum aminopeptidase 1	<i>Rattus norvegicus</i>	3.66E-43	-3.33
adi_EST_assem_9934	Q96JA1	leucine-rich repeats and immunoglobulin-like domains 1	<i>Homo sapiens</i>	7.72E-21	-5.19
adi_EST_assem_28915	Q9DE07	nibrin	<i>Gallus gallus</i>	9.80E-23	-4.66
adi_EST_assem_17218	P16621	Leukocyte-antigen-related-like	<i>Drosophila melanogaster</i>	1.81E-26	-2.79
adi_EST_assem_2273	O18738	dystroglycan 1 (dystrophin-associated glycoprotein 1)	<i>Bos taurus</i>	3.44E-21	-2.33
adi_EST_assem_858	Q95218	Dmbt1 deleted in malignant brain tumors 1	<i>Oryctolagus cuniculus</i>	5.64E-112	-2.32
adi_EST_assem_26531	Q8N6G6	ADAMTS-like 1	<i>Homo sapiens</i>	2.31E-25	-2.31
adi_EST_assem_2070	Q3UG20	myeloid/lymphoid or mixed-lineage leukemia 5 Histone-lysine N-methyltransferase 2E	<i>Mus musculus</i>	1.22E-71	-2.26
adi_EST_assem_3633	Q9UGM3	deleted in malignant brain tumors 1 Glycoprotein 340	<i>Homo sapiens</i>	4.32E-81	-2.25

adi_EST_assem_2961	Q05BQ1	immunoglobulin superfamily, member 9	<i>Mus musculus</i>	1.10E-48	-2.14
adi_EST_assem_8762	P58022	lysyl oxidase-like 2	<i>Mus musculus</i>	8.67E-75	-2.14
adi_EST_assem_683	Q4VGL6	RING CCCH (C3H) domains 1 Roquin-1	<i>Mus musculus</i>	5.58E-173	-2.01
adi_EST_assem_4797	Q8WWQ8	stabilin 2	<i>Homo sapiens</i>	9.75E-145	-2.01
adi_EST_assem_638	Q10741	ADAM metallopeptidase domain 10	<i>Bos taurus</i>	1.23E-161	-1.99
adi_EST_assem_7004	B4F6N6	lysyl oxidase-like 2	<i>Xenopus (Silurana) tropicalis</i>	3.04E-100	-1.94
adi_EST_assem_5267	P46531	Notch homolog 1, translocation-associated (Drosophila)	<i>Homo sapiens</i>	7.82E-32	-1.84
adi_EST_assem_3259	Q62077	phospholipase C, gamma 1	<i>Mus musculus</i>	0	-1.70
adi_EST_assem_4854	Q05B92	transcription factor binding to IGHM enhancer 3	<i>Bos taurus</i>	4.23E-54	-1.65
adi_EST_assem_4981	P31266	recombination signal binding protein for immunoglobulin kappa J region	<i>Mus musculus</i>	0	-1.62
adi_EST_assem_7676	Q91035	sonic hedgehog homolog (Drosophila)	<i>Gallus gallus</i>	6.31E-97	-1.59
adi_EST_assem_26435	O13076	adenosine A2b receptor	<i>Gallus gallus</i>	1.10E-17	-1.50
adi_EST_assem_580	Q9UHD2	TANK-binding kinase 1 (TBK1)	<i>Homo sapiens</i>	3.05E-144	-1.45
adi_EST_assem_3427	A6H7G2	drebrin-like	<i>Bos taurus</i>	1.26E-43	-1.32
adi_EST_assem_1987	P42232	signal transducer and activator of transcription 5B	<i>Mus musculus</i>	6.98E-108	-1.31
adi_EST_assem_5314	Q5R9T9	guanylate binding protein family, member 6	<i>Pongo abelii</i>	1.29E-49	-1.31
adi_EST_assem_1695	Q7ZW34	zgc:55318 Contactin-5	<i>Danio rerio</i>	2.59E-18	-1.23
adi_EST_assem_1190	O54928	suppressor of cytokine signaling 5 Cytokine-inducible SH2-containing protein 5	<i>Mus musculus</i>	5.93E-57	-1.20
adi_EST_assem_3877	Q9PVW8	sal_silasrhamnose-binding lectin	<i>Silurus asotus</i>	1.62E-08	-1.54
adi_EST_assem_1432	Q66S03	lecg_galactose-specific lectin nattectin	<i>Thalassophryne nattereri</i>	1.49E-12	-1.98
adi_EST_assem_1463	Q0V8S9	cntp5_contactin-associated 5	<i>Gallus gallus</i>	7.24E-14	-1.90
adi_EST_assem_35905	P17336	cata_catalase	<i>Drosophila melanogaster</i>	6.04E-70	-2.44
adi_EST_assem_3532	aug_v2a.12354	SODC_Superoxide dismutase [Cu-Zn]	<i>Acropora digitifera</i>	4.70E-38	-2.24
adi_EST_assem_3531	aug_v2a.01713	SODC1_Superoxide dismutase [Cu-Zn] 1	<i>Acropora digitifera</i>	0.00E+00	-1.26
adi_EST_assem_4182	P12527	lox5_arachidonate 5-lipoxygenase	<i>Rattus norvegicus</i>	4.90E-46	-1.04
adi_EST_assem_2688	P48999	lox5_arachidonate 5-lipoxygenase	<i>Mus musculus</i>	1.35E-104	-1.47

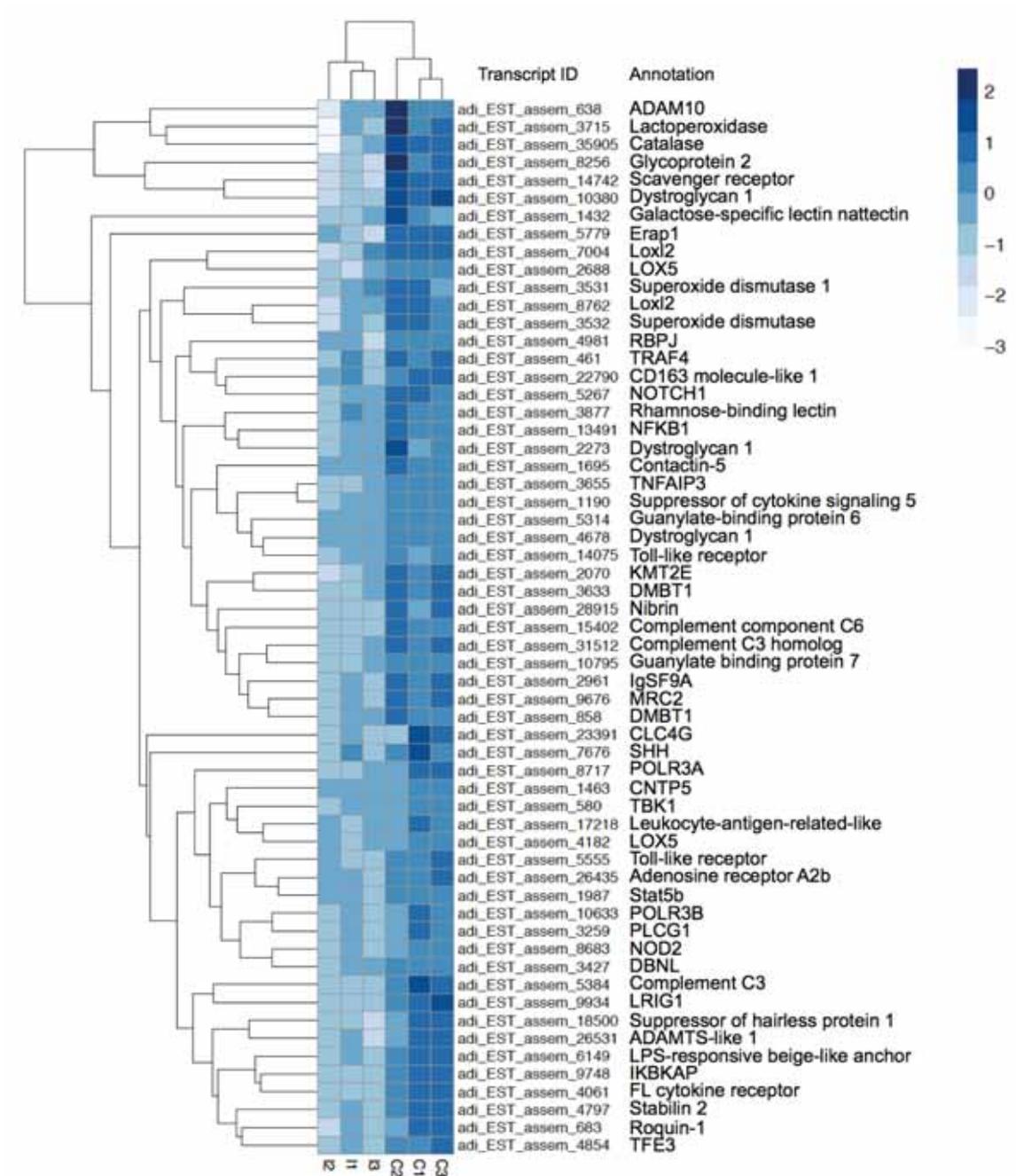


Figure 3.13 Heat map of differentially expressed genes likely involved in immunity, inflammatory and stress responses in *Chromera*-infected (I1, I2, I3) and control larvae (C1, C2, C3) at the 48 h time point. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against the control at 48h post-infection. Expression values were \log_2 -transformed and median-centered by transcript. The blue scale represents the relative expression values.

3.4.5.2 Genes involved in phagocytosis and the host endocytic pathway, and in particular phagosome maturation, were affected

Genes implicated in phagocytosis were down-regulated in the late response to *Chromera* including AP1 gamma subunit binding protein 1, intersectin 2, adaptor-related protein complex AP-4 beta 1, adaptor protein complex AP-1 gamma 1 subunit, adaptor-related protein complex 2 alpha 2 subunit, very low density lipoprotein receptor, clathrin interactor 1, stonin 2, and myosin VI, which are involved in endocytosis, clathrin-dependent endocytosis, formation of clathrin coat and clathrin-coated vesicles (CCPs). Only one gene encoding clathrin, light chain A was up-regulated with 1.08-fold increase in expression (Table 3.14).

Table 3.14 Differential expression of *A. digitifera* clusters likely involved in phagocytosis in *Chromera*- infected larvae at 48 h post-infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the \log_2 fold change values

Cluster ID	Protein ID	Protein name	Species	E-value	logFC
adi_EST_assem_14553	Q5SV85	AP1 gamma subunit binding protein 1	<i>Mus musculus</i>	3.02E-17	-3.88
adi_EST_assem_5204	Q9NZM3	Intersectin 2	<i>Homo sapiens</i>	0	-2.50
adi_EST_assem_6352	Q9WV76	Adaptor-related protein complex AP-4, beta 1	<i>Mus musculus</i>	1.14E-145	-2.28
adi_EST_assem_286	P18484	Adaptor-related protein complex 2, alpha 2 subunit	<i>Rattus norvegicus</i>	0	-2.04
adi_EST_assem_6885	P98155	Very low density lipoprotein receptor	<i>Homo sapiens</i>	8.13E-162	-1.81
adi_EST_assem_4258	P22892	Adaptor protein complex AP-1, gamma 1 subunit	<i>Mus musculus</i>	0	-1.79
adi_EST_assem_14154	Q8WXE9	Stonin 2	<i>Homo sapiens</i>	3.65E-62	-1.74
adi_EST_assem_918	Q9I8D1	Myosin VI	<i>Gallus gallus</i>	0	-1.50
adi_EST_assem_4580	Q14677	Clathrin interactor 1	<i>Homo sapiens</i>	3.26E-82	-1.49
adi_EST_assem_749	P98156	Very low density lipoprotein receptor	<i>Mus musculus</i>	6.74E-155	-1.37
adi_EST_assem_2285	O08585	Clathrin, light polypeptide (Lca)	<i>Mus musculus</i>	2.63E-34	1.08

Eleven genes implicated in early phagosome formation including phosphoinositide-3-kinase class 2 and 3, phosphatidylinositol 4-kinase, inositol polyphosphate phosphatase-like 1, the early endosome markers; early endosome antigen 1 (EEA1), amyotrophic lateral sclerosis 2 (ALS-2) were down-regulated in the late response to *Chromera*. Specifically ALS-2 is a Rab5 effector protein, which localizes with Rab5 on early endosomal compartments. On the other hand, eleven genes

implicated in late phagosome formation and phagosomal maturation including the late phagosome markers (Ras-related protein 7A (Rab7A), CD63 molecule, and lysosomal-associated membrane protein 1 (LAMP1)), TBC1 domain family member 5 (which can displace Rab7A), subunits of lysosomal ATPase proton transporting and the late endosomal/ lysosomal adaptor and MAPK and MTOR activator 2-B protein were up-regulated (Table 3.15 and Figures 3.14, 3.15). In addition, genes implicated in autophagy and lysosome function were differentially expressed; including up-regulation of lysosomal hydrolases thioesterase PPT2-A, iduronidase alpha-L and proteins forming autophagosomal vacuoles and down-regulation of lysosomal lipase A and two mannosidases beta A (Table 3.16).

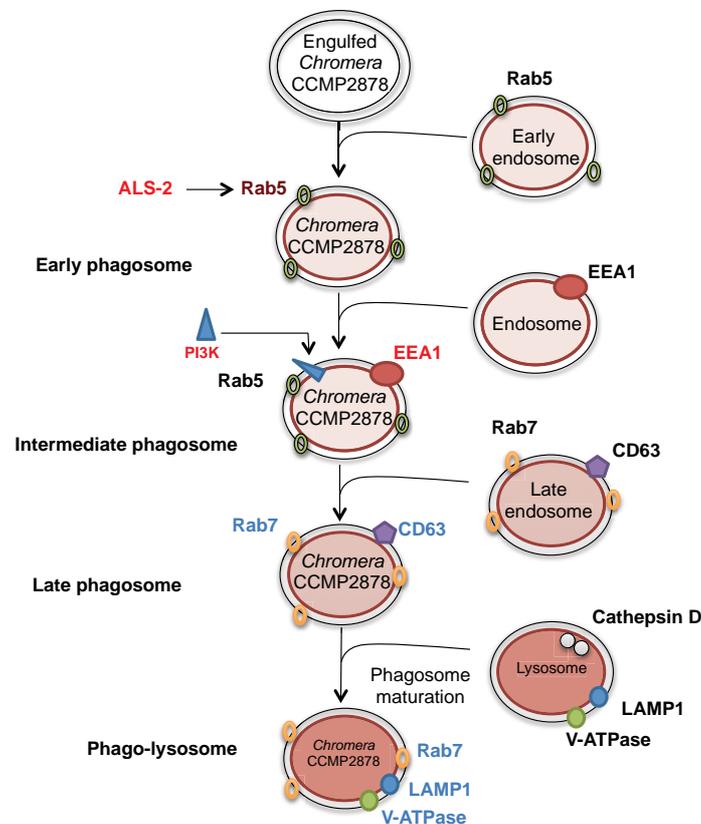


Figure 3.14 Modulation of the coral host endocytic pathway and enhancement of phagosome maturation and lysosome fusion in response to *Chromera*. Upon phagocytosis, the phagosome acquires the Rab5 GTPase via fusion with early endosomes. During *Chromera* infection, the Rab5 effector ALS2 was up-regulated. Rab5 acts to recruit phosphatidylinositol-3-kinase (PI3K) to generate phosphatidylinositol-3-phosphate (PI3P) and recruit the early endosomal antigen (EEA1) from endosomes. EEA1 is a Rab5 effector protein and was up-regulated to trigger fusion of phagosome with late endosome. During the phagosomal maturation process, Rab7 replaces Rab5 and the intermediate phagosome fuses with late endosomal vesicles and acquires proteins such as the proton-ATPase pump (V-ATPase), lysosome-associated membrane glycoprotein 1 (LAMP1), CD63, and lysosomal hydrolases. Vacuoles containing

Chromera fused with late endosomes/lysosomes as indicated by the up-regulation of genes encoding Rab7, LAMP1, and CD63 (plus late endosomal/ lysosomal adaptor, SNAP-associated protein, vesicle-associated membrane protein 7 (VAMP7); not shown). Moreover, the lysosome V-ATPase was up-regulated, indicating that lysosome acidification was activated in order to kill *Chromera* (the phagolysosome is rich in hydrolytic enzymes and has a very low pH). Genes highlighted in red are down-regulated, while those in blue are up-regulated.

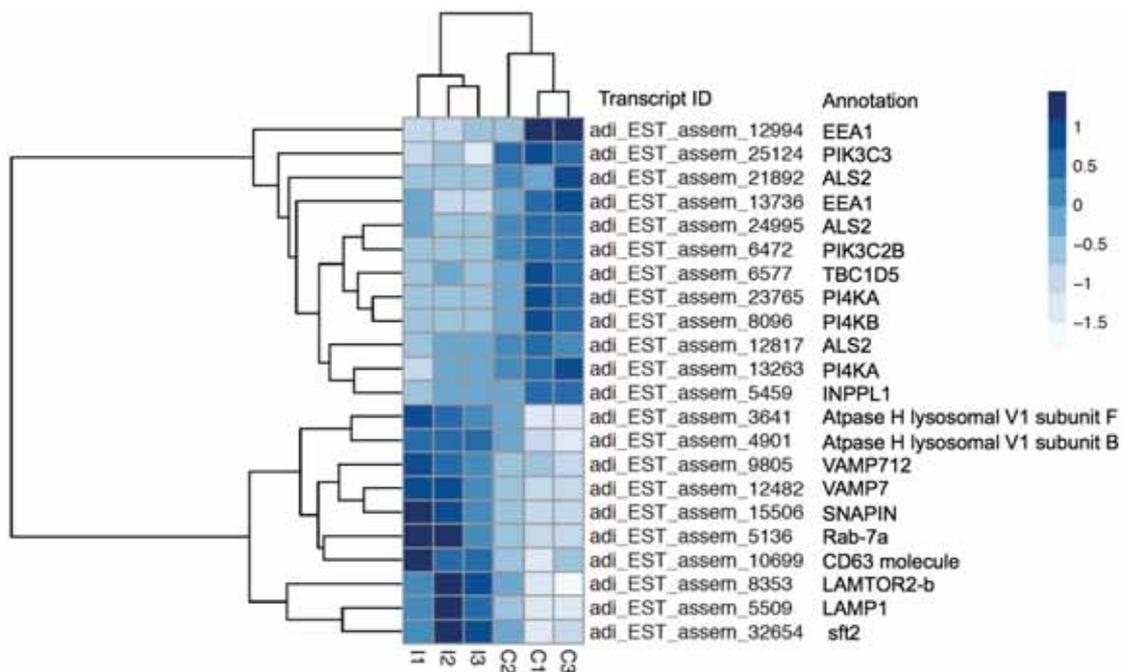


Figure 3.15 Heat map of differentially expressed genes involved in phagosome maturation in *Chromera*-infected (I1, I2, I3) and control larvae (C1, C2, C3) at the 48-h time point. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against the control at 48 h post-infection. Expression values were \log_2 -transformed and median-centered by transcript. The blue scale represents the relative expression values.

Table 3.15 Differential expression of *A. digitifera* transcripts likely involved in early and/or late endosome formation and phagosomal maturation in *Chromera*-infected larvae at 48 h post infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the log₂fold change values.

Cluster ID	UniProt/gene model ID	Protein name	Species	E value	logFC
adi_EST_assem_12994	aug_v2a.14153	Zinc finger FYVE domain-containing protein, early endosome antigen 1 (EEA1)	<i>Acropora digitifera</i>	5.00E-18	-4.11
adi_EST_assem_13736	aug_v2a.14153	Zinc finger FYVE domain-containing protein, early endosome antigen 1 (EEA1)	<i>Acropora digitifera</i>	5.00E-18	-3.50
adi_EST_assem_24995	Q96Q42	Amyotrophic lateral sclerosis 2 (juvenile)	<i>Homo sapiens</i>	1.76E-14	-3.19
adi_EST_assem_21892	Q920R0	Amyotrophic lateral sclerosis 2 (juvenile) homolog (human)	<i>Mus musculus</i>	1.20E-29	-5.53
adi_EST_assem_12817	P0C5Y8	Amyotrophic lateral sclerosis 2 (juvenile) homolog (human)	<i>Rattus norvegicus</i>	0	-1.53
adi_EST_assem_13263	P42356	Phosphatidylinositol 4-kinase, catalytic, alpha	<i>Homo sapiens</i>	4.25E-63	-2.35
adi_EST_assem_23765	O08662	Phosphatidylinositol 4-kinase, catalytic, alpha	<i>Rattus norvegicus</i>	7.42E-93	-2.40
adi_EST_assem_8096	A9X1A0	Phosphatidylinositol 4-kinase, catalytic, beta	<i>Papio anubis</i>	0	-1.69
adi_EST_assem_6472	O00750	Phosphoinositide-3-kinase, class 2, beta polypeptide	<i>Homo sapiens</i>	2.83E-25	-5.20
adi_EST_assem_25124	O88763	Phosphoinositide-3-kinase, class 3	<i>Rattus norvegicus</i>	3.98E-164	-3.58
adi_EST_assem_5459	Q9WVR3	Inositol polyphosphate phosphatase-like 1	<i>Rattus norvegicus</i>	8.17E-169	-1.17
adi_EST_assem_6577	Q92609	TBC1 domain family, member 5	<i>Homo sapiens</i>	2.43E-130	-1.67
adi_EST_assem_5136	P51149	Ras-related protein Rab-7a	<i>Homo sapiens</i>	4.00E-33	1.80
adi_EST_assem_10699	P28648	CD63 molecule Mast cell antigen AD1	<i>Rattus norvegicus</i>	3.38E-25	1.38
adi_EST_assem_12482	Q5ZL74	Vesicle-associated membrane protein 7 (VAMP7)	<i>Gallus gallus</i>	3.61E-87	1.34
adi_EST_assem_5509	aug_v2a.01938	Lysosomal-associated membrane protein 1 LAMP1	<i>Acropora digitifera</i>	1.70E-24	1.31
adi_EST_assem_3641	P50408	Atpase, H transporting, lysosomal V1 subunit F	<i>Rattus norvegicus</i>	2.00E-54	1.22
adi_EST_assem_4901	Q2TA24	Atpase, H+ transporting, lysosomal 21kda, V0 subunit b	<i>Bos taurus</i>	9.55E-70	1.00
adi_EST_assem_9805	Q9SIQ9	Vesicle-associated membrane protein 712 (VAMP712)	<i>Arabidopsis thaliana</i>	1.33E-44	1.05
adi_EST_assem_32654	Q9P6K1	Protein transport protein sft2	<i>Schizosaccharomyces pombe</i>	4.10E-18	1.25
adi_EST_assem_15506	O95295	SNAP-associated protein	<i>Homo sapiens</i>	4.71E-36	1.37
adi_EST_assem_8353	Q7ZXB7	Late endosomal/lysosomal adaptor and MAPK and MTOR activator 2-B	<i>Xenopus laevis</i>	4.34E-54	1.59

Table 3.16 Differential expression of *A. digitifera* transcripts likely involved in autophagy and lysosomal functions in *Chromera*-infected larvae at 48 h post-infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the log₂fold change values.

Cluster ID	UniProt ID	Protein name	E-value	logFC
adi_EST_assem_2028	P60517	GABA(A) receptor-associated protein <i>Rattus norvegicus</i>	1.36E-75	1.21
adi_EST_assem_24964	Q2TBJ5	ATG12 autophagy related 12 homolog (S. cerevisiae) <i>Rattus norvegicus</i>	2.44E-32	1.57
adi_EST_assem_4839	Q9CPX6	autophagy-related 3 (yeast) <i>Mus musculus</i>	1.32E-140	1.14
adi_EST_assem_22677	Q99J83	autophagy-related 5 (yeast) <i>Mus musculus</i>	1.07E-116	1.77
adi_EST_assem_4610	Q2HJ23	microtubule-associated protein 1 light chain 3 alpha <i>Bos taurus</i>	1.03E-46	1.17
adi_EST_assem_17647	Q6PCJ9	Lysosomal thioesterase PPT2-A <i>Xenopus laevis</i>	2.31E-93	1.14
adi_EST_assem_8111	P48441	iduronidase, alpha-L- <i>Mus musculus</i>	7.85E-27	1.19
adi_EST_assem_1447	Q86YJ5	E3 ubiquitin-protein ligase MARCH9 <i>Homo sapiens</i>	6.34E-23	1.35
adi_EST_assem_1823	P38571	lipase A, lysosomal acid, cholesterol esterase <i>Homo sapiens</i>	1.21E-31	-1.79
adi_EST_assem_14969	Q29444	mannosidase, beta A, lysosomal <i>Bos taurus</i>	2.62E-163	-1.66
adi_EST_assem_25764	O00462	mannosidase, beta A, lysosomal <i>Homo sapiens</i>	1.35E-31	-6.45

Thirty seven genes involved in endosomal trafficking including members of the Rab proteins and vacuolar protein sorting proteins were differentially expressed (Supplementary Table S3.7). Rab proteins required for protein transport from the endoplasmic reticulum to the Golgi complex, transport between the plasma membrane and early endosomes, and transport between the endosomes and the trans-Golgi network including Rab18B, Rab9A, Rab36, Rab10, Rab28, Rab4-like, Rab2, Rab3 as well as the protein transport protein SFT2 were up- regulated. Moreover one gene encoding SNAP-associated protein that is involved in SNARE-mediated membrane fusion was also up-regulated. However, Rab GTPase activating protein 1-like, Rab11 family interacting protein 3, Rab3 GTPase activating protein subunits 1, 2 and Rab3A interacting protein were down- regulated.

Genes encoding vacuolar protein sorting proteins were down-regulated. Specifically the vacuolar protein sorting protein 18 that is required for membrane docking/fusion reactions of late endosomes/lysosomes had 5.11-fold decrease. One gene encoding Golgi associated gamma-adaptin-related protein 1 that plays a role in protein sorting and trafficking between the trans-Golgi network (TGN) and endosomes had 1.33-fold decrease. 7 members of the TBC1 domain family of proteins were also down regulated including proteins that act as GTPase-activating protein for Rab1, 2 and other Rab family proteins (Supplementary Table S3.7).

3.4.5.3 Genes involved in host apoptosis were affected

Gene encoding proteins with both pro- and anti-apoptotic functions and proteins involved in regulation of intrinsic and extrinsic apoptotic pathways were differentially expressed in the late response to *Chromera*. More details about genes implicated in apoptosis are provided in Appendix III, the supplementary material of this Chapter.

Genes implicated in the extrinsic apoptotic pathway with pro-apoptotic functions were down-regulated in the late response to *Chromera* including: death-domain containing protein (CRADD) (which acts as an apoptotic adapter for caspase-2 by recruiting it to the TNFR-1 signaling complex), a homolog of FEM-1 protein (which acts as a death receptor-associated protein thus mediating apoptosis), TNF ligand superfamily member 10 (a TNF-related apoptosis inducing ligand), TRAF7, SH3-domain kinase binding protein 1 and MAP-kinase activating death domain (which both play important regulatory roles in TNF-mediated apoptosis) (Figure 3.16 and Table 3.17). Genes implicated in the intrinsic apoptotic pathway with pro-apoptotic functions were down-regulated in the late response to *Chromera* including: APAF1 (which acts to activate caspase-3), MAPK9, Bcl-2/ adenovirus E1B 19 kDa protein-interacting protein 3 (which acts as apoptosis-inducing protein that can overcome B-cl2 suppression), and mitochondrial protein 18 ((Figure 3.16 and Table 3.17).

Some other genes implicated in the extrinsic apoptotic pathway with anti-apoptotic functions were differentially expressed the late response to *Chromera* including down-regulation of TNFR superfamily member 23, MKL protein, and baculoviral IAP repeat-containing 3 (which inhibits apoptosis by binding to TNF receptor-associated factors TRAF1 and TRAF2) and up-regulation of the TNF receptor-associated factors TRAF3 and TRAF5. TRAF3 plays a central role in the regulation of B-cell survival through activation of NF-kappa-B and MAP kinases where as TRAF5 acts as adapter protein and signal transducer that links members of the TNF receptor family to different signaling pathways (Figure 3.17 and Table 3.18). Also, other genes implicated in the intrinsic apoptotic pathway with anti-apoptotic functions were differentially expressed in the late response to *Chromera* including down-regulation of inhibitor of p53-induced apoptosis-beta, Bcl-2 associated athanogene 4 (BAG family molecular chaperone regulator 4, which acts as silencer of death domains), and two

baculoviral IAP repeat-containing 6 proteins, which act as inhibitor of the caspases-3, 7 and 9 and up-regulation of the apoptosis regulator Bcl-2, BAG family molecular chaperone regulator 1, and apoptosis regulator R1. Specifically Bcl-2 suppresses apoptosis via inhibiting caspase activity either by preventing cytochrome c release from mitochondria and/or by APAF1 binding. Moreover, 28 genes implicated in modulation of apoptotic pathway including programmed cell death proteins and proteins involved in phagocytosis of apoptotic cells were differentially expressed (Figure 3.17 and Table 3.18).

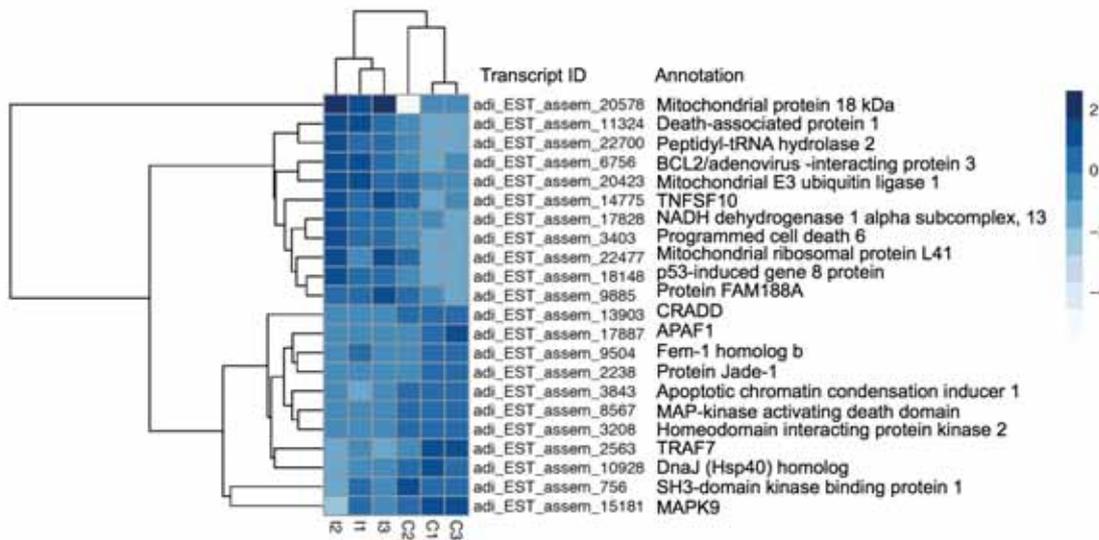


Figure 3.16 Heat map of differentially expressed genes involved in pro-apoptosis in *Chromera*-infected (I1, I2, I3) and control larvae (C1, C2, C3) at the 48-h time point. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against the control at 48 h post-infection. Expression values were \log_2 -transformed and median-centered by transcript. The blue scale represents the relative expression values.

Table 3.17 Differential expression of *A. digitifera* clusters likely have pro-apoptotic functions in *Chromera*-infected larvae at 48 h post-infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the log₂fold change values

Cluster ID	Protein ID	Protein Name	Species	E-value	LogFC	
adi_EST_assem_13903	aug_v2a.0919 1	Death-domain containing protein (CRADD)	<i>Acropora digitifera</i>	1.10E-13	-1.20	Extrinsic
adi_EST_assem_2563	Q6Q0C0	TNF receptor-associated factor 7	<i>Homo sapiens</i>	0.00E+00	-2.44	Extrinsic
adi_EST_assem_756	Q96B97	SH3-domain kinase binding protein 1	<i>Homo sapiens</i>	3.34273E-25	-1.41	Extrinsic
adi_EST_assem_9504	Q5ZM55	fem-1 homolog b (C. elegans)	<i>Gallus gallus</i>	6.1464E-151	-2.11	Extrinsic
adi_EST_assem_8567	Q80U28	MAP-kinase activating death domain	<i>Mus musculus</i>	3.963E-164	-2.06	Extrinsic
adi_EST_assem_17887	O14727	apoptotic peptidase activating factor 1 (APAF1)	<i>Homo sapiens</i>	2.30E-56	-2.39	Intrinsic
adi_EST_assem_15181	Q9WTU6	mitogen-activated protein kinase 9	<i>Mus musculus</i>	2.27E-40	-1.87	Intrinsic
adi_EST_assem_2238	Q6GQJ2	MGC79115 protein Protein Jade-1	<i>Xenopus laevis</i>	6.85E-138	-1.12	
adi_EST_assem_3843	Q9UKV3	apoptotic chromatin condensation inducer 1	<i>Homo sapiens</i>	4.03E-28	-1.65	
adi_EST_assem_10928	Q96EY1	DnaJ (Hsp40) homolog, subfamily A, member 3	<i>Homo sapiens</i>	8.47E-128	-1.81	
adi_EST_assem_3208	Q9H2X6	homeodomain interacting protein kinase 2	<i>Homo sapiens</i>	0.00E+00	-1.56	
adi_EST_assem_14775	P50591	Tumor necrosis factor ligand superfamily member 10 (TNF-related apoptosis inducing ligand)	<i>Homo sapiens</i>	4.12E-11	1.09	Extrinsic
adi_EST_assem_6756	Q12983	BCL2/adenovirus E1B 19 kDa protein-interacting protein 3	<i>Homo sapiens</i>	1.17E-38	1.26	Intrinsic
adi_EST_assem_20578	Q9UDX5	mitochondrial protein 18 kDa	<i>Homo sapiens</i>	1.59E-18	2	Intrinsic
adi_EST_assem_17828	Q95KV7	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 13	<i>Bos taurus</i>	7.94E-38	1.19	
adi_EST_assem_11324	P51397	death-associated protein 1	<i>Homo sapiens</i>	1.21E-17	1.78	
adi_EST_assem_18148	O14681	p53-induced gene 8 protein	<i>Homo sapiens</i>	4.41E-33	1.16	
adi_EST_assem_22477	Q6DJI4	mitochondrial ribosomal protein L41	<i>Xenopus laevis</i>	5.45E-15	1.17	
adi_EST_assem_22700	Q9Y3E5	peptidyl-tRNA hydrolase 2	<i>Homo sapiens</i>	1.93E-56	1.84	
adi_EST_assem_9885	A0AUR5	Protein FAM188A	<i>Danio rerio</i>	5.95E-48	1.09	
adi_EST_assem_20423	Q969V5	mitochondrial E3 ubiquitin ligase 1	<i>Homo sapiens</i>	1.30E-64	1.27	
adi_EST_assem_3403	P12815	programmed cell death 6	<i>Mus musculus</i>	3.39E-76	1.27	

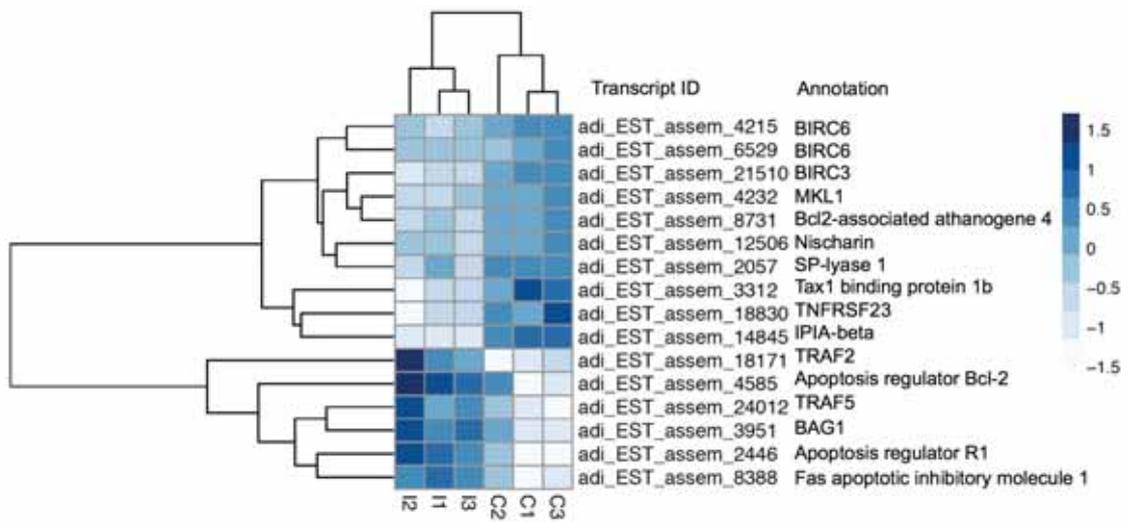


Figure 3.17 Heat map of differentially expressed genes involved in anti-apoptosis in *Chromera*-infected (I1, I2, I3) and control larvae (C1, C2, C3) at the 48-h time point. The hierarchical clustering shown was obtained by comparing the expression values (Fragments Per Kilobase of transcript per Million; FPKM) for *Chromera*-infected samples compared against the control at 48h post infection. Expression values were \log_2 -transformed and median-centered by transcript. The blue scale represents the relative expression values.

Table 3.18 Differential expression of *A. digitifera* transcripts that likely have anti-apoptotic functions in *Chromera*-infected larvae at 48 h post-infection with corrected $P \leq 0.05$. Columns correspond to coral cluster ID, annotated protein ID and name, species, E-value and the log₂fold change values

Cluster ID	Protein ID	Protein Name	Species	E-value	logFC	
adi_EST_assem_18830	Q9ER63	Tumor necrosis factor receptor superfamily member 23	<i>Mus musculus</i>	4.19E-21	-2.73	Extrinsic
adi_EST_assem_4232	Q8K4J6	MKL (megakaryoblastic leukemia)/myocardin-like 1	<i>Mus musculus</i>	7.27E-26	-1.16	Extrinsic
adi_EST_assem_21510	O08863	baculoviral IAP repeat-containing 3	<i>Mus musculus</i>	9.79E-30	-2.13	Extrinsic
adi_EST_assem_14845	P56597	Inhibitor of p53-induced apoptosis-beta	<i>Homo sapiens</i>	9.34E-13	-5.41	Intrinsic
adi_EST_assem_8731	O95429	BCL2-associated athanogene 4 Silencer of death domains	<i>Homo sapiens</i>	4.28E-13	-1.29	Intrinsic
adi_EST_assem_4215	O88738	baculoviral IAP repeat-containing 6	<i>Mus musculus</i>	1.40E-22	-2.88	Intrinsic
adi_EST_assem_6529	Q9NR09	baculoviral IAP repeat-containing 6	<i>Homo sapiens</i>	4.00E-179	-1.23	Intrinsic
adi_EST_assem_3312	Q6P132	Tax1 (human T-cell leukemia virus type I) binding protein 1b	<i>Danio rerio</i>	2.02E-37	-2.05	
adi_EST_assem_12506	Q4G017	nischarin	<i>Rattus norvegicus</i>	6.03E-23	-1.26	
adi_EST_assem_2057	Q8CHN6	sphingosine-1-phosphate lyase 1	<i>Rattus norvegicus</i>	0	-1.02	
adi_EST_assem_18171	Q12933	TNF receptor-associated factor 2	<i>Homo sapiens</i>	8.64E-11	1.60	Extrinsic
adi_EST_assem_24012	P70191	TNF receptor-associated factor 5	<i>Mus musculus</i>	5.61E-15	1.30	Extrinsic
adi_EST_assem_4585	Q00709	Apoptosis regulator Bcl-2 B-cell CLL/lymphoma 2	<i>Gallus gallus</i>	1.33E-18	1.31	Intrinsic
adi_EST_assem_3951	Q99933	BAG family molecular chaperone regulator 1	<i>Homo sapiens</i>	2.33E-30	1.25	Intrinsic
adi_EST_assem_2446	Q91827	Apoptosis regulator R1	<i>Xenopus laevis</i>	2.42E-26	1.55	Intrinsic
adi_EST_assem_8388	Q8R5H8	Fas apoptotic inhibitory molecule	<i>Rattus norvegicus</i>	2.44E-84	1.11	

3.5 DISCUSSION

In the present Chapter, the transcriptomic response of *Acropora digitifera* larvae challenged with *Chromera* was investigated. A distinct coral transcriptomic signature was identified in response to *Chromera* infection compared to the control. Massive changes in the coral transcriptome were detected at 48 h post *Chromera*-infection. The transcriptomic response of *A. digitifera* during infection with *Chromera* therefore differs fundamentally to that observed with competent *Symbiodinium* (in chapter 2), where significant changes in the transcriptome (predominantly down-regulation), were detected only at the 4-h time point, whereas no differential expression could be detected at the 12- and 48-h time points.

3.5.1 *Coral responses common to Chromera and Symbiodinium infection*

One gene encoding pancreatic secretory granule membrane major glycoprotein GP2 was found in the early response (4 h time point) to *Chromera* infection. While the specific function of the GP2 protein in corals is unknown, GP2 was down regulated in *Chromera* as well as in competent *Symbiodinium* infections (see chapter 2). Hase *et al.* (2009) have found that, in mammals, GP2 is exclusively expressed on membranous (M) cells where it acts as a bacterial uptake receptor. GP2 recognizes FimH, which is a major component of the type1 pilus on the outer membrane of some gram-negative enterobacilli such as *Escherichia coli* and *Salmonella enterica* (Hase *et al.* 2009; Yu & Lowe 2009). The interaction between GP2 and FimH is necessary for efficient uptake of enterobacteria by M cells and subsequent specific mucosal immune responses (Ohno & Hase 2010). Thus GP2 acts as an integral membrane protein that recognizes and binds pathogenic enterobacteria in humans and it has an important role in innate immunity. Consequently, suppressing the coral GP2 appears to be a key component of host-algal recognition and subsequent signaling pathways.

3.5.2 *The late response of coral to Chromera infection*

The late response of coral to *Chromera* was complex, and most of the significant DEGs were down-regulated. The GO molecular function GTPase regulator activity was highly over-represented in the late response, implying modulation of the host endocytic pathway, which is a key player in host-pathogen interactions including *Mycobacteria tuberculosis* (Koul *et al.* 2004). On the other hand, the cellular component mitochondrion and the molecular function structural constituent of ribosome (ribosomal proteins) were highly over-represented among the up-regulated genes, implying activation of protein synthesis and metabolism. Those results contrast markedly with the response of coral larvae to a competent *Symbiodinium* strain, where *A. digitifera* larvae showed temporary suppression of protein synthesis and metabolism (chapter 2). Up regulation of ribosomal and mitochondrial functions implies activation of protein synthesis, metabolism, and energy production. Such responses have been reported in the sea fan, *Gorgonia ventalina* exposed to the *Aplanochytrium* parasite (Burge *et al.* 2013) and in other invertebrates in response to bacterial pathogens (Gestal *et al.* 2007; Travers *et al.* 2010). This might indicate an increased demand to energy in order to tolerate the

presence of parasites/ pathogens and mounting appropriate downstream responses to get rid of the invading microbes.

3.5.3 *Suppression of the host immune response during Chromera infection*

It is now recognized that animals use the immune system not only to recognize and respond to pathogenic or infectious microbes, but also to identify beneficial bacteria. During *Chromera* infection, many PRRs, including C-type lectins, scavenger receptors, complement, TLRs, NLRs, and GP2 were strongly down regulated. Suppression of the expression of these PRRs in response to *Chromera* infection implies roles for these in identifying and recognizing invading microbes through PRR-MAMP signaling. PRR-MAMP interactions have been described in the context of cnidarian-dinoflagellate symbioses (Weis *et al.* 2008). As well, there are many examples of PRR-MAMP involvement in mutualistic host-microbe interactions such as symbiotic squid-bacteria and plant-nitrogen fixing bacteria associations (Cullimore & Denarie 2003; Nyholm & McFall-Ngai 2004). In the anemone *Aiptasia pulchella* and larvae of the coral *Fungia scutaria*, glycan-lectin interactions has been implicated in recognition and symbiosis establishment (Lin *et al.* 2000; Wood-Charlson *et al.* 2006). Moreover, a C-type lectin, mannose receptor 2 (MRC2), was highly up regulated in *A. digitifera* larvae infected with a competent *Symbiodinium* strain (see Chapter 2), suggesting a function in host-symbiont recognition during the establishment of symbiosis. Other PRRs that might have roles in cnidarian-dinoflagellate recognition processes are complement C3 and scavenger receptors. Kvennefors *et al.* (2010) localized a C3 homolog near the resident symbionts in the coral *A. millepora*, suggesting that C3 could be coating the symbionts as an opsonin, thus playing a role in host-symbiont communication. A scavenger receptor homolog was up-regulated during symbiosis in the sea anemone *Anthopleura elegantissima*, suggesting a function in host-symbiont communication (Rodriguez-Lanetty *et al.* 2006).

On the other hand, PRR-MAMP interactions are involved in recognizing pathogenic microbes; for example, TLRs, NLRs, C-type lectin receptors and scavenger receptors are involved in recognition of pathogenic *Mycobacteria* (Killick *et al.* 2013). TLRs and NLRs are involved in sensing the malaria parasite *Plasmodium spp.* in different vertebrate hosts (Gazzinelli *et al.* 2014), and activation of some TLRs results

in the production of antimicrobial peptides (Davies *et al.* 2008) in response to pathogen challenge. Suppression of these receptors in response to *Chromera* infection might underlie the observed down-regulation of genes encoding antibacterial proteins such as guanylate binding protein 7 and lactoperoxidase, suggesting that the invading *Chromera* inactivates the host defense mechanisms in order to survive inside the host. The transcriptomic results presented here are reminiscent of the interaction of mammalian hosts with pathogenic strains of *Mycobacteria* strains. In the latter, the mannose-capped lipoglycan, mannosylated lipoarabinomannan is produced in order to inhibit TLR signaling, thus inhibiting dendritic cell maturation and enhancing the production of immunosuppressive cytokines, leading to impairment of T cell activation. By contrast, in the Caribbean Sea fan, *Gorgonia ventalina*, genes encoding PRRs and antibacterial proteins, including the guanylate binding protein, were up regulated in during infection with the parasite *Aplanochytrium* (Burge *et al.* 2013).

Functional and genomic analyses have demonstrated the presence of key innate immune components, including pattern recognition proteins, signaling cascades, and potential effectors (Augustin & Bosch 2010; Miller *et al.* 2007) and, in some cases, conservation of function has been demonstrated (Moya *et al.* 2016; Sakamaki *et al.* 2015). Genes implicated in immune responses, included PRRs, TLR/NLR signaling pathways components, antibacterial proteins, ROS/ inflammatory response proteins and fibrinogen-domain and immunoglobulin-domain containing proteins, were highly down-regulated in response to *Chromera* infection, suggesting suppression of the coral immune system. The down regulation of catalase and superoxide dismutase observed in response to *Chromera* infection implies inactivation of host protective systems by the infective organism. The observed suppression of the immune response is contrasted to what have been shown in corals infected with white syndrome and white band diseases (Libro *et al.* 2013; Wright *et al.* 2015). Moreover, activation of the coral immune system has been reported in response to viral and bacterial mimics (Weiss *et al.* 2013), infectious coral diseases and challenge with pathogens such as *Vibrio coralliilyticus* and *Alteromonas sp.* (Brown *et al.* 2013).

3.5.4 Modulation of the endocytic pathway and phagosome maturation in the coral host response to *Chromera* infection

Phagocytosis is well understood in the case of, for example, vertebrate macrophages, which engulf and destroy non-self cells or microbial invaders. Many reviews have discussed pathogen strategies for subverting host defense responses. In the case of *Chromera* infection, many coral genes encoding proteins involved in phagocytosis were down regulated, suggesting that *A. digitifera* larvae may attempt to restrict further uptake of *Chromera* cells. I interpret other aspects of the data as reflecting suppression of host immune responses by *Chromera*. Down-regulation of genes involved in actin remodeling and enrichment of the KEGG pathways regulation of actin cytoskeleton and focal adhesion among down-regulated genes suggests that actin remodelling is required at the site of phagocytosis in order to prevent the formation and extension of pseudopods. Immediately after phagocytosis, the Rab5 GTPase is recruited to the phagosome to direct fusion of the phagosome with an early endosome (Vieira *et al.* 2002). Rab5 recruits Phosphatidylinositol-3-phosphate kinase (PI3K), and the phosphoinositol 3-phosphate PI(3)P generated by this activity recruits early endosome antigen 1 (EEA1) from endosomes. EEA1 is a Rab5 effector protein that triggers fusion of phagosome with late endosome (Koul *et al.* 2004). The phagolysosome is a microbicidal environment that contains antimicrobial peptides and hydrolases. Pathogenic microbes have evolved many strategies for survival inside host phagocytes. For example, *Mycobacterium tuberculosis* arrests phagosome maturation, whereas *Trypanosoma cruzi* escapes from the phagosome by making a pore. Others such as in *Leshmania mexicana* and *Coxiella brunetii*, are well adapted to the highly acidic environment of the phagolysosomes (Flannagan *et al.* 2009; Koul *et al.* 2004; Sacks & Sher 2002).

During *Chromera* infection, genes involved in early endosome formation such as Rab5 effectors (EAA1 and ALS2) and phosphoinositide-3- and 4-kinases (PI3K and PI4K) were down regulated. The response to *Chromera* therefore differs dramatically to that to a competent *Symbiodinium* strain, where ALS2, Ras GTPases and Rab5 effector proteins involved in early phagosome formation were up-regulated (Chapter 2). Moreover, localization of orthologs of Rab5 and Rab7 on *Aiptasia pulchella* phagosomes containing *Symbiodinium* provides evidence for modulation of the host endocytic pathway by the symbiont (Chen *et al.* 2004; Chen *et al.* 2003). During the process of phagosomal maturation, Rab7 replaces Rab5 on the late phagosomes, which acquire lysosomal markers such as lysosomal-associated membrane protein 1 (LAMP1), CD63, and acid hydrolases through fusion with lysosomes (Clemens & Horwitz 1995).

Acquisition of the vacuolar proton-ATPase molecules is also required for maturation of phagosomes, and this results in acidifying the phagolysosome medium (Sturgill-Koszycki *et al.* 1994). During *Chromera* infection, genes involved in late phagosome formation and phagosome maturation were up-regulated including the late endosome markers (Rab7a, LAMP1, VAMP molecules, CD63 molecule) and proteins involved in vesicle fusion (late endosomal/ lysosomal adaptor, SNAP-associated protein, VAMP7). Two subunits of the lysosomal proton-ATPase were also up regulated. Moreover several genes encoding vacuolar protein sorting components, TBC1 domain members, other Rab proteins, and lysosomal hydrolases had altered expression compared to the control, indicating the dynamic nature of the endosomal pathway. My previous findings imply that the host enhances phagosome maturation in order to promote fusion with late endosomes and/or lysosomes, thus destroying invading parasites. During the infection of the liver, there is a strong interaction between the *Plasmodium* parasite and late endosome/ lysosome markers, including Rab7a, LAMP1, and CD63 (Lopes da Silva *et al.* 2012).

The phagosome maturation step of the endosomal pathway has been a target for manipulation by many pathogens (Gruenberg & van der Goot 2006). Pathogenic bacteria, including *M. tuberculosis* (Koul *et al.* 2004) and parasitic apicomplexans, such as *Plasmodium* spp. (Lopes da Silva *et al.* 2012) utilize the phagosome maturation arrest strategy in order to survive and replicate in the phagolysosome. Pathogenic *M. tuberculosis* maintains Rab5, but inhibits Rab7 acquisition to the phagosome (Sun *et al.* 2007). In addition, its phagosome remains in mildly acidic medium, possibly via inhibition of the recruitment of proton-ATPase to the phagosomal membrane (Sturgill-Koszycki *et al.* 1994). However non-pathogenic strains of *M. tuberculosis* lack the ability to arrest the phagosome maturation (Koul *et al.* 2004).

3.5.5 Apoptosis as a double-edged sword for both coral host and Chromera survival

A key defense mechanism of multi-cellular organisms is killing infected cells in order to contain an infection, so the apoptotic pathway is critical in defense against many pathogens (Ashida *et al.* 2011; Lamkanfi & Dixit 2010). For example, macrophages that engulf potential pathogens frequently activate their apoptotic pathways in order to resolve the infection. However, many pathogenic bacteria, such as *Helicobacter pylori* (Mimuro *et al.* 2007) and pathogenic *M. tuberculosis* strains (Briken & Miller 2008), are able to manipulate host apoptotic pathways to enable their

survival. During *Chromera* infection, genes encoding both anti- and pro-apoptotic activities and involved in intrinsic and extrinsic apoptotic pathways were differentially expressed implying equilibrium between host cell survival and death.

Down-regulation of proteins with pro-apoptotic functions, including death receptors and APAF1 and up-regulation of Bcl-2, TRAF2, and TRAF5, which have anti apoptotic functions, suggests suppression of both extrinsic and intrinsic apoptotic pathways occurs during *Chromera* infection as is also seen in *M. tuberculosis* infection (Briken & Miller 2008). On the other hand, caspase inhibitors, suppressors of TNF-induced cell death, and other genes with anti-apoptotic activities were down regulated in *Chromera* infection, and genes implicated in apoptosis-inducing functions, for example, programmed cell death 6, death associated protein 1 and peptidyl-tRNA hydrolase 2 were up-regulated during *Chromera* infection. The simplest interpretation of these data is that *Chromera* infection provokes apoptosis of some host cells, whereas others up-regulate pro-survival genes in order to “ride-out” the clearance of the infection.

3.5.6 Conclusion

Since the discovery in 2008 of *Chromera* and its association with Australian corals, great interest has developed around this alga, primarily because of its unique position in the phylogenetic tree. Despite this interest, very little is known about the relationship of this novel alga with corals. The work described in this chapter aimed to address for the first time the nature of the relationship between corals and *Chromera* by challenging *A. digitifera* larvae with *Chromera* and looking at the whole transcriptome response of the host. Interestingly, the response to *Chromera* was found to differ fundamentally to that during infection with a competent *Symbiodinium* strain. The characteristics of the coral response to *Chromera* resemble those of mammalian hosts to parasites, as many key host immune responses are suppressed. During the infection, the host endocytic pathway is activated, presumably in order to kill the invading *Chromera* through phagosome maturation and lysosomal degradation. As has been reported in other host-parasite interactions, the host apoptotic pathways reacted in complex ways, reflecting attempts by the host to clear the infection by killing infected cells but also protect uninfected cells. In summary, the reaction of *A. digitifera* larvae to *Chromera* is similar to the responses of mammalian hosts to pathogens such as *Mycobacterium* and

parasites such as *Plasmodium*. Based on these results, I suggest that, rather than being a symbiont, *Chromera* should be considered a facultative parasite of corals.

3.6 REFERENCES

- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**, R106.
- Ashida H, Mimuro H, Ogawa M, *et al.* (2011) Cell death and infection: a double-edged sword for host and pathogen survival. *Journal of Cell Biology* **195**, 931-942.
- Augustin R, Bosch TC (2010) Cnidarian immunity: a tale of two barriers. *Adv Exp Med Biol* **708**, 1-16.
- Briken V, Miller JL (2008) Living on the edge: inhibition of host cell apoptosis by *Mycobacterium tuberculosis*. *Future Microbiol* **3**, 415-422.
- Brown T, Bourne D, Rodriguez-Lanetty M (2013) Transcriptional activation of c3 and hsp70 as part of the immune response of *Acropora millepora* to bacterial challenges. *Plos One* **8**, e67246.
- Burge CA, Mouchka ME, Harvell CD, Roberts S (2013) Immune response of the Caribbean sea fan, *Gorgonia ventalina*, exposed to an Aplanochytrium parasite as revealed by transcriptome sequencing. *Front Physiol* **4**, 180.
- Chen MC, Cheng YM, Hong MC, Fang LS (2004) Molecular cloning of Rab5 (ApRab5) in *Aiptasia pulchella* and its retention in phagosomes harboring live zooxanthellae. *Biochemical and Biophysical Research Communications* **324**, 1024-1033.
- Chen MC, Cheng YM, Sung PJ, Kuo CE, Fang LS (2003) Molecular identification of Rab7 (ApRab7) in *Aiptasia pulchella* and its exclusion from phagosomes harboring zooxanthellae. *Biochem Biophys Res Commun* **308**, 586-595.
- Clemens DL, Horwitz MA (1995) Characterization of the *Mycobacterium tuberculosis* phagosome and evidence that phagosomal maturation is inhibited. *J Exp Med* **181**, 257-270.
- Cullimore J, Denarie J (2003) Plant sciences. How legumes select their sweet talking symbionts. *Science* **302**, 575-578.
- Cumbo VR, Baird AH, Moore RB, *et al.* (2013) *Chromera velia* is Endosymbiotic in Larvae of the Reef Corals *Acropora digitifera* and *A. tenuis*. *Protist* **164**, 237-244.
- Davies D, Meade KG, Herath S, *et al.* (2008) Toll-like receptor and antimicrobial peptide expression in the bovine endometrium. *Reprod Biol Endocrinol* **6**, 53.
- Davy SK, Allemand D, Weis VM (2012) Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiol Mol Biol Rev* **76**, 229-261.
- DeSalvo MK, Sunagawa S, Voolstra CR, Medina M (2010) Transcriptomic responses to heat stress and bleaching in the elkhorn coral *Acropora palmata*. *Marine Ecology Progress Series* **402**, 97-113.
- Flannagan RS, Cosio G, Grinstein S (2009) Antimicrobial mechanisms of phagocytes and bacterial evasion strategies. *Nature Reviews Microbiology* **7**, 355-366.
- Gazzinelli RT, Kalantari P, Fitzgerald KA, Golenbock DT (2014) Innate sensing of malaria parasites. *Nat Rev Immunol* **14**, 744-757.
- Gestal C, Costa M, Figueras A, Novoa B (2007) Analysis of differentially expressed genes in response to bacterial stimulation in hemocytes of the carpet-shell clam *Ruditapes decussatus*: identification of new antimicrobial peptides. *Gene* **406**, 134-143.

- Gruenberg J, van der Goot FG (2006) Mechanisms of pathogen entry through the endosomal compartments. *Nat Rev Mol Cell Biol* **7**, 495-504.
- Guillard RR, Ryther JH (1962) Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (Cleve) Gran. *Canadian Journal of Microbiology* **8**, 229-239.
- Hase K, Kawano K, Nochi T, *et al.* (2009) Uptake through glycoprotein 2 of FimH(+) bacteria by M cells initiates mucosal immune response. *Nature* **462**, 226-230.
- Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57.
- Janouskovec J, Horak A, Barott KL, Rohwer FL, Keeling PJ (2012) Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. *Current Biology* **22**, R518-R519.
- Jenner RG, Young RA (2005) Insights into host responses against pathogens from transcriptional profiling. *Nature Reviews Microbiology* **3**, 281-294.
- Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30.
- Killick KE, Ni Cheallaigh C, O'Farrelly C, *et al.* (2013) Receptor-mediated recognition of mycobacterial pathogens. *Cell Microbiol* **15**, 1484-1495.
- Koul A, Herget T, Klebl B, Ullrich A (2004) Interplay between mycobacteria and host signalling pathways. *Nature Reviews Microbiology* **2**, 189-202.
- Kvennefors EC, Leggat W, Kerr CC, *et al.* (2010) Analysis of evolutionarily conserved innate immune components in coral links immunity and symbiosis. *Dev Comp Immunol* **34**, 1219-1229.
- Lamkanfi M, Dixit VM (2010) Manipulation of host cell death pathways during microbial infections. *Cell Host Microbe* **8**, 44-54.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25.
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.
- Libro S, Kaluziak ST, Vollmer SV (2013) RNA-seq Profiles of Immune Related Genes in the Staghorn Coral *Acropora cervicornis* Infected with White Band Disease. *Plos One* **8**.
- Lin KL, Wang JT, Fang LS (2000) Participation of glycoproteins on zooxanthellal cell walls in the establishment of a symbiotic relationship with the sea anemone, *Aiptasia pulchella*. *Zoological Studies* **39**, 172-178.
- Lopes da Silva M, Thieleke-Matos C, Cabrita-Santos L, *et al.* (2012) The host endocytic pathway is essential for *Plasmodium berghei* late liver stage development. *Traffic* **13**, 1351-1363.
- Miller DJ, Hemmrich G, Ball EE, *et al.* (2007) The innate immune repertoire in cnidaria--ancestral complexity and stochastic gene loss. *Genome Biol* **8**, R59.
- Mimuro H, Suzuki T, Nagai S, *et al.* (2007) *Helicobacter pylori* dampens gut epithelial self-renewal by inhibiting apoptosis, a bacterial strategy to enhance colonization of the stomach. *Cell Host Microbe* **2**, 250-263.
- Mohamed AR, Cumbo V, Harii S, *et al.* (2016) The transcriptomic response of the coral *Acropora digitifera* to a competent *Symbiodinium* strain: the symbiosome as an arrested early phagosome. *Molecular Ecology*.
- Moore RB, Obornik M, Janouskovec J, *et al.* (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**, 959-963.

- Moya A, Sakamaki K, Mason BM, *et al.* (2016) Functional conservation of the apoptotic machinery from coral to man: the diverse and complex Bcl-2 and caspase repertoires of *Acropora millepora*. *Bmc Genomics* **17**, 62.
- Nyholm SV, McFall-Ngai MJ (2004) The winnowing: establishing the squid-vibrio symbiosis. *Nature Reviews Microbiology* **2**, 632-642.
- Ocampo ID, Zarate-Potes A, Pizarro V, *et al.* (2015) The immunotranscriptome of the Caribbean reef-building coral *Pseudodiploria strigosa*. *Immunogenetics*.
- Ohno H, Hase K (2010) Glycoprotein 2 (GP2): grabbing the FimH bacteria into M cells for mucosal immunity. *Gut Microbes* **1**, 407-410.
- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140.
- Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25.
- Rodriguez-Lanetty M, Phillips WS, Weis VM (2006) Transcriptome analysis of a cnidarian-dinoflagellate mutualism reveals complex modulation of host gene expression. *Bmc Genomics* **7**, 23.
- Sacks D, Sher A (2002) Evasion of innate immunity by parasitic protozoa. *Nat Immunol* **3**, 1041-1047.
- Sakamaki K, Imai K, Tomii K, Miller DJ (2015) Evolutionary analyses of caspase-8 and its paralogs: Deep origins of the apoptotic signaling pathways. *Bioessays* **37**, 767-776.
- Shinzato C, Shoguchi E, Kawashima T, *et al.* (2011) Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* **476**, 320-323.
- Sturgill-Koszycki S, Schlesinger PH, Chakraborty P, *et al.* (1994) Lack of acidification in Mycobacterium phagosomes produced by exclusion of the vesicular proton-ATPase. *Science* **263**, 678-681.
- Sun J, Deghmane AE, Soualhine H, *et al.* (2007) Mycobacterium bovis BCG disrupts the interaction of Rab7 with RILP contributing to inhibition of phagosome maturation. *J Leukoc Biol* **82**, 1437-1445.
- Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192.
- Travers MA, Meistertzheim AL, Cardinaud M, *et al.* (2010) Gene expression patterns of abalone, *Haliotis tuberculata*, during successive infections by the pathogen *Vibrio harveyi*. *Journal of Invertebrate Pathology* **105**, 289-297.
- Vidal-Dupiol J, Dheilily NM, Rondon R, *et al.* (2014) Thermal stress triggers broad *Pocillopora damicornis* transcriptomic remodeling, while *Vibrio coralliilyticus* infection induces a more targeted immuno-suppression response. *Plos One* **9**, e107672.
- Vieira OV, Botelho RJ, Grinstein S (2002) Phagosome maturation: aging gracefully. *Biochem J* **366**, 689-704.
- Weis VM, Davy SK, Hoegh-Guldberg O, Rodriguez-Lanetty M, Pringle JR (2008) Cell biology in model systems as the key to understanding corals. *Trends in Ecology & Evolution* **23**, 369-376.
- Weiss Y, Foret S, Hayward DC, *et al.* (2013) The acute transcriptional response of the coral *Acropora millepora* to immune challenge: expression of GiMAP/IAN genes links the innate immune responses of corals with those of mammals and plants. *Bmc Genomics* **14**, 400.

- Wood-Charlson EM, Hollingsworth LL, Krupp DA, Weis VM (2006) Lectin/glycan interactions play a role in recognition in a coral/dinoflagellate symbiosis. *Cell Microbiol* **8**, 1985-1993.
- Wright RM, Aglyamova GV, Meyer E, Matz MV (2015) Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. *Bmc Genomics* **16**, 371.
- Yu S, Lowe AW (2009) The pancreatic zymogen granule membrane protein, GP2, binds *Escherichia coli* Type 1 fimbriae. *BMC Gastroenterol* **9**, 58.

Chapter 4.0 *Chromera* transcriptomics: a functional genomic resource for a chromerid alga isolated from the Great Barrier Reef and comparative transcriptomic analyses with parasitic and photosynthetic relatives

The content of the Chapter will be submitted as manuscript entitled:

Deciphering the poorly understood coral-*Chromera* symbiosis using comparative transcriptomics by Mohamed AR, Cumbo VR, Chan CX, Ragan MA, and Miller DJ. 2016.

Target journal: Molecular Biology and Evolution

Amin Mohamed wrote the entire chapter, with co-authors providing intellectual guidance in the design and implementation of the research and editorial contributions to the paper. Amin Mohamed conducted the experiments, collected and analysed data and produced all of the tables and figures.

4.1 ABSTRACT

Chromera is a newly identified apicomplexan-related alga that associates with reef-building corals in Australia including on the Great Barrier Reef (GBR). Here I used Illumina sequence data to construct a *de novo* transcriptome assembly of a strain of *Chromera* isolated from *Montipora digitata* from Magnetic Island (central GBR), referred to here as “GBR *Chromera*”. I obtained more than 39,000 contigs (N50 of 2,220 bases), 7,644 of which were annotated using the Swiss-Prot protein database, with less than 0.1 % of sequences having matches against bacterial genomes. The *de novo* transcriptome clusters from the GBR *Chromera* strain were compared with the coding sequences (CDS) of the reference *Chromera* strain isolated from *Plesiastrea versipora* from Sydney harbor (referred to here as “Sydney *Chromera*”) in order to identify orthologous sequences and genes under selection. This comparison led to the identification of 664 pairs of putative orthologs that share high levels of sequence identity to known proteins in the Swiss-Prot database. All of the ortholog pairs were under negative selection (with Ka/Ks ratio < 1). Only one gene, encoding a tetratricopeptide protein, was found to be under positive selection (with Ka/Ks > 1). CDS and transcriptome datasets of both Sydney and GBR *Chromera* strains were compared with data from the symbiont *Symbiodinium kawaguti* and the parasite *Plasmodium falciparum* in order to identify common functional categories and pathways shared between *Chromera* and its photosynthetic and parasitic relatives. 2,400 genes were common to both *Chromera* strains and were involved in the GO terms DNA repair, cellular response to stress, plastid, organelle envelope, transport. Two KEGG pathways involved in inositol phosphate metabolism and phosphatidylinositol signaling were enriched amongst the 507 genes common to *Chromera* and *Symbiodinium*. 241 genes were common to *Chromera* and *Plasmodium*, among them one KEGG pathway “proteasome” was significantly enriched. Mapping the transcriptomes to the KEGG server revealed similar overall pathway distributions in both *Chromera* and *Symbiodinium* that is quite different to that of *Plasmodium*. In *Chromera* and *Symbiodinium*, most of the genes were assigned to metabolism (specially carbon and amino acids metabolisms). On the other hand, in *Plasmodium* most of the genes were assigned to human diseases. Comparing KEGG pathways involved in glycan biosynthesis and transcription machinery the four datasets revealed the genetic uniqueness of the symbiotic dinoflagellate *Symbiodinium*. This chapter shows that the

Chromera gene catalogue includes some key genes that may be implicated in coral-algal symbiosis, such as those involved in photosynthesis, transport and cellular response to stress. More importantly, the similarity of the proteasome machinery in *Chromera* and *Plasmodium* is significant in the context of the potential role of *Chromera* as a model organism to develop anti-malarial drugs targeting the proteasome.

4.2 INTRODUCTION

Chromerida is a newly defined phylum of photoautotrophic alveolates that so far includes only two described species *Chromera velia* (Moore *et al.* 2008) and *Vitrella brassicaformis* (Obornik *et al.* 2012) isolated from corals in Australia. These novel algae are the closest free-living relatives of the parasitic Apicomplexa (Moore *et al.* 2008), however they are associated with corals and the relationship has been assumed to be symbiotic (Cumbo *et al.* 2013). *Chromera* has been isolated from and detected in hard corals of Sydney Harbor (Moore *et al.* 2008) and the Great Barrier Reef, Australia (Cumbo *et al.* 2013; Slapeta & Linares 2013). *Chromera* sequences were also detected in the soft coral *Lobophytum pauciflorum* from Orpheus Island on the GBR using *Chromera*-specific PCR primers and confirmed by Sanger sequencing (Appendix II). *Chromera* has also been isolated from the Caribbean coral *Agaricia agaricites* (Visser *et al.* 2012). *Chromera* has been experimentally demonstrated to successfully infect larvae of *Acropora digitifera* and *A. tenuis* (Cumbo *et al.* 2013), and can be cultured in the laboratory. The plastid, mitochondrial and nuclear genomes of *Chromera* have been described (Flegontov *et al.* 2015; Janouskovec *et al.* 2010; Moore *et al.* 2008; Woo *et al.* 2015). Results presented in Chapter 3 suggest that the relationship between *Chromera* and corals may be parasitic rather than symbiotic, as the transcriptomic responses of *Acropora digitifera* larvae to *Chromera* infection showed pathways similar to that found in host-pathogen interactions such as phagosome maturation, immune suppression and apoptosis. Recent genome sequencing data for the only two described Chromerids (*Chromera* and *Vitrella*) (Woo *et al.* 2015) have reconfirmed the close relatedness of these photosynthetic algae to the apicomplexans, and suggest the loss of thousands of genes during evolution from free-living algae into the apicomplexan ancestor. The *Chromera* “CCMP2878” strain (originally isolated from Sydney harbor; Moore *et al.*, 2008) has an estimated a genome size of approximately 193.6 Mb and has 26,112 predicted protein-encoding genes (Woo *et al.* 2015). In this Chapter, a

Chromera strain isolated from a different coral host from a distinct geographic location was used. The *Chromera* “Mdig03 strain” was originally isolated from *Montipora digitata* on the Great Barrier Reef (Cumbo *et al.* 2013). RNA was isolated from *Chromera* cultures grown under different conditions and sequence data obtained using an Illumina Hi-Seq 2000 platform. I assembled a *de novo* transcriptome, providing a novel sequence resource for *Chromera*. I conducted large-scale comparative analyses to explore the presence of conserved and/or fast-evolving genes in the two *Chromera* strains. In addition, to explore the possible roles of the *Chromera* genes in symbiosis, gene ontology and KEGG pathway analyses were performed on genes shared between *Chromera* strains and between *Chromera* and *Symbiodinium* and/or *Plasmodium*. Finally, the two *Chromera* datasets were compared with transcriptomes of symbiotic *Symbiodinium* and parasitic *Plasmodium* in terms of KEGG pathways.

4.3 MATERIALS AND METHODS

4.3.1 *Chromera velia* culture

A culture of *Chromera* (Mdig3 strain) was obtained from the University of Technology Sydney. This strain was used in this Chapter and referred to as “GBR *Chromera*”). It was originally isolated from the stony coral *M. digitata* (Cnidaria: Acroporidae) from Nelly Bay, Magnetic Island (Lat 19°09'44. 39''S, Long 146°51'14. 90''E) on the inner central part of the Great Barrier Reef (Cumbo *et al.*, 2013). Cultures were maintained at 25 °C in Guillard’s f/2 medium on 12 h/12 h, day and night regime. The medium used for culturing was prepared by adding 20 ml of Guillard’s f/2 Marine Water Enrichment Solution 50X (G0154, Sigma-Aldrich) to 980 mL 0.2 µm sterile FSW.

4.3.2 *Chromera* culturing conditions for transcriptome construction

Chromera cultures were inspected with microscopy for any bacterial and/or protest contamination and *Chromera* identity was checked using specific PCR primers (see Appendix IV). Cultures growing in the mid exponential (log) phase (+11 days after inoculation) were harvested at the middle of the cultures’ daytime phase and labeled as

“control”. In order to maximize the variety of expressed genes, the cultures were subjected to variety of treatments before RNA isolation and cDNA libraries preparation. Hence, cultures were subjected to dark stress (24 hour dark period), cold shock (4 °C for 4 hours), and heat shock (36 °C for 4 hours). Cultures growing in the control conditions +8 days after inoculation cultures were harvested at the middle of the cultures’ daytime phase and labeled as “motile” as cultures showed both *Chromera* life forms. In addition cultures were also grown in f/2 media autotrophically while supplemented with exogenous organic compounds at final concentration of total 0.1%(w/v) including; Galactose (D+) (D00201; Sigma-Aldrich), sodium acetate (D00385; ICN Biomedicals) and Glycerol (D00217; Sigma-Aldrich) and labeled as “mixotrophic”. In all cases, exponentially growing and cultures were separated and subjected to the treatment condition and harvested at the end of the experimental treatment. During the culturing no antibiotics were used to exclude any potential contribution of the antibiotic treatment to the mRNA expression in the cultures.

4.3.3 RNA isolation and high-throughput next generation sequencing

50 ml of *Chromera* cultures were pelleted by spinning the cultures at 3,000 x g for 5 min. Pellets were suspended in 1 ml 0.2 µm sterile FSW and centrifuged at 3,000 x g for 5 min. Pellets were snap frozen in liquid nitrogen and stored at -80 °C until further treatment. Total RNA was isolated from ~ 80 mg of the frozen *Chromera* pellets using the RNAqueous® Total RNA Isolation Kit (Ambion). The pellets were lysed twice for 20 s at 4.0 ms⁻¹ in Lysing Matrix D tubes (MP Biomedicals, Australia) containing 960 µL of lysis/binding solution plus 80 µL of the Plant RNA Isolation Aid (Ambion, USA) on a FastPrep®-24 Instrument (MP Biomedicals, Australia). RNA was bound to filter cartridges supplied with the kit and washed three times, finally RNA was eluted in 40 µL of the elution solution. RNA quantity and quality were assessed using NanoDrop ND-1000 spectrometer, a Qubit fluorometer and Agilent 2100 bioanalyzer.

Messenger RNA (mRNA) was isolated from 1 µg of total RNA and 6 RNA-Seq libraries were prepared using the TruSeq RNA Sample Preparation Kit (Illumina). Libraries were sequenced on Illumina HiSeq 2000 platform at the Australian Genome

Research Facility (AGRF) in Melbourne, Australia. Sequencing produced a total of 189.5 million individual 100-bp paired-end reads.

4.3.4 RNA-Seq data quality control, processing and quality filtering

Illumina raw reads from each paired end files were visualized using FastQC version 0.11.2 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), a quality control tool for high throughput sequencing data in order to determine the quality of the data. In addition, the raw reads were inspected for adapters contamination by searching for the illumina universal and indexed adapters sequences. The raw illumina reads were cleaned and filtered using Trimmomatic software version 0.32 (Bolger *et al.* 2014) (<http://www.usadellab.org/cms/index.php?page=trimmomatic>) based on quality and size. Trimmomatic was used to clip both universal and indexed illumina adapters (sequences are provided in the supplementary file Appendix IV). Quality trimming was also performed; leading and trailing bases with Phred quality score < 25 were removed, average Phred quality score was calculated in 4 bp sliding windows. Bases were trimmed from the point in the read where average Phred quality score dropped below 20 (which means that the chances that a base is called incorrectly is 1 in 100). Following trimming reads that had < 50 bp were also excluded.

The following bash script was used to run Trimmomatic:

```
java -jar PATH/trimmomatic-0.32.jar PE -phred33 -trimlog  
trim.log R1.fastq.gz R2.fastq.gz R1_paired.fastq  
R1_unpaired.fastq trimo_R2_paired.fastq trimo_R2_unpaired.fastq  
ILLUMINACLIP: adapter.txt: 2: 30: 10 LEADING: 25 TRAILING: 25  
SLIDINGWINDOW: 4: 20 MINLEN: 50
```

4.3.5 Transcriptome de novo assembly using the TRINITY software

The trimmed/ filtered illumina reads were used for *de novo* transcriptome assembly using the Trinity RNA-Seq assembly software (<http://sourceforge.net/projects/trinityrnaseq>). The assembly was carried out with the recommended protocol described in (Haas *et al.* 2013) and using options appropriate for *de novo* transcriptome assembly of strand specific RNA-Seq libraries, a minimum contig length of 500 and reads normalization were specified.

The following bash script was used to run the trinity assembly:

```
Trinity.pl --seqType fq \  
--left trimmed_R1_paired.fastq 6 files --right \  
trimmed_R1_paired.fastq 6 files \  
--JM 200G --CPU 6 --SS_lib_type FR --min_contig_length 500 \  
--normalize_reads --output Assembly_Chromera_Trinity_full
```

The trinity platform represents a novel algorithm developed specifically for *de novo* reconstruction of transcriptomes from RNA-Seq data. It combines three independent software modules: 1) Inchworm, 2) Chrysalis and 3) Butterfly, applied sequentially to process large volumes of RNA-Seq data in order to make linear contigs from RNA-Seq reads, generate and expand de Bruijn graphs and finally output both transcripts and isoforms in a FASTA format as the fully assembled transcriptome (Haas *et al.* 2013) (Figure 4.1). A flowchart for steps of generating the *de novo* transcriptome is depicted in Figure 4.1.

4.3.6 Assessing the quality of the de novo transcriptome assembly

In order to validate the accuracy of the *de novo* assembly, reads were mapped to the *de novo*-assembled transcriptome using BOWTIE aligner version 0.12.7 (Langmead *et al.* 2009) using the default parameters. Bowtie was used to map reads back to the transcriptome assembly separately, after which the read pairs were grouped into properly paired reads. The percent of the mapped reads as proper pairs was used to assess the assembly quality. Moreover, BLASTN (E-value of $\leq 10^{-10}$) was performed against bacterial genomes downloaded from the GenBank at NCBI to determine if there were bacterial contamination in the used *Chromera* cultures.

4.3.7 Transcriptome functional annotations

4.3.7.1 Gene Ontology (GO) terms

The longest isoform per gene was selected using custom perl script from the trinity output “assembled transcriptome” for the purpose of annotation. *Chromera* contigs were annotated by similarity search using BLAST software. Batch BLASTX searches were conducted locally against the Swiss-Prot protein database downloaded in

September 2014 (E-value cut off of $1.0E^{-3}$ and maximum 20 hits). Blast2GO suite (version 2.6.5) (<http://blast2go.com/b2ghome>) was used for transcriptome functional annotation and analysis of *Chromera* assembled contigs according to molecular function, cellular component, and biological process categories. GO annotations were generated through 3-step process: local blast, mapping and annotation. BLASTX output files (in xml format) were imported into Blast2GO software in order to map and annotate GO terms to *Chromera* contigs with BLAST hits searches against the Swiss-Prot. GO terms associated with the Swiss-Prot hits were extracted and modulated. Combined graphs were generated with a level 2 cut-off. Moreover, in order to identify protein domains InterPro terms were obtained from InterProScan database and merged to GOs. The GO distributions were statistically analyzed.

4.3.7.2 KEGG pathways

KEGG analysis was performed using the online KEGG Automatic Annotation Server (KAAS) (Moriya *et al.* 2007) (http://www.genome.jp/kaas-bin/kaas_main) in order to obtain an overview of gene pathways. The bi-directional best hit (BBH) method was carried out by BLAST searching against the KEGG database in order to obtain KEGG orthology (KO) assignments.

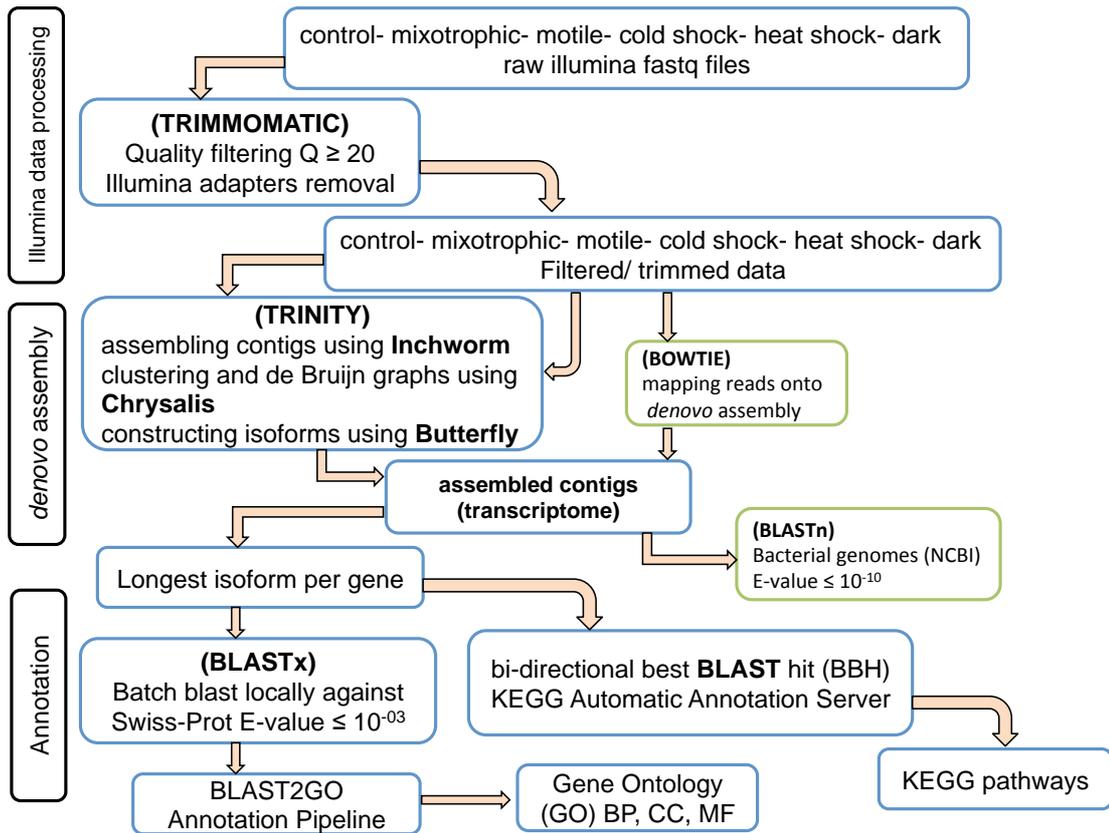


Figure 4.1 *Chromera* transcriptome construction workflow. The schematic flowchart shows the three-steps process of raw data processing, *de novo* transcriptome assembly and functional annotation.

4.3.8 *Chromera* stains comparative transcriptome analysis

4.3.8.1 Identification of the orthologous genes between two *Chromera* strains

Chromera (CCMP2878) called “Sydney *Chromera*” was originally isolated from *Plesiastrea versipora* (Faviidae) from Sydney Harbor (Moore *et al.* 2008). Genome and transcriptome are available for that strain (Woo *et al.* 2015) and the coding sequence (CDS) data was downloaded from the CryptoDB database (<http://cryptodb.org/cryptodb/>). The Reciprocal Best Blast Hit (RBBH) approach was followed to identify pairs of putative orthologs between GBR- and Sydney- *Chromera* sequences. The BRBH approach has been widely used to identify orthologous genes between closely related species. Briefly, all-against-all BLASTN was conducted using $E\text{-value} \leq 10^{-15}$. A custom python script (provided in the supplementary material Appendix IV) was used to identify the reciprocal best hits and considered putative

orthologs. In order to filter those putative orthologs (to get rid of paralogs), BLASTx against Swiss-Prot database was performed so that each putative ortholog pair should only hit the same Swiss-Prot ID. Batch BLASTX was conducted locally against the Swiss-Prot protein database downloaded in September 2015 (E-value cut off $\leq 10^{-15}$ and maximum 20 hits).

4.3.8.2 Chromera orthologs functional profile

In order to infer the function of the list of *Chromera* orthologs, GO and KEGG pathway enrichment analyses were performed using the database for annotation, visualization and integrated discover (DAVID) (Huang *et al.* 2009). The UNIPROT ACCESSION IDs of the annotated genes were used as identifiers. The same BLASTX was performed to the whole transcriptomes and all the Swiss-Prot annotated genes contributed to the background gene set for the enrichment analysis. DAVID uses the Fisher's exact test to ascertain statistically significant pathway enrichment among differentially expressed clusters relative to the background. The Benjamini-corrected *P*-value was applied with a cutoff of 0.05 to filter the significantly enriched KEGG pathways and GO categories.

4.3.8.3 Estimation of substitution rates between Chromera strains

In order to identify the translated proteins for each of the 664 ortholog pairs, sequences were translated using EMBOSS Transeq (http://www.ebi.ac.uk/Tools/st/emboss_transeq/). The Sydney *Chromera* orthologs were translated at frame +1 (as they are coding sequences) while the GBR *Chromera* orthologs were translated at all six frames. A *K*-mer ($k=3$) matching was used to compare each of the six protein sequences (in the case of GBR *Chromera*); the protein that has the highest percentage *K*-mer match against the Sydney *Chromera* protein was considered in the correct reading frame and was used as the GBR *Chromera* protein ortholog. Pairwise alignments of each protein pair were generated using ClustalW2 via an array of bash scripting. The obtained pairwise protein alignments were translated into codon alignments using Pal2Nal (version14) (Suyama *et al.* 2006) and the perl script parseFastaIntoAXT.pl (distributed with the KaKs_Calculator) was used to convert the alignment file into the AXT format required for KaKs-calculator. The codon

alignments were used to calculate substitution rates for non-synonymous (Ka) and synonymous (Ks) sites using a 14-model averaging method implemented in KaKs_Calculator (Wang *et al.* 2010). Ka/Ks ratios were used to determine whether genes were under positive, neutral or negative selection. Ka/Ks ratio > 1 (positive selection) implies adaptive evolution, Ka/Ks ratio = 1 (neutral selection) implies neutral evolution and Ka/Ks < 1 (purifying selection) implies conservation.

4.3.9 Chromera, Symbiodinium and Plasmodium comparative transcriptome analyses

4.3.9.1 Symbiodinium and Plasmodium sequences and functional annotation

Coding sequences (CDS) of *Symbiodinium kawagutii* (clade F) were downloaded from the Symka Genome Database (http://web.malab.cn/symka_new/) and transcriptome data for *Plasmodium falciparum* were downloaded from the PlasmoDB (<http://plasmodb.org/plasmo/>). GBR/Sydney *Chromera*, *Symbiodinium* and *Plasmodium* sequences were aligned to Swiss-Prot using BLASTX (E-value $\leq 10^{-15}$ and maximum 20 hits).

4.3.9.2 GO enrichment and KEGG pathways analyses based on shared genes

The GBR/ Sydney *Chromera*, *Symbiodinium*, and *Plasmodium* Swiss-Prot hits were used to obtain a four-way Venn diagram using the online software VENNY (<http://bioinfogp.cnb.csic.es/tools/venny/>) in order to identify shared hits among *Chromera*, *Symbiodinium* and *Plasmodium*. GO and KEGG pathway enrichment analyses were performed on the list of shared genes using DAVID (tool as described above) to identify enriched functional categories and pathways. Sydney *Chromera*, *Symbiodinium* and *Plasmodium* sequences were also mapped against the KEGG database as previously performed for GBR *Chromera*. Genes mapped to different KEGG categories were calculated and compared. Moreover, genes involved in N-glycan biosynthesis, transcription and proteasome pathways were compared.

4.4 RESULTS

4.4.1 *Illumina sequencing and de novo assembly of the Chromera transcriptome*

Illumina RNA-Seq data was utilized to generate a reference transcriptome data for the *Chromera velia* isolated from *M. digitata* on the GBR. *Chromera* cultures were grown under a wide variety of conditions in an attempt to trigger expression of most *Chromera* genes. In each case, cultures were harvested at the mid exponential phase of growth, total RNA isolated and cDNA libraries (6) were generated and sequenced using the Illumina HiSeq 2000 platform. Sequencing yielded 49.1 GB of data, from a total of 189.5 million Illumina 100bp paired end (PE) reads and an average of 31.5 million PE reads per library (Table 4.1). After quality filtering and adaptor trimming, 166.5 million PE processed reads were assembled *de novo* with the Trinity assembler, which makes large consensus sequences called “contigs” from overlapping sequencing reads (using Inchworm), clusters the Inchworm contigs into a de Bruijn graph component (using Chrysalis) and extracts all possible sets of isoforms or transcripts (using Butterfly). Trinity generated a transcriptome assembly consisting of 79,842 contigs with an N50 value of 2,289 bases, GC content of 53.42% and minimum contig length of 500 bases. The longest isoform per gene was extracted from the full transcriptome, resulting in 39,457 genes with an N50 of 2,220 bases. Trinity assembly statistics and length distribution of contigs are shown in Table 4.2 and Figure 4.2.

Table 4.1 Raw and processed Illumina reads after quality filtering and trimming

Library	Raw reads	Processed reads	Yield (Gb)
Control	31355987.00	27593316.00	7.10
Cold shock	31750562.00	28034742.00	7.00
Heat shock	31491195.00	27467388.00	7.00
Dark	31804840.00	27901593.00	7.00
Motile	30174194.00	26474964.00	7.30
Mixotrophic	32917403.00	28932188.00	6.60
Total	189494181.00	166404191.00	42.00

Table 4.2 Summary statistics of *Chromera de novo* transcriptome assembly using Trinity software

Trinity Outputs:	
Total trinity 'genes'	39457.00
Total trinity transcripts	79842.00
Percent GC	53.42
Statistics based on ALL transcript contigs:	
Contig N10	5429.00
Contig N20	3997.00
Contig N30	3230.00
Contig N40	2703.00
Contig N50	2289.00
Median contig length	1461.00
Average contig	1838.03
Total assembled bases	146752195.00
Statistics based on ONLY LONGEST ISOFORM per 'GENE':	
Contig N10	5212.00
Contig N20	3867.00
Contig N30	3127.00
Contig N40	2624.00
Contig N50	2220.00
Median contig length	1403.00
Average contig	1764.26
Total assembled bases	69612282.00

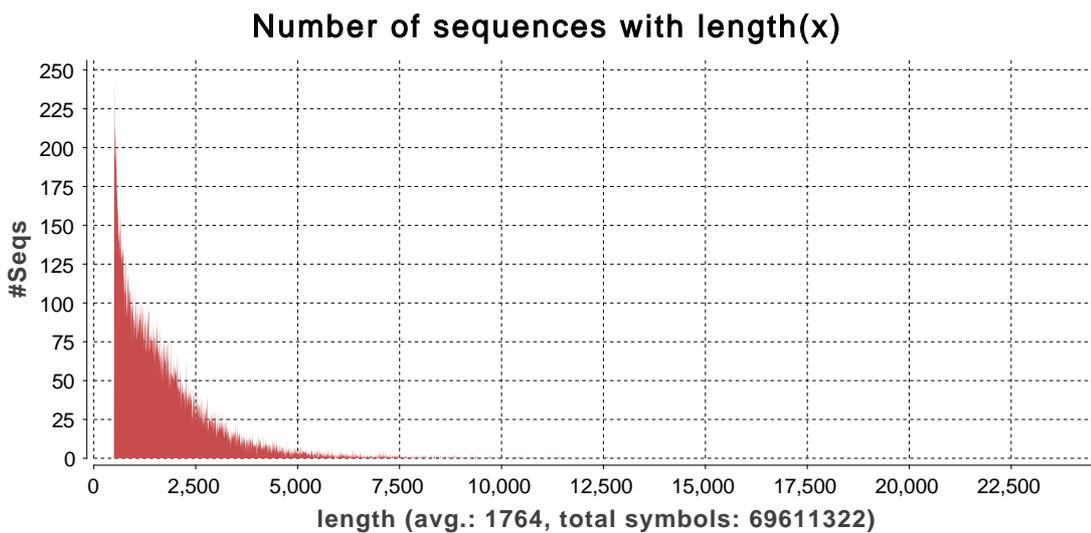


Figure 4.2 Distribution of contig length (≥ 500 nt) of GBR *Chromera de novo* transcriptome assembly.

4.4.2 Transcriptome annotation

BLASTX searches of list of the longest isoform per gene (39,457 genes) against well-annotated proteins in the Swiss-Prot database were conducted with E-value cut off of $1.0E^{-3}$. 7,644 genes (19.4%) had hits to the Swiss-Prot database. The low percentage of annotated genes might be attributed to limited information about protein sequences of phylogenetically closely related organisms in the database. Moreover some of these sequences might represent novel genes unique to chromerid algae. BLASTX top-hit species distribution of Swiss-Prot annotated genes showed high level of identity to human (*Homo sapiens*), followed by the flowering plant *Arabidopsis thaliana*, then the house mouse (*Mus musculus*) and slime mold (*Dictyostelium discoideum*). In addition, some *Chromera* genes had greater similarity to genes from parasites and pathogens, such as the malaria parasite (*Plasmodium falciparum*), the tuberculosis pathogen (*Mycobacterium tuberculosis*), *Acanthamoeba polyphaga* and *Haemophilus influenza* (Figure 4.3).

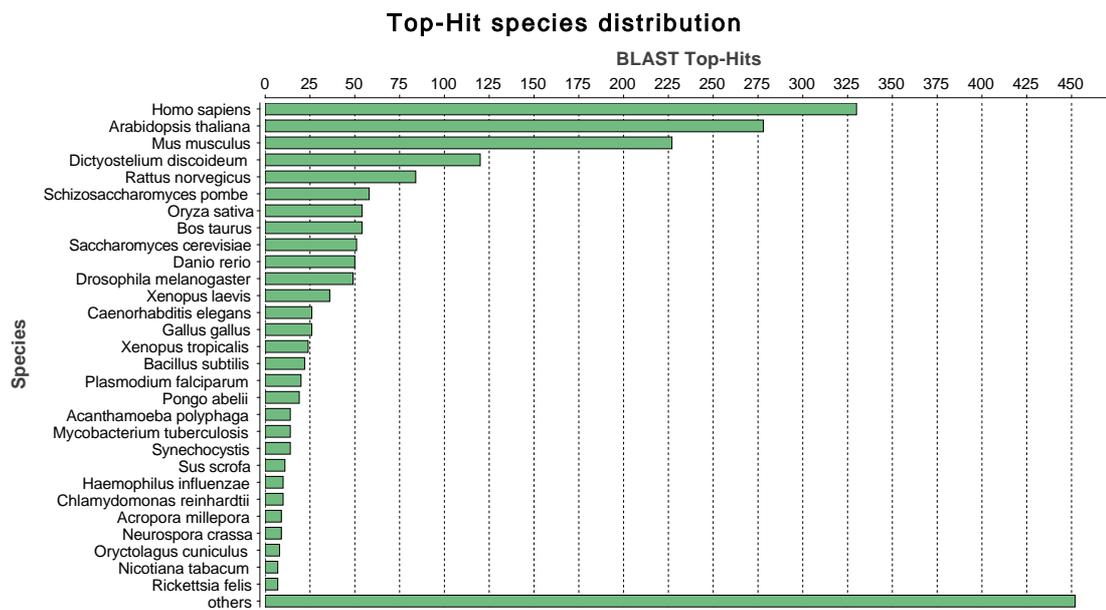


Figure 4.3 BLASTX top-hit species distribution of gene annotations with high homologies to species with known protein sequences in the Swiss-Prot database with E-value cut off of $1.0E^{-3}$.

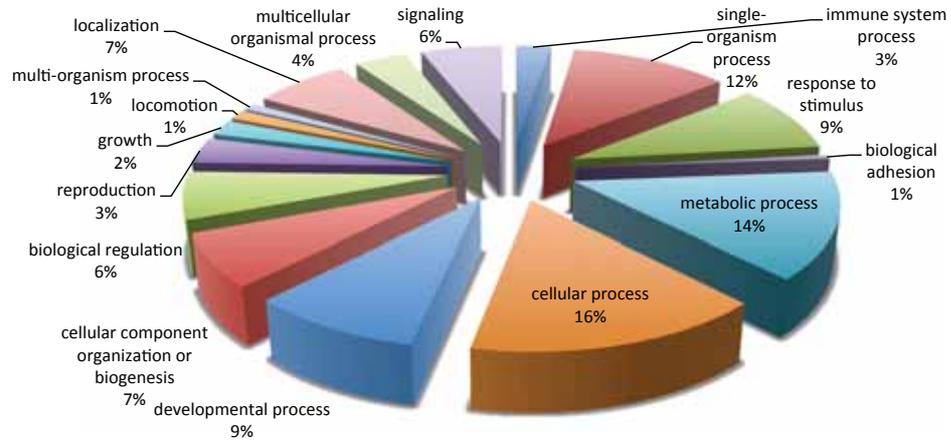
4.4.2.1 GO terms distribution

In order to classify *Chromera* genes based on their putative function, gene ontology (GO) analysis was performed where genes were assigned to appropriate biological process (GO-BP), molecular function (GO-MF) and cellular component (GO-CC) functions using the Blast2GO suite. The Swiss-Prot BLAST results were used to retrieve the associated GO terms in the three ontologies. Of the 7,644 genes that had Swiss-Prot annotation, 5,225 (68.35%) were assigned to GO terms, yielding a total of 38,271 GO assignments. Biological process accounted for the majority of GO assignments (22,205 counts, 58.02%), followed by cellular component (11,149 counts, 29.1%) and molecular function (4,917 counts, 12.8%). Genes involved in the cellular process “GO:0009987” and the metabolic process “GO:0008152” categories (16 and 14 % respectively) were highly represented among the biological process category. In the molecular function, the most represented GO categories were the GO:0003824 catalytic activity (46%) followed by the GO:0005488 binding (36%). Regarding cellular component, the GO:0005623 cell and GO:0043226 organelle were highly represented with 39% and 33%, respectively (Figure 4.4).

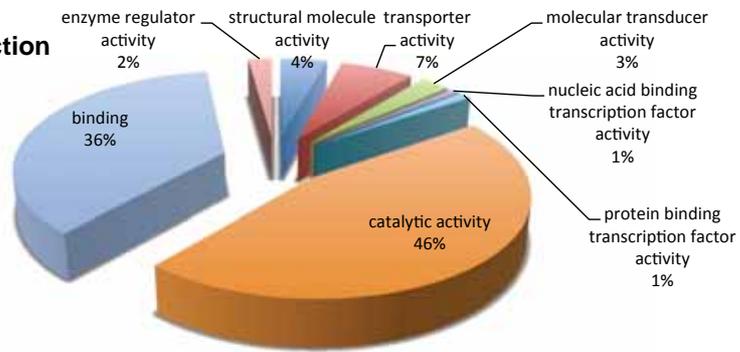
4.4.2.2 KEGG pathway analysis

KEGG analysis was performed using the KEGG Automatic Annotation Server (KAAS), and resulting in 4,220 genes involved in 336 different pathways being assigned with KEGG annotations (Table 4.3). Of the 4,220 genes with KEGG annotation, 34% (1,442) were classified into the metabolism category and these were primarily associated with “carbohydrate metabolism” (275 genes) and “amino acid metabolism” (266 genes). 18% of the KEGG-annotated genes were classified as associated with human disease pathways (Figure 4.5), of which pathways of infectious diseases were well represented (41% of human diseases pathways). Of the genes classified into genetic information processing pathways (17% of the KEGG-annotated genes), most were associated with translation and folding, sorting, and degradation rather than with transcriptional or other processes. Signal transduction and infectious diseases were the most well-represented pathways. Pathway data within the main KEGG categories are summarized in Figure 4.6.

A) Biological Process



B) Molecular Function



C) Cellular Components

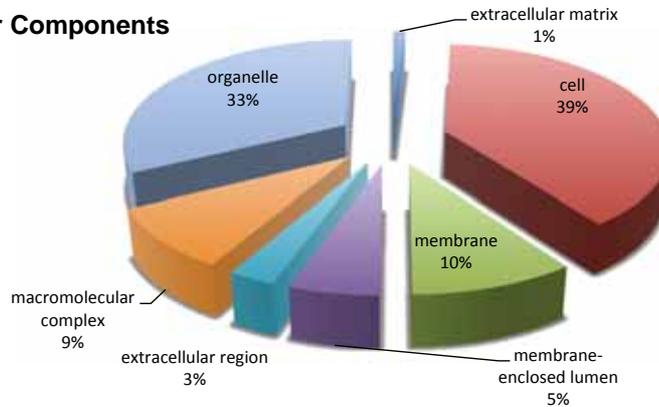


Figure 4.4 Gene Ontology (GO) assignment (2nd level GO terms) of transcripts from the GBR *Chromera* strain. Biological processes (A) constituted that majority of GO assignment of contigs (22,205 counts, 58.02%), followed by cellular components (C) (11,149 counts, 29.1%) and molecular function (B) (4,917 counts, 12.8%).

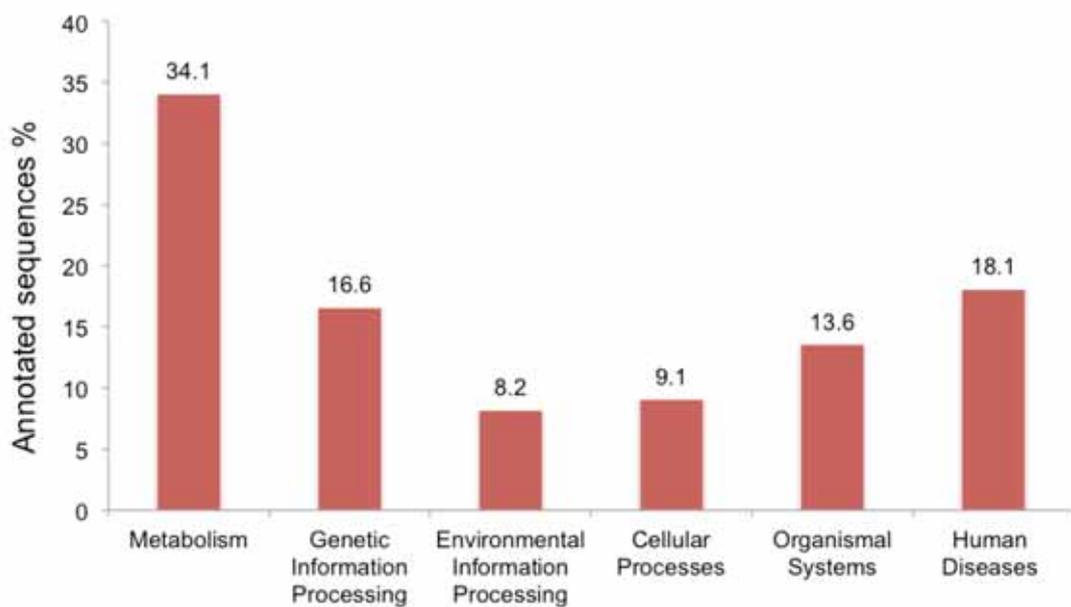


Figure 4.5 Main KEGG pathway category representation and percentages in the case of the GBR *Chromera* strain. Numbers above the bars give percent of annotated sequences in each category.

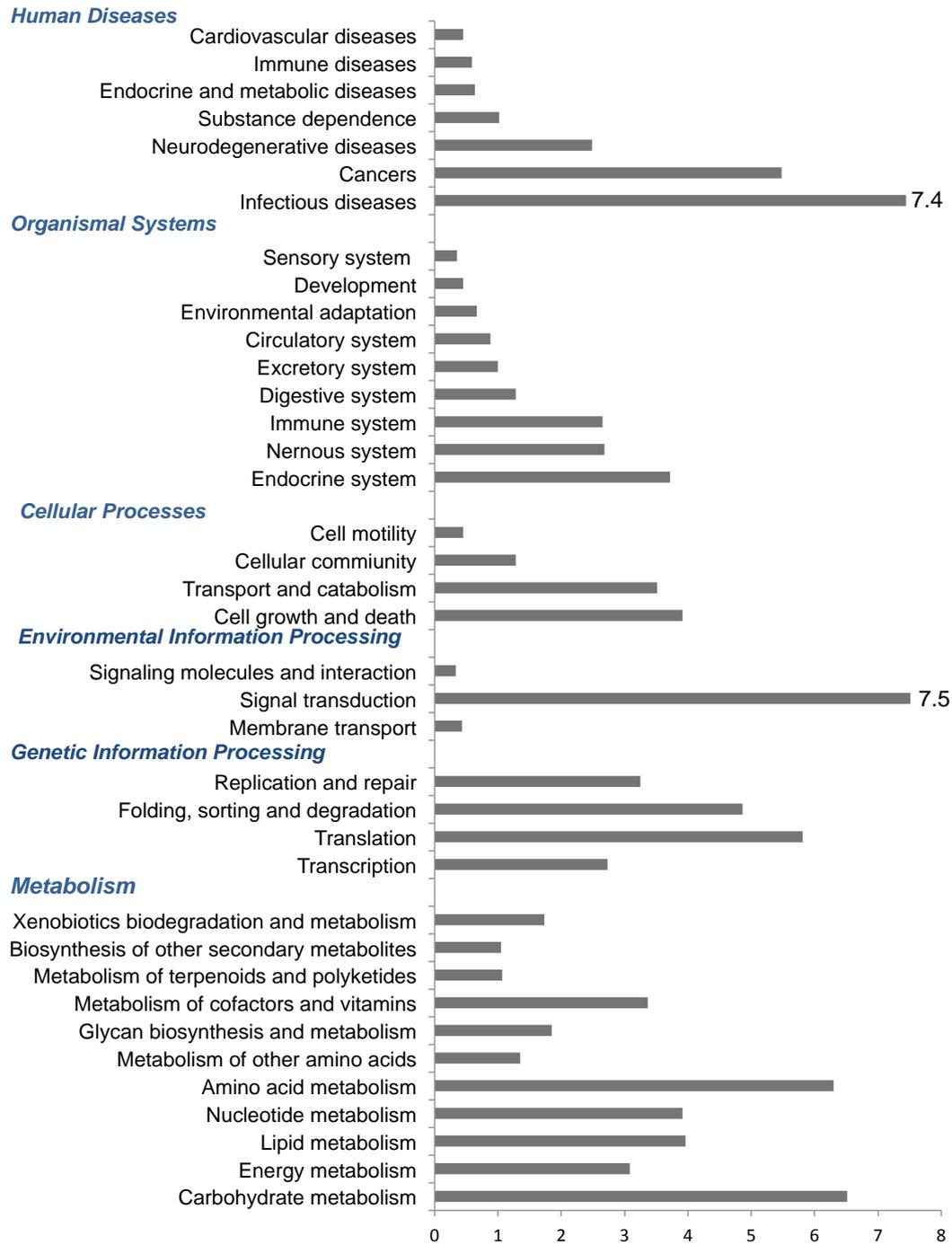


Figure 4.6 Distribution of KEGG pathways in transcriptome of the GBR *Chromera* strain. The charts show the percentages of the sequences assigned with each category.

Table 4.3 Summary of KEGG orthology data for the GBR *Chromera* strain

KEGG categories	No. of annotated sequences (%)	No. of pathways
Metabolism	1442	127
Carbohydrate metabolism	275 (19.07)	15
Energy metabolism	130 (9.01)	8
Lipid metabolism	167 (11.58)	17
Nucleotide metabolism	165 (11.44)	2
Amino acid metabolism	266 (18.44)	13
Metabolism of other amino acids	57 (3.95)	6
Glycan biosynthesis and metabolism	78 (5.40)	12
Metabolism of cofactors and vitamins	142 (9.84)	12
Metabolism of terpenoids and polyketides	45 (3.12)	13
Biosynthesis of other secondary metabolites	44 (3.05)	13
Xenobiotics biodegradation and metabolism	73 (5.06)	16
Genetic Information Processing	702	22
Translation	245 (35)	5
Folding, sorting and degradation	205 (29.1)	7
Replication and repair	137 (19.5)	7
Transcription	115 (16.4)	3
Environmental Information Processing	349	33
Signal transduction	317 (90.8)	27
Membrane transport	18 (5.2)	2
Signaling molecules and interaction	14 (4)	4
Cellular Processes	386	20
Cell growth and death	165 (42.7)	7
Transport and catabolism	148 (38.4)	5
Cellular community	54 (13.9)	5
Cell motility	19 (4.9)	3
Organismal Systems	577	69
Endocrine system	157 (27.2)	14
Nervous system	113 (19.6)	10
Immune system	112 (19.4)	15
Digestive system	54 (9.4)	9
Excretory system	42 (7.3)	5
Circulatory system	37 (6.4)	3
Environmental adaptation	28 (4.9)	5
Development	19 (3.3)	3
Sensory system	15 (2.5)	5
Human Diseases	764	65
Infectious diseases	314 (41)	24
Cancers	231 (30.3)	20
Neurodegenerative diseases	105 (13.7)	5
Substance dependence	43 (5.7)	5
Endocrine and metabolic diseases	27 (3.5)	3
Immune diseases	25 (3.3)	4
Cardiovascular diseases	19 (2.5)	4

4.4.3 Assessing the quality of the de novo transcriptome assembly

To assess the quality of the transcriptome assembly, reads from each RNA-Seq library were mapped to the assembly, with the result that an average of 84 % of the reads mapped successfully (Table 4.4). The assembled *Chromera* sequences (contigs) had 0.1% of sequences with BLASTN hits to bacterial databases (E-value $\leq 10^{-12}$), indicating minimal bacterial contamination. In order to estimate the completeness of the assembled transcriptome, the KEGG annotation was searched for core protein complexes/ pathways, with the result that for each pathway the majority of genes were found (Table 4.5; for more detailed results, see also Supplementary Figures S4.2, S4.3, S4.4, S4.5 and S4.6).

Table 4.4 Assessing the read content of the transcriptome using the percentages of the mapped reads.

RNA-Seq library	Control	Cold shock	Heat shock	Dark	Motile	Mixotrophic
Mapped paired reads %	85.2	83.9	86.9	83.6	81.3	82.5

Table 4.5 Selected KEGG pathways/protein complexes identified in the GBR *Chromera* transcriptome

Pathway/protein complex	Pathway ID	Known genes	Identified genes
Ribosome biogenesis in eukaryotes	ko03008	82	48
Ribosome	ko03010	143	88
RNA polymerase	ko03020	32	20
Spliceosome	ko03040	121	86
Proteasome	ko03050	48	29

4.4.4 Comparative transcriptomics of *Chromera* stains

4.4.4.1 Identification of orthologous genes between Sydney and GBR *Chromera* strains

The Reciprocal Best Blast Hit (RBBH) method was used to compare the 31,799 coding sequences identified in the Sydney *Chromera* strain to the 39,457 contigs assembled for the GBR *Chromera*. More than 19,000 pairs of putative orthologs were identified (BLASTN; E-value $\leq 10^{-15}$). For the next stage of analysis, the sequences were compared with the Swiss-Prot database using BLASTX with a cutoff of $E \leq 10^{-15}$,

only sequence pairs that hit the same Swiss-Prot ID were considered as true orthologs. The 664 ortholog pairs identified using this approach were subjected to further analyses (Figure 4.7).

4.4.4.2 *Chromera* orthologs functional profile

GO analysis of the 664 *Chromera* ortholog pairs revealed significant enrichment of specific GO categories (using the Benjamini correction P -value ≤ 0.05 as cutoff). Enriched GO categories included 8 cellular component categories related to plastid and organelle envelope. In the case of the molecular function category, there were 11 enriched categories related to ATP binding and ATPase activity. There were no enriched biological process categories (Table 4.6).

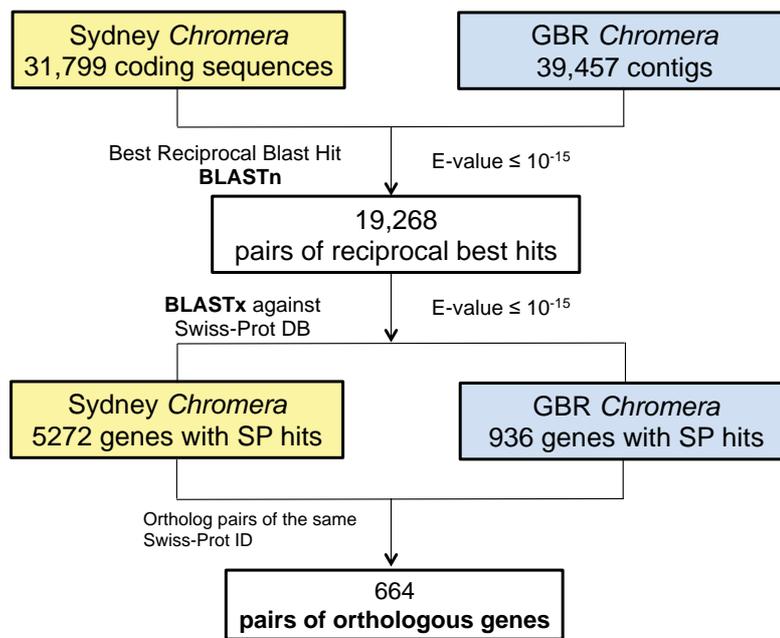


Figure 4.7 Strategy for identification of orthologous gene pairs between Sydney and GBR *Chromera* strains. The best reciprocal blast hit method was used to identify putative orthologs. BLASTX with a threshold of 10^{-15} against the Swiss-Prot database was used to filter paralogs.

Table 4.6 GO categories enriched amongst the *Chromera* ortholog pairs (Benjamini-corrected P -value ≤ 0.05)

Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_CC	GO:0031967~organelle envelope	15	3.749047705	0.003737509
	GO:0044434~chloroplast part	17	3.318077511	0.00187588
	GO:0031975~envelope	15	3.71909526	0.001362614
	GO:0044435~plastid part	17	3.216916612	0.001366639
	GO:0009941~chloroplast envelope	11	4.655067568	0.002510198
	GO:0009526~plastid envelope	11	4.443014598	0.003065569
	GO:0009507~chloroplast	31	1.808359581	0.008368664
	GO:0009536~plastid	31	1.771182505	0.010593495
GOTERM_MF	GO:0005524~ATP binding	31	2.228952991	0.003516512
	GO:0032559~adenyl ribonucleotide binding	31	2.202597128	0.002231321
	GO:0030554~adenyl nucleotide binding	31	2.063192247	0.005315981
	GO:0001883~purine nucleoside binding	31	2.063192247	0.005315981
	GO:0001882~nucleoside binding	31	2.056683754	0.004236857
	GO:0032555~purine ribonucleotide binding	31	1.989984739	0.006298849
	GO:0032553~ribonucleotide binding	31	1.989984739	0.006298849
	GO:0017076~purine nucleotide binding	31	1.872800718	0.015749736
	GO:0000166~nucleotide binding	34	1.716321632	0.028795719
	GO:0042623~ATPase activity, coupled	9	4.351293103	0.030886664
	GO:0016887~ATPase activity	10	3.602783726	0.045284842

4.4.4.3 Analysis of K_a/K_s , a test for selection

In order to identify genes likely to be under positive selection, K_a and K_s values were calculated for the 664 pairs of putative orthologs from the two *Chromera* strains. 321 ortholog pairs had significant K_a/K_s ratios at P -value ≤ 0.05 . The distribution of K_a and K_s values amongst these 321 ortholog pairs is shown in Figure 4.8. All but one of the ortholog pairs showed a K_a/K_s ratio < 1 , with the majority having a K_a/K_s ratio of close to 0 (Figure 4.9). The single ortholog pair that had a $K_a/K_s > 1$ (1.18) (Figure 4.9 and Table 4.7) encodes a *Chromera* homolog of the human tetratricopeptide repeat protein 21B (Table 4.8).

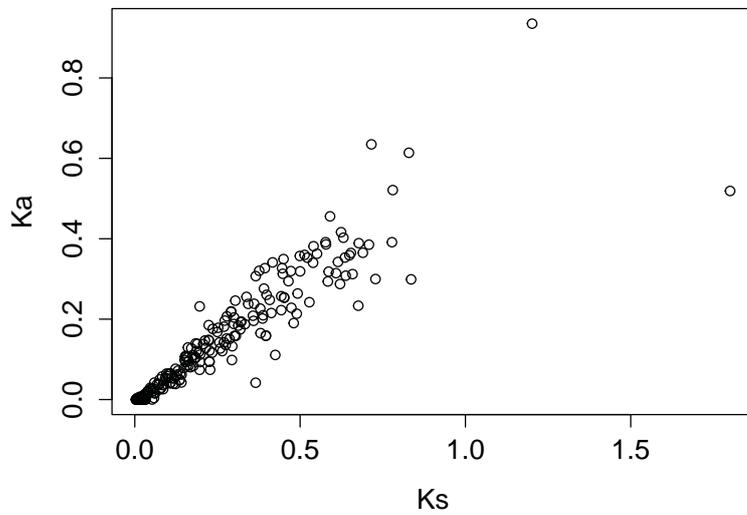


Figure 4.8 Distribution of nonsynonymous (K_a) and synonymous (K_s) substitution rates in 321 Sydney and GBR *Chromera* orthologous pairs. The threshold of $K_a/K_s < 1$ indicates negative selection at $P\text{-value} \leq 0.05$. The analysis was performed using the KaKs calculator software (Zhang *et al.* 2006).

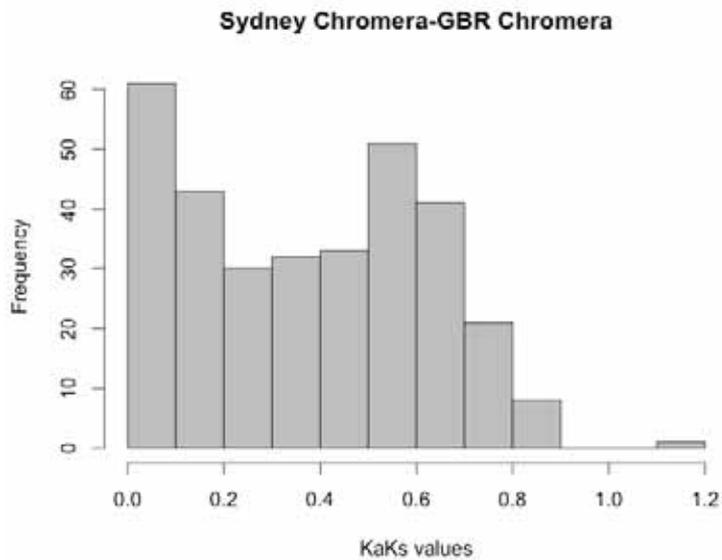


Figure 4.9 K_a/K_s distribution in ortholog pairs between the Sydney and GBR *Chromera* strains. The frequency distribution of K_a/K_s rates is shown on the x-axis while the K_a/K_s values are shown on the y-axis.

Table 4.7 Candidate gene under positive selection in *Chromera*

Orthologs pair	Method	Ka	Ks	Ka/Ks	P-Value (Fisher)
Cvel_12441-c19864_g1_i2	MA	0.23	0.19	1.18	0.0441

Table 4.8 Functional annotation of *Chromera* ortholog pair under positive selection

Ortholog	Length	E-value	Swiss-Prot BLAST hit description
c19864_g1_i2	4,742	2.40E-128	TT21B- Human Tetratricopeptide repeat protein 21B (Q7Z4L5)
Cvel_12441	3,630	0	

4.4.5 Comparative analyses of *Chromera*, *Symbiodinium* and *Plasmodium* transcriptome data

In order to better understand lifestyle evolution in this group of organisms, the transcriptomes of the *Chromera* strains were compared with that of *Plasmodium falciparum* and the gene predictions (CDSs) for *Symbiodinium kawagutii*. To select sequences for comparative analyses, the datasets were compared to the Swiss-Prot database (BLASTX; E-value $\leq 10^{-15}$), resulting in 4,750, 4,452, 3,075 and 2,383 non-redundant Swiss-Prot hits were obtained for the Sydney and GBR *Chromera* strains, *Symbiodinium* and *Plasmodium* respectively. Comparative data from this four-way analysis are presented as a Venn diagram (Figure 4.8).

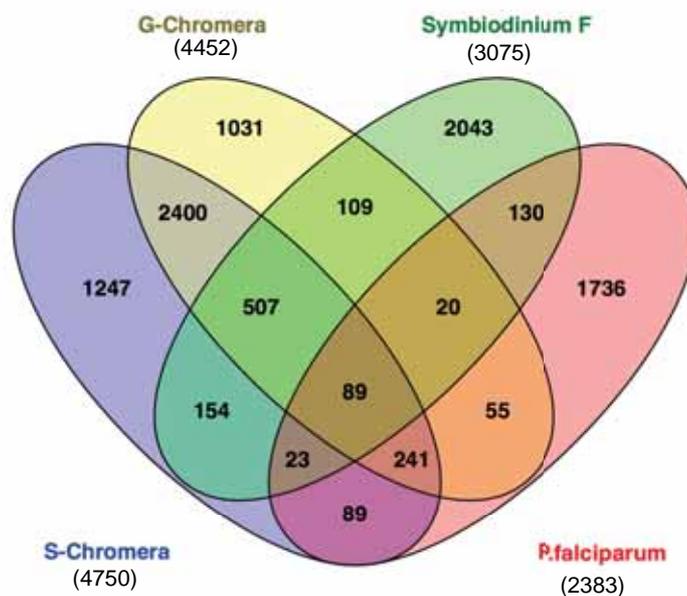


Figure 4.10 Venn diagram showing overlap between Sydney/ GBR *Chromera*, *Symbiodinium* and *Plasmodium* Swiss-Prot annotated genes at BLASTX E-value $\leq 10^{-15}$. Note the number of non-redundant hits for each species is shown in parentheses.

4.4.5.1 Functional profile for the set of genes shared between Sydney and GBR *Chromera*

GO enrichment analysis of the 2,400 genes shared between the two *Chromera* strains revealed significant enrichment of specific GO categories (using the Benjamini correction P-value ≤ 0.05 as cutoff), including 8 biological process categories (GO-BP) relating to DNA repair and cellular response to stress and DNA damage stimulus and DNA repair. In the case of the cellular component category (GO-CC), there were 15 categories related to the plastid and organelle envelope were enriched. For the molecular function category (GO-MF), there was enrichment in genes related to 30 categories related to ATP binding, ATPase activity and transmembrane transporters (Supplementary Table S4.1).

Functional annotation clustering analysis revealed 25 significant clusters (Enrichment score ≥ 1.5 ; Benjamini-corrected P value ≤ 0.05) of which plastid and transport clusters were highly enriched (enrichment scores 7.7 and 7.2 respectively; Figures 4.11, 4.12 and Supplementary Table S4.2). Other significant clusters were involved in the cellular response to stress and sugar transport activity (Figures 4.13, 4.14 and Supplementary Table S4.2).

Housekeeping genes (HKGs) are genes that are expressed by cell in order to maintain basic cellular functions and are expressed at constant levels under a range of conditions. The list of genes shared between Sydney and GBR *Chromera* included some HKGs (actin, calmodulin, tubulin, glyceraldehyde-3-phosphate dehydrogenase), several ribosomal genes, cytochrome genes (cytochrome c and P450), and photosynthetic genes (Ribulose-1, 5-bisphosphate carboxylase/oxygenase “RuBisCO”, fucoxanthin-chlorophyll a-c binding protein, chlorophyll synthase, Photosystem II stability/assembly factor HCF136) (Supplementary Table S4.3). Genes involved in vesicular trafficking identified included several RAB proteins, one RAB GTPase activating protein 1, two members of TBC1 domain family, several vacuolar protein sorting proteins, Phosphatidylinositol-4-phosphate 5-kinase, and several lysosomal cathepsins. Genes implicated in symbiotic interactions were identified including genes involved in stress response. Common antioxidant genes important in stress responses were identified including catalase-peroxidase 2, superoxide dismutase (SOD), several peroxiredoxins, Thioredoxin, genes involved in glutathione metabolism. A gene encoding the ultraviolet-b receptor, a key player in UV-stress and protective response was identified. A gene encoding dehydroquinate synthase (DHQS) enzyme that may be involved in the biosynthetic pathway for the photoprotective mycosporine-like amino acid (MAA) shinorine was identified. Some Calcium dependent protein kinases that are important in intercellular signaling were also identified. Some genes involved in nutrient/metabolite transport were identified including several ABC transporters, a glutamine synthetase, four glucose transporters, a nitrate transporter and a potential ammonium transporter (Supplementary Table S4.3).



Figure 4.11 Genes-to-terms heat map of the significant FAC “plastid” showing the plastid-related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.

	Probable mitochondrial-processing peptidase subunit alpha-1
	Protein transport protein Sec24-like CDP
	AT3C25438
	Customer subunit alpha-1
	AT3D04000
	AT3E11945
	AT3E20038
	AT3E28278
	Calcium-dependent protein kinase 15
	Calcium-dependent protein kinase 14
	Calcium-dependent protein kinase 39
	Aspartic proteinase-like protein 1
	AT3G511738
	AT4G21138
	AT3G21478
	AT4G20468
	AT3G04338
	AT3G09748
	AT3G24148
	Mitochondrial import receptor subunit TOM48 homolog 1
	AT3G27528
	AT4G24078
	Probable polyol transporter 4
	Cellulose synthase 7
	PGR5-like protein 18, chloroplastic
	ATPass 6, plasma membrane-type
	ATPass 7, plasma membrane-type
	Probable plastidic glucose transporter 1
	Putative copper-transporting ATPase 3
	Cyclooxylase cyclinase
	Putative respiratory burst oxidase homolog protein 8
	Wall-associated receptor kinase-like 11
	Probable inorganic phosphate transporter 1-6
	AT3G51588
	Cytochrome P450 86A1
	Ergosterol biosynthetic protein 28
	AT3G51138
	Calcium-transporting ATPase 1, endoplasmic reticulum-type
	AT3G01138
	7-dehydrocholesterol reductase
	Inner membrane protein ALK103, chloroplastic
	ADP-ATP carrier protein 2, chloroplastic
	Peptide transporter PTR1
	Sodium/hydrogen exchanger 4
	CAMK prenyl protease 1 homolog
	5-acyltransferase 1171
	Glycoprotein 3-alpha-L-fucosyltransferase A
	Metal transporter Wramp2
	Probable steroid reductase DET2
	Putative phospholipid-transporting ATPase 4
	AT3G29648
	Probable 5-acyltransferase At1g4188
	Putative phospholipid-transporting ATPase 4
	AT3G06138
	Two pore calcium channel protein 1
	Probable 5-acyltransferase At1g04078
	Sodium/hydrogen exchanger 7
	Vacuolar-sorting receptor 7
	AT3G48898
	1-acyl-sn-glycerol-3-phosphate acyltransferase 1, chloroplastic
	Plastidic glucose transporter 4
	Putative polyol transporter 1
	Respiratory burst oxidase homolog protein 8
	Diacylglycerol kinase 1
	Probable 5-acyltransferase At1g04038
	Uncharacterized protein At1g03988
	AT3G51128
	Alternative oxidase 4, chloroplastic/chromoplastic
	Glycine-rich RNA-binding protein 7
	Probable anion transporter 1, chloroplastic
	Sodium/hydrogen exchanger 6
	Probable inositol transporter 1
	AT3G41708
	ALA-interacting subunit 1
	Sodium-dependent phosphate transport protein 1, chloroplastic
Transmembrane region	
Transmembrane	
GO:0016021-Integral to membrane	
GO:0016022-Integral to membrane	
GO:0016023-Integral to membrane	
GO:0016024-Integral to membrane	

Figure 4.12 Genes-to-terms heat map of the significant FAC “transport” showing the membrane genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively (previous page).

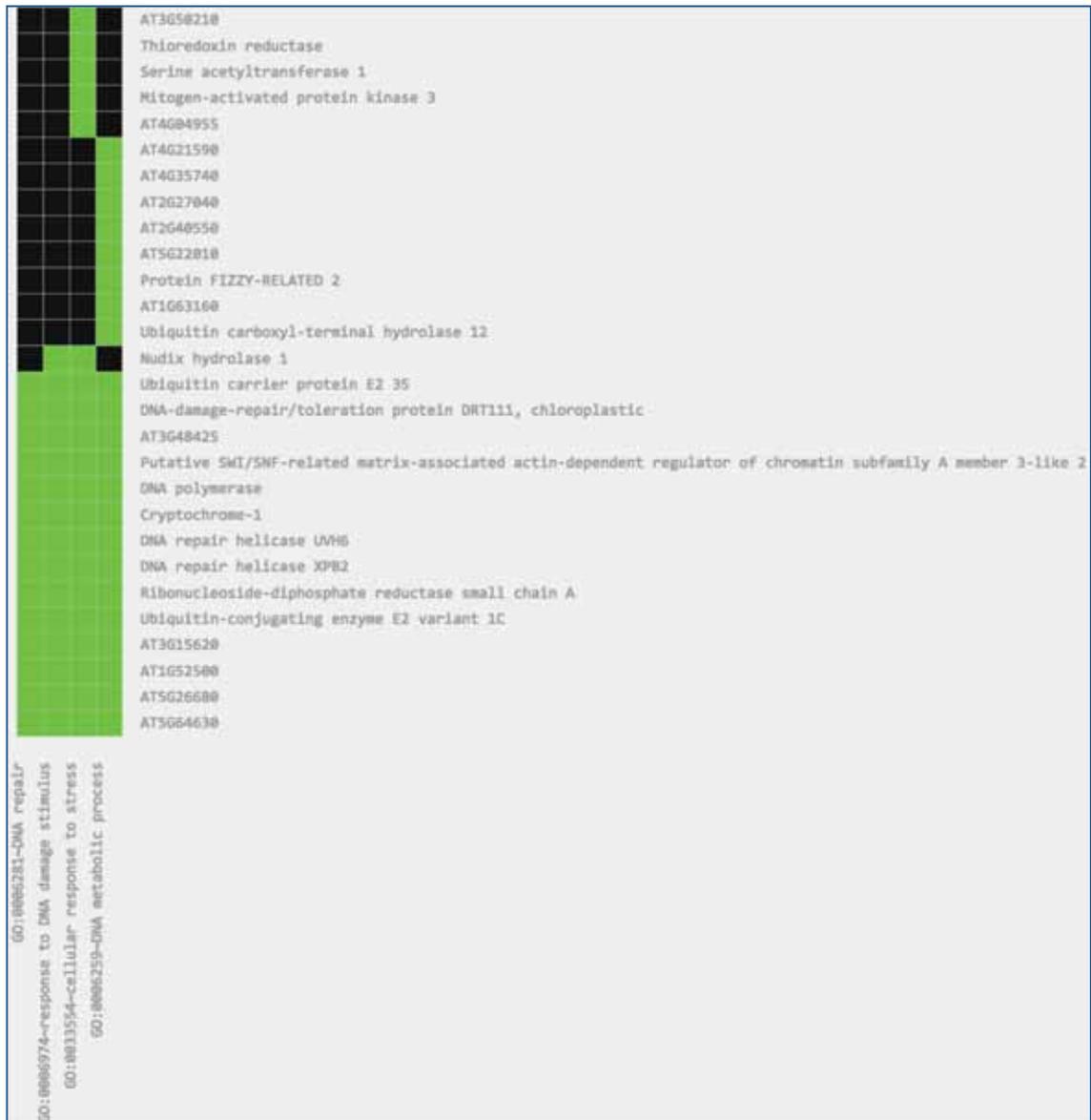


Figure 4.13 Genes-to-terms heat map of the significant FAC “cellular response to stress” showing the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.

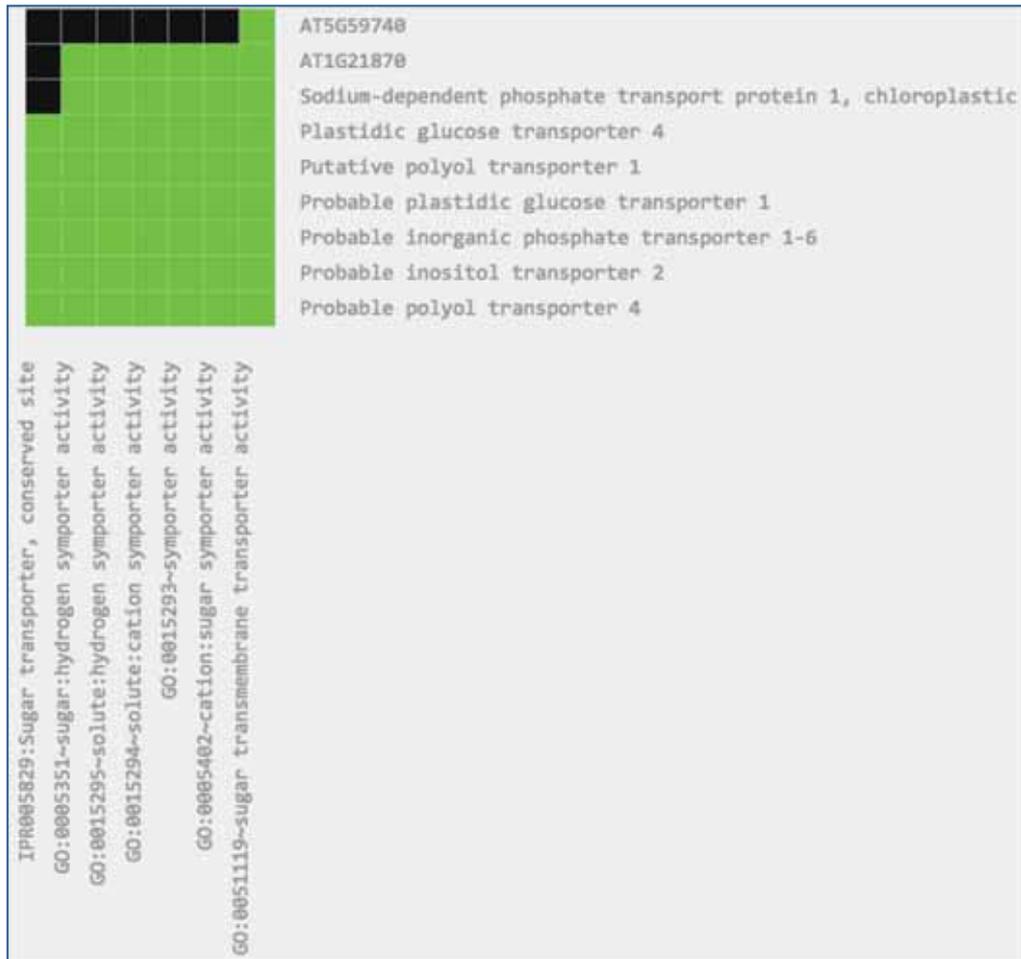


Figure 4.14 Genes-to-terms heat map of the significant FAC “sugar transport” showing the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.

4.4.5.2 Functional profile for the set of genes shared between Chromera and Symbiodinium

Using the Benjamini correction P -value ≤ 0.05 as cutoff, GO enrichment analysis of the 507-shared genes between *Chromera* and *Symbiodinium* revealed GO categories and KEGG pathways that were significantly enriched (Supplementary Tables S4.4, S4.5). Enriched GO categories included 15 biological process categories (GO-BP) related to phosphatidylinositol metabolic process, glycerophospholipid metabolic process, chlorophyll biosynthetic process and pigment biosynthetic process. In the case of the cellular component category (GO-CC), there were 21 enriched categories related to the plastid and organelle envelope. For the molecular function category (GO-MF), there were enrichment in genes related to 14 categories related to phosphatidylinositol phosphate/ lipid kinases activities and ATP binding. Amongst all enriched categories, the GO-MFs 1-phosphatidylinositol-4-phosphate 5-kinase activity (GO:0016308) and phosphatidylinositol phosphate kinase activity (GO:0016307) were the most highly over-represented (52.5- and 49.2- fold, respectively). In addition, the KEGG pathways inositol phosphate metabolism and phosphatidylinositol signaling were significantly enriched (Table S4.5).

Functional annotation clustering analysis revealed 7 significant clusters (Enrichment score ≥ 1.5 ; Benjamini-corrected P value ≤ 0.05) of which the plastid cluster was highly enriched with 12.7 enrichment score (Figure 4.15). Other significant clusters were involved in organelle envelope (Figure 4.16), 1-phosphatidylinositol-4-phosphate 5-kinase activity (Figure 4.17), chlorophyll metabolic process and thylakoid (Supplementary Table S4.6).



Figure 4.15 Genes-to-terms heat map of the significant FAC “plastid” showing the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.

4.4.5.3 Functional profile for the set of genes shared between *Chromera* and *Plasmodium*

GO categories and KEGG pathways that were significantly enriched amongst the 241 genes shared between *Chromera* and *Plasmodium* and summarized in Supplementary Table S4.7. Enriched GO categories included 9 biological process categories (GO-BP) related to proteolysis and protein catabolic processes. In the case of the cellular component category (GO-CC), there were 7 enriched categories related to the proteasome. For the molecular function category (GO-MF), there was enrichment in genes related to 10 categories related to ATP binding. Amongst all enriched categories, GO-CCs proteasome accessory complex (GO:0022624) and proteasome regulatory particle (GO:0005838) were the most highly over-represented with 70.6 fold enrichment. The KEGG pathway proteasome (ath03050) was significantly enriched (Figure 4.18). In addition, functional annotation clustering analysis revealed 4 significant clusters related to proteasome (Figure 4.19) and proteolysis (Figure 4.20) (Enrichment score ≥ 1.5 ; Benjamini-corrected P value ≤ 0.05) (Supplementary Table S4.8).

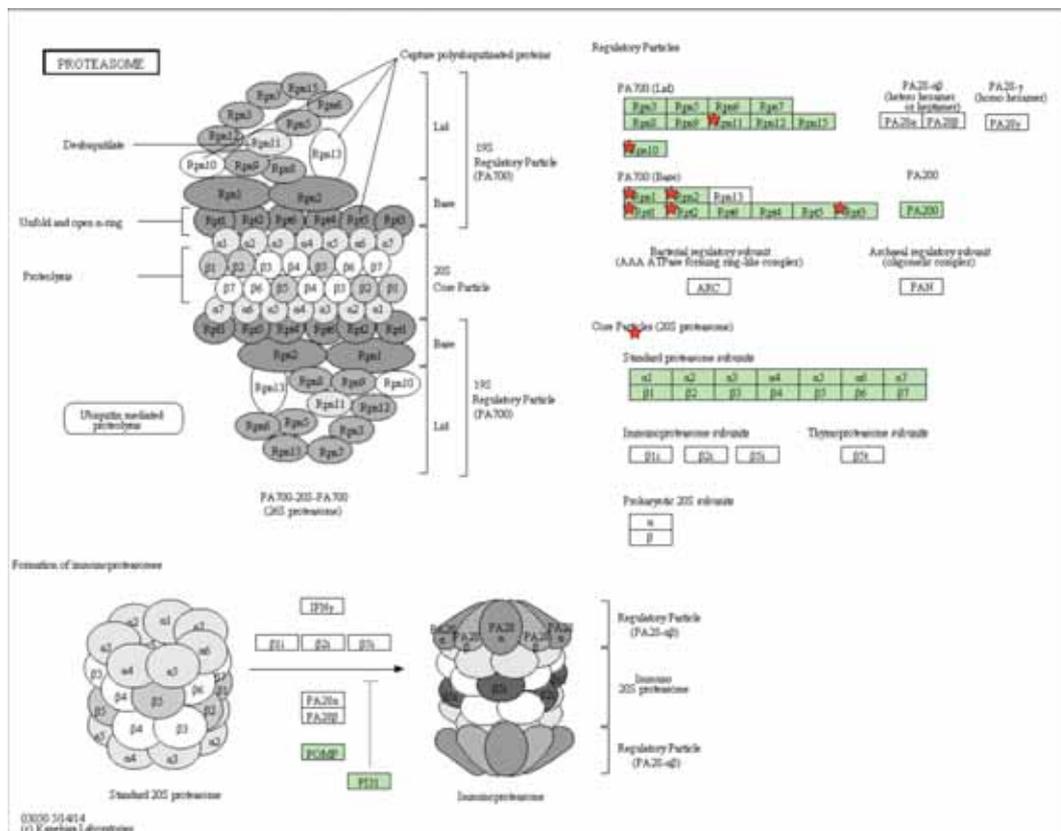


Figure 4.18 KEGG pathway Proteasome “ath03050” enriched in the list of 240 genes shared between *Chromera* and *Plasmodium*. The red stars represent similar components found in both datasets.

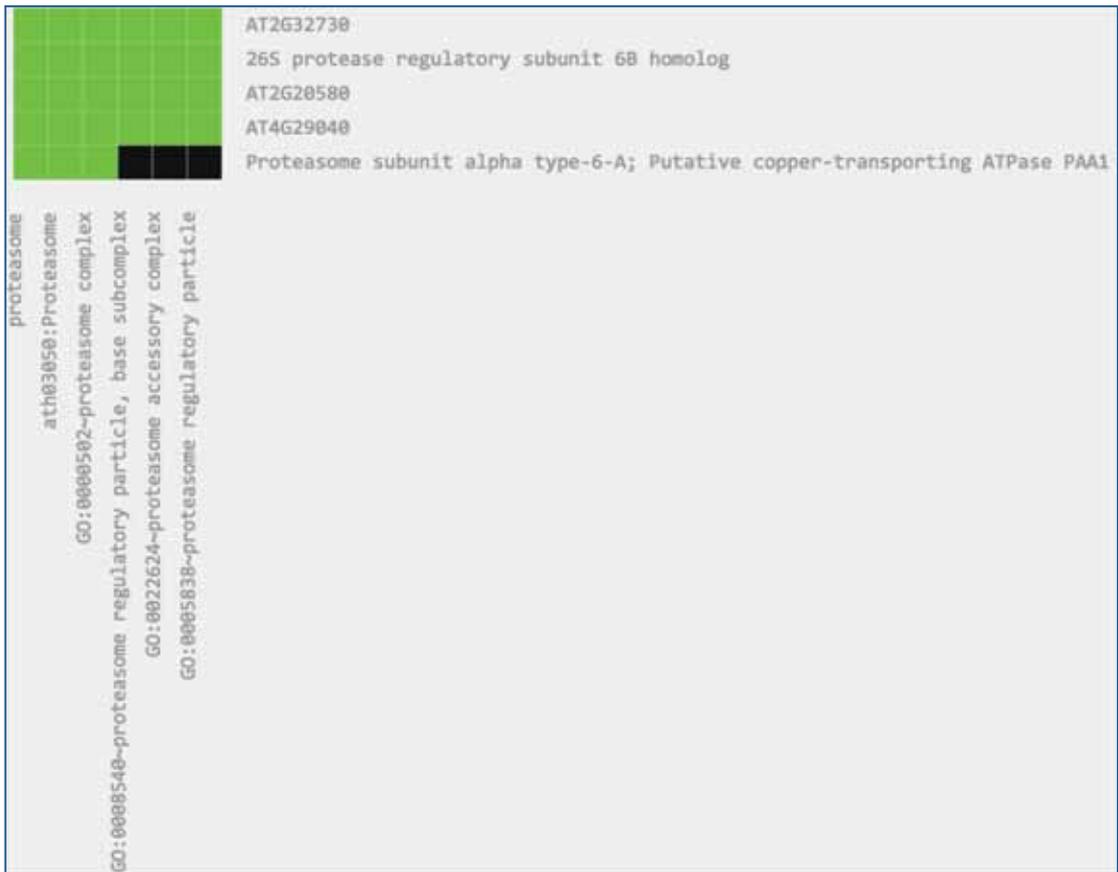


Figure 4.19 Genes-to-terms heat map of the significant FAC “proteasome” showing the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.

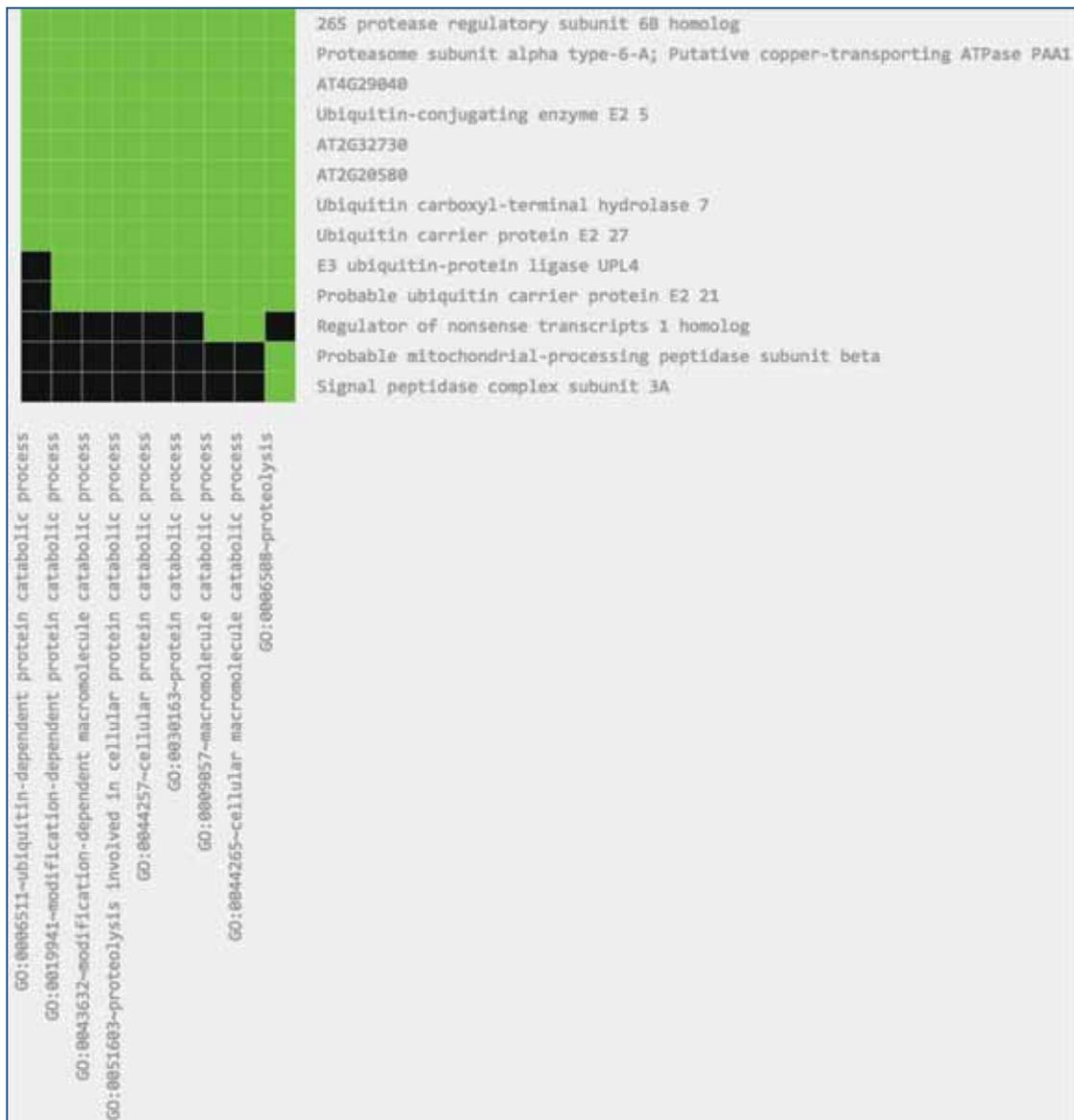


Figure 4.20 Genes-to-terms heat map of the significant FAC “proteolysis the related genes and their associated annotation terms. The green and black colors represent the positive and negative association between the gene-term respectively.

4.4.6 KEGG pathways comparative analysis

The KEGG automatic annotation server (KAAS) was used to map the *Chromera*, *Symbiodinium* and *Plasmodium* sequences as described previously in the current Chapter (section 4.3.7 Transcriptome functional annotations), KEGG-annotated genes classified and the organisms compared at the level of pathway category representation. In the GBR and Sydney *Chromera* strains and *Symbiodinium* the overall distribution with respect to the six main KEGG categories was similar, approximately one third of annotated genes in each case being assigned to metabolism (34%, 33% and 33% in GBR/ Sydney *Chromera* and *Symbiodinium* respectively). On the other hand, in *Plasmodium* the largest proportion of the annotated genes were assigned to human diseases (30%) (Figure 4.21). The pathways involved in the main KEGG categories are summarized in Figure 4.22, translation and infectious diseases being the most well-represented pathways in *Plasmodium*, reflecting the importance of these categories in the life style of an obligate parasite. Signal transduction was also well represented in Sydney *Chromera*, *Symbiodinium* and GBR *Chromera* (with 9%, 8% and 7.5% respectively) consistent with the need of a wide range of receptors in order to sense the surrounding environment. Carbohydrate metabolism was also among the well-represented pathways with 8%, 6.6% and 6.5% in *Symbiodinium*, Sydney and GBR *Chromera* respectively. Amino acid metabolism was well represented with 6.5%, 6.3% and 5.8% in *Symbiodinium*, GBR and Sydney *Chromera* respectively (Figure 4.22 and Supplementary Table S4.9). Despite the fact that metabolism was the most highly-ranked category in the *Chromera* strains and *Symbiodinium* with higher representation for carbohydrate and amino acid metabolism, the sub-categories energy metabolism and metabolism of other amino acids were higher in *Plasmodium* (Figure 4.22 and Supplementary Table S4.9) indicating the importance of these categories in the parasitic life style.

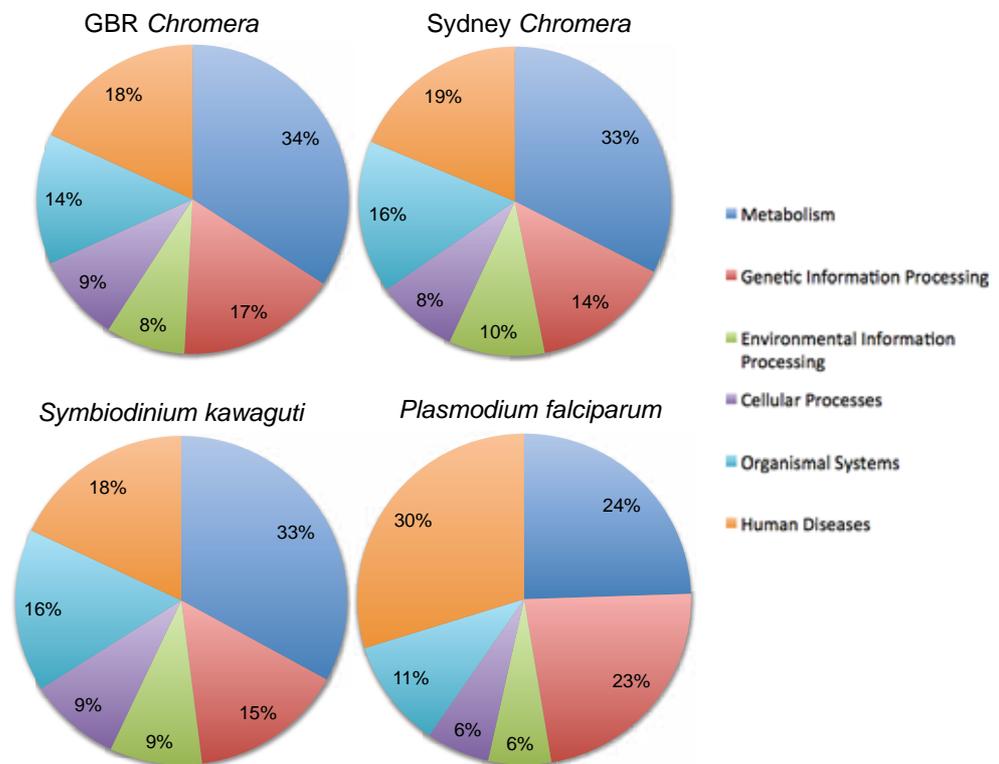


Figure 4.21 Overall distribution of the main KEGG categories in Sydney/ GBR *Chromera*, *Symbiodinium* and *Plasmodium*. The pie charts show the percentages of the sequences assigned with the six categories; metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems and human diseases.

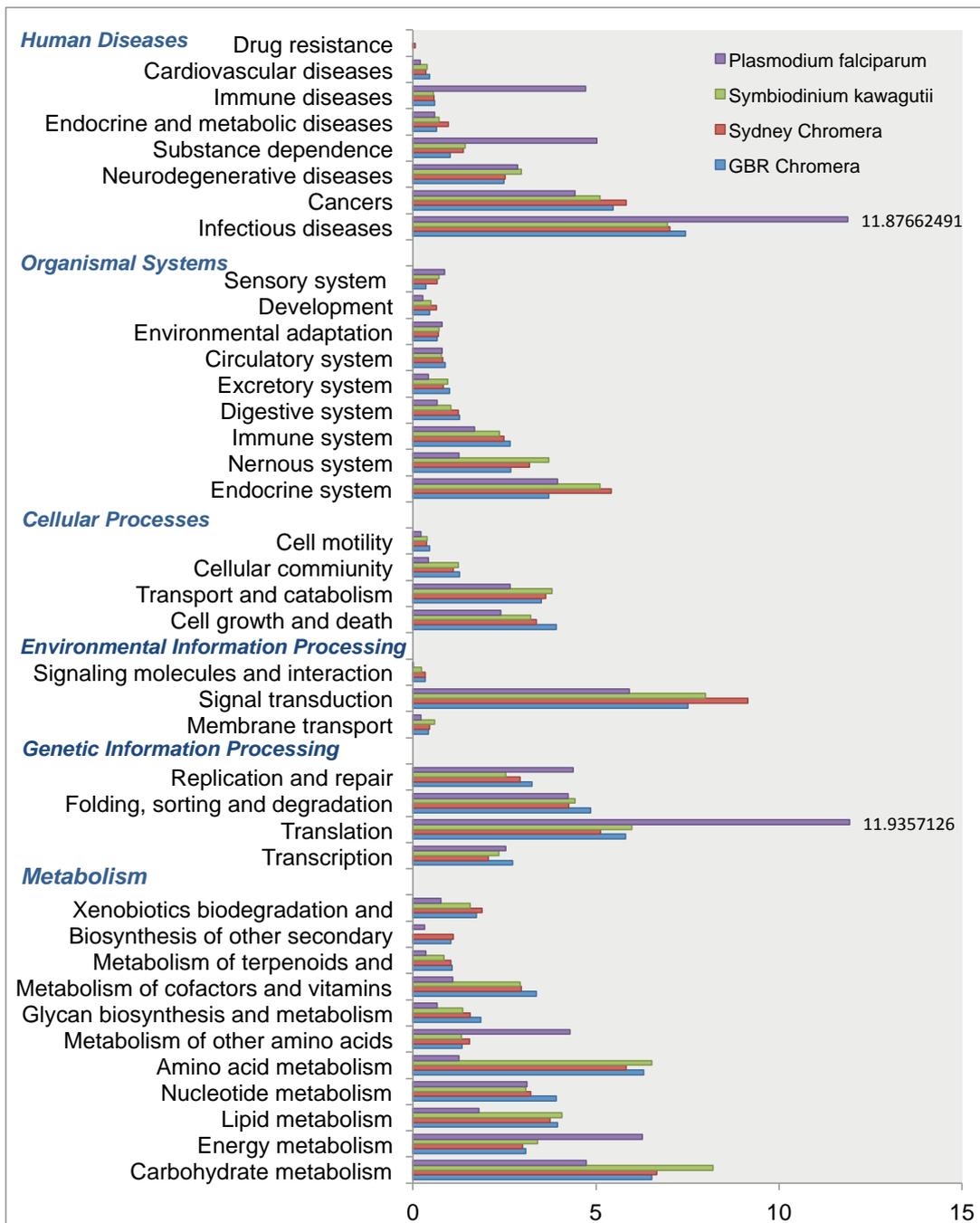


Figure 4.22 Distribution of the KEGG pathways in Sydney/ GBR *Chromera*, *Symbiodinium* and *Plasmodium*. The charts show the percentages of the sequences assigned with each category.

4.4.6.1 Glycan biosynthesis in *Chromera*, *Symbiodinium* and *Plasmodium*

Animal cells use pattern recognition receptors (PPRs) in order to identify the microorganism-associated molecular patterns (MAMPs) on microbes and that is true in the case of cnidarian-algal interactions, where host lectins bind to glycans on the symbiont cell wall (Wood-Charlson *et al.* 2006). The N-Glycan biosynthesis pathway was identified in Sydney-, GBR *Chromera*, *Symbiodinium*, and *Plasmodium* datasets and compared to the reference N-Glycan biosynthesis pathway ko00510 (Figure 4.23; Table 4.9). Both *Chromera* strains showed the highest representation of the pathway with 53.5% and 48.4% in Sydney and GBR *Chromera* respectively. On the other hand, *Plasmodium* and *Symbiodinium* showed the lowest representation of the N-Glycan biosynthesis pathway with only 16.3% and 23.3% respectively. Enzymes involved in mannose-rich glycan biosynthesis are quite similar in the GBR and Sydney *Chromera*. Ten enzymes were identified only in both *Chromera* and not in *Symbiodinium* or *plasmodium* including beta-1, 4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (MGAT3), Glycoprotein 6-alpha-L-fucosyltransferase (FUT8), oligosaccharyltransferase complex alpha (ribophorin I) and gamma subunits, and many mannosyl transferases. Alpha-1, 6-mannosyl-glycoprotein beta-1, 2-N-acetylglucosaminyltransferase (MGAT2) was identified only in Sydney *Chromera* and not in GBR *Chromera*, *Symbiodinium* or *Plasmodium* datasets. Four enzymes identified in both *Chromera* and *Symbiodinium*, but not in *Plasmodium* including Mannosyl-oligosaccharide alpha-1, 2 mannosidase (MAN1) and alpha 1, 3-glucosidase (GANAB). Three enzymes were present only in *Symbiodinium* including alpha-1,3-mannosyl-glycoprotein beta-1, 2-N-acetylglucosaminyltransferase (MGAT1).

Table 4.9 Genes mapped to the KEGG pathway N-Glycan biosynthesis “ko00510” in Sydney, GBR *Chromera*, *Symbiodinium* and *Plasmodium*

N-Glycan biosynthesis [PATH:ko00510]	Sydney <i>Chromera</i>	GBR <i>Chromera</i>	<i>Symbiodinium</i>	<i>Plasmodium</i>
K00902 E2.7.1.108; dolichol kinase [EC:2.7.1.108]	Cvel_1999.2; Cvel_1999.1	NA	NA	NA
K01001 ALG7; UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminophosphotransferase [EC:2.7.8.15]	Cvel_14903	c23122_g1_i5	Skav223256	CA854828
K00729 ALG5; dolichyl-phosphate beta-glucosyltransferase [EC:2.4.1.117]	Cvel_2662	c8929_g1_i3	NA	NA
K07432 ALG13; beta-1,4-N-acetylglucosaminyltransferase [EC:2.4.1.141]	Cvel_19651	c14397_g1_i3	Skav207516	BQ577226
K07441 ALG14; beta-1,4-N-acetylglucosaminyltransferase [EC:2.4.1.141]	Cvel_14157	c21264_g1_i10	NA	PFMC428TF; PFMC428TR
K00721 DPM1; dolichol-phosphate mannosyltransferase [EC:2.4.1.83]	Cvel_21980	c8514_g1_i1	NA	BQ739533
K03842 ALG1; beta-1,4-mannosyltransferase [EC:2.4.1.142]	Cvel_2426	c15711_g1_i5	NA	NA
K03843 ALG2; alpha-1,3/alpha-1,6-mannosyltransferase [EC:2.4.1.132 2.4.1.257]	Cvel_14961	c20956_g1_i2	NA	NA
K03844 ALG11; alpha-1,2-mannosyltransferase [EC:2.4.1.131]	Cvel_7259	c12638_g1_i2	NA	NA
K03845 ALG3; alpha-1,3-mannosyltransferase [EC:2.4.1.258]	Cvel_3783	c15542_g1_i1	NA	NA
K03846 ALG9; alpha-1,2-mannosyltransferase [EC:2.4.1.259 2.4.1.261]	Cvel_34074	c22006_g1_i4	NA	NA
K03848 ALG6; alpha-1,3-glucosyltransferase [EC:2.4.1.267]	Cvel_26907	c19441_g1_i2	Skav210156	NA
K03849 ALG8; alpha-1,3-glucosyltransferase [EC:2.4.1.265]	Cvel_3282	c3087_g1_i2	Skav229367	NA
K07151 STT3; dolichyl-diphosphooligosaccharide--protein glycosyltransferase [EC:2.4.99.18]	Cvel_16233	c20485_g2_i1	Skav228589; Skav220878; Skav234489	EL504165; PFRCB11TF
K12666 OST1, RPN1; oligosaccharyltransferase complex subunit alpha (ribophorin I)	Cvel_24038	c7427_g1_i2	NA	NA
K12668 OST2, DAD1; oligosaccharyltransferase complex subunit epsilon	Cvel_21041	c393_g1_i2	NA	BM275916; BM275927
K12669 OST3, OST6; oligosaccharyltransferase complex subunit gamma	Cvel_6758	c24028_g6_i8	NA	NA
K12670 WBP1; oligosaccharyltransferase complex subunit beta	Cvel_18634	c14484_g1_i1	NA	BI815817
K05546 GANAB; alpha 1,3-glucosidase [EC:3.2.1.84]	Cvel_13655	c15195_g1_i3	Skav200484	NA
K01230 MAN1; mannosyl-oligosaccharide alpha-1,2-mannosidase [EC:3.2.1.113]	Cvel_12472; Cvel_35120	c22314_g1_i3	Skav234315; Skav222590	NA
K00736 MGAT2; alpha-1,6-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase [EC:2.4.1.143]	Cvel_5421	NA	NA	NA
K00717 FUT8; glycoprotein 6-alpha-L-fucosyltransferase [EC:2.4.1.68]	Cvel_9340	c20791_g2_i1	NA	NA
K00737 MGAT3; beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase [EC:2.4.1.144]	Cvel_9710	c25923_g1_i1	NA	NA
K03847 ALG12; alpha-1,6-mannosyltransferase [EC:2.4.1.260]	NA	NA	Skav221206	NA
K03850 ALG10; alpha-1,2-glucosyltransferase [EC:2.4.1.256]	NA	NA	Skav232461	NA
K00726 MGAT1; alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase [EC:2.4.1.101]	NA	NA	Skav222983; Skav223442;	NA

4.4.6.2 Transcription machinery in *Chromera*, *Symbiodinium*, and *Plasmodium*

Dinoflagellates (including *Symbiodinium*) have some unique genomic characteristics including gene regulation, which are thought to be controlled primarily at the post-transcriptional or -translational level, rather than at the point of transcription. Transcription machinery pathways such as RNA polymerase, basal transcription factors and spliceosome (Tables 4.9, 4.10, 4.11) were identified in Sydney-, GBR *Chromera*, *Symbiodinium*, and *Plasmodium* datasets and compared to the reference KEGG pathways RNA polymerase “ko03020”, basal transcription factors “ko03022”, and spliceosome “ko03040”. *Symbiodinium* showed the lowest representation of RNA polymerase, basal transcription factors and spliceosome pathways with 17.6%, 11.4% and 47.1% recovery, respectively. Many subunits of (DNA-directed) RNA polymerases I, II and III and were not identified in *Symbiodinium* (Table 4.10). Regarding basal transcription factors, subunits of the transcription initiation factors TFIIB, TFIIF, TFIIH and TFIID were not identified in *Symbiodinium* including the TFIID TATA-box binding protein (TBP). Moreover, transcription initiation factor TFIID subunit 1 was uniquely identified in *Symbiodinium* (and not in *Chromera* and *Plasmodium*) (Table 4.11). For the spliceosome, many of the spliceosome components were not identified in *Symbiodinium* including small nuclear ribonucleoproteins U1, U2, and U4/U6 snRNPs, ATP-dependent RNA helicase subunits, pre-mRNA splicing factors, splicing factors 3A and 3B subunits, and poly U-binding splicing factor. Moreover, some splicing components involved in pre-mRNA processing such as nuclear cap-binding protein subunit 1 (NCBP1), pre-mRNA-splicing factor (SYF2), pre-mRNA-splicing factor ATP-dependent RNA helicase (DHX15), and heterogeneous nuclear ribonucleoprotein U (hnRNPU) were uniquely identified in *Symbiodinium* (and not in *Chromera* and *Plasmodium*) (Table 4.12).

Table 4.10 Genes mapped to the KEGG pathway RNA polymerase “ko03020” in Sydney, GBR *Chromera*, *Symbiodinium* and *Plasmodium*

RNA polymerase [PATH:ko03020]	Sydney <i>Chromera</i>	GBR <i>Chromera</i>	<i>Symbiodinium</i>	<i>Plasmodium</i>
K03040 rpoA; DNA-directed RNA polymerase subunit alpha [EC:2.7.7.6]	Cvel_8681	c21742_g2_i8	Skav220749	BM274786; CA855512; EL496238; EL504130
K03043 rpoB; DNA-directed RNA polymerase subunit beta [EC:2.7.7.6]	NA	c25323_g1_i1	Skav222794	EL505301
K03010 RPB2, POLR2B; DNA-directed RNA polymerase II subunit RPB2 [EC:2.7.7.6]	Cvel_11008	c23655_g2_i11	Skav204808; Skav204809	CA856446; EL501718; EL505255
K03006 RPB1, POLR2A; DNA-directed RNA polymerase II subunit RPB1 [EC:2.7.7.6]	Cvel_26433; Cvel_25301	c23635_g1_i5	NA	EL504214; DK891270; PFOCA49TF
K03011 RPB3, POLR2C; DNA-directed RNA polymerase II subunit RPB3	Cvel_9916	c24594_g1_i1	NA	CA856369
K03008 RPB11, POLR2J; DNA-directed RNA polymerase II subunit RPB11	Cvel_20935	c14874_g1_i3	Skav213209	EL500412; EL500040
K03012 RPB4, POLR2D; DNA-directed RNA polymerase II subunit RPB4	NA	NA	NA	BQ739732
K03017 RPB9, POLR2I; DNA-directed RNA polymerase II subunit RPB9	Cvel_5989	c12824_g1_i2	NA	BI815002; BM275778; BU498622
K03013 RPB5, POLR2E; DNA-directed RNA polymerases I, II, and III subunit RPABC1	Cvel_10624	c17940_g1_i4	NA	DK887949
K03014 RPB6, POLR2F; DNA-directed RNA polymerases I, II, and III subunit RPABC2	Cvel_15670	c20880_g1_i3	Skav208741	BQ596725; BU498415; BU495745; PFMC833TF; PFMC833TR
K03016 RPB8, POLR2H; DNA-directed RNA polymerases I, II, and III subunit RPABC3	Cvel_9667	NA	NA	BQ450982
K03007 RPB10, POLR2L; DNA-directed RNA polymerases I, II, and III subunit RPABC5	Cvel_9127	NA	NA	EL498767
K03021 RPC2, POLR3B; DNA-directed RNA polymerase III subunit RPC2 [EC:2.7.7.6]	Cvel_14119	c22496_g1_i1	NA	EL493366; DK892431; PFRCA52TR
K03018 RPC1, POLR3A; DNA-directed RNA polymerase III subunit RPC1 [EC:2.7.7.6]	Cvel_27010	c16890_g1_i1	Skav210477	BQ576893
K03027 RPC40, POLR1C; DNA-directed RNA polymerases I and III subunit RPAC1	Cvel_9564	c22248_g1_i4	NA	BQ633455
K03022 RPC8, POLR3H; DNA-directed RNA polymerase III subunit RPC8	Cvel_11091	c17664_g1_i2	NA	BI670788; EL504263
K03002 RPA2, POLR1B; DNA-directed RNA polymerase I subunit RPA2 [EC:2.7.7.6]	Cvel_15870	c22754_g2_i2	NA	EL505728; EL500233
K02999 RPA1, POLR1A; DNA-directed RNA polymerase I subunit RPA1 [EC:2.7.7.6]	Cvel_8831	c19998_g1_i2 c21019_g1_i3	NA	BU495729; EL500190
K03004 RPA43; DNA-directed RNA polymerase I subunit RPA43	NA	NA	NA	BQ633589
K03015 RPB7, POLR2G; DNA-directed RNA polymerase II subunit RPB7	Cvel_4163	c26987_g1_i1	NA	NA
K03020 RPC19, POLR1D; DNA-directed RNA polymerases I and III subunit RPAC2	Cvel_13599	NA	NA	NA
K14721 RPC5, POLR3E; DNA-directed RNA polymerase III subunit RPC5	Cvel_13555	c16686_g1_i5	NA	NA
K03019 RPC11, POLR3K; DNA-directed RNA polymerase III subunit RPC11	Cvel_2030	c11567_g1_i1	Skav204856	NA
K03025 RPC6, POLR3F; DNA-directed RNA polymerase III subunit RPC6	Cvel_9357	c7235_g1_i1	NA	NA
K03046 rpoC; DNA-directed RNA polymerase subunit beta' [EC:2.7.7.6]	NA	c157_g1_i1	Skav208138	NA
K03009 RPB12, POLR2K; DNA-directed RNA polymerases I, II, and III subunit RPABC4	NA	NA	Skav236032	NA

Table 4.11 Genes mapped to the KEGG pathway basal transcription factors “ko03022” in Sydney, GBR *Chromera*, *Symbiodinium* and *Plasmodium*

Basal transcription factors [PATH:ko03022]	Sydney <i>Chromera</i>	GBR <i>Chromera</i>	<i>Symbiodinium</i>	<i>Plasmodium</i>
K03124 TFIIB, GTF2B, SUA7, tfb; transcription initiation factor TFIIB	Cvel_28182	c1959_g1_i1	NA	NA
K03120 TBP, tbp; transcription initiation factor TFIID TATA-box-binding protein	Cvel_17608	c26721_g1_i1	NA	NA
K03132 TAF7; transcription initiation factor TFIID subunit 7	Cvel_12302	c16841_g3_i2	NA	NA
K03130 TAF5; transcription initiation factor TFIID subunit 5	Cvel_19215	NA	NA	NA
K03142 TFIH2, GTF2H2, SSL1; transcription initiation factor TFIH subunit 2	Cvel_11181	c22407_g1_i2	NA	DK892055
K03143 TFIH3, GTF2H3, TFB4; transcription initiation factor TFIH subunit 3	Cvel_25942	NA	NA	EL502808
K03144 TFIH4, GTF2H4, TFB2; transcription initiation factor TFIH subunit 4	Cvel_17986	c21127_g1_i4	Skav202753	DK895988
K10843 ERCC3, XPB; DNA excision repair protein ERCC-3 [EC:3.6.4.12]	Cvel_16950	c21650_g1_i1	Skav231294; Skav225079	PFRCE84TR
K10844 ERCC2, XPD; DNA excision repair protein ERCC-2 [EC:3.6.4.12]	Cvel_6598	c23577_g3_i6	Skav225559; Skav225566	EL507572; DK893076
K02202 CDK7; cyclin-dependent kinase 7 [EC:2.7.11.22 2.7.11.23]	Cvel_30093	c12515_g1_i1	NA	BM276045; K02202
K10842 MNAT1; CDK-activating kinase assembly factor MAT1	Cvel_27769	c3146_g1_i1	NA	NA
K10845 TTDA, GTF2H5, TFB5; TFIH basal transcription factor complex TTD-A subunit	NA	NA	NA	BM276544
K06634 CCNH; cyclin H	NA	NA	NA	BQ451636; BQ633187; BU498041; BU495466
K03125 TAF1; transcription initiation factor TFIID subunit 1	NA	NA	Skav212472	NA

Table 4.12 Genes mapped to the KEGG pathway spliceosome “ko03040” in Sydney, GBR *Chromera*, *Symbiodinium* and *Plasmodium*

Spliceosome [PATH:ko03040]	Sydney <i>Chromera</i>	GBR <i>Chromera</i>	<i>Symbiodinium</i>	<i>Plasmodium</i>
K12811 DDX46, PRP5; ATP-dependent RNA helicase DDX46/PRP5 [EC:3.6.4.13]	Cvel_22090	c13885_g1_i3	NA	CA856425
K12812 UAP56, BAT1, SUB2; ATP-dependent RNA helicase UAP56/SUB2 [EC:3.6.4.13]	Cvel_12637	c19379_g1_i2	Skav207203;Skav207201	DK888894; DK890179; XPF2n4443F
K12813 DHX16; pre-mRNA-splicing factor ATP-dependent RNA helicase DHX16 [EC:3.6.4.13]	Cvel_18648	NA	NA	NA
K12815 DHX38, PRP16; pre-mRNA-splicing factor ATP-dependent RNA helicase DHX38/PRP16 [EC:3.6.4.13]	Cvel_30788	c21339_g1_i6	NA	CA856497; PFOCD40TF
K12816 CDC40, PRP17; pre-mRNA-processing factor 17	Cvel_13898	c24235_g2_i2	NA	EL496693; PFOC973TF
K12817 PRPF18, PRP18; pre-mRNA-splicing factor 18	Cvel_18822	c15696_g1_i1	NA	BI816027; BM273361; CA855352
K12818 DHX8, PRP22; ATP-dependent RNA helicase DHX8/PRP22 [EC:3.6.4.13]	Cvel_14851	c22762_g2_i4	Skav20063 Skav233864	PFMC818TF
K12819 SLU7; pre-mRNA-processing factor SLU7	Cvel_25941	c17642_g1_i1	Skav211916	BI815549; BU495355; DK889372
K11086 SNRPB, SMB; small nuclear ribonucleoprotein B and B'	Cvel_24547	c22242_g1_i4	Skav212187	NA
K11087 SNRPD1, SMD1; small nuclear ribonucleoprotein D1	Cvel_32284	c7181_g1_i1	Skav226633	NA
K11096 SNRPD2, SMD2; small nuclear ribonucleoprotein D2	NA	c8550_g1_i1	Skav212166Ska v210534	NA
K11088 SNRPD3, SMD3; small nuclear ribonucleoprotein D3	Cvel_781	c13714_g2_i3	Skav209609Ska v236236	BU496691; EL502997
K11097 SNRPE, SME; small nuclear ribonucleoprotein E	Cvel_3966	c21053_g2_i1	Skav218188Ska v233219	BI814671; BU495835
K11098 SNRPF, SMF; small nuclear ribonucleoprotein F	Cvel_6682	NA	NA	BQ577114; BQ451259; BQ451128
K11099 SNRPG, SMG; small nuclear ribonucleoprotein G	NA	c17086_g1_i1	Skav203133	NA
K11093 SNRP70; U1 small nuclear ribonucleoprotein 70kDa	Cvel_18305	c19232_g1_i1	NA	EL503160;
K11091 SNRPA; U1 small nuclear ribonucleoprotein A	NA	c21457_g3_i6	NA	NA
K11095 SNRPC; U1 small nuclear ribonucleoprotein C	Cvel_4366	c16277_g2_i1	NA	BQ596003; BU496101
K12821 PRPF40, PRP40; pre-mRNA-processing factor 40	Cvel_34464	c24408_g2_i2	Skav224145 Skav217608	DK891334
K12822 RBM25, S164; RNA-binding protein 25	Cvel_3245	c16894_g1_i3	Skav216278Ska v216279	NA
K12823 DDX5, DBP2; ATP-dependent RNA helicase DDX5/DBP2 [EC:3.6.4.13]	Cvel_10178; Cvel_14217	c13652_g1_i2 c21212_g1_i1 c21793_g2_i3	Skav204722	DK888058; PFMC391TF
K11092 SNRPA1; U2 small nuclear ribonucleoprotein A'	Cvel_7793	c5367_g1_i1	Skav205488	NA
K11094 SNRPB2; U2 small nuclear ribonucleoprotein B''	Cvel_26064	c3364_g1_i1	NA	BU494621
K12825 SF3A1, SAP114; splicing factor 3A subunit 1	Cvel_14969	c22102_g1_i8	Skav201266	BQ596521
K12826 SF3A2, SAP62; splicing factor 3A subunit 2	Cvel_6397	c15796_g1_i1	NA	BQ633260
K12827 SF3A3, SAP61, PRP9; splicing factor 3A subunit 3	Cvel_2734	c19580_g1_i2	NA	BI814632
K12828 SF3B1, SAP155; splicing factor 3B subunit 1	Cvel_17506	c18097_g1_i4	NA	EL509072; BQ597044; EL509073; PFRCC49TF; PFRCC49TR
K12829 SF3B2, SAP145, CUS1; splicing factor 3B subunit 2	Cvel_20419	c13407_g2_i3	NA	EL505786; EL505809; EL501042; DK887671
K12830 SF3B3, SAP130, RSE1; splicing factor 3B subunit 3	Cvel_24695	c23646_g3_i1	Skav230556	EL508243; EL508243; PFRCF81TR
K12831 SF3B4, SAP49; splicing factor 3B subunit 4	Cvel_26166	c24110_g2_i2	NA	EL496991
K12832 SF3B5, SF3B10; splicing factor 3B subunit 5	Cvel_29606	NA	Skav204388	NA

K12833 SF3B14; pre-mRNA branch site protein p14	Cvel_9822	c12290_g1_i2	NA	BQ596962; BM275926; BU495633; PFRCD58TR
K12834 PHF5A; PHD finger-like domain-containing protein 5A	Cvel_6893	c5252_g1_i1	Skav230592	BM276535; CA854815; EL501612
K12835 DDX42, SF3B125; ATP-dependent RNA helicase DDX42 [EC:3.6.4.13]	Cvel_9638	c10176_g1_i2	NA	NA
K12836 U2AF1; splicing factor U2AF 35 kDa subunit	Cvel_593	c20656_g2_i1	Skav216448	PFRCD079TF; PFRCD079TR
K12837 U2AF2; splicing factor U2AF 65 kDa subunit	Cvel_12665	c17058_g1_i2 c23134_g2_i8 c23743_g1_i2	Skav228815	EL504195; DK888643
K12838 PUF60; poly(U)-binding-splicing factor PUF60	Cvel_20568	c20710_g5_i1	NA	NA
K12839 SMNDC1, SPF30; survival of motor neuron-related-splicing factor 30	Cvel_20027	c11162_g1_i2	NA	NA
K12840 RBM17, SPF45; splicing factor 45	Cvel_13104	c6273_g1_i1	Skav227235	CA856841
K12842 SR140; U2-associated protein SR140	Cvel_26610	c22865_g1_i7	NA	DK887330; K12842 SR140; U2-associated protein SR140
K12621 LSM2; U6 snRNA-associated Sm-like protein LSm2	Cvel_15476	c10466_g1_i1	Skav226701	NA
K12622 LSM3; U6 snRNA-associated Sm-like protein LSm3	Cvel_6713	c2840_g1_i1	Skav206167	NA
K12623 LSM4; U6 snRNA-associated Sm-like protein LSm4	Cvel_18065	c19564_g1_i1	Skav213286	BI814313
K12624 LSM5; U6 snRNA-associated Sm-like protein LSm5	Cvel_9066	c27244_g1_i1	Skav223437	BU498691; EL499495; EL501130; EL500209; PFOCD51TF; PFOCD51TR
K12625 LSM6; U6 snRNA-associated Sm-like protein LSm6	Cvel_3839	NA	Skav200559	NA
K12626 LSM7; U6 snRNA-associated Sm-like protein LSm7	Cvel_18116	c25926_g1_i1	Skav218233	PFMCB67TF
K12627 LSM8; U6 snRNA-associated Sm-like protein LSm8	Cvel_19726	c28945_g1_i1	Skav235922	BI816112; BQ577062
K12843 PRPF3, PRP3; U4/U6 small nuclear ribonucleoprotein PRP3	Cvel_17696	c23657_g16_i2; c23657_g6_i4	NA	BM273888; BM275386; CA855699
K12662 PRPF4, PRP4; U4/U6 small nuclear ribonucleoprotein PRP4	Cvel_6746	c22049_g1_i1	NA	DK896669
K09567 PPIH, CYPH; peptidyl-prolyl isomerase H (cyclophilin H) [EC:5.2.1.8]	Cvel_14214	c28508_g1_i1	Skav204969	BI815114
K12844 PRPF31; U4/U6 small nuclear ribonucleoprotein PRP31	Cvel_1363	c20410_g2_i3	NA	PFOC941TR
K12845 SNU13, NHP2L; U4/U6 small nuclear ribonucleoprotein SNU13	Cvel_21038	c12451_g1_i3	Skav232992	BU496374
K12847 USP39, SAD1; U4/U6.U5 tri-snRNP-associated protein 2	Cvel_5742	c20666_g1_i2	Skav202971	DK896104; XPF2n1259
K11984 SART1, HAF, SNU66; U4/U6.U5 tri-snRNP-associated protein 1	Cvel_11594	c22858_g2_i2	Skav230080	DK888977
K12848 SNU23; U4/U6.U5 tri-snRNP component SNU23	Cvel_801	c29809_g1_i1	Skav222309	NA
K12849 PRPF38A; pre-mRNA-splicing factor 38A	Cvel_13638	c13625_g1_i1	Skav218433	BI936066; PFOC053TF
K12850 PRPF38B; pre-mRNA-splicing factor 38B	Cvel_18055	c19671_g7_i1	Skav224713	PFMC443TF
K12852 EFTUD2; 116 kDa U5 small nuclear ribonucleoprotein component	Cvel_20700	c19582_g1_i3	NA	NA
K12854 SNRNP200, BRR2; pre-mRNA-splicing helicase BRR2 [EC:3.6.4.13]	Cvel_20444	c17392_g1_i3	Skav208897; Skav208900	AU087864; AU088551; EL503763; DK895431; DK893521; XPF2n6480; XPF2n5257; XPFn4704; XPF2n3872
K12855 PRPF6, PRP6; pre-mRNA-processing factor 6	Cvel_21488	c22318_g1_i5	Skav212249	EL501640; PFMC496TF; PFMC841TF; PFMC841TR
K12856 PRPF8, PRP8; pre-mRNA-processing factor 8	Cvel_16090	c21834_g1_i2	Skav217821; Skav217822	EL508438; EL508439; PFMC879TF; PFOC074TF
K12857 SNRNP40, PRP8BP; Prp8 binding protein	Cvel_11991	c19625_g1_i2	NA	BU497660

K12858 DDX23, PRP28; ATP-dependent RNA helicase DDX23/PRP28 [EC:3.6.4.13]	Cvel_1397	c13318_g1_i1	NA	BI815965; BM276154
K12859 TXNL4A, DIB1; U5 snRNP protein, DIM1 family	Cvel_4012	c9671_g1_i1	Skav232636	BI936144
K10599 PRPF19, PRP19; pre-mRNA-processing factor 19 [EC:2.3.2.27]	Cvel_4688	c29547_g1_i1	Skav223223	BI816324; BQ597299; BU496803; DK894720; XPF2n5491
K12860 CDC5L, CDC5, CEF1; pre-mRNA-splicing factor CDC5/CEF1	Cvel_16891	c30169_g1_i1	Skav228992	BQ596608; BM275823; BU496476
K12861 BCAS2; pre-mRNA-splicing factor SPF27	Cvel_5388	c22667_g3_i1	Skav235723	PFRC471TF
K12862 PLRG1, PRL1, PRP46; pleiotropic regulator 1	Cvel_12526	c18375_g1_i1	Skav200354	EL494504; DK889712
K12863 CWC15; protein CWC15	Cvel_9841	c20017_g1_i2	NA	BU496358; PFMC140TR
K12864 CTNBL1; beta-catenin-like protein 1	Cvel_11014	c22241_g1_i2	NA	BU495294; XPFn5552
K03283 HSPA1_8; heat shock 70kDa protein 1/8	Cvel_19711; Cvel_14986	c23978_g1_i9	Skav234833; Skav203944	EL501664; EL501041; DK891623; DK888839; DK891720; DK888611; DK888910; DK891753; DK890256; DK8889976; DK887723; DK889060; DK889522; DK889106; DK891112; DK888386; DK891400; DK891511; DK888123; DK888145; DK887953; DK888243
K12865 PQBP1, NPW38; polyglutamine-binding protein 1	Cvel_32502	c12025_g1_i1	NA	NA
K06063 SNW1, SKIIP, SKIP; SNW domain-containing protein 1	Cvel_729	NA	NA	BM276436; CA854721; BU495458
K12867 SYF1, XAB2; pre-mRNA-splicing factor SYF1	Cvel_763	c12732_g1_i1	NA	EL501051
K12869 CRN, CRNKL1, CLF1, SYF3; crooked neck	Cvel_473	c13336_g2_i1	Skav233808; Skav233809	CA855375; EL493568; DK894620
K12870 ISY1; pre-mRNA-splicing factor ISY1	Cvel_19216	c20510_g1_i2	Skav229738	NA
K12733 PPIL1; peptidyl-prolyl cis-trans isomerase-like 1 [EC:5.2.1.8]	Cvel_22133	c17246_g1_i2	NA	NA
K12871 CCDC12; coiled-coil domain-containing protein 12	Cvel_32010	NA	Skav211695; Skav224143	NA
K12872 RBM22, SLT11; pre-mRNA-splicing factor RBM22/SLT11	Cvel_19299	c7525_g1_i3	Skav224504	CA856286; EL506120; EL506121;
K12873 BUD31, G10; bud site selection protein 31	Cvel_25805	c226_g1_i1	Skav229818	BQ596784; PFMC110TF
K12874 AQR; intron-binding protein aquarius	Cvel_15273	c15679_g1_i2	Skav230329	EL504449
K13025 EIF4A3, FAL1; ATP-dependent RNA helicase [EC:3.6.4.13]	Cvel_15615	c16536_g1_i4	NA	DK895302
K12876 RBM8A, Y14; RNA-binding protein 8A	Cvel_24799	c15143_g1_i2	NA	CA854954; CA856219
K12877 MAGOH; protein mago nashi	Cvel_18540	c21667_g6_i2	Skav228232	BU495233
K12879 THOC2; THO complex subunit 2	Cvel_6001	c19906_g1_i2	Skav232238Ska v232241	NA
K12881 THOC4, ALY; THO complex subunit 4	NA	c8905_g1_i1	NA	BI815717; BQ577230
K12883 NCBP2, CBP20; nuclear cap-binding protein subunit 2	Cvel_685	c14009_g1_i3	NA	BI815527; BU497617
K12741 HNRNPA1_3; heterogeneous nuclear ribonucleoprotein A1/A3	Cvel_20842; Cvel_8129	c22794_g3_i3	NA	NA
K12885 RBMX, HNRNPG; heterogeneous nuclear ribonucleoprotein G	Cvel_26729	NA	NA	NA
K12887 HNRNPM; heterogeneous nuclear ribonucleoprotein M	NA	c8178_g1_i1	NA	NA
K12890 SFRS1_9; splicing factor, arginine/serine-rich 1/9	Cvel_3508; Cvel_9867; Cvel_20360	c20946_g2_i1 c22386_g1_i1	NA	N97998; BQ451355; DK887637; DK890492; DK890421
K12891 SFRS2; splicing factor, arginine/serine-rich 2	Cvel_13429	NA	Skav217099	BI815872
K12892 SFRS3; splicing factor, arginine/serine-rich 3	Cvel_15547	NA	NA	NA

K12900 FUSIP1; FUS-interacting serine-arginine-rich protein 1	Cvel_3780; Cvel_19171	c24414_g2_i2 c5078_g1_i2	NA	NA
K12896 SFRS7; splicing factor, arginine/serine-rich 7	NA	c12930_g1_i2	NA	NA
K12820 DHX15, PRP43; pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15/PRP43 [EC:3.6.4.13]	NA	NA	Skav221160	NA
K12824 TCERG1, CA150; transcription elongation regulator 1	NA	NA	Skav207286	NA
K12868 SYF2; pre-mRNA-splicing factor SYF2	NA	NA	Skav209465	NA
K12882 NCBP1, CBP80; nuclear cap-binding protein subunit 1	NA	NA	Skav210737Ska v223247	NA
K12888 HNRNPU; heterogeneous nuclear ribonucleoprotein U	NA	NA	Skav219545	NA

4.5 DISCUSSION

4.5.1 Transcriptome assembly and completeness

In this Chapter I describe the *de novo* assembly and characteristics of the transcriptome of a chromerid alga isolated from *M. digitata* on the GBR. Illumina reads generated on a Hi-Seq 2000 platform) were assembled in trinity *de novo* using a custom assembly pipeline. A total of 79,842 contigs (≥ 500 bp) were obtained after assembling 166 million 100bp PE reads derived from *Chromera* cultured under a variety of conditions. Of these contigs, 39,457 (49%) were identified as trinity genes (longest isoform per gene). This figure is in the same range as numbers of genes predicted for various *Symbiodinium* genes (30,000-46,000) based on transcriptome and genome sequencing (Bayer *et al.* 2012; Rosic *et al.* 2015; Shoguchi *et al.* 2013).

Only 19.4% of the 39,457-trinity genes had significant BLAST hits against the Swiss-Prot protein database, thus the vast majority of *Chromera* genes are un-annotated at this stage. This level of novelty is to be expected for organisms that are evolutionarily distant from well-characterized species, and has previously also been observed with *Symbiodinium* and other dinoflagellates (Bayer *et al.* 2012; Lin *et al.* 2010; Voolstra *et al.* 2009). Most of the KEGG-based annotations (34% of all assignments) were assigned to the metabolic pathway category, followed by the human disease category (18% of all assignments). Moreover, signal transduction and infectious diseases were the most highly represented pathways. The assembled transcriptome was judged to be relatively comprehensive on the basis of high percentages of reads mapping to it and very low

bacterial contamination. In addition, the majority of conserved pathways and protein complexes were present and well covered in the transcriptome assembly.

4.5.2 Identification of Chromera orthologs and genes under selection

The numbers of predicted genes in the GBR and Sydney *Chromera* strains (39,457 and 31,799, respectively) were similar based on coding sequences (CDS) regions (Woo *et al.* 2015). In order to enable Ka/Ks analysis, pairs of putative orthologs from Sydney and GBR *Chromera* were identified using the RBBH method. The large number (>19,000) of putative orthologs initially identified was dramatically reduced (to 664) by requiring both genes to have the same hit (protein ID) against the Swiss-Prot database. Amongst the 664 ortholog pairs the GO cellular components “plastid” and “organelle envelope” were enriched, implying the importance of photosynthesis and transport to *Chromera* as in *Symbiodinium*, the photosynthetic dinoflagellate symbiont in corals. In this context, in *Symbiodinium* 1053 pairs of orthologous genes were identified among four *Symbiodinium* clades. Those orthologs showed enrichment to many cellular components including chloroplast (Rosic *et al.* 2015). Moreover, Voolstra *et al.* (2009) used RBBH to find orthologous gene pairs between EST libraries of cultured and symbiotic *Symbiodinium* spp. and revealed the presence of 132 putative orthologs similarly including HKGs and photosynthesis-related genes.

Amongst the ortholog pairs identified, there was high degree of conservation between the sequences from the Sydney and GBR *Chromera* strains, as shown by analyses of synonymous to non-synonymous substitution data (Ka/Ks). For the vast majority of ortholog pairs the Ka/Ks values were significantly less than 1, implying that they are under varying levels of purifying (negative) selection. Positive selective (Darwinian) change was apparent for only a single ortholog pair (with a Ka/Ks value greater than 1). The fact that only one gene under positive selection was detected using this approach may be a consequence of the use of a relatively stringent E-value ($\leq 10^{-15}$) to identify pairs of orthologs, and it is likely that relaxing the stringency somewhat would allow detection of more genes under positive selection. That gene under positive selection identified here encodes a homolog of the tetratricopeptide protein 21B (TTC21B), which is a component of the intraflagellar transport complex A (IFT-A). For

many protists, flagella play important roles in processes such as locomotion, cell division and signal transduction.

4.5.3 Functional profile of shared genes among Chromera strains

Functional categories enriched in the set of genes shared between both Sydney and GBR *Chromera* included cellular response to stress, plastid, organelle envelope, transmembrane transporters and sugar transport. Pathways and processes related to nutrient transfer and oxidative stress have been suggested as key physiological processes required for a symbiotic life style in *Symbiodinium* within the coral holobiont (Weber & Medina 2012). Some of the HKGs such as actin, calmodulin, tubulin, GAPDH were among the shared *Chromera* genes. Finding such genes to be expressed in both *Chromera* strains is expected taking into account the key roles of their encoded proteins in essential cell processes. Such genes were also expressed in four different *Symbiodinium* clades (Huggett *et al.* 2005; Rosic *et al.* 2015). Photosynthesis-related genes shared among *Chromera* genes included RuBisCO, fucoxanthin-chlorophyll a-c binding protein, chlorophyll synthase, Photosystem II stability/assembly factor HCF136. RuBisCO and other photosynthesis-related genes were also common in four different *Symbiodinium* clades (Rosic *et al.* 2015). Other genes implicated in the symbiotic interaction were also identified in both *Chromera* including antioxidant genes and genes involved in UV protection, intercellular signaling, nutrient and metabolite transport. Several antioxidant genes were identified including catalase, superoxide dismutase, thioredoxin, and glutathione metabolism genes. Those genes are known to be involved in the oxidative stress response and reactive oxygen species (ROS) scavenging (Lesser 2006) and were identified in many symbiotic organisms including the coral algal symbiont *Symbiodinium* (Bayer *et al.* 2012; Rosic *et al.* 2015) and in legume roots where thioredoxin plays a key role in nodule formation as they show increased expression in order to lower ROS levels (Lee *et al.* 2005).

4.5.4 Functional profiles of genes shared between Chromera and Symbiodinium or Chromera and Plasmodium

Functional categories enriched in the set of genes shared between *Chromera* and *Symbiodinium* to the exclusion of *Plasmodium* included glycerophospholipid metabolic processes, plastid and organelle envelope. Genes from the enriched GO-MFs “Phosphatidylinositol phosphate kinase activity” and “1-phosphatidylinositol-4-phosphate 5-kinase activity” were the most highly represented genes common in *Chromera* and *Symbiodinium*. Moreover, KEGG pathways enrichment analysis showed that phosphate metabolism and phosphatidylinositol signaling are significantly enriched. This finding might be significant in context of *Chromera* being a potential symbiont in corals whereas the KEGG pathway “phosphatidylinositol signaling system” was enriched among four *Symbiodinium* clades (Rosic *et al.* 2015) including phosphatidylinositol-4-phosphate 5-kinases. Moreover, legume-rhizobacteria symbiotic interactions and signaling are mediated by phosphatidylinositide-regulated endocytosis (Peleg-Grossman *et al.* 2007). In the previous paper, inoculation of *Sinorhizobium meliloti* induced expression of PI3K and other vesicle trafficking genes in *Medicago truncatula*. In addition, inhibition of PI3K activity resulted in blocking endocytic internalization of the host plasma membrane. Whereas glycosylphosphatidylinositol is important for cell recognition and microbes detection (Davy *et al.* 2012), the previous result supports the possible involvement of the conserved “Phosphatidylinositol signaling system” pathway in symbiotic interactions.

On the other hand, functional categories enriched in the set of genes shared between *Chromera* and *Plasmodium* included the KEGG pathway proteasome and proteolysis-related GO terms. The proteasome is a protein complex of multi-component proteases and serves as a regulated protein degradation apparatus in eukaryotes, hence responsible for regulating key cellular processes such as cell cycle, signal transduction and stress response. Moreover, the proteasome has an important role in pathogenicity of some parasites, including apicomplexans, where the proteasome plays an important role during switching between different life cycle (morphological) stages. In the malaria causative agent *Plasmodium falciparum*, transcriptome sequencing of different stages of the parasite shows up-regulation of genes related to ubiquitin-proteasomal degradation during transforming from late trophozoite into schizont stage (Bozdech *et al.* 2003),

indicating the particular importance of the proteasome pathway in that transition. The previous finding is of high significance in the context of the potential of using *Chromera* as a model organism to develop anti-malaria treatments. Many reports indicate the suitability of the proteasome of *Plasmodium* as a promising drug target (Aminake *et al.* 2012; Gantt *et al.* 1998; Li *et al.* 2016).

4.5.5 KEGG pathway comparisons reveal similarity between Chromera and Symbiodinium, suggesting that Chromera may be a coral symbiont

The overall similarity of the GBR and Sydney *Chromera* strains and *Symbiodinium* in terms of main KEGG categories presumably reflects broadly similar lifestyle characteristics with metabolism being the dominant category. The transcriptomes of other *Symbiodinium* strains are also dominated by genes assigned to metabolism (Bayer *et al.* 2012; Voolstra *et al.* 2009). In both *Chromera* and *Symbiodinium*, the second largest KEGG category of genes was “human disease”, whereas in *Plasmodium* the category “human disease” had the highest number of genes assigned and this was followed by the “metabolism” category. This previous result is expected because *Plasmodium* is an obligate parasite (well-adapted to the parasitic life), whereas *Chromera* and *Symbiodinium* are photosynthetic and can be cultured autotrophically ex-host (except for some specific clades of *Symbiodinium*). In this context, Bayer *et al.* (2012) reported a similar distribution of KEGG categories in *Symbiodinium* and *Plasmodium* where metabolism followed by organismal systems and human disease categories had the highest number of genes. However, (surprisingly) the data for these comparisons were not shown in the previous study.

4.5.6 Variation in N-Glycan biosynthesis enzymes might explain the coral host recognition specificity

Recognition of algal symbionts by coral hosts is presumably mediated by interactions between pattern recognition receptors (PRRs) of the coral and microorganism-associated molecular patterns (MAMP) on the symbiont, and there is considerable evidence that lectins present on the host cell surface bind to high-mannose glycans present on the symbiont cell wall (Wood-Charlson *et al.* 2006). In this context, the KEGG pathway “N-Glycan biosynthesis” is likely to be of particular importance

during the initial interactions between corals and algal symbionts. Comparing the N-Glycan biosynthesis pathways of *Chromera*, *Symbiodinium* and *Plasmodium* revealed similar repertoires of glycan biosynthesis enzymes in the two *Chromera* strains but distinct complements in *Symbiodinium* and *Plasmodium*. The presence of some enzymes involved in the regulation of the biosynthesis of glycoprotein oligosaccharides only in the two *Chromera* strains (for example, MGAT3) and three enzymes unique to *Symbiodinium* (for example, MGAT1) is consistent with the idea that variability in glycoprotein structure on the algal surface might explain how the coral host recognizes and distinguishes different potential symbionts. On the basis of the data available, *Symbiodinium* and *Plasmodium* appear to lack a number of key glycan biosynthesis enzymes and the fact that some of the missing enzymes catalyse terminal steps of the common glycan biosynthesis pathway is particularly interesting. In this context, Lin *et al.* (2015) reported that mannose-rich glycan biosynthesis enzymes even differ between different clades of *Symbiodinium* (in that cases clades B and F) and suggested that this variation in the glycoprotein structure is required for the host recognition specificity.

4.5.7 Reduction of the transcription machinery in Symbiodinium; a unique feature of dinoflagellates

Symbiodinium is a dinoflagellate alga, and these latter have some unique characteristics, such as very large genome sizes, unusual chromatin structure, and permanently condensed chromosomes (Bayer *et al.* 2012; Shoguchi *et al.* 2013). Those unique features of dinoflagellate chromosomes are thought to affect gene transcription and its regulation (Erdner & Anderson 2006; Moustafa *et al.* 2010), thus transcriptional regulation may be of less significance in *Symbiodinium* than in most other eukaryotes, and post-transcriptional or -translational regulation may be more important. Bayer *et al.* (2012) identified a small number of putative transcription factors (proteins with sequence specific DNA binding domains) in *Symbiodinium*. In this context, the absence of many RNA polymerase subunits, basal transcription factors and spliceosome components as well as the presence of some spliceosome components in *Symbiodinium* that were not detected in *Chromera* or *Plasmodium* is significant and provides further evidence that the major coral algal symbiont *Symbiodinium* is a typical dinoflagellate

4.5.8 Conclusion

This Chapter describes the *de novo* assembly and annotation of a transcriptome for a strain of *Chromera* associated with *M. digitata* on the GBR, with the aim of enabling exploration of the genes and pathways in this unique photosynthetic apicomplexan alga. The transcriptomes of two different *Chromera* strains, the dinoflagellate *Symbiodinium kawagutii* and the apicomplexan parasite *Plasmodium falciparum* were compared in order to better understand the metabolic capabilities and likely lifestyle of *Chromera*. Focusing on the conserved features and pathways shared between these related organisms revealed the uniqueness of dinoflagellates in terms of transcriptional machinery but the otherwise overall similarity of the *Chromera* and *Symbiodinium* pathway data is consistent with the idea that *Chromera* is a coral symbiont. The similarity of *Chromera* and *Plasmodium* in terms of proteasome complement suggests that the former could be useful in developing anti-apicomplexan drugs targeting the parasite proteasome.

4.6 REFERENCES

- Aminake MN, Arndt HD, Pradel G (2012) The proteasome of malaria parasites: A multi-stage drug target for chemotherapeutic intervention? *Int J Parasitol Drugs Drug Resist* **2**, 1-10.
- Bayer T, Aranda M, Sunagawa S, *et al.* (2012) Symbiodinium Transcriptomes: Genome Insights into the Dinoflagellate Symbionts of Reef-Building Corals. *Plos One* **7**.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.
- Bozdech Z, Llinas M, Pulliam BL, *et al.* (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *Plos Biology* **1**, E5.
- Cumbo VR, Baird AH, Moore RB, *et al.* (2013) *Chromera velia* is Endosymbiotic in Larvae of the Reef Corals *Acropora digitifera* and *A. tenuis*. *Protist* **164**, 237-244.
- Davy SK, Allemand D, Weis VM (2012) Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiol Mol Biol Rev* **76**, 229-261.
- Erdner DL, Anderson DM (2006) Global transcriptional profiling of the toxic dinoflagellate *Alexandrium fundyense* using massively parallel signature sequencing. *Bmc Genomics* **7**.
- Flegontov P, Michalek J, Janouskovec J, *et al.* (2015) Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. *Molecular Biology and Evolution* **32**, 1115-1131.
- Gantt SM, Myung JM, Briones MR, *et al.* (1998) Proteasome inhibitors block development of *Plasmodium* spp. *Antimicrob Agents Chemother* **42**, 2731-2738.

- Haas BJ, Papanicolaou A, Yassour M, *et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512.
- Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57.
- Huggett J, Dheda K, Bustin S, Zumla A (2005) Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun* **6**, 279-284.
- Janouskovec J, Horak A, Obornik M, Lukes J, Keeling PJ (2010) A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci U S A* **107**, 10949-10954.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25.
- Lee MY, Shin KH, Kim YK, *et al.* (2005) Induction of thioredoxin is required for nodule development to reduce reactive oxygen species levels in soybean roots. *Plant Physiology* **139**, 1881-1889.
- Lesser MP (2006) Oxidative stress in marine environments: biochemistry and physiological ecology. *Annual Review of Physiology* **68**, 253-278.
- Li H, O'Donoghue AJ, van der Linden WA, *et al.* (2016) Structure- and function-based design of Plasmodium-selective proteasome inhibitors. *Nature* **530**, 233-236.
- Lin S, Cheng S, Song B, *et al.* (2015) The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* **350**, 691-694.
- Lin S, Zhang H, Zhuang Y, Tran B, Gill J (2010) Spliced leader-based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates. *Proc Natl Acad Sci U S A* **107**, 20033-20038.
- Moore RB, Obornik M, Janouskovec J, *et al.* (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**, 959-963.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35**, W182-185.
- Moustafa A, Evans AN, Kulis DM, *et al.* (2010) Transcriptome Profiling of a Toxic Dinoflagellate Reveals a Gene-Rich Protist and a Potential Impact on Gene Expression Due to Bacterial Presence. *Plos One* **5**.
- Obornik M, Modry D, Lukes M, *et al.* (2012) Morphology, ultrastructure and life cycle of *Vitrella brassicaformis* n. sp., n. gen., a novel chromerid from the Great Barrier Reef. *Protist* **163**, 306-323.
- Peleg-Grossman S, Volpin H, Levine A (2007) Root hair curling and Rhizobium infection in *Medicago truncatula* are mediated by phosphatidylinositide-regulated endocytosis and reactive oxygen species. *J Exp Bot* **58**, 1637-1649.
- Rosic N, Ling EY, Chan CK, *et al.* (2015) Unfolding the secrets of coral-algal symbiosis. *Isme Journal* **9**, 844-856.
- Shoguchi E, Shinzato C, Kawashima T, *et al.* (2013) Draft Assembly of the Symbiodinium minutum Nuclear Genome Reveals Dinoflagellate Gene Structure. *Current Biology* **23**, 1399-1408.
- Slapeta J, Linares MC (2013) Combined Amplicon Pyrosequencing Assays Reveal Presence of the Apicomplexan "type-N" (cf. *Gemmocystis cylindrus*) and *Chromera velia* on the Great Barrier Reef, Australia. *Plos One* **8**.
- Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**, W609-612.

- Voolstra CR, Sunagawa S, Schwarz JA, *et al.* (2009) Evolutionary analysis of orthologous cDNA sequences from cultured and symbiotic dinoflagellate symbionts of reef-building corals (Dinophyceae: Symbiodinium). *Comparative Biochemistry and Physiology D-Genomics & Proteomics* **4**, 67-74.
- Wang D, Zhang Y, Zhang Z, Zhu J, Yu J (2010) KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* **8**, 77-80.
- Weber MX, Medina M (2012) The Role of Microalgal Symbionts (Symbiodinium) in Holobiont Physiology. *Genomic Insights into the Biology of Algae* **64**, 119-140.
- Woo YH, Ansari H, Otto TD, *et al.* (2015) Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *Elife* **4**, e06974.
- Wood-Charlson EM, Hollingsworth LL, Krupp DA, Weis VM (2006) Lectin/glycan interactions play a role in recognition in a coral/dinoflagellate symbiosis. *Cell Microbiol* **8**, 1985-1993.
- Zhang Z, Li J, Fau - Zhao X-Q, Zhao Xq, Fau - Wang J, *et al.* (2006) KaKs_Calculator: calculating Ka and Ks through model selection and model averaging.

Chapter 5.0 General discussion, major thesis findings and future research

5.1 General discussion

Coral reefs are exposed to a variety of stressors and corals are very sensitive to increases in sea surface temperatures (SSTs) that lead to breakdown of the symbiotic relationship between corals and algae and cause coral bleaching and death globally. Corals live in multi-partite relationships with a range of microorganisms, amongst which algal symbionts of the genus *Symbiodinium* are particularly important. Consequently, understanding the interactions between corals and algae at the molecular level is important to better understand different stages of the symbiotic relationship (onset, establishment and breakdown). One of the objectives of this thesis was to use NGS technology to better understand the molecular processes underlying coral-*Symbiodinium* interactions during establishment of the symbiosis. The main objective of the thesis, however, was to establish the nature of the relationship between corals and *Chromera*, a newly described photosynthetic alga closely related to apicomplexans, by analyzing the response of the coral transcriptome to *Chromera* infection. And further, to provide a functional genomic resource (a *de novo* transcriptome assembly) for *Chromera* isolated from *M. digitata* from the GBR and using transcriptomic data to compare the gene complement of *Chromera* with those of its symbiotic and parasitic relatives. My PhD research aims to improve current knowledge of the molecular bases of interactions occurring during the onset and establishment of the coral-algal symbiosis based on analyses of changes in the coral transcriptome upon initial interactions between coral larvae and competent *Symbiodinium*. The state of thinking about these issues prior to my work is elegantly reviewed by Davy *et al.* (2012) and Meyer and Weis (2012), and these papers provided a starting point for interpreting the transcriptomic data obtained during the course of my investigations.

Before the commencement of this study, the consensus was that the establishment of coral-*Symbiodinium* symbiosis was contingent on the ability of potential symbionts to evade host recognition, and that the host transcriptome response to a compatible symbiont was minimal. Previous attempts to understand the coral

response during symbiont uptake were based on cDNA microarray technologies Voolstra *et al.* (2009) and Schnitzler and Weis (2010) reported that coral larvae showed either no or very few transcriptional changes following infection with a competent strain of *Symbiodinium*. It was this inability to detect host signals that led to the assumption that establishment of symbiosis was dependent on the ability of symbionts to evade host recognition. However, the microarrays available at that time represented only a subset of the transcriptome, and were relatively “noisy” – without high levels of technical (as well as biological) replication, detection of differentially expressed genes was challenging. The authors of those papers (Voolstra *et al.* (2009), Schnitzler and Weis (2010) recognized the limitations of the technology and also the potential for having missed transcriptional responses by not sampling over a longer time period. In Chapter 2, I used the Illumina RNA-Seq technology to follow changes across the whole transcriptome during infection of *Acropora digitifera* larvae with a competent strain of *Symbiodinium*, and I sampled at time points (4 and 12 h post-infection) that had not previously been investigated. Interestingly, the coral host transcriptome showed significant changes (1073 genes) only early in the infection process (4 h post-infection), whereas no changes were detected at later time points. This is the first report to reveal differential expression of significant numbers of genes during the infection of coral larvae with a competent strain of symbiont. The transcriptomic data from this experiment enabled a model for genes and pathways operating during establishment of coral-*Symbiodinium* symbiosis to be proposed, and this is presented in Chapter 2.

The novel microalga *Chromera* was isolated during surveys of the algal associates and symbionts of Australian corals (Moore *et al.* 2008). Since its discovery, great interest has been developed in *Chromera* because of its unique position in the phylogenetic tree, between the photosynthetic dinoflagellates and the parasitic apicomplexans (Moore *et al.* 2008). In terms of morphology, *Chromera* resembles the coral symbiont *Symbiodinium* but is the closest photosynthetic relative to the Apicomplexa (Moore *et al.* 2008). *Chromera* has been isolated and detected from phylogenetically diverse corals at widely separated locations in Australia (Cumbo *et al.* 2013; Moore *et al.* 2008; Slapeta & Linares 2013) as well as in the Caribbean (Visser *et al.* 2012). It has been shown that *Chromera* can infect and form stable associations with larvae of the corals *A. digitifera* and *A. tenuis* (Cumbo *et al.* 2013), however the nature of the interaction is not clear and its close relationship to apicomplexans suggests that

Chromera might actually be a facultative parasite. Immune challenge, stress and disease have distinct transcriptomic signatures in corals (Libro *et al.* 2013; Pinzon *et al.* 2015; Seneca & Palumbi 2015; Weiss *et al.* 2013) as does interaction with a competent *Symbiodinium* strain (Chapter 2). Consequently, investigation of the impact of *Chromera* infection on the coral transcriptome might be expected to provide insights into the nature of the relationship between coral and *Chromera*. In Chapter 3, I used Illumina RNA-Seq analyses to follow whole changes in gene expression occurring in *A. digitifera* larvae during infection with *Chromera*. While the coral transcriptome responded rapidly during infection with a competent *Symbiodinium* strain (Chapter 2), in the case of *Chromera* the coral response took place on a longer time scale and involved larger numbers of genes (Chapter 3).

The initial interactions between corals and microbes involve recognition systems on both sides of the interaction. The coral has a wide range of PRRs that recognize MAMPs on the surface of microbes (Davy *et al.* 2012). In Chapters 2 and 3 I showed that during infection with competent *Symbiodinium* some elements of the coral immune system are suppressed (only 5 genes), while the immune suppression observed on *Chromera* infection was much more extensive involving a wide range of recognition receptors including (lectins, scavenger receptors, GP2 and toll-like) and their downstream pathways (NF-Kappa-B and MAPK cascades, TLR and NLR signaling pathway, complement system components) and the production of anti-microbial peptides (AMPs). In this context, the up-regulation of a C-type lectin (MRC2) in *Acropora digitifera* larvae during infection with *Symbiodinium* (Chapter 2) and the down-regulation of the same gene (and other lectins) in response to *Chromera* infection (Chapter 3) is particularly interesting. Results reported in Chapters 2 and 3 include down-regulation of a gene encoding the human pancreatic secretory granule membrane major glycoprotein (GP2) during early interactions with both competent *Symbiodinium* and *Chromera*. Down-regulating GP2 is very interesting in the context of coral-microbes interactions as in man, GP2 acts as an integral membrane protein that binds pathogenic enterobacteria (Hase *et al.* 2009), suggesting that the coral homolog may play a role in interactions not only with algae, but perhaps also other microbes. These coral genes (MRC2 and GP2) therefore merit further investigation.

Both *Symbiodinium* and *Chromera* presumably gain entry into coral host cells

via phagocytosis, it is interesting to compare aspects of post-phagocytic processes likely to occur in the two cases. In Chapter 2, I showed that a competent *Symbiodinium* strain appears to arrest phagosome maturation at an early stage. Rab5 associates with early endosomes immediately after phagocytosis, but would normally be replaced by Rab7 during the phagosome maturation process (Fig 5.1). In the case of *Symbiodinium* infection, the transcriptomic data imply that the interaction of Rab5 with phagosomes is stabilized, interrupting phagosomal maturation at an early stage and thus enabling the symbiosome to be formed and maintained. Consistent with this idea, the sea anemone (*Aiptasia pulchella*) homolog of Rab5 was detected on vesicles containing *Symbiodinium* (Chen *et al.* 2004). On the other hand, in Chapter 3 I showed that, in the case of *Chromera* infection, expression of genes involved in phagosome maturation and phago-lysosome fusion were up-regulated, presumably in an attempt to destroy the engulfed *Chromera* cells (Figure 5.1)

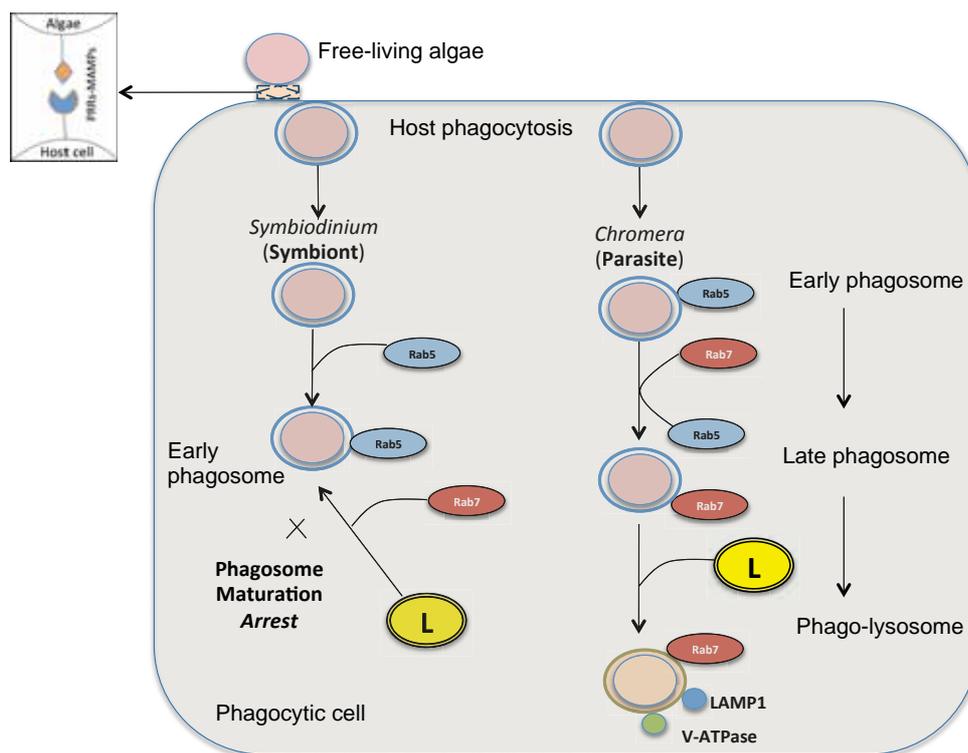


Figure 5.1 Diagram showing phagosome maturation and its arrest in coral larvae during interaction with a parasite (*Chromera*; right side) and symbiont (competent *Symbiodinium*; left side). Initially a stage of recognition occurs between the host and invading microbes and immediately after phagocytosis the host endocytic pathway is responsible for either killing invading microbes or protecting beneficial ones from host degradation enzymes. In case of *Chromera*, coral larvae showed increased expression of genes implicated in phagosome maturation, late phagosome formation (specially RAB7) and phago-lysosome fusion in order to

kill *Chromera*. On the other hand, in case of competent *Symbiodinium* phagosome maturation is arrested primarily via increased expression of RAB5-activating genes, consequently preventing the acquisition of RAB7. .

The life cycle of *Chromera* includes a free-living motile stage (Obornik *et al.* 2011) similar to that seen in some clades of the dinoflagellate *Symbiodinium*. The motile stage is significant in terms of the ability to infect corals, as the majority (85%) of corals acquire symbionts from the environment at each generation (horizontal mode of transmission) (Stat *et al.* 2006). Although generally referred to as “symbiosis”, the coral-*Symbiodinium* relationship more likely represents a range of interactions that span symbiotic to parasitic associations (Lesser *et al.* 2013). Some *Symbiodinium* strains are obligately associated with corals, as they have no free-living stage. In these cases the algal associates are maternally inherited (horizontal mode of transmission), and it has been speculated that this type of obligate association reflects a more tight integration of the coral and alga that represents symbiosis in the strict sense of the word (i.e. a mutualistic relationship) (Lesser *et al.*, 2013). In terms of similarity of KEGG pathway presence/absence, *Chromera* and *Symbiodinium* were very similar (Chapter 4). However, *Chromera* showed higher growth rates in culture and higher thermal tolerance than did *Symbiodinium* clade B7 (Visser *et al.* 2012). Although *Chromera* strains have a full repertoire of genes required for autotrophic growth, the transcriptomic response of coral larvae to *Chromera* infection suggests that *Chromera* might actually be parasitic in, rather than symbiotic with, corals (Chapter 3). Overall, I suggest that *Chromera* be regarded as an opportunistic parasite in corals, as is probably also the case for some strains of *Symbiodinium* (Ragni *et al.* 2010; Stat *et al.* 2008; Suggett *et al.* 2008). Consequently coral hosts harboring such parasitic algal “symbionts” might show lower overall fitness (Mieog *et al.* 2009; Stat *et al.* 2008).

Overall, the achievements of the research described in this PhD thesis are: 1) the identification of a window of gene expression in the coral host during infection with a competent *Symbiodinium* strain; 2) the use of next-gen sequencing (Illumina RNA-Seq) for the first time to explore changes in host transcriptome during infection with a competent *Symbiodinium* strain; 3) insights into the molecular mechanisms underpinning the onset and establishment of coral-*Symbiodinium* symbiosis; 4)

exploration for the first time of the poorly-understood relationship between corals and novel apicomplexan-related alga *Chromera*; 5) suggesting that *Chromera* might be parasitic in corals, based on the coral transcriptomic response to *Chromera* infection; 6) providing novel insights into genes and pathways involved in interactions between a coral host and microbes; 7) provided a functional genomic resource for *Chromera* isolated from coral at the GBR that can potentially improve our understanding of chromerid biology and 8) highlighted the potential of using *Chromera* as a model organism for the development of anti-malarial drugs targeting the *Plasmodium* proteasome.

5.2 Major thesis findings

- 1- The use of next generation sequencing allowed the detection of coral signals during initial interaction with competent *Symbiodinium* (Chapter 2).
- 2- The establishment of coral-*Symbiodinium* symbiosis between planula larvae and competent *Symbiodinium* involves rapid, transient and subtle changes in the host transcriptome. Oxidative metabolism, translation and protein transport are suppressed in the coral host during the onset of coral-*Symbiodinium* interactions (Chapter 2).
- 3- Uptake of competent *Symbiodinium* by the coral host involves similar mechanisms to those used to identify microbial pathogens (Chapter 2).
- 4- The symbiosome is likely to represent an arrested early phagosome in corals, as is also likely to be the case in symbiotic sea anemones, suggesting common mechanisms in corals and sea anemones (Chapter 2).
- 5- Both competent *Symbiodinium* and the coral host are involved in the establishment of the symbiosis; recognition of the symbiont is likely to be an active process on the part of the coral, whereas the apparent suppression of host immune responses may be initiated by the symbiont (Chapter 2).
- 6- Apoptosis, transcription control, the cell cycle and ROS defense might be key processes involved in onset and establishment of coral-*Symbiodinium* symbiosis.
- 7- The coral-*Chromera* association might be parasitic rather than symbiotic and *Chromera* infection involves major changes in the expression of suites of genes involved in phagosome maturation, apoptosis and immunity (Chapter 3).

- 8- Down-regulation of the coral homolog of the human pancreatic secretory granule membrane major glycoprotein (GP2) gene might be crucial for establishment of symbiosis (Chapters 2 and 3).
- 9- Comparative transcriptome analysis of *Chromera*, *Symbiodinium* and *Plasmodium* revealed high overall similarity between *Chromera* and *Symbiodinium* in terms of the presence / absence of KEGG pathways (Chapter 4).
- 10- *Chromera* might be a useful model organism for the development of anti-malarial drugs targeting the proteasome (Chapter 4).

5.3 Future research

In Chapters 2 and 3, gene expression landscapes were profiled in *Acropora digitifera* larvae after infection with either *Symbiodinium* or *Chromera*, which is an artificial situation. In order to replicate a more natural situation, after coral spawning *A. digitifera* larvae were simultaneously exposed to both *Chromera* and *Symbiodinium* clade B1 in order to investigate the response of the larvae in a mixed infection scenario. This experiment was carried out during the same coral spawning event, at the same location and using the same batches of larvae as those described in Chapters 2 and 3. In Chapter 2, I examined the transcriptomic impact on coral larvae of infection with competent *Symbiodinium* during the early stages of the interaction. An advance on this would be to simultaneously examine the transcriptomic responses of both coral and symbiont during the infection process. To enable this, after the coral spawning event on the Great Barrier Reef in early summer 2014, *Acropora tenuis* larvae were exposed to clade C1 *Symbiodinium*, which is the natural symbiont of *A. tenuis*, and sampled at 4, 12, 48 and 72 h post-infection. The intention is to obtain transcriptomic data at sufficient depth to enable the responses of the host and symbiont to be determined from the same RNA preparation.

In Chapter 3, I examined the transcriptomic impact on coral larvae of infection with the culture collection *Chromera* strain (CCMP2878; isolated from *Plesiastrea versipora* from Sydney harbor). I was interested in investigating the interaction of coral larvae with another *Chromera* strain (Mdig03; isolated from *M. digitata* at the GBR). During the 2014 coral spawning event on the GBR, larvae of *A. tenuis* and *A. millepora* were obtained exposed to cultured *Chromera* Mdig03, and sampled at 4, 12, 48 and 72

h post-infection. As in the case of the *Symbiodinium* experiment described in the previous paragraph, the intention is to use a dual RNA-Seq approach to simultaneously investigate the responses of both coral and *Chromera* during infection.

In Chapter 4, I assembled a *de novo* transcriptome for *Chromera* isolated from *M. digitata* on the GBR. The availability of this reference assembly will allow profiling whole gene expression landscapes in *Chromera* to further understand basic biology of these unique organisms and their responses to different environmental and/or physiological conditions. *Chromera* has been shown to grow under mixotrophic conditions (Foster *et al.* 2014), which resemble those encountered within coral hosts. To explore *Chromera* transcriptome-wide gene expression while growing under mixotrophic conditions, *Chromera* was cultured under normal light and temperature conditions in F/2 medium supplemented with organic compounds and RNA-Seq data obtained. Comparisons between mixotrophically grown and control cultures indicate that more than 2000 genes were differentially expressed. Hopefully, analyses of these genes will provide insights into how *Chromera* metabolism differs in the free-living (autotrophic) and “symbiotic” (mixotrophic) states. One important initial implication of this result is that *Chromera* is a typical eukaryote in terms of gene regulation occurring at the transcription level, whereas *Symbiodinium* is quite different in this respect. *Chromera* grows at temperatures of up to 35 °C (Visser *et al.* 2012), and showed higher thermal tolerance (personal observations) than did *Symbiodinium* (clade B strain) when in culture. In order to investigate mechanisms underlying *Chromera* thermotolerance, cultures were exposed to both heat stress and heat shock. Illumina RNA-Seq will be used to investigate *Chromera* transcriptome-wide gene expression during thermal challenges.

Attempts were made towards detecting *Chromera* in a wide variety of corals at multiple locations on the Great Barrier Reef (Appendix II), with little success. Reasons for this are not clear – perhaps the coral-*Chromera* association is highly host-specific or seasonal. Given the higher thermal tolerance of *Chromera* in cultures, it is possible that *Chromera* may increase in absolute or relative abundance in corals experiencing bleaching (where *Symbiodinium* density is minimal). This was observed with the unidentified apicomplexan type-N (Toller *et al.* 2002). I also detected *Chromera* sequences in the soft coral *Lobophytum pauciflorum*, implying that it can infect a broad

range of hosts, so it is worth looking for chromerids and apicomplexans in other marine invertebrates - for example sea anemones and mollusks. Hence molecular prospecting for Chromerid-like sequences in invertebrates is worthwhile.

5.4 Conclusion and impact of the work

Coral reefs are subjected to a variety of natural and anthropogenic stressors globally. Despite attempts for coral reef monitoring and assessment, these protocols depend mainly on macroscopic signs of mortality as a key indicator of stress. In order for efficient reef management and conservation, it is crucial that stress can be identified at sub-lethal levels. New technologies in molecular biology (-omics approaches; genomics, transcriptomics, proteomics) using high throughput sequencing can be of a great benefit to coral reef conservation. Understanding the molecular mechanisms underlying the biology reef-building corals is one of the important issues for effective coral reef management and conservation. Engineering the coral symbioses might hold a great promise to increase coral thermotolerance. Corals hosting thermotolerant symbionts show high resilience thus inoculating corals with thermotolerant symbionts might be a means of coral survival against current and future climate changes. Understanding the symbiont infection process therefore is very important in this context. The PhD research provided novel insights into the molecular events in the coral host during infection with 2 different symbionts (*Symbiodinium* and the newly-discovered *Chromera*). Consequently, the thesis contributes to better understanding the symbiont infection process. Research in this thesis should initiate more work to understand the symbiont infection process in different host/symbiont models, thus helping engineering coral symbioses to increase thermotolerance and more effective reef management and conservation. In addition, the thesis provided insights into host-pathogen/parasite interactions and showed that coral responses to *Chromera* have similarities to the responses of vertebrates to parasites. Moreover, the thesis provided a genomic resource for a *Chromera* strain that can be used as a reference for large-scale gene expression and comparative analyses to better understand the biology of these newly discovered algae and suggested the potential use of *Chromera* as a model organism in developing anti-malarial drugs.

5.5 References

- Chen MC, Cheng YM, Hong MC, Fang LS (2004) Molecular cloning of Rab5 (ApRab5) in *Aiptasia pulchella* and its retention in phagosomes harboring live zooxanthellae. *Biochem Biophys Res Commun* **324**, 1024-1033.
- Cumbo VR, Baird AH, Moore RB, *et al.* (2013) *Chromera velia* is Endosymbiotic in Larvae of the Reef Corals *Acropora digitifera* and *A. tenuis*. *Protist* **164**, 237-244.
- Davy SK, Allemand D, Weis VM (2012) Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiol Mol Biol Rev* **76**, 229-261.
- Foster C, Portman N, Chen M, Slapeta J (2014) Increased growth and pigment content of *Chromera velia* in mixotrophic culture. *Fems Microbiology Ecology* **88**, 121-128.
- Hase K, Kawano K, Nochi T, *et al.* (2009) Uptake through glycoprotein 2 of FimH(+) bacteria by M cells initiates mucosal immune response. *Nature* **462**, 226-230.
- Kvennefors ECE, Leggat W, Hoegh-Guldberg O, Degnan BM, Barnes AC (2008) An ancient and variable mannose-binding lectin from the coral *Acropora millepora* binds both pathogens and symbionts. *Developmental and Comparative Immunology* **32**, 1582-1592.
- Lesser MP, Stat M, Gates RD (2013) The endosymbiotic dinoflagellates (*Symbiodinium* sp.) of corals are parasites and mutualists. *Coral Reefs* **32**, 603-611.
- Libro S, Kaluziak ST, Vollmer SV (2013) RNA-seq Profiles of Immune Related Genes in the Staghorn Coral *Acropora cervicornis* Infected with White Band Disease. *Plos One* **8**.
- Meyer E, Weis VM (2012) Study of cnidarian-algal symbiosis in the "omics" age. *Biological Bulletin* **223**, 44-65.
- Mieog JC, Olsen JL, Berkelmans R, *et al.* (2009) The Roles and Interactions of Symbiont, Host and Environment in Defining Coral Fitness. *Plos One* **4**.
- Moore RB, Obornik M, Janouskovec J, *et al.* (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**, 959-963.
- Obornik M, Vancova M, Lai DH, *et al.* (2011) Morphology and Ultrastructure of Multiple Life Cycle Stages of the Photosynthetic Relative of Apicomplexa, *Chromera velia*. *Protist* **162**, 115-130.
- Pinzon JH, Kamel B, Burge CA, *et al.* (2015) Whole transcriptome analysis reveals changes in expression of immune-related genes during and after bleaching in a reef-building coral. *R Soc Open Sci* **2**, 140214.
- Ragni M, Airs RL, Hennige SJ, *et al.* (2010) PSII photoinhibition and photorepair in *Symbiodinium* (Pyrrhophyta) differs between thermally tolerant and sensitive phylotypes. *Marine Ecology Progress Series* **406**, 57-70.
- Schnitzler CE, Weis VM (2010) Coral larvae exhibit few measurable transcriptional changes during the onset of coral-dinoflagellate endosymbiosis. *Mar Genomics* **3**, 107-116.
- Seneca FO, Palumbi SR (2015) The role of transcriptome resilience in resistance of corals to bleaching. *Molecular Ecology* **24**, 1467-1484.
- Slapeta J, Linares MC (2013) Combined Amplicon Pyrosequencing Assays Reveal Presence of the Apicomplexan "type-N" (cf. *Gemmocystis cylindrus*) and *Chromera velia* on the Great Barrier Reef, Australia. *Plos One* **8**.

- Stat M, Carter D, Hoegh-Guldberg O (2006) The evolutionary history of Symbiodinium and scleractinian hosts - Symbiosis, diversity, and the effect of climate change. *Perspectives in Plant Ecology Evolution and Systematics* **8**, 23-43.
- Stat M, Morris E, Gates RD (2008) Functional diversity in coral-dinoflagellate symbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9256-9261.
- Suggett DJ, Warner ME, Smith DJ, *et al.* (2008) Photosynthesis and production of hydrogen peroxide by Symbiodinium (Pyrrhophyta) phylotypes with different thermal tolerances. *Journal of Phycology* **44**, 948-956.
- Toller WW, Rowan R, Knowlton N (2002) Genetic evidence for a protozoan (phylum Apicomplexa) associated with corals of the *Montastraea annularis* species complex. *Coral Reefs* **21**, 143-146.
- Voolstra CR, Schwarz JA, Schnetzer J, *et al.* (2009) The host transcriptome remains unaltered during the establishment of coral-algal symbioses. *Molecular Ecology* **18**, 1823-1833.
- Weiss Y, Foret S, Hayward DC, *et al.* (2013) The acute transcriptional response of the coral *Acropora millepora* to immune challenge: expression of GiMAP/IAN genes links the innate immune responses of corals with those of mammals and plants. *Bmc Genomics* **14**, 400.

Appendix I

The transcriptomic response of the coral *Acropora digitifera* to a competent *Symbiodinium* strain: the symbiosome as an arrested early phagosome

Chapter 2 published as an original article:

Mohamed, A.R., Cumbo, V., Harii, S., Shinzato, C., Chan, C.X., Ragan, M.A., Bourne, D.G., Willis, B.L., Ball, E.E., Satoh, N. and Miller, D.J. (2016). The transcriptomic response of the coral *Acropora digitifera* to a competent *Symbiodinium* strain: the symbiosome as an arrested early phagosome. *Molecular Ecology* 25, 3127-3141.
doi:10.1111/mec.13659

The transcriptomic response of the coral *Acropora digitifera* to a competent *Symbiodinium* strain: the symbiosome as an arrested early phagosome

A. R. MOHAMED,*†‡§ V. CUMBO,*¶ S. HARI,** C. SHINZATO,†† C. X. CHAN,‡‡
M. A. RAGAN,‡‡ D. G. BOURNE,§§ B. L. WILLIS,*¶¶ E. E. BALL,**** N. SATOH†† and
D. J. MILLER*†

*ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Qld 4811, Australia, †Comparative Genomics Centre and Department of Molecular and Cell Biology, James Cook University, Townsville, Qld 4811, Australia, ‡Zoology Department, Faculty of Science, Benha University, Benha 13518, Egypt, §AIMS@JCU, Department of Molecular and Cell Biology, Australian Institute of Marine Science, James Cook University, Townsville, Qld 4811, Australia, ¶Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia, **Sesoko Station, Tropical Biosphere Research Center, University of the Ryukyus, 3422 Sesoko, Motobu Okinawa 905-0227, Japan, ††Marine Genomics Unit, Okinawa Institute of Science and Technology Promotion Corporation, Onna, Okinawa 904-0412, Japan, ‡‡ARC Centre of Excellence in Bioinformatics and Institute for Molecular Bioscience, The University of Queensland, Brisbane, Qld 4072, Australia, §§Australian Institute for Marine Science, PMB 3, Townsville, Qld 4811, Australia, ¶¶Department of Marine Ecosystems and Impacts, James Cook University, Townsville, Qld 4811, Australia, ****Evolution, Ecology and Genetics, Research School of Biology, Australian National University, Canberra, ACT 0200, Australia

Abstract

Despite the ecological significance of the relationship between reef-building corals and intracellular photosynthetic dinoflagellates of the genus *Symbiodinium*, very little is known about the molecular mechanisms involved in its establishment. Indeed, microarray-based analyses point to the conclusion that host gene expression is largely or completely unresponsive during the establishment of symbiosis with a competent strain of *Symbiodinium*. In this study, the use of Illumina RNA-Seq technology allowed detection of a transient period of differential expression involving a small number of genes (1073 transcripts; <3% of the transcriptome) 4 h after the exposure of *Acropora digitifera* planulae to a competent strain of *Symbiodinium* (a clade B strain). This phenomenon has not previously been detected as a consequence of both the lower sensitivity of the microarray approaches used and the sampling times used. The results indicate that complex changes occur, including transient suppression of mitochondrial metabolism and protein synthesis, but are also consistent with the hypothesis that the symbiosome is a phagosome that has undergone early arrest, raising the possibility of common mechanisms in the symbiotic interactions of corals and symbiotic sea anemones with their endosymbionts.

Keywords: *Acropora*, *Symbiodinium*, symbiosis, symbiosome, transcriptome

Received 31 December 2015; revision received 4 April 2016; accepted 14 April 2016

Content has been removed
due to copyright restrictions

Appendix II

Detecting the presence of *Chromera* in some corals of the Great Barrier Reef, Australia using PCR assay

The aim of this work was to detect *Chromera* specific sequences in different coral species in different locations in the GBR using polymerase chain reaction (PCR).

Sampling

Coral species were chosen to represent common species in the GBR. The sampled coral species list included *Montipora digitata*, *Acropora tenuis*, *A. millepora*, *Pocillopora damicornis*, massive *Porites* and *Fungia concina* from reefs around Heron Island at the southern GBR. *M. digitata* was collected from nelly bay at the central GBR Magnetic Island. *M. digitata*, *A. tenuis*, *A. millepora*, massive *Porites*, *Fungia concina*, and the soft coral *Lopophytum pauciflorum* were collected from Orpheus Island at the central GBR (Table 1). For each species, 3 coral nubbins or fragments from 3 different colonies, which are 10-m spaced were collected. Coral nubbins were sampled using SCUBA and snorkeling equipment and placed in Zip Lock bag under water and transferred to the laboratory. Samples collected from Heron Island were fixed in 95% Ethanol and stored at 4 °C, while samples collected from Magnetic and Orpheus Islands were fixed in liquid nitrogen and stored at -80 °C until further treatment.

Genomic DNA (gDNA) extraction from coral tissues

Fixed samples were crushed into fine powder using mortar and pestle. The resulting coral powder was lysed using the following lysing buffer (20mM Tris pH 8.0, 5mM EDTA, 1% SDS (w/v), 400mM NaCl, 400 $\mu\text{g mL}^{-1}$ Proteinase K) and incubated overnight at 55 °C with agitation. The lysate was centrifuged at 15,000 g for 15min at 4 °C. The resulting supernatant was transferred to an equal volume of phenol-chloroform mixture (1:1) and mixed by inverting the tube followed by centrifugation at 12,000 g and 4 °C for 10 min. The aqueous phase was transferred into a new tube and DNA was

precipitated by adding equal volume of isopropanol and centrifuged for 15 min at 9,000 g and 4 °C. The resulting pellet was washed with 2 mL of 70% ethanol and centrifuged for 5 min at 15,000 g and 4 °C. The pellet was air dried for 15 min at room temperature and suspended in 100 µL water and stored at -20 °C. Nanpdrop was used to determine DNA concentration and/or quality.

Table 1. Information for the coral species sampled in the study from the GBR.

N= 9 per species (3 nubbins x 3 colonies).

Coral Species	Reef	Location	Depth	Date
<i>Montipora digitata</i>	Heron Reef Flat	Heron Island	2 m	Apr-13
<i>Acropora tenuis</i>	Coral Gardens	Southern GBR	6 m	
<i>Acropora millepora</i>	Heron Reef Flat		3 m	
<i>Pocillopora damicornis</i>	Coral Gardens		6 m	
<i>Massive Porites</i>	Coral Canyons		5 m	
<i>Fungia concina</i>	Coral Canyons		5 m	
<i>Montipora digitata</i>	Nelly Bay	Magnetic Island Central GBR	2 m	May-13
<i>Montipora digitata</i>	Little Pioneer Bay	Orpheus Island	2 m	Nov-13
<i>Acropora tenuis</i>	Cattle Bay	Central GBR	4 m	
<i>Acropora millepora</i>	Cattle Bay		4 m	
<i>Massive Porites</i>	Pioneer Bay		6 m	
<i>Fungia concina</i>	Cattle Bay		5 m	
<i>Lobophytum pauciflorum</i>	Pioneer Bay		3 m	

Design of Chromera-specific PCR primers

Chromera large subunit ribosomal RNA gene partial sequence (GenBank: EU106870.1) (Moore *et al.* 2008) and *Chromera* clone JS497 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence (GenBank: JN935835.1) (Morin-Adeline *et al.* 2012) were retrieved from the GenBank and used as templates to specifically design PCR primers for *Chromera*. Specific PCR primers were designed using the NCBI primer BLAST tool (Ye *et al.* 2012) (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). The following primers (Table 1) were used to and successfully amplify *Chromera* sequences from gDNA isolated from pure *Chromera* cultures.

Table (1): List of *Chromera* specific primers used and their target regions.

Primer Pair	Target gene/region	Product length	Primer Sequence
1st	LSU RNA 28S region	755 bp	Forward Primer AGCCTAAGTGGGAGATCCGT Reverse Primer ACAAAGAAAGCTGCGTGCTG
2nd	LSU RNA 28S region	416 bp	Forward Primer GTTTTGGAAAGCTTCGGCGT Reverse Primer ACGGATCTCCCACTTAGGCT
3rd	SSU RNA 18S region	778 bp	Forward Primer CCGACTAGAGATTGGCGGTC Reverse Primer CTGACGGACTGTCGTGTGAA
4th	SSU RNA 18S region	482 bp	Forward Primer TTCACACGACAGTCCGTCAG Reverse Primer CAGCACTGCAAACACATGCT

Presence/ absence screening for *Chromera* sequences in coral tissues

PCR was used to amplify the *Chromera* ribosomal genes using the above primers, a positive control reaction was applied using the following *Symbiodinium* general primers (in order to check that gDNA is free from PCR inhibitors) and a negative control reaction was applied using MQ water. PCR reaction conditions

PCR reaction was conducted in 50 µl reaction using 1 µl of gDNA (approx. 100ng of DNA) as template. 1 µl of GoTaq® DNA polymerase and 2X GoTaq® reaction buffer were used, 5 µl of each primer were used and finally sterile MQW was added to the reaction mixture to make a total volume of 50 µl. The PCR profile was one cycle for 2 min at 94 °C for initial denaturation followed by 34 cycles of 30 sec at 94 °C, annealing for 30 sec at 47 °C/ 51 °C and extension for 2 min at 72 °C. The final extension was at 72 °C for 10 min. The obtained amplicons were run on 1.5% Agarose gel and visualized using UV trans-illuminator

Findings

Amplifying *Chromera* sequences from coral holobiont DNA was very hard as the samples scored negative for the majority of the coral species selected (see for example Figure 1), except for *Montipora digitata* collected from Magnetic Island (Figure 2) and the soft coral *Lobophytum pauciflorum* collected from Orpheus Island (Figure 3). In addition, the presence of *Chromera* in the soft coral was confirmed by sequencing the amplified bands using Sanger sequencer. The sequences were trimmed and assembled using the sequencer software. BLASTN was utilized to identify the sequences. *Chromera* sequences were deposited to the GenBank and accession numbers will be assigned to them.

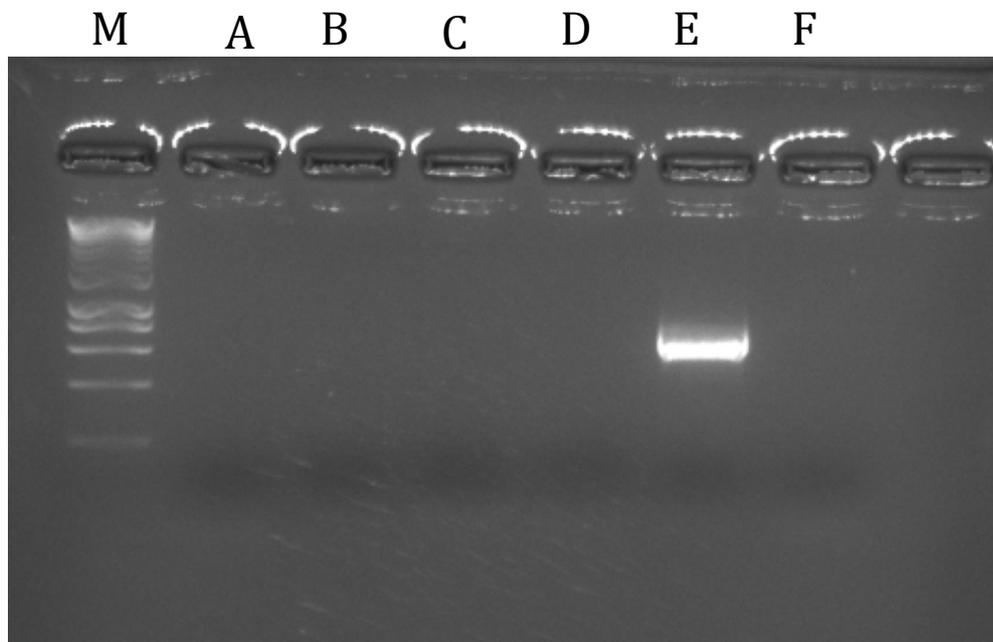


Figure 1. Gel picture showing the absence of *Chromera* sequenced using *Acropora millepora* (Heron Island) using 4 sets of *Chromera* primers (lanes: A, B, C and D). Whereas *Symbiodinium* sequences were easily detected using the same *A. millepora* gDNA and *Symbiodinium* primers (lane E). Lane F shows a negative control reaction.

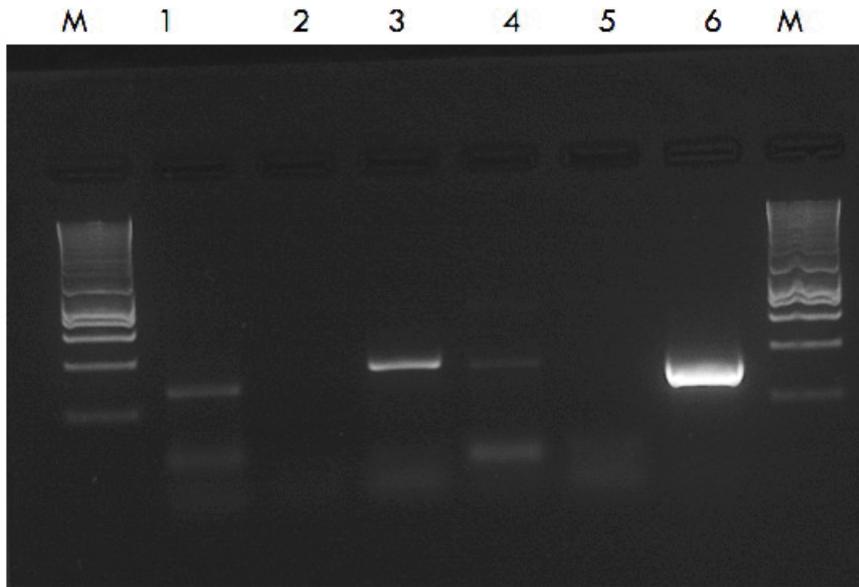


Figure 2. Gel picture showing the presence of *Chromera* sequenced using *Montipora digitata* (Magnetic Island) using 3 sets of *Chromera* primers (lanes: 1, 3 and 4). Lane 5 shows negative results using *A. millepora* gDNA with the 3rd *Chromera* primer pair. Whereas *Symbiodinium* sequences were easily detected using the same *M. digitata* gDNA and *Symbiodinium* primers (lane 6). Lane F shows a negative control reaction.

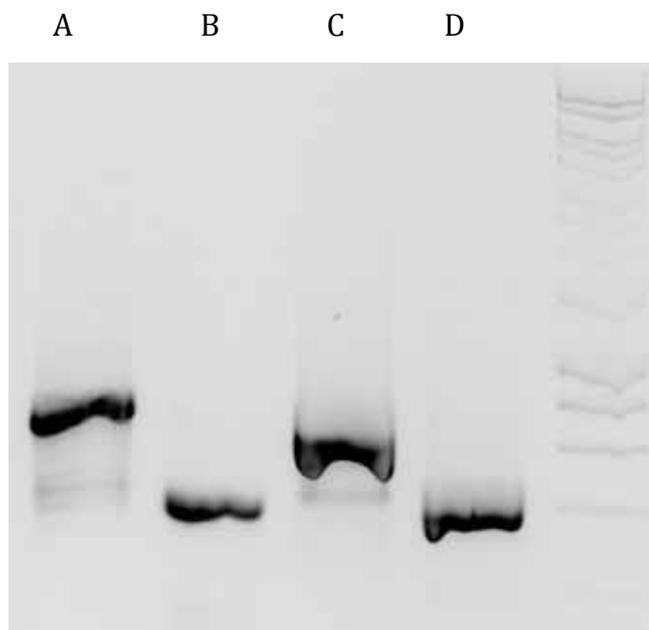


Figure 3. Gel picture showing the presence of *Chromera* sequenced using the soft coral *Lobophytum pauciflorum* using 4 sets of *Chromera* primers (lanes: A, B, C and D).

>Seq1 [organism= *Chromera velia*] large subunit ribosomal RNA gene (28S rDNA), partial sequence

GTTTAATTAACAAAGCATTGCGATGGTGAAAATTCGTGTTGACGCAAT
GTGATTTCTGCCAGTGCCCTGAATGTCAAAGTGATGAAATTCAAACAAG
CGCGGGTAAACGGCGGGAGTAACTATGACTCTCTTAAGGTAGCCAAATGC
CTCGTCATCTAATTAGTGACGCGCATGAATGGATTAACGAGATTCCTACT
GTCCCTATCTACTATCTAGCGAAACCACAGCCAAGGGAACGGGCTTGAC
TAATCAGCGGGGAAAGAAGACCCTGTTGAGCTTGACTCTAGTCCGACTTT
GTGAAATGACC

>Seq2 [organism= *Chromera velia*] large subunit ribosomal RNA gene (28S rDNA), partial sequence

TTAGTCCAAGCCCGTTCCCTTGGCTGTGGTTTCGCTAGATAGTAGATAGG
GACAGTAGGAATCTCGTTAATCCATTCATGCGCGTCACTAATTAGATGAC
GAGGCATTTGGCTACCTTAAGAGAGTCATAGTTACTCCCGCCGTTTACCC
GCGCTTGTTTGAATTCATCACTTTGACATTCAGGGCACTGGGCAGAAAT
CACATTGCGTCAACACGAATTTTCACCATCGCAATGCTTTGTTTTAATTA
AACAGTCGGATTCCCCTTGTCCGCTCCAGTTCTGAGATGACCGTTTGATG
CAGAGAGAACGCCG

>Seq3 [organism= *Chromera velia*] small subunit ribosomal RNA gene (18S rDNA), partial sequence

GTTCTGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATTGATGGA
AGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGA
AAACTCACCAGGTCCAGACATAGTGAGGATTGACAGATTGATAGCTCTTT
CTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGAT
TTGTCTACTTAATTGTGATAATGAACGAGACCTAACCTGCTAAATAGTC
GGTCGAATCTTTCGATTCGGCATGGACTTTTTAGAGGGACTTTGCGTGTC
TAACGCAAGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGT
TCTGGGCTGCACGCGCGCTACACTGATGCAGTCAGCGAGTTTTTCCTGTT
CCGGAAGGATCGGGTAATCTTCTGAACTGCATCGTGATGGGGATAGATT
ATTGCAATTATTAATCTTGAACGAGGAATTCCTAGTAAGTGCAAGTTATC
AGCTTGTACTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTCC
TACCGATTGAATGATCCGGTGAATTATTTGGACCGTGCTCGATTCTCTTG
AAGAGAGCGTGCGAAATTTTGTGAACCTTATCATTTAGAGGAAGGAGAAG

TCGTAACAAGGTTTCCGTAGGTGAACCTGCAGAAGGATCA

>Seq4 [organism= *Chromera velia*] small subunit ribosomal RNA gene (18S rDNA),
partial sequence

GACTTCTCCTTCTCTAAATGATAAGGTTACAAAATTTTCGCACGCTCTC
TTCAAGAGAATCGAGCACGGTCCAAATAATTCACCGGATCATTCAATCGG
TAGGAGCGACGGGCGGTGTGTACAAAGGGCAGGGACGTAATCAGTACAAG
CTGATAACTTGCACCTACTAGGAATTCCTCGTTCAAGATTAATAATTGCA
ATAATCTATCCCCATCACGATGCAGTTTCAGAAGATTACCCGATCCTTCC
GGAACAGGAAAACTCGCTGACTGCATCAGTGTAGCGCGCGTGCAGCCCA
GAACATCTAAGGGCATCACAGACCTGTTATTGCCTCAAACCTTCCTTGCGT
TAGACACGCAAAGTCCCTCTAAAAGTCCATGCCGAATCGAAAGATTCGA
CCGACTATTTAGCAGGTTAAGGTCTCGTTCATTATCACAATTAAGTAGAC
AAATCACTCCACCAACTAAGAACGGCCATGCACCACCACCCATAGAATCA
AGAAAGAGCTATCAATCTGTCAATCCTCACTATGTCTGGACCTGGTGAGT
TTTCCCGTGTGAGTCAAATTAAGCCGCAGGCTCCACTCCTGGTGGTGCC
CTTCCATCAATTCCTTTAAGTTTCAGCCTTGCAGCATACTCCCCCAGA
ACCCAAAACTTTGATTTCTCATAAGGTGCTGAAGGTGTCGTAATGGAAC

>Seq5 [organism= *Chromera velia*] small subunit ribosomal RNA gene (18S rDNA),
partial sequence

GGACTTGACTTGAACATAGGAACTCGGCCTGGTGACCTCACAGGTGTTCT
GAAGCAGTGTGCCTAGCACTCTACTCTACCTAGAAACGTTTTAAAGAAC
TTAAGACTTTCGGCGATGGATGTCTTGGTTCACACAACGATGAAGGACGC
GGCCAACTGTGATACTCAGTGTGAATTGCAGATTTTCAGTGAATCATCAGA
CAGCTGAACGCGCTAGGTTCCCCTTCGGGGGAAGGTTGCAGTCAGTACCT
TTTGTGATTTTGCTAAAGTACTGAGTTGTTTCATCAACCTAAGGCTTAGCT
CTTTGCCCTGTGAATACAG

>Seq6 [organism= *Chromera velia*] small subunit ribosomal RNA gene (18S rDNA),
partial sequence

AGCAAATCACAAAAGGTACTGACTGCAACCTTCCCCGAAGGGGAACCT
AGCGCGTTCAGCTGTCTGATGATTCAGTAAATCTGCAATTCACACTGAG
TATCACAGTTGGCCGCGTCCTTCATCGTTGTGGGAACCAAGACATCCATC
GCCGAAAGTCTTAAGTTCTTTTAAAAC

Discussion

Chromera has been isolated from different hard corals (at least three species) (Cumbo *et al.* 2013; Moore *et al.* 2008) and detected in tissue the soft coral *Lobophytum pauciflorum* (above). Attempts have been made to detect *Chromera* sequences in multiple corals, but it was very difficult to detect *Chromera* sequences in coral holobiont DNA using PCR. The reason for the inability to amplify *Chromera* sequences is unclear. *Chromera* might be present but at low abundances that are undetectable by PCR. Also seasonality should be taken into consideration as the association might vary in different seasons making the *Chromera* abundance too low to be detected from coral holobiont gDNA. As *Chromera* showed high thermal tolerance and increase in growth rate at higher temperatures up to (36°C) (Visser *et al.* 2012), hence *Chromera* might be more abundant in the summer months or during bleaching and could be detected in bleached coral tissues using PCR. Finally, it is worth looking for *Chromera* associations with different marine species such as other cnidarians, porifera and mollusca.

References

- Moore RB, Obornik M, Janouskovec J, *et al.* (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**, 959-963.
- Morin-Adeline V, Foster C, Slapeta J (2012) Identification of *Chromera* velia by fluorescence in situ hybridization. *Fems Microbiology Letters* **328**, 144-149.
- Ye J, Coulouris G, Zaretskaya I, *et al.* (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134.

Appendix III

Chapter 3: Deciphering the nature of the coral-*Chromera* association via next generation sequencing. Supplementary Information

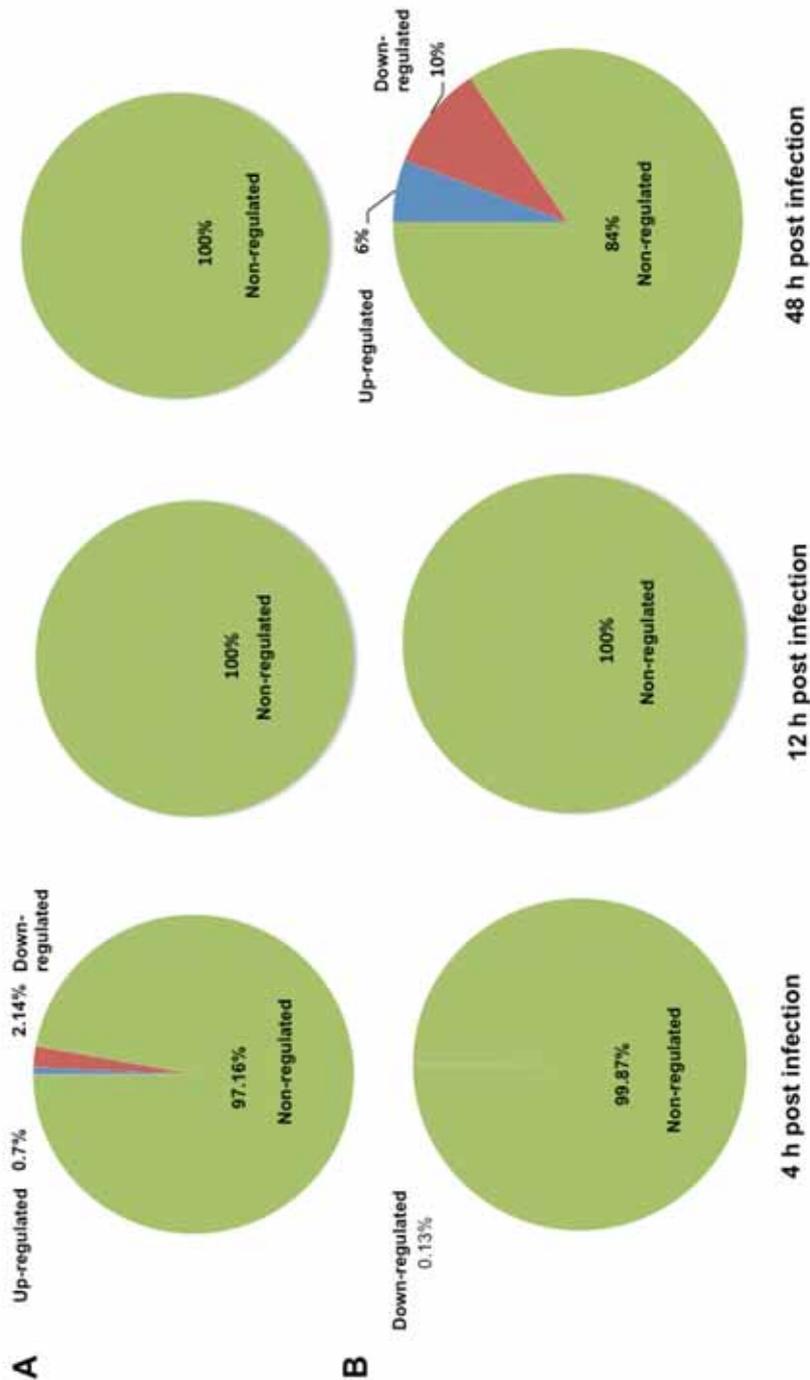


Figure S3.1 A. *digitifera* transcriptome changes during *Symbiodinium* and *Chromera* infections. Part A shows the transcriptome changes in *Symbiodinium*-infected larvae compared to control. Part B shows the transcriptome changes in *Chromera*-infected larvae compared to control.

Transcriptome profiling was investigated at 4, 12 and 48 h post infection in both cases and a P -value ≤ 0.05 was used to filter significant differentially expressed genes.

Table S3.1 Genes involved in the KEGG pathway ECM receptor interaction

Cluster ID	Swiss-Prot ID	Best BLAST Hit	Species	E-value	logFC
adi_EST_assem_13417	O00468	Agrin	<i>Homo sapiens</i>	8.92E-45	-1.09
adi_EST_assem_5774	A6NMZ7	Collagen type VI alpha 6	<i>Homo sapiens</i>	2.87E-67	-2.88
adi_EST_assem_2378	P12107	Collagen, type XI, alpha 1	<i>Homo sapiens</i>	1.30E-48	-2.34
adi_EST_assem_4678	Q14118	Dystroglycan 1 (dystrophin-associated glycoprotein 1)	<i>Homo sapiens</i>	2.89E-27	-2.37
adi_EST_assem_5356	P53708	Integrin, alpha 8	<i>Homo sapiens</i>	7.89E-82	-2
adi_EST_assem_128	Q13797	Integrin, alpha 9	<i>Homo sapiens</i>	1.83E-98	-1.5
adi_EST_assem_2138	P05556	Integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	<i>Homo sapiens</i>	1.91E-158	-1.43
adi_EST_assem_53	P25391	Laminin, alpha 1	<i>Homo sapiens</i>	2.73E-45	-3.31
adi_EST_assem_574	P07942	Laminin, beta 1	<i>Homo sapiens</i>	0.00E+00	-1.89
adi_EST_assem_576	P55268	Laminin, beta 2 (laminin S)	<i>Homo sapiens</i>	8.13E-78	-2.53

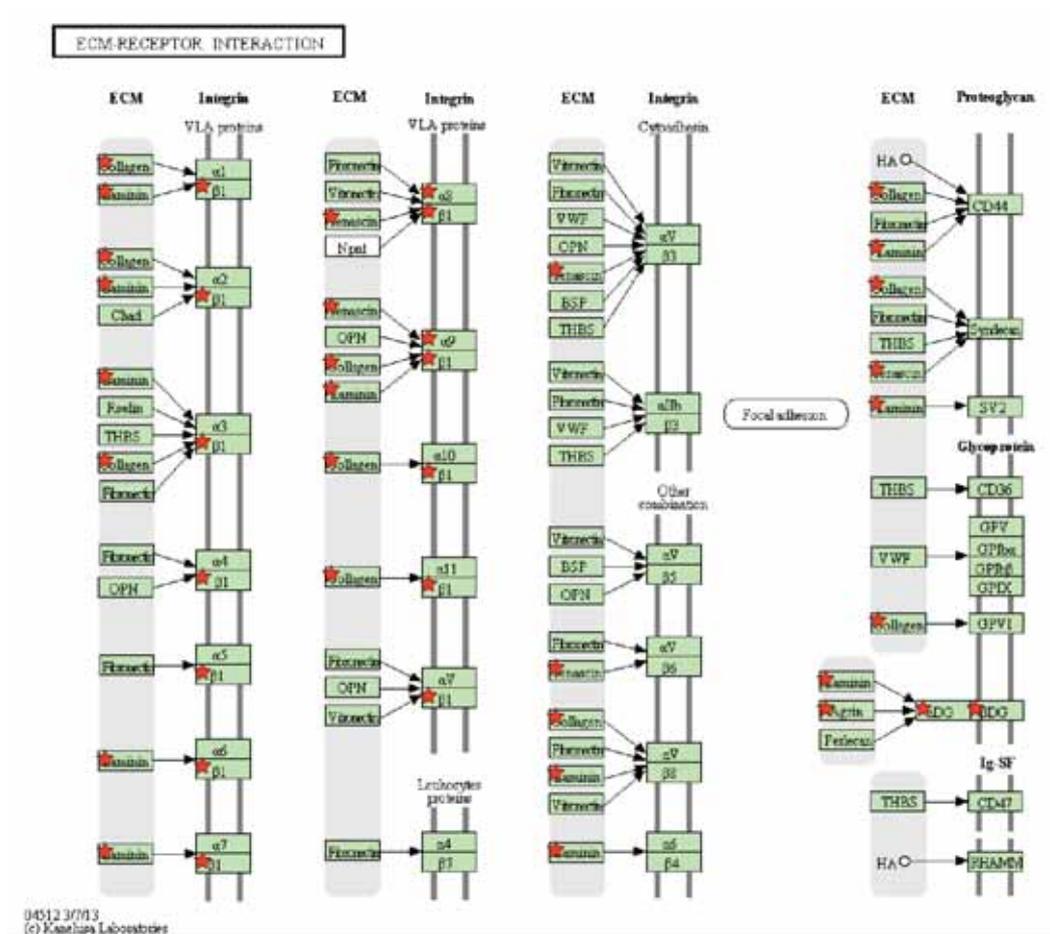


Figure S3.2 ECM-receptor interaction (KEGG pathway, hsa04512) significantly enriched in the set of down regulated in *Chromera*-infected larvae at 48 h post infection. Red stars highlight proteins present in our dataset.

Table S3.3 Genes involved in the KEGG pathway focal adhesion

Cluster ID	Swiss-Prot ID	Best BLAST Hit	Species	E-value	logFC
adi_EST_assem_8781	Q13905	Rap guanine nucleotide exchange factor (GEF) 1	<i>Homo sapiens</i>	1.50E-16	-1.8
adi_EST_assem_5250	O75116	Rho-associated, coiled-coil containing protein kinase 2	<i>Homo sapiens</i>	0	-2.82
adi_EST_assem_5774	A6NMZ7	Collagen type VI alpha 6	<i>Homo sapiens</i>	2.87E-67	-2.88
adi_EST_assem_2378	P12107	Collagen, type XI, alpha 1	<i>Homo sapiens</i>	1.30E-48	-2.34
adi_EST_assem_12712	Q14185	Dedicator of cytokinesis 1	<i>Homo sapiens</i>	0.00E+00	-2.17
adi_EST_assem_5356	P53708	Integrin, alpha 8	<i>Homo sapiens</i>	7.89E-82	-2
adi_EST_assem_128	Q13797	Integrin, alpha 9	<i>Homo sapiens</i>	1.83E-98	-1.5
adi_EST_assem_2138	P05556	Integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	<i>Homo sapiens</i>	1.91E-158	-1.43
adi_EST_assem_53	P25391	Laminin, alpha 1	<i>Homo sapiens</i>	2.73E-45	-3.31
adi_EST_assem_574	P07942	Laminin, beta 1	<i>Homo sapiens</i>	0.00E+00	-1.89
adi_EST_assem_576	P55268	Laminin, beta 2 (laminin S)	<i>Homo sapiens</i>	8.13E-78	-2.53
adi_EST_assem_3616	P27986	Phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	<i>Homo sapiens</i>	5.02E-87	-1.21
adi_EST_assem_659	O14974	Protein phosphatase 1, regulatory (inhibitor) subunit 12A	<i>Homo sapiens</i>	2.00E-80	-1.43
adi_EST_assem_5223	Q07889	Son of sevenless homolog 1 (<i>Drosophila</i>)	<i>Homo sapiens</i>	0.00E+00	-2.21
adi_EST_assem_175	Q9Y490	Talin 1	<i>Homo sapiens</i>	2.34E-161	-1.79
adi_EST_assem_10072	Q9UKW4	Vav 3 guanine nucleotide exchange factor	<i>Homo sapiens</i>	5.43E-32	-2.89

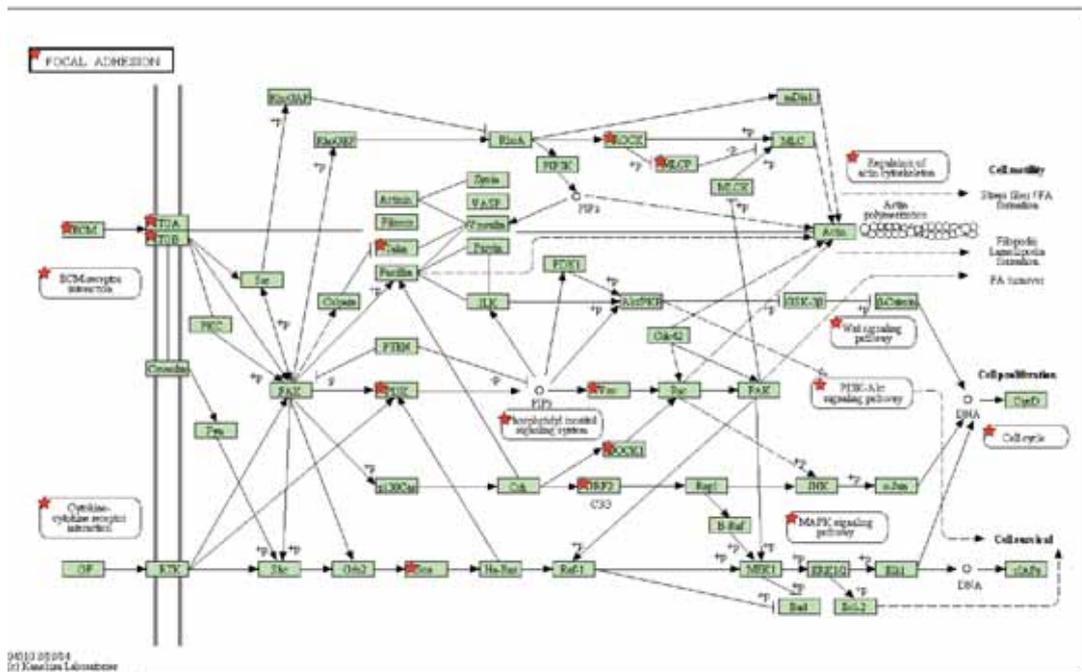


Figure S3.4 Focal adhesion (KEGG pathway, hsa04510) significantly enriched in the set of down regulated in *Chromera*-infected larvae at 48 h post infection. Red stars highlight proteins present in our dataset.

Table S3.4 Genes involved in the KEGG pathway regulation of actin cytoskeleton

Cluster ID	Swiss-Prot ID	Best BLAST Hit	Species	E-value	logFC
adi_EST_assem_13215	Q9Y2X7	G protein-coupled receptor kinase interacting arfgap 1	<i>Homo sapiens</i>	0	-2.26
adi_EST_assem_3391	P46940	IQ motif containing gtpase activating protein 1	<i>Homo sapiens</i>	0	-2.67
adi_EST_assem_5250	O75116	Rho-associated, coiled-coil containing protein kinase 2	<i>Homo sapiens</i>	0	-2.83
adi_EST_assem_8750	Q13009	T-cell lymphoma invasion and metastasis 1	<i>Homo sapiens</i>	9.40E-148	-2.32
adi_EST_assem_8735	P42768	Wiskott-Aldrich syndrome (eczema-thrombocytopenia)	<i>Homo sapiens</i>	1.38E-56	-1.4
adi_EST_assem_4349	P25054	Adenomatous polyposis coli	<i>Homo sapiens</i>	3.40E-114	-1.64
adi_EST_assem_12712	Q14185	Dedicator of cytokinesis 1	<i>Homo sapiens</i>	0.00E+00	-2.17
adi_EST_assem_5341	O60879	Diaphanous homolog 2 (Drosophila)	<i>Homo sapiens</i>	2.06E-106	-2.4
adi_EST_assem_7497	Q8N8S7	Enabled homolog (Drosophila)	<i>Homo sapiens</i>	1.37E-43	-1.58
adi_EST_assem_900	P11362	Fibroblast growth factor receptor 1	<i>Homo sapiens</i>	5.07E-95	-1.79
adi_EST_assem_898	P22607	Fibroblast growth factor receptor 3	<i>Homo sapiens</i>	1.65E-99	-2.3
adi_EST_assem_2436	P15311	Hypothetical protein LOC100129652; ezrin	<i>Homo sapiens</i>	4.03E-156	-2.05
adi_EST_assem_5356	P53708	Integrin, alpha 8	<i>Homo sapiens</i>	7.89E-82	-2
adi_EST_assem_128	Q13797	Integrin, alpha 9	<i>Homo sapiens</i>	1.83E-98	-1.5
adi_EST_assem_2138	P05556	Integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	<i>Homo sapiens</i>	1.91E-158	-1.4
adi_EST_assem_3616	P27986	Phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	<i>Homo sapiens</i>	5.02E-87	-1.21
adi_EST_assem_659	O14974	Protein phosphatase 1, regulatory (inhibitor) subunit 12A	<i>Homo sapiens</i>	2.00E-80	-1.43
adi_EST_assem_5223	Q07889	Son of sevenless homolog 1 (Drosophila)	<i>Homo sapiens</i>	0	-2.21
adi_EST_assem_10072	Q9UKW4	Vav 3 guanine nucleotide exchange factor	<i>Homo sapiens</i>	5.43E-32	-2.89

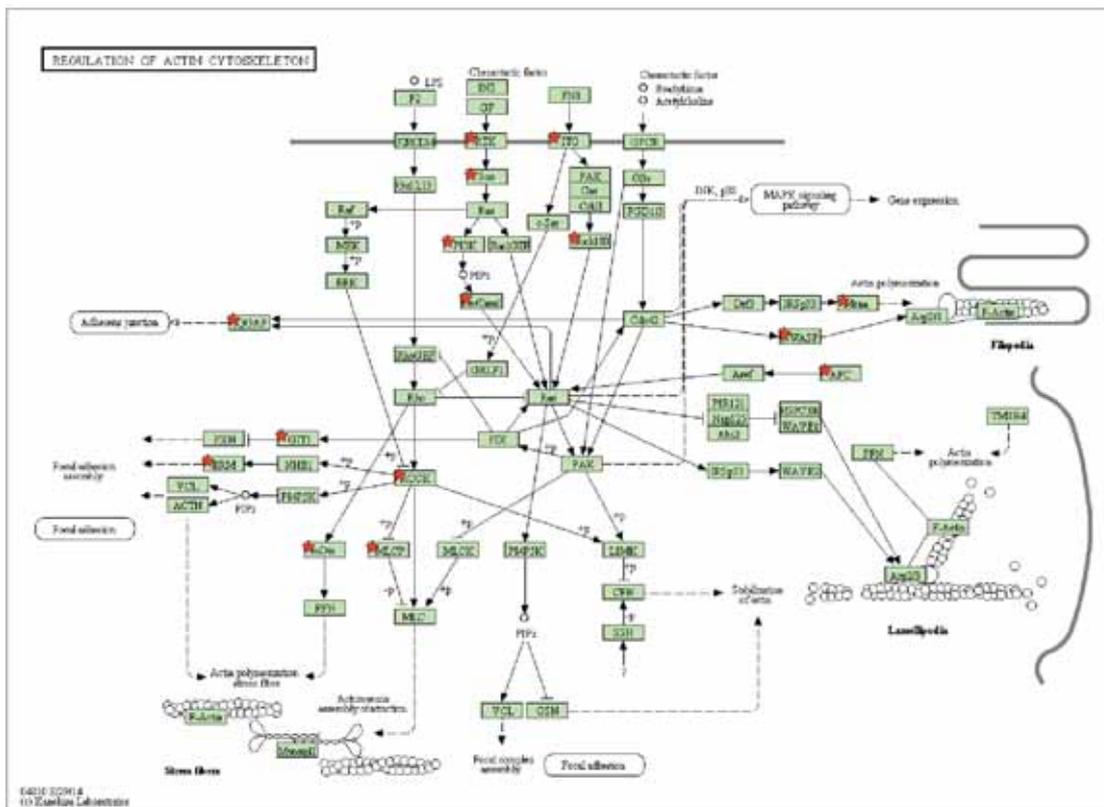


Figure S3.5 Regulation of actin cytoskeleton (KEGG pathway, hsa04810) significantly enriched in the set of down regulated genes in *Chromera*-infected larvae at 48 h post infection. Red stars highlight proteins present in our dataset.

Table S3.5 Annotations of the up regulated *A. digitifera* clusters involved in ribosome functions and translation in *Chromera*- infected larvae at 48 h post infection with corrected $P \leq 0.05$

Uni-Prot ID	Protein Name	Species
Q6NTS3	28S ribosomal protein S24-B, mitochondrial	<i>Xenopus laevis</i>
A6U882	30S ribosomal protein S11	<i>Sinorhizobium medicae</i>
Q9FNP8	40S ribosomal protein S19-3	<i>Arabidopsis thaliana</i>
B3CRZ0	50S ribosomal protein L20	<i>Orientia tsutsugamushi</i>
A0LFC4	50S ribosomal protein L35	<i>Syntrophobacter fumaroxidans</i>
O14464	54S ribosomal protein YPL183W-A, mitochondrial	<i>Saccharomyces cerevisiae</i>
Q95WA0	60S ribosomal protein L26	<i>Littorina littorea</i>
Q9NB33	60S ribosomal protein L44	<i>Ochlerotatus triseriatus</i>
C4KZM8	Eat1b_1615	<i>Exiguobacterium sp. AT1b</i>
A7SGZ5	Eukaryotic translation initiation factor 3 subunit K	<i>Nematostella vectensis</i>
A7SPW6	NEMVEDRAFT_v1g215604	<i>Nematostella vectensis</i>
P42678	Protein translation factor SUI1 homolog	<i>Anopheles gambiae</i>
P49180	Ribosomal Protein, Large subunit	<i>Caenorhabditis elegans</i>
P36241	Ribosomal protein L19	<i>Drosophila melanogaster</i>
Q5ZJ39	Density-regulated protein	<i>Gallus gallus</i>
O70251	Eukaryotic translation elongation factor 1 beta 2	<i>Mus musculus</i>
P70541	Eukaryotic translation initiation factor 2B, subunit 3 gamma	<i>Rattus norvegicus</i>
Q13542	Eukaryotic translation initiation factor 4E binding protein 2	<i>Homo sapiens</i>
O60573	Eukaryotic translation initiation factor 4E family member 2	<i>Homo sapiens</i>
P67985	Heparin binding protein	<i>Sus scrofa</i>
Q8R035	Immature colon carcinoma cluster 1	<i>Mus musculus</i>
Q3TBW2	Mitochondrial ribosomal protein L10	<i>Mus musculus</i>
Q9D1P0	Mitochondrial ribosomal protein L13	<i>Mus musculus</i>
Q9D1I6	Mitochondrial ribosomal protein L14	<i>Mus musculus</i>
Q2TBI6	Mitochondrial ribosomal protein L32	<i>Bos taurus</i>
Q9CQP0	Mitochondrial ribosomal protein L33	<i>Mus musculus</i>
Q9DCU6	Mitochondrial ribosomal protein L4	<i>Mus musculus</i>
Q6DJI4	Mitochondrial ribosomal protein L41	<i>Xenopus laevis</i>
Q08DT6	Mitochondrial ribosomal protein L47	<i>Bos taurus</i>
Q9CQ40	Mitochondrial ribosomal protein L49; similar to mitochondrial ribosomal protein L49	<i>Mus musculus</i>
Q9VE04	Mitochondrial ribosomal protein L55	<i>Drosophila melanogaster</i>
Q9VFB2	Mitochondrial ribosomal protein S10	<i>Drosophila melanogaster</i>
O35680	Mitochondrial ribosomal protein S12	<i>Mus musculus</i>
Q9CR88	Mitochondrial ribosomal protein S14; similar to mitochondrial ribosomal protein S14	<i>Mus musculus</i>
Q9V6Y3	Mitochondrial ribosomal protein S16	<i>Drosophila melanogaster</i>
Q99N85	Mitochondrial ribosomal protein S18A	<i>Mus musculus</i>
Q767K8	Mitochondrial ribosomal protein S18B	<i>Sus scrofa</i>

Q8R2L5	Mitochondrial ribosomal protein S18C	<i>Mus musculus</i>
P82920	Mitochondrial ribosomal protein S21	<i>Bos taurus</i>
Q9VZD5	Mitochondrial ribosomal protein S6	<i>Drosophila melanogaster</i>
Q9HBH1	Peptide deformylase (mitochondrial); component of oligomeric golgi complex 8	<i>Homo sapiens</i>
Q9Y3E5	Peptidyl-trna hydrolase 2	<i>Homo sapiens</i>
P51410	Predicted gene 10117; similar to ribosomal protein L9; ribosomal protein L9	<i>Mus musculus</i>
O09167	Predicted gene 12618; predicted gene 8724; predicted gene 10155; predicted gene 3355; predicted gene 3713; predicted gene 3201; predicted gene 13641; similar to ribosomal protein L21	<i>Mus musculus</i>
O35972	Predicted gene 13671; mitochondrial ribosomal protein L23	<i>Mus musculus</i>
Q7ZWJ4	Ribosomal protein I18a	<i>Danio rerio</i>
P47830	Ribosomal protein L22	<i>Xenopus laevis</i>
Q8JGR4	Ribosomal protein L24	<i>Danio rerio</i>
P17078	Ribosomal protein L35; similar to 60S ribosomal protein L35	<i>Rattus norvegicus</i>
P62282	Ribosomal protein S11	<i>Rattus norvegicus</i>
P84175	Ribosomal protein S12	<i>Gallus gallus</i>
P62268	Ribosomal protein S23; similar to ribosomal protein S23	<i>Rattus norvegicus</i>
P62247	Ribosomal protein S8	<i>Danio rerio</i>
P47826	Ribosomal protein, large, P0	<i>Gallus gallus</i>
P23358	Similar to 60S ribosomal protein L12; ribosomal protein L12	<i>Rattus norvegicus</i>
P24049	Similar to 60S ribosomal protein L17 (L23); similar to 60S ribosomal protein L17 (L23) (Amino acid starvation-induced protein) (ASI); ribosomal protein L17; hypothetical gene supported by X60212	<i>Rattus norvegicus</i>
Q767K8	Similar to mitochondrial ribosomal protein S18-2	<i>Sus scrofa</i>
Q767K8	Similar to mitochondrial ribosomal protein S18-2	<i>Sus scrofa</i>
P62278	Similar to ribosomal protein S13; ribosomal protein S13	<i>Rattus norvegicus</i>
P23403	Similar to ribosomal protein S20	<i>Xenopus laevis</i>

Table S3.6 Annotations of the up regulated *A. digitifera* clusters involved in mitochondria functions in *Chromera*- infected larvae at 48 h post infection with corrected $P \leq 0.05$

ID	Protein Name	Species
Q6NTS3	28S ribosomal protein S24-B, mitochondrial	<i>Xenopus laevis</i>
O14464	54S ribosomal protein YPL183W-A, mitochondrial	<i>Saccharomyces cerevisiae</i>
Q8VZF6	AT5G45560	<i>Arabidopsis thaliana</i>
O75964	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit G	<i>Homo sapiens</i>
P22027	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit s (factor B)	<i>Bos taurus</i>
Q00709	B-cell CLL/lymphoma 2	<i>Gallus gallus</i>
P0C7P0	CDGSH iron sulfur domain 3	<i>Homo sapiens</i>
A0JNC1	CDP-diacylglycerol synthase (phosphatidate cytidyltransferase) 2	<i>Bos taurus</i>
P06197	CDP-diacylglycerol--inositol 3-phosphatidyltransferase	<i>Saccharomyces cerevisiae</i>
Q3SZM6	COX assembly mitochondrial protein homolog (S. cerevisiae)	<i>Bos taurus</i>
Q2NKS2	COX16 cytochrome c oxidase assembly homolog (S. cerevisiae)	<i>Bos taurus</i>
Q6J3Q7	COX17 homolog, cytochrome c oxidase assembly protein (yeast)	<i>Canis lupus</i>
Q8VC74	COX18 cytochrome c oxidase assembly homolog (S. cerevisiae)	<i>Mus musculus</i>
Q5FVL2	COX4 neighbor	<i>Rattus norvegicus</i>
B0XK69	CpipJ_CPIJ019830	<i>Culex quinquefasciatus</i>
Q6QLW4	Cytochrome c	<i>Pectinaria gouldii</i>
Q3ZBN8	DnaJ (Hsp40) homolog, subfamily C, member 19	<i>Bos taurus</i>
P37193	Ferredoxin	<i>Drosophila melanogaster</i>
Q5RDW1	GTP binding protein 5 (putative)	<i>Pongo abelii</i>
Q9N121	H protein	<i>Oryctolagus cuniculus</i>
Q9D7P6	IscU iron-sulfur cluster scaffold homolog (E. coli); similar to nitrogen fixation cluster-like	<i>Mus musculus</i>
O15091	KIAA0391	<i>Homo sapiens</i>
Q8K215	LYR motif containing 4	<i>Mus musculus</i>
Q91V16	LYR motif containing 5	<i>Mus musculus</i>
Q6GR66	MGC78819 protein	<i>Xenopus laevis</i>
Q6DFN1	MGC79777 protein	<i>Xenopus (Silurana) tropicalis</i>
Q68EV6	MGC84279 protein	<i>Xenopus laevis</i>
Q6DDX7	MGC84796 protein	<i>Xenopus laevis</i>
Q66L32	MGC85218 protein	<i>Xenopus laevis</i>
Q4P4Y2	Mitochondrial genome maintenance protein MGM101	<i>Ustilago maydis</i>
Q9Y1A3	Mitochondrial import inner membrane translocase subunit Tim8	<i>Drosophila melanogaster</i>
Q39056	Molybdopterin biosynthesis protein CNX3	<i>Arabidopsis thaliana</i>
Q2KIN6	MpV17 mitochondrial inner membrane protein	<i>Bos taurus</i>
Q95KV7	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 13	<i>Bos taurus</i>
O43678	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 2, 8kDa	<i>Homo sapiens</i>
Q6PBH5	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4; hypothetical LOC799961	<i>Danio rerio</i>
P23935	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5, 13kDa	<i>Bos taurus</i>
Q05752	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 7, 14.5kDa	<i>Bos taurus</i>

Q0MQB1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8, 19kDa	<i>Pan troglodytes</i>
Q5BK63	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9	<i>Rattus norvegicus</i>
Q02365	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa	<i>Bos taurus</i>
Q0MQF0	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9, 22kDa	<i>Pan troglodytes</i>
P23709	NADH dehydrogenase (ubiquinone) Fe-S protein 3, 30kDa (NADH-coenzyme Q reductase)	<i>Bos taurus</i>
Q5XIF3	NADH dehydrogenase (ubiquinone) Fe-S protein 4	<i>Rattus norvegicus</i>
P52504	NADH dehydrogenase (ubiquinone) Fe-S protein 6	<i>Rattus norvegicus</i>
Q9M9B4	NADH dehydrogenase [ubiquinone] iron-sulfur protein 6	<i>Arabidopsis thaliana</i>
Q86UD5	Na ⁺ /H ⁺ exchanger domain containing 2	<i>Homo sapiens</i>
Q24439	Oligomycin sensitivity-conferring protein	<i>Drosophila melanogaster</i>
Q9DAK2	PARK2 co-regulated	<i>Mus musculus</i>
Q90673	PRELI domain containing 1	<i>Gallus gallus</i>
Q86BN8	PTEN-like phosphatase	<i>Drosophila melanogaster</i>
Q8UW59	Parkinson disease (autosomal recessive, early onset) 7	<i>Gallus gallus</i>
Q9VAI1	Probable complex I intermediate-associated protein 30, mitochondrial	<i>Drosophila melanogaster</i>
A7S1H9	Protein ACN9 homolog, mitochondrial	<i>Nematostella vectensis</i>
Q66GV0	Protein Mpv17	<i>Xenopus laevis</i>
A8WGF7	Protein spinster homolog 1	<i>Xenopus (Silurana) tropicalis</i>
Q9CR10	RIKEN cDNA 1810049H13 gene	<i>Mus musculus</i>
Q8C1Q6	RIKEN cDNA 2010107H07 gene	<i>Mus musculus</i>
Q7TNS2	RIKEN cDNA 2310028O11 gene	<i>Mus musculus</i>
Q9CWB7	RIKEN cDNA C330018D20 gene	<i>Mus musculus</i>
Q5SUC9	SCO cytochrome oxidase deficient homolog 1 (yeast)	<i>Mus musculus</i>
P60924	SEL1 domain containing protein RGD735029	<i>Rattus norvegicus</i>
Q95KK4	Sjogren syndrome antigen B (autoantigen La)	<i>Oryctolagus cuniculus</i>
Q9P7Q5	Uncharacterized protein C1834.10c	<i>Schizosaccharomyces pombe</i>
Q9CYF5	Williams-Beuren syndrome chromosome region 16 homolog (human)	<i>Mus musculus</i>
Q5R833	Acyl-coa thioesterase 13	<i>Pongo abelii</i>
O95881	Apoptosis-inducing factor, mitochondrion-associated, 1	<i>Homo sapiens</i>
O95563	Brain protein 44	<i>Homo sapiens</i>
P63031	Brain protein 44-like; similar to brain protein 44-like	<i>Rattus norvegicus</i>
Q9Y259	Choline kinase beta; carnitine palmitoyltransferase 1B (muscle)	<i>Homo sapiens</i>
Q9NZJ6	Coenzyme Q3 homolog, methyltransferase (S. Cerevisiae)	<i>Homo sapiens</i>
Q63ZK1	Coiled-coil-helix-coiled-coil-helix domain containing 4	<i>Xenopus laevis</i>
Q8K2Q5	Coiled-coil-helix-coiled-coil-helix domain containing 7	<i>Mus musculus</i>
O35796	Complement component 1, q subcomponent binding protein	<i>Rattus norvegicus</i>
P50613	Cyclin-dependent kinase 7	<i>Homo sapiens</i>
P10606	Cytochrome c oxidase subunit Vb	<i>Homo sapiens</i>
P56391	Cytochrome c oxidase, subunit vib polypeptide 1	<i>Mus musculus</i>
Q8WNV7	Dehydrogenase/reductase (SDR family) member 4	<i>Sus scrofa</i>
Q6AY55	Dephospho-coa kinase domain containing	<i>Rattus norvegicus</i>
Q80Y81	Elac homolog 2 (E. Coli)	<i>Mus musculus</i>
P38117	Electron-transfer-flavoprotein, beta polypeptide	<i>Homo sapiens</i>
Q16595	Frataxin	<i>Homo sapiens</i>

Q6PBM1	Glutaredoxin 5 homolog (S. Cerevisiae)	<i>Danio rerio</i>
Q9Y2Q3	Glutathione S-transferase kappa 1	<i>Homo sapiens</i>
P28799	Granulin	<i>Homo sapiens</i>
Q9W6X3	Heat shock protein 10	<i>Oryzias latipes</i>
Q9WU63	Heme binding protein 2	<i>Mus musculus</i>
Q6P963	Hydroxyacylglutathione hydrolase	<i>Danio rerio</i>
O02691	Hydroxysteroid (17-beta) dehydrogenase 10	<i>Bos taurus</i>
Q5U4U5	Hypothetical LOC495431	<i>Xenopus laevis</i>
Q28ED6	Hypothetical LOC496604	<i>Xenopus (Silurana) tropicalis</i>
Q5M8Z2	Hypothetical LOC496649	<i>Xenopus (Silurana) tropicalis</i>
Q2M2S5	Hypothetical LOC768072	<i>Bos taurus</i>
Q8R035	Immature colon carcinoma cluster 1	<i>Mus musculus</i>
Q5ZJ74	Iron-sulfur cluster assembly 1 homolog (S. Cerevisiae)	<i>Gallus gallus</i>
Q9DCB8	Iron-sulfur cluster assembly 2 homolog (S. Cerevisiae)	<i>Mus musculus</i>
Q0VBY0	Mature T-cell proliferation 1	<i>Bos taurus</i>
Q8IVH4	Methylmalonic aciduria (cobalamin deficiency) cbla type	<i>Homo sapiens</i>
Q9D273	Methylmalonic aciduria (cobalamin deficiency) type B homolog (human)	<i>Mus musculus</i>
Q9D1I5	Methylmalonyl coa epimerase	<i>Mus musculus</i>
Q969V5	Mitochondrial E3 ubiquitin ligase 1	<i>Homo sapiens</i>
Q5HZI9	Mitochondrial carrier triple repeat 1	<i>Mus musculus</i>
Q9UDX5	Mitochondrial protein 18 kda	<i>Homo sapiens</i>
Q3TBW2	Mitochondrial ribosomal protein L10	<i>Mus musculus</i>
Q9D1P0	Mitochondrial ribosomal protein L13	<i>Mus musculus</i>
Q9D1I6	Mitochondrial ribosomal protein L14	<i>Mus musculus</i>
Q2TBI6	Mitochondrial ribosomal protein L32	<i>Bos taurus</i>
Q9CQP0	Mitochondrial ribosomal protein L33	<i>Mus musculus</i>
Q9DCU6	Mitochondrial ribosomal protein L4	<i>Mus musculus</i>
Q6DJI4	Mitochondrial ribosomal protein L41	<i>Xenopus laevis</i>
Q08DT6	Mitochondrial ribosomal protein L47	<i>Bos taurus</i>
Q9CQ40	Mitochondrial ribosomal protein L49; similar to mitochondrial ribosomal protein L49	<i>Mus musculus</i>
Q96EL3	Mitochondrial ribosomal protein L53	<i>Homo sapiens</i>
Q9VE04	Mitochondrial ribosomal protein L55	<i>Drosophila melanogaster</i>
Q9VFB2	Mitochondrial ribosomal protein S10	<i>Drosophila melanogaster</i>
O35680	Mitochondrial ribosomal protein S12	<i>Mus musculus</i>
Q9CR88	Mitochondrial ribosomal protein S14; similar to mitochondrial ribosomal protein S14	<i>Mus musculus</i>
Q9V6Y3	Mitochondrial ribosomal protein S16	<i>Drosophila melanogaster</i>
Q99N85	Mitochondrial ribosomal protein S18A	<i>Mus musculus</i>
Q767K8	Mitochondrial ribosomal protein S18B	<i>Sus scrofa</i>
Q8R2L5	Mitochondrial ribosomal protein S18C	<i>Mus musculus</i>
P82920	Mitochondrial ribosomal protein S21	<i>Bos taurus</i>
Q9D125	Mitochondrial ribosomal protein S25	<i>Mus musculus</i>
Q9CY16	Mitochondrial ribosomal protein S28	<i>Mus musculus</i>
Q9VZD5	Mitochondrial ribosomal protein S6	<i>Drosophila melanogaster</i>
Q9CR24	Nudix (nucleoside diphosphate linked moiety X)-type motif 8	<i>Mus musculus</i>

Q9HBH1	Peptide deformylase (mitochondrial); component of oligomeric golgi complex 8	<i>Homo sapiens</i>
Q9Y3E5	Peptidyl-trna hydrolase 2	<i>Homo sapiens</i>
Q07066	Peroxisomal membrane protein 2	<i>Rattus norvegicus</i>
Q61907	Phosphatidylethanolamine N-methyltransferase	<i>Mus musculus</i>
O35972	Predicted gene 13671; mitochondrial ribosomal protein L23	<i>Mus musculus</i>
Q9CQL5	Predicted gene 13675; mitochondrial ribosomal protein L18	<i>Mus musculus</i>
O55003	Predicted gene 14506; BCL2/adenovirus E1B interacting protein 3; predicted gene 6532; similar to E1B 19K/Bcl-2-binding protein homolog	<i>Mus musculus</i>
Q9D7J4	Predicted gene 15683; RIKEN cdna 2310005N03 gene; similar to RIKEN cdna 2310005N03	<i>Mus musculus</i>
Q9EP80	Protein interacting with PRKCA 1	<i>Rattus norvegicus</i>
Q15119	Pyruvate dehydrogenase kinase, isozyme 2	<i>Homo sapiens</i>
O46504	Pyruvate dehydrogenase phosphatase regulatory subunit	<i>Bos taurus</i>
P31399	Similar to ATP synthase D chain, mitochondrial; ATP synthase, H ⁺ transporting, mitochondrial F0 complex, subunit d	<i>Rattus norvegicus</i>
Q3MIE0	Similar to enoyl Coenzyme A hydratase domain containing 3	<i>Rattus norvegicus</i>
P21571	Similar to mitochondrial ATP synthase coupling factor 6; ATP synthase, H ⁺ transporting, mitochondrial F0 complex, subunit F6	<i>Rattus norvegicus</i>
Q767K8	Similar to mitochondrial ribosomal protein S18-2	<i>Sus scrofa</i>
Q767K8	Similar to mitochondrial ribosomal protein S18-2	<i>Sus scrofa</i>
Q9UBX3	Solute carrier family 25 (mitochondrial carrier; dicarboxylate transporter), member 10	<i>Homo sapiens</i>
Q08DK7	Solute carrier family 25, member 29	<i>Bos taurus</i>
Q8BGF9	Solute carrier family 25, member 44	<i>Mus musculus</i>
Q9Y6N5	Sulfide quinone reductase-like (yeast)	<i>Homo sapiens</i>
Q9WVJ4	Synaptojanin 2 binding protein	<i>Rattus norvegicus</i>
Q95108	Thioredoxin 2	<i>Bos taurus</i>
O95881	Thioredoxin domain containing 12 (endoplasmic reticulum)	<i>Homo sapiens</i>
Q32LD4	Transcription factor B2, mitochondrial	<i>Bos taurus</i>
P62074	Translocase of inner mitochondrial membrane 10 homolog (yeast)	<i>Rattus norvegicus</i>
Q9WV98	Translocase of inner mitochondrial membrane 9 homolog (yeast)	<i>Mus musculus</i>
Q5RA31	Translocase of outer mitochondrial membrane 20 homolog (yeast)	<i>Pongo abelii</i>
Q9P0U1	Translocase of outer mitochondrial membrane 7 homolog (yeast)	<i>Homo sapiens</i>
P50637	Translocator protein	<i>Mus musculus</i>
Q9CQN6	Transmembrane protein 14C	<i>Mus musculus</i>
P00129	Ubiquinol-cytochrome c reductase binding protein	<i>Bos taurus</i>
Q9UDW1	Ubiquinol-cytochrome c reductase complex (7.2 kd)	<i>Homo sapiens</i>
Q5ZLR5	Ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide-like 1	<i>Gallus gallus</i>
P40337	Von Hippel-Lindau tumor suppressor	<i>Homo sapiens</i>
Q499R4	Yrdc domain containing (E.coli)	<i>Rattus norvegicus</i>
A3KP37	Zgc:162919	<i>Danio rerio</i>
Q6DGJ3	Zgc:92895	<i>Danio rerio</i>
Q8BGC4	Zinc binding alcohol dehydrogenase, domain containing 2	<i>Mus musculus</i>
Q5ZJ74	Zinc finger, CCHC domain containing 6	<i>Gallus gallus</i>

Table S3.7 Annotations of the down regulated *A. digitifera* clusters involved in GTPase regulator activity in *Chromera*- infected larvae at 48 h post infection with corrected $P \leq 0.05$

ID	Protein Name	Species
Q12802	A kinase (PRKA) anchor protein 13	<i>Homo sapiens</i>
Q9FN03	At5g63860	<i>Arabidopsis thaliana</i>
P52594	Arfgap with FG repeats 1	<i>Homo sapiens</i>
Q4LDD4	Arfgap with rhogap domain, ankyrin repeat and PH domain 1	<i>Mus musculus</i>
O97902	Arfgap with SH3 domain, ankyrin repeat and PH domain 1	<i>Bos taurus</i>
Q96P50	Arfgap with coiled-coil, ankyrin repeat and PH domains 3	<i>Homo sapiens</i>
O54874	CDC42 binding protein kinase alpha	<i>Rattus norvegicus</i>
Q8TDJ6	Dmx-like 2	<i>Homo sapiens</i>
Q96EY1	Dnaj (Hsp40) homolog, subfamily A, member 3	<i>Homo sapiens</i>
Q8IUD2	ELKS/RAB6-interacting/CAST family member 1	<i>Homo sapiens</i>
Q91VS8	FERM, rhogef and pleckstrin domain protein 2	<i>Mus musculus</i>
P52734	FYVE, rhogef and PH domain containing 1	<i>Mus musculus</i>
Q96M96	FYVE, rhogef and PH domain containing 4	<i>Homo sapiens</i>
Q6ZV73	FYVE, rhogef and PH domain containing 6	<i>Homo sapiens</i>
Q9Y2X7	G protein-coupled receptor kinase interacting arfgap 1	<i>Homo sapiens</i>
P81274	G-protein signaling modulator 2 (AGS3-like, <i>C. Elegans</i>)	<i>Homo sapiens</i>
Q5VVW2	Gtpase activating Rap/rangap domain-like 3	<i>Homo sapiens</i>
P33277	Gtpase-activating protein	<i>Schizosaccharomyces pombe</i>
Q8R0S2	IQ motif and Sec7 domain 1	<i>Mus musculus</i>
P46940	IQ motif containing gtpase activating protein 1	<i>Homo sapiens</i>
Q86X10	Kiaa1219	<i>Homo sapiens</i>
Q5TH69	Kiaa1244	<i>Homo sapiens</i>
Q80U28	MAP-kinase activating death domain	<i>Mus musculus</i>
O15068	MCF.2 cell line derived transforming sequence-like	<i>Homo sapiens</i>
Q6GPD0	MGC80493 protein	<i>Xenopus laevis</i>
Q28C33	Novel protein containing TBC domain domain	<i>Xenopus tropicalis</i>
Q5ZJ17	RAB gtpase activating protein 1-like	<i>Gallus gallus</i>
O75154	RAB11 family interacting protein 3 (class II)	<i>Homo sapiens</i>
Q15042	RAB3 gtpase activating protein subunit 1 (catalytic)	<i>Homo sapiens</i>
Q8BMG7	RAB3 gtpase activating protein subunit 2	<i>Mus musculus</i>
Q9H2M9	RAB3 gtpase activating protein subunit 2 (non-catalytic)	<i>Homo sapiens</i>
Q62739	RAB3A interacting protein (rabin3)	<i>Rattus norvegicus</i>
P47736	RAP1 gtpase activating protein	<i>Homo sapiens</i>
Q8IV61	RAS guanyl releasing protein 3 (calcium and DAG-regulated)	<i>Homo sapiens</i>
P28818	RAS protein-specific guanine nucleotide-releasing factor 1	<i>Rattus norvegicus</i>
A1IGU4	RIKEN cdna 4933429F08 gene	<i>Mus musculus</i>
Q8BQZ4	RIKEN cdna B230339M05 gene	<i>Mus musculus</i>
Q6QI06	RPTOR independent companion of MTOR, complex 2	<i>Mus musculus</i>
Q13905	Rap guanine nucleotide exchange factor (GEF) 1	<i>Homo sapiens</i>
Q8TEU7	Rap guanine nucleotide exchange factor (GEF) 6	<i>Homo sapiens</i>
Q8JZL7	Rasgef domain family, member 1B; hypothetical protein LOC100044232	<i>Mus musculus</i>

Q8IWW6	Rho gtpase activating protein 12	<i>Homo sapiens</i>
Q8K0Q5	Rho gtpase activating protein 18	<i>Mus musculus</i>
Q6REY9	Rho gtpase activating protein 20	<i>Rattus norvegicus</i>
Q5T5U3	Rho gtpase activating protein 21	<i>Homo sapiens</i>
Q7Z5H3	Rho gtpase activating protein 22	<i>Homo sapiens</i>
Q8N264	Rho gtpase activating protein 24	<i>Homo sapiens</i>
Q9UNA1	Rho gtpase activating protein 26	<i>Homo sapiens</i>
P97393	Rho gtpase activating protein 5	<i>Mus musculus</i>
O43182	Rho gtpase activating protein 6	<i>Homo sapiens</i>
Q9V9S7	Rho gtpase-activating protein 100F	<i>Drosophila melanogaster</i>
Q8C033	Rho guanine nucleotide exchange factor (GEF) 10	<i>Mus musculus</i>
O15085	Rho guanine nucleotide exchange factor (GEF) 11	<i>Homo sapiens</i>
Q96PE2	Rho guanine nucleotide exchange factor (GEF) 17	<i>Homo sapiens</i>
Q8N5H7	SH2 domain containing 3C	<i>Homo sapiens</i>
Q13009	T-cell lymphoma invasion and metastasis 1	<i>Homo sapiens</i>
Q5F361	TBC domain-containing protein kinase-like	<i>Gallus gallus</i>
Q86TI0	TBC1 (tre-2/USP6, BUB2, cdc16) domain family, member 1	<i>Homo sapiens</i>
Q8C9V1	TBC1 domain family, member 10c	<i>Mus musculus</i>
A6H7I8	TBC1 domain family, member 14	<i>Bos taurus</i>
Q9D9I4	TBC1 domain family, member 20	<i>Mus musculus</i>
Q9NU19	TBC1 domain family, member 22B	<i>Homo sapiens</i>
Q9UPU7	TBC1 domain family, member 2B	<i>Homo sapiens</i>
Q92609	TBC1 domain family, member 5	<i>Homo sapiens</i>
P83510	TRAF2 and NCK interacting kinase	<i>Mus musculus</i>
Q92738	USP6 N-terminal like	<i>Homo sapiens</i>
Q6NXY1	WD repeat domain 67	<i>Mus musculus</i>
P42768	Wiskott-Aldrich syndrome (eczema-thrombocytopenia)	<i>Homo sapiens</i>
A6QNS3	Active BCR-related gene	<i>Bos taurus</i>
Q96Q42	Amyotrophic lateral sclerosis 2 (juvenile)	<i>Homo sapiens</i>
P0C5Y8	Amyotrophic lateral sclerosis 2 (juvenile) homolog (human)	<i>Rattus norvegicus</i>
Q920R0	Amyotrophic lateral sclerosis 2 (juvenile) homolog (human)	<i>Mus musculus</i>
Q3UMR0	Ankyrin repeat domain 27 (VPS9 domain)	<i>Mus musculus</i>
P11274	Breakpoint cluster region	<i>Homo sapiens</i>
Q2PPJ7	Chromosome 20 open reading frame 74	<i>Homo sapiens</i>
O14578	Citron (rho-interacting, serine/threonine kinase 21)	<i>Homo sapiens</i>
Q14185	Dedicator of cytokinesis 1	<i>Homo sapiens</i>
Q8BUR4	Dedicator of cytokinesis 1	<i>Mus musculus</i>
Q8R1A4	Dedicator of cytokinesis 7	<i>Mus musculus</i>
Q8BIK4	Dedicator of cytokinesis 9	<i>Mus musculus</i>
Q9BZ29	Dedicator of cytokinesis 9	<i>Homo sapiens</i>
Q6XZF7	Dynamin binding protein	<i>Homo sapiens</i>
Q8CHW4	Eukaryotic translation initiation factor 2B, subunit 5 epsilon	<i>Mus musculus</i>
Q8TBA6	Golgi autoantigen, golgin subfamily a, 5	<i>Homo sapiens</i>
Q15751	Hect (homologous to the E6-AP (UBE3A) carboxyl terminus) domain and RCC1 (CHC1)-like domain (RLD) 1	<i>Homo sapiens</i>
O95714	Hect domain and RLD 2	<i>Homo sapiens</i>
Q92619	Histocompatibility (minor) HA-1	<i>Homo sapiens</i>

Q9PT60	Hypothetical protein MGC81374	<i>Xenopus laevis</i>
O00410	Importin 5	<i>Homo sapiens</i>
O95373	Importin 7	<i>Homo sapiens</i>
Q9NZM3	Intersectin 2	<i>Homo sapiens</i>
Q9Z0R6	Intersectin 2	<i>Mus musculus</i>
Q76NI1	Kinase non-catalytic C-lobe domain (KIND) containing 1	<i>Homo sapiens</i>
Q5S007	Leucine-rich repeat kinase 2	<i>Homo sapiens</i>
Q8BPM2	Mitogen-activated protein kinase kinase kinase kinase 5	<i>Mus musculus</i>
B2RTY4	Myosin IXA	<i>Homo sapiens</i>
Q13459	Myosin IXB	<i>Homo sapiens</i>
P70569	Myosin Vb	<i>Rattus norvegicus</i>
P97526	Neurofibromin 1	<i>Rattus norvegicus</i>
Q5BKC9	Neuronal guanine nucleotide exchange factor	<i>Rattus norvegicus</i>
Q9ULL1	Pleckstrin homology domain containing, family G (with rhogef domain) member 1	<i>Homo sapiens</i>
Q58EX7	Pleckstrin homology domain containing, family G (with rhogef domain) member 4	<i>Homo sapiens</i>
O94827	Pleckstrin homology domain containing, family G (with rhogef domain) member 5	<i>Homo sapiens</i>
Q60695	Ral guanine nucleotide dissociation stimulator,-like 1	<i>Mus musculus</i>
Q9JIS1	Regulating synaptic membrane exocytosis 2	<i>Rattus norvegicus</i>
Q8CGE9	Regulator of G-protein signaling 12	<i>Mus musculus</i>
Q92834	Retinitis pigmentosa gtpase regulator	<i>Homo sapiens</i>
Q5RHR6	Si:dkey-233p4.1	<i>Danio rerio</i>
P69735	Similar to RAB3 gtpase-activating protein	<i>Rattus norvegicus</i>
O43147	Small G protein signaling modulator 2	<i>Homo sapiens</i>
Q96HU1	Small G protein signaling modulator 3	<i>Homo sapiens</i>
Q07889	Son of sevenless homolog 1 (Drosophila)	<i>Homo sapiens</i>
Q9R0X5	Sushi-repeat-containing protein; retinitis pigmentosa gtpase regulator	<i>Mus musculus</i>
Q96C24	Synaptotagmin-like 4	<i>Homo sapiens</i>
Q8WZ42	Titin	<i>Homo sapiens</i>
O75962	Triple functional domain (PTPRF interacting)	<i>Homo sapiens</i>
Q0KL02	Triple functional domain (PTPRF interacting)	<i>Mus musculus</i>
P49815	Tuberous sclerosis 2	<i>Homo sapiens</i>
Q07912	Tyrosine kinase, non-receptor, 2	<i>Homo sapiens</i>
Q8R5L3	Vacuolar protein sorting 39 (yeast)	<i>Mus musculus</i>
Q60992	Vav 2 oncogene	<i>Mus musculus</i>
Q9UKW4	Vav 3 guanine nucleotide exchange factor	<i>Homo sapiens</i>

Table S3.8 Differential expression of *A. digitifera* clusters likely involved in endosomal trafficking in *Chromera*-infected larvae at 48 h post infection with corrected $P \leq 0.05$. Column corresponds to coral cluster ID, annotated protein ID and name, species, E-value and the \log_2 fold change values

Cluster ID	UniProt ID	Protein name	E value	logFC
adi_EST_assem_16252	P59015	Vacuolar protein sorting protein 18 <i>Danio rerio</i>	3.53E-87	-5.11
adi_EST_assem_4366	Q91W86	Vacuolar protein sorting 11 (yeast) <i>Mus musculus</i>	0	-2.26
adi_EST_assem_15180	Q920Q4	Vacuolar protein sorting 16 (yeast) <i>Mus musculus</i>	0	-1.94
adi_EST_assem_11058	Q8R5L3	Vacuolar protein sorting 39 (yeast) <i>Mus musculus</i>	0	-1.86
adi_EST_assem_3691	Q8R0H9	Golgi associated, gamma adaptin ear containing, ARF binding protein 1 <i>Mus musculus</i>	1.86E-87	-1.33
adi_EST_assem_7197	A2RSQ0	DENN/MADD domain containing 5B <i>Mus musculus</i>	0	-1.53
adi_EST_assem_2060	Q5ZJ17	RAB gtpase activating protein 1-like <i>Gallus gallus</i>	1.14E-169	-2.04
adi_EST_assem_8022	O75154	RAB11 family interacting protein 3 (class II) <i>Homo sapiens</i>	9.14E-29	-2.14
adi_EST_assem_12934	Q15042	RAB3 gtpase activating protein subunit 1 (catalytic) <i>Homo sapiens</i>	5.64E-164	-1.99
adi_EST_assem_528	Q8BMG7	RAB3 gtpase activating protein subunit 2 <i>Mus musculus</i>	1.47E-74	-1.96
adi_EST_assem_13048	Q9H2M9	RAB3 gtpase activating protein subunit 2 (non-catalytic) <i>Homo sapiens</i>	0	-3.46
adi_EST_assem_3833	Q62739	RAB3A interacting protein (rabin3) <i>Rattus norvegicus</i>	3.67E-64	-2.24
adi_EST_assem_948	P47736	RAP1 gtpase activating protein <i>Homo sapiens</i>	8.98E-113	-1.22
adi_EST_assem_9191	Q8IV61	RAS guanyl releasing protein 3 (calcium and DAG-regulated) <i>Homo sapiens</i>	3.81E-134	-1.36
adi_EST_assem_6077	P28818	RAS protein-specific guanine nucleotide-releasing factor 1 <i>Rattus norvegicus</i>	4.11E-15	-2.89
adi_EST_assem_22365	Q5F361	TBC domain-containing protein kinase-like <i>Gallus gallus</i>	0	-1.05
adi_EST_assem_1683	Q86TI0	TBC1 (tre-2/USP6, BUB2, cdc16) domain family, member 1 <i>Homo sapiens</i>	1.57E-117	-2.28
adi_EST_assem_23011	Q8C9V1	TBC1 domain family, member 10c <i>Mus musculus</i>	2.71E-34	-2.66
adi_EST_assem_1755	A6H7I8	TBC1 domain family, member 14 <i>Bos taurus</i>	1.45E-169	-1.11
adi_EST_assem_5379	Q9D9I4	TBC1 domain family, member 20 <i>Mus musculus</i>	1.63E-99	-1.45
adi_EST_assem_4543	Q9NU19	TBC1 domain family, member 22B <i>Homo sapiens</i>	1.38E-162	-1.25
adi_EST_assem_20484	Q9UPU7	TBC1 domain family, member 2B <i>Homo sapiens</i>	2.50E-37	-2.25
adi_EST_assem_26789	Q0VCJ7	RAS-like, estrogen-regulated, growth inhibitor <i>Bos taurus</i> (<i>Bos taurus</i>)	1.41E-15	2.04
adi_EST_assem_1935	P22125	Ras-related protein ORAB-1 <i>Discopyge ommata</i> (<i>Discopyge ommata</i>)	1.86E-89	1.89
adi_EST_assem_1446	Q05975	Ras-related protein Rab-2 <i>Lymnaea stagnalis</i> (<i>Lymnaea stagnalis</i>)	3.00E-136	1.62
adi_EST_assem_1323	Q9UII4	Rab acceptor 1 (prenylated) <i>Homo sapiens</i> (<i>Homo sapiens</i>)	5.37E-29	1.48
adi_EST_assem_1748	Q6DHC1	RAB18B, member RAS oncogene family <i>Danio rerio</i> (<i>Danio rerio</i>)	1.19E-108	1.28
adi_EST_assem_17049	Q99P75	RAB9A, member RAS oncogene family <i>Rattus norvegicus</i> (<i>Rattus norvegicus</i>)	1.34E-69	1.23

adi_EST_assem_11824	O95755	RAB36, member RAS oncogene family Homo sapiens (Homo sapiens)	7.16E-79	1.08
adi_EST_assem_12773	Q5RAV6	RAB6A, member RAS oncogene family Pongo abelii Pongo abelii	2.03E-123	1.05
adi_EST_assem_452	Q5R5U1	RAB10, member RAS oncogene family Pongo abelii Pongo abelii	2.79E-106	1.00
adi_EST_assem_6154	A4IHM6	Rab-like protein 3 Xenopus (Silurana) tropicalis	6.18E-69	1.81
adi_EST_assem_12084	Q5RFI2	RAB28, member RAS oncogene family Pongo abelii	6.18E-93	1.76
adi_EST_assem_20522	Q5M7D1	Rab and dnaj domain-containing protein B Xenopus laevis	1.70E-26	1.71
adi_EST_assem_11749	Q6IMK3	Rab and dnaj domain-containing protein Danio rerio	1.60E-07	1.35
adi_EST_assem_22058	Q0VCN3	RAB, member of RAS oncogene family- like 4 Bos taurus	2.16E-66	1.30
adi_EST_assem_4227	P25228	Rab-protein 3 Drosophila melanogaster	3.70E-131	1.13

Table S3.9 Differential expression of *A. digitifera* clusters likely involved in regulation of apoptosis in *Chromera*-infected larvae at 48 h post infection with corrected $P \leq 0.05$. Column corresponds to coral cluster ID, annotated protein ID and name, species, E-value and the \log_2 fold change values

Cluster ID	Protein ID	Protein Name	Species	E-value	logFC
adi_EST_assem_1890	Q9P289	Serine/threonine protein kinase MST4	<i>Homo sapiens</i>	3.54E-149	-1.27
adi_EST_assem_5478	Q923E4	Sirtuin 1 (silent mating type information regulation 2, homolog) 1 (<i>S. Cerevisiae</i>)	<i>Mus musculus</i>	8.98E-150	-1.84
adi_EST_assem_8857	Q0IHU9	Hypothetical protein MGC145921	<i>Xenopus (Silurana) tropicalis</i>	1.28E-18	-1.60
adi_EST_assem_587	Q5F499	Optic atrophy 1 (autosomal dominant)	<i>Gallus gallus</i>	0	-1.96
adi_EST_assem_1205	Q6NS46	Programmed cell death 11	<i>Mus musculus</i>	0	-1.93
adi_EST_assem_2986	Q8WUM4	Programmed cell death 6 interacting protein	<i>Homo sapiens</i>	0	-1.20
adi_EST_assem_4472	P21127	Similar to cell division cycle 2-like 1 (PITSLRE proteins); cell division cycle 2-like 1 (PITSLRE proteins); cell division cycle 2-like 2 (PITSLRE proteins)	<i>Homo sapiens</i>	0	-1.43
adi_EST_assem_4761	P05625	V-raf-1 murine leukemia viral oncogene homolog 1	<i>Gallus gallus</i>	8.16E-177	-1.54
adi_EST_assem_11971	Q9ESK9	RB1-inducible coiled-coil 1	<i>Mus musculus</i>	5.39E-49	-2.48
adi_EST_assem_1233	Q8N201	Integrator complex subunit 1	<i>Homo sapiens</i>	6.27E-116	-3.67
adi_EST_assem_1941	Q6P4S8	Integrator complex subunit 1	<i>Mus musculus</i>	2.83E-88	-1.52
adi_EST_assem_4220	P51111	Huntingtin	<i>Rattus norvegicus</i>	0	-1.90
adi_EST_assem_3436	P00519	C-abl oncogene 1, receptor tyrosine kinase	<i>Homo sapiens</i>	0	-1.76
adi_EST_assem_469	Q5ZIU3	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2	<i>Gallus gallus</i>	1.09E-94	-2.02
adi_EST_assem_6133	P05696	Protein kinase C, alpha	<i>Rattus norvegicus</i>	0	-2.47
adi_EST_assem_4877	P09215	Protein kinase C, delta	<i>Rattus norvegicus</i>	0	-1.35
adi_EST_assem_8760	Q02156	Protein kinase C, epsilon	<i>Homo sapiens</i>	0	-1.51
adi_EST_assem_1629	Q5PQS4	GULP, engulfment adaptor PTB domain containing 1	<i>Rattus norvegicus</i>	2.38E-49	-1.30
adi_EST_assem_12712	Q14185	Dedicator of cytokinesis 1	<i>Homo sapiens</i>	0	-2.17
adi_EST_assem_7899	Q8BUR4	Dedicator of cytokinesis 1	<i>Mus musculus</i>	0	-2.84
adi_EST_assem_4797	Q8WWQ8	Stabilin 2	<i>Homo sapiens</i>	9.75E-145	-2.02
adi_EST_assem_18171	Q13114	TNF receptor-associated factor 3	<i>Homo sapiens</i>	1.97E-04	1.6
adi_EST_assem_24012	P70191	TNF receptor-associated factor 5	<i>Mus musculus</i>	5.61E-15	1.3
adi_EST_assem_22957	Q5JPI3	Uncharacterized protein c3orf38	<i>Homo sapiens</i>	6.81E-45	1.43
adi_EST_assem_4841	Q6DF07	Programmed cell death 10	<i>Xenopus tropicalis</i>	9.77E-44	1.47
adi_EST_assem_7217	Q2YDC9	Programmed cell death 2	<i>Bos taurus</i>	8.88E-39	1.3
adi_EST_assem_11984	Q8BKD6	E3 ubiquitin-protein ligase RNF144B	<i>Mus musculus</i>	1.92E-73	1.43

adi_EST_assem_3525	Q58CU4	Probable palmitoyltransferase ZDHHC16	<i>Bos taurus</i>	1.27E- 06	2.46
adi_EST_assem_22677	Q99J83	Autophagy-related 5 (yeast)	<i>Mus musculus</i>	1.07E- 116	1.76
adi_EST_assem_17635	Q8JHF0	Presenilin enhancer 2 homolog	<i>Danio rerio</i>	6.96E- 45	1.19

Appendix IV

Chapter 4.0 *Chromera* transcriptomics: Functional genomic resource for a chromerid alga isolated from *Montipora digitata* at the Great Barrier Reef and comparative transcriptomic analyses with parasitic and photosynthetic relatives. Supplementary Information

The starting *Chromera* culture was checked with inverted microscope in order to check for Protista and Bacteria contamination before small aliquots were subjected to genetic identification, growth and further application of the experimental treatments.

Genetic Identification

***Chromera* gDNA extraction**

gDNA was extracted from 50 ml of exponentially growing culture. Cultures were centrifuged at 9000 rpm for 5 minutes at 4 °C, the *Chromera* pellet was re-suspended in 1ml fresh f/2 medium, centrifuged at maximum speed for 5 minutes at 4°C and stored at -80 °C until further treatment. The ISOLATE II Plant DNA Kit (BIOLINE) was used for DNA extraction according to manufacturer instructions. DNA was eluted in 100 µl of elution buffer in 1.5 ml tube. DNA was checked by running onto agarose gel and Nanodrop ® ND-100 Spectrophotometer (Wilmington, U.S.A) was used to estimate the concentration and quality of the DNA obtained from the DNA extractions. Milli-Q water was used to blank the instrument. 1.5 µl of sample was placed directly onto a fibre optic measurement surface where a retention system using surface tension, held the sample in place. DNA concentrations, absorbance at 230 (λ_{230}) and the ratio 260/280 were recorded.

Amplification of *Chromera* ribosomal genes using polymerase chain reaction (PCR)

Amplification of *Chromera* ribosomal genes was undertaken using specific primers (Appendix 2) to obtain a PCR product ranging between 416 to 778 bp in size. PCR reaction was conducted in 50 µl reaction using 1 µl of *Chromera* gDNA (approx. 100ng of DNA) as template. 1 µl of GoTaq® DNA polymerase and 2X GoTaq® reaction buffer were used, 5 µl of each primer were used and finally sterile MQW was added to the reaction mixture to make a total volume of 50 µl. The PCR profile was one cycle for

2 min at 94 °C for initial denaturation followed by 34 cycles of 30 sec at 94 °C, annealing for 30 sec at 47 °C/ 51 °C and extension for 2 min at 72 °C. The final extension was at 72 °C for 10 min. The obtained amplicons were run on 1.5% Agarose gel and visualized using UV trans-illuminator (Figure S4.1).

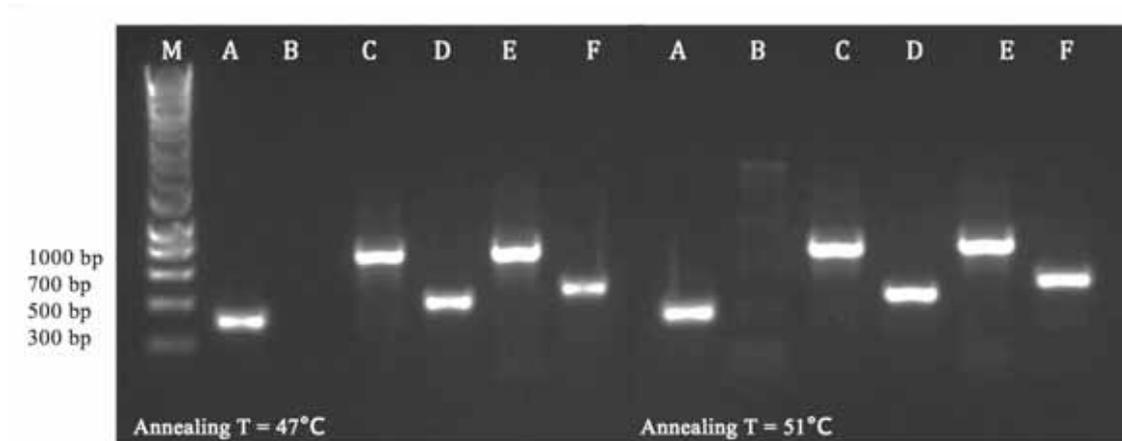


Figure S4.1 Gel photographs of successful amplification of *Chromera* ribosomal genes using specific PCR primers. M refers to the marker or DNA ladder. A refers to positive control reaction (*Symbiodinium* gDNA and *Symbiodinium*-specific primers), while B refers to negative control reaction (MQ water as a template). C, D, E and F are *Chromera* gDNA tested with the four primer pairs at two different annealing temperatures 47°C and 51°C.

6 libraries Illumina TruSeq Adapters

TruSeq Universal Adapter >>>> For R2 reads

5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT.

The reverse complementary will be used below:

AGATCGGAAGAGCGTCTGTAGGGAAAGAGTGTAGATCTCGGTGGTCCCGTATCATT

TruSeq Indexed Adapter, Index N >>> For R1 reads

5'GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNNATCTCGTATGCCGTCTTCTGCTTG

1- Chromera_cold-shock_C5GLGACXX_CTTGTA_L008_R1

5'GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTTGTAATCTCGTATGCCGTCTTCTGCTTG

2- Chromera_control_C5GLGACXX_CAGATC_L008_R1

5'GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCATCTCGTATGCCGTCTTCTGCTTG

3- Chromera_dark_C5GLGACXX_ACTTGA_L008_R1

5'GATCGGAAGAGCACACGTCTGAACTCCAGTCACACTTGAAATCTCGTATGCCGTCTTCTGCTTG

4- Chromera_heat-shock_C5GLGACXX_GGCTAC_L008_R1

5'GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGCTACATCTCGTATGCCGTCTTCTGCTTG

5- Chromera_mixotroph_C5GLGACXX_TAGCTT_L008_R1

5'GATCGGAAGAGCACACGTCTGAACTCCAGTCACTAGCTTATCTCGTATGCCGTCTTCTGCTTG

6- Chromera_motile_C5GLGACXX_GATCAG_L008_R1

5'GATCGGAAGAGCACACGTCTGAACTCCAGTCACGATCAGATCTCGTATGCCGTCTTCTGCTTG

The perl script "*A1_get_genes_from_trinity.pl*"

This raw data has been removed

A1_split_query_generate_blastx_sh.pl

This raw data has been removed

This raw data has been removed

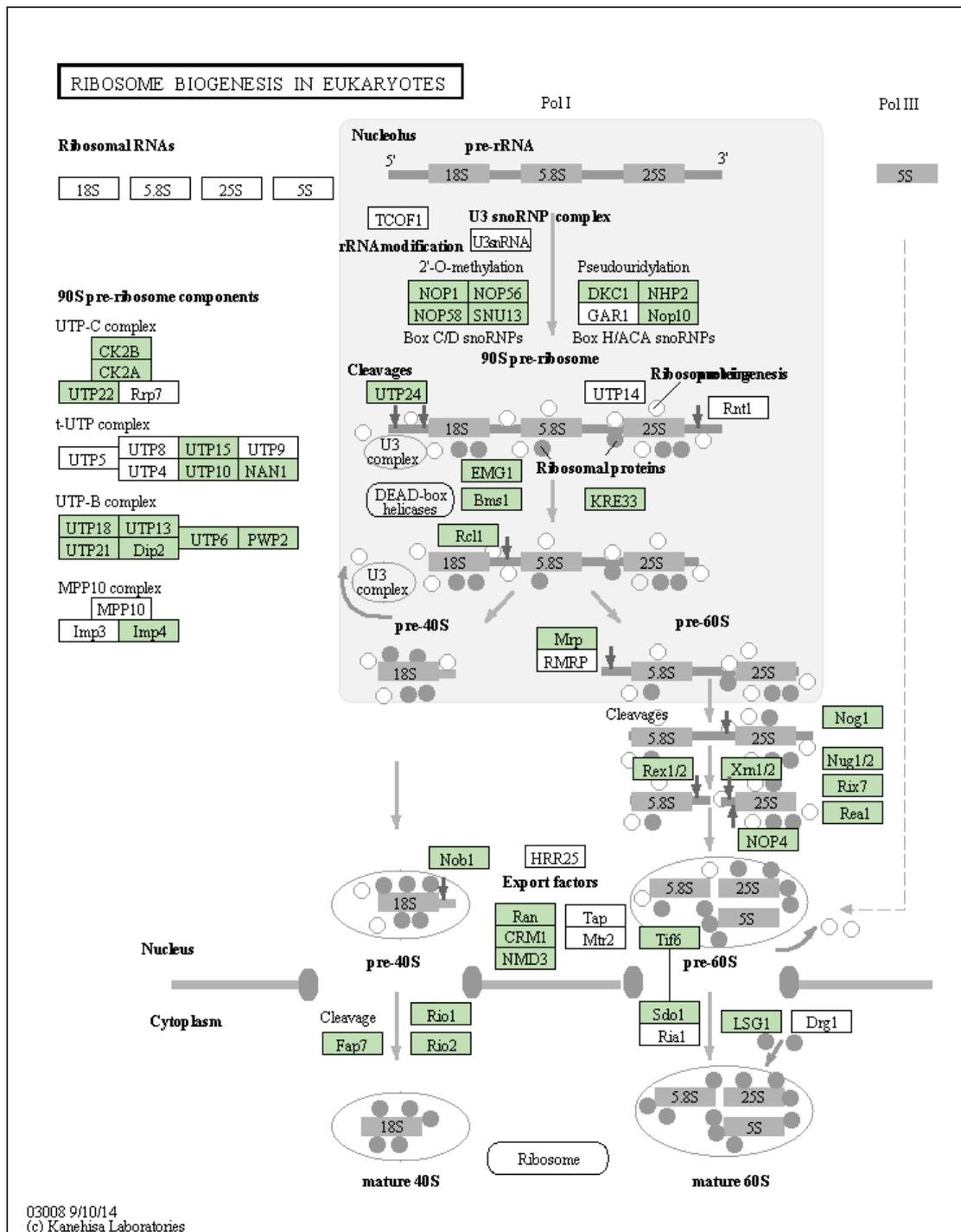


Figure S4.2 Ribosome biogenesis (eukaryotes) pathway (ko03008) identified in the *Chromera* transcriptome. KEGG pathways analysis shows *Chromera* orthologs involved in ribosome biogenesis (highlighted in green).

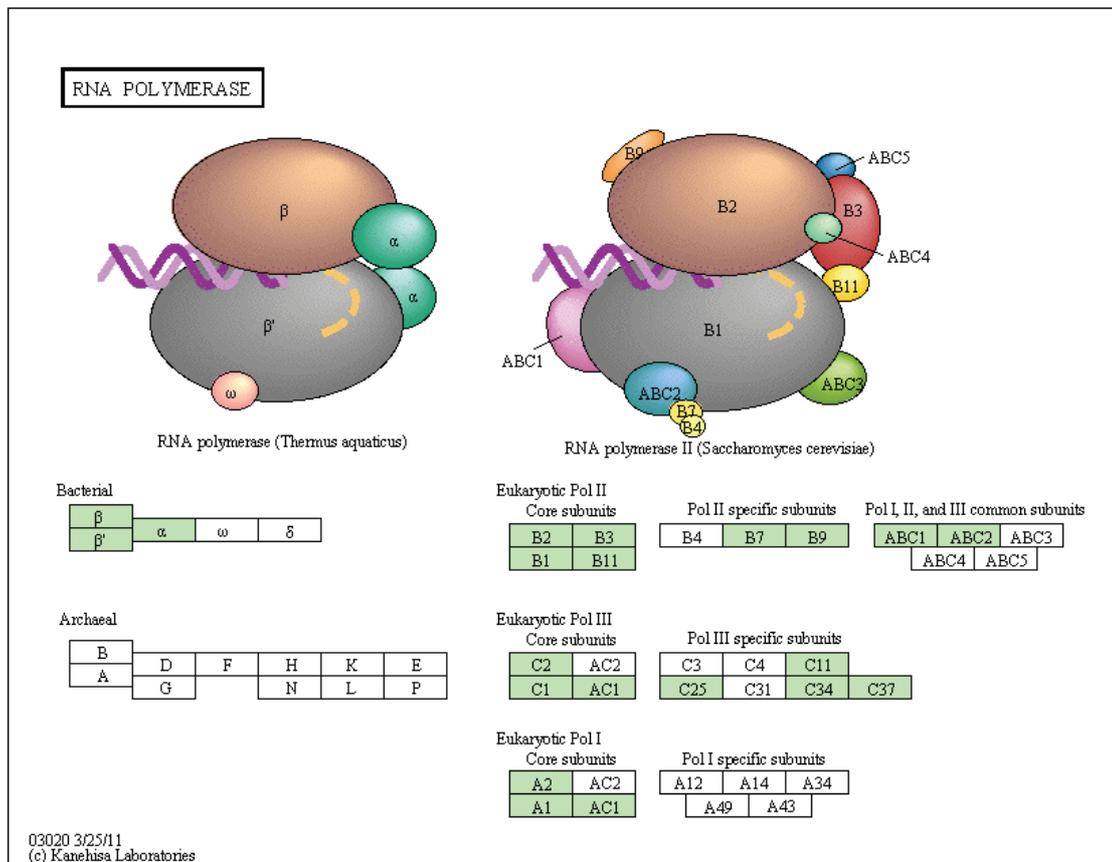


Figure S4.4 RNA polymerase pathway (ko03020) identified in the *Chromera* transcriptome. KEGG pathways analysis shows *Chromera* orthologs involved in RNA polymerase (highlighted in green).

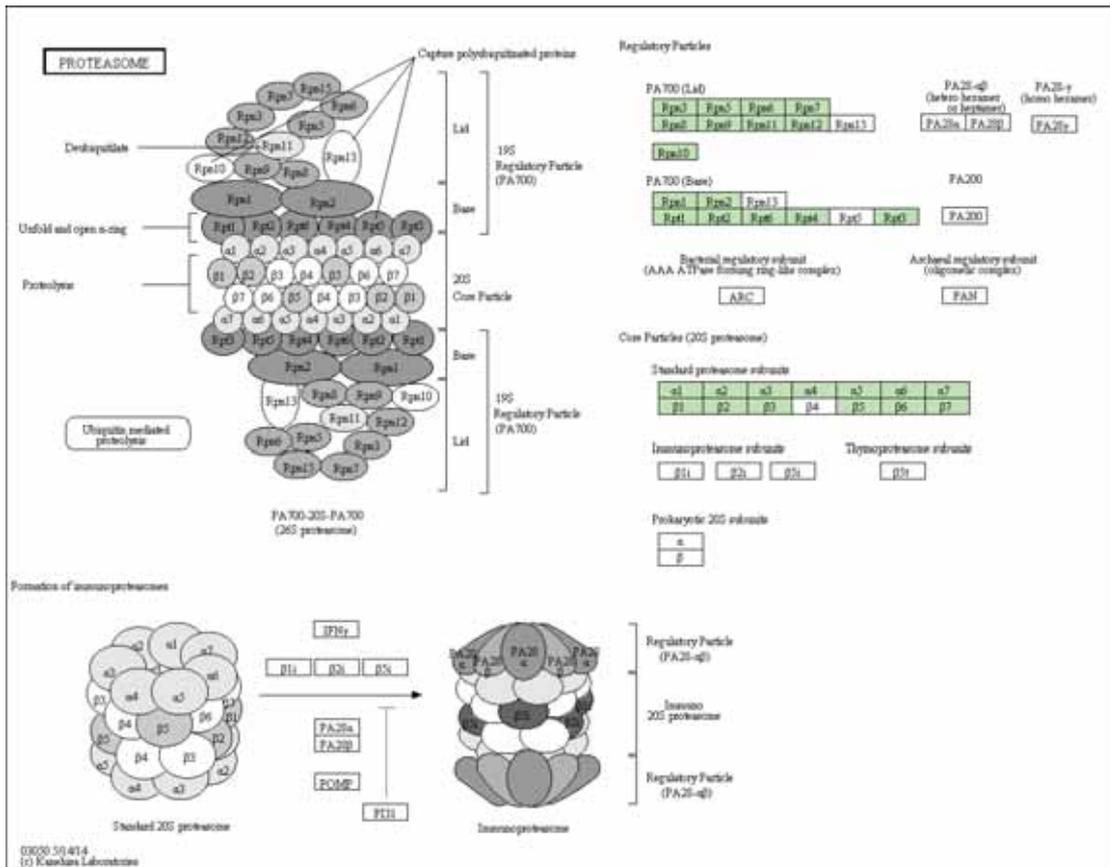


Figure S4.6 Proteasome pathway (hsa03050) identified in the *Chromera* transcriptome. KEGG pathways analysis shows *Chromera* orthologs involved in proteasome (highlighted in green).

Table 4S.1 GO categories (BP, CC, MF) enriched in the list of 2400 shared genes between the 2 *Chromera* with corrected *P*-value ≤ 0.05

Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_BP	GO:0006259~DNA metabolic process	22	3.437008929	0.001728089
	GO:0051188~cofactor biosynthetic process	14	4.814563679	0.003319904
	GO:0006974~response to DNA damage stimulus	15	4.339657738	0.003070993
	GO:0008610~lipid biosynthetic process	22	2.898682229	0.005904789
	GO:0006281~DNA repair	14	4.229368094	0.005343136
	GO:0051186~cofactor metabolic process	16	3.135752688	0.029817288
	GO:0009108~coenzyme biosynthetic process	9	5.46796875	0.031873711
	GO:0033554~cellular response to stress	20	2.567121479	0.03643285
GOTERM_CC	GO:0044434~chloroplast part	46	2.711821332	2.31E-07
	GO:0044435~plastid part	46	2.629143853	3.05E-07
	GO:0031090~organelle membrane	37	2.451013197	6.04E-05
	GO:0016021~integral to membrane	66	1.692610067	5.61E-04
	GO:0009941~chloroplast envelope	23	2.939860853	4.74E-04
	GO:0009507~chloroplast	86	1.515260089	6.89E-04
	GO:0009526~plastid envelope	23	2.805940945	6.99E-04
	GO:0009536~plastid	87	1.501365763	6.40E-04
	GO:0031967~organelle envelope	30	2.264730859	0.001246046
	GO:0031975~envelope	30	2.246637137	0.001287624
	GO:0031224~intrinsic to membrane	70	1.481135118	0.007132372
	GO:0010287~plastoglobule	7	6.561428571	0.010875563
	GO:0009579~thylakoid	19	2.176324868	0.043782152
	GO:0009570~chloroplast stroma	17	2.276413994	0.047672177
	GO:0044436~thylakoid part	15	2.383085438	0.055906218
GOTERM_MF	GO:0042623~ATPase activity, coupled	25	3.910475828	1.31E-05
	GO:0016887~ATPase activity	26	3.030576899	4.09E-04
	GO:0000166~nucleotide binding	93	1.518855562	0.001556927
	GO:0000287~magnesium ion binding	20	3.240108543	0.001767239
	GO:0030554~adenyl nucleotide binding	75	1.614927518	0.001437774
	GO:0001883~purine nucleoside binding	75	1.614927518	0.001437774
	GO:0001882~nucleoside binding	75	1.609833109	0.001340161
	GO:0015405~P-P-bond-hydrolysis-driven transmembrane transporter activity	16	3.674857285	0.00226523
	GO:0015399~primary active transmembrane transporter activity	16	3.659416708	0.002086886
	GO:0043492~ATPase activity, coupled to movement of substances	14	4.119316375	0.00203178
	GO:0042626~ATPase activity, coupled to transmembrane movement of substances	14	4.119316375	0.00203178
	GO:0048037~cofactor binding	25	2.592086835	0.001845515
	GO:0005524~ATP binding	69	1.605099925	0.002102995
	GO:0016820~hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	14	3.989913767	0.00212288
	GO:0032559~adenyl ribonucleotide binding	69	1.586120703	0.002592514
	GO:0017076~purine nucleotide binding	77	1.504992607	0.004303067

GO:0032555~purine ribonucleotide binding	71	1.474552259	0.015587752
GO:0032553~ribonucleotide binding	71	1.474552259	0.015587752
GO:0005506~iron ion binding	27	2.027190669	0.026085277
GO:0051119~sugar transmembrane transporter activity	9	4.494535888	0.025453114
GO:0008026~ATP-dependent helicase activity	9	4.335437272	0.030343231
GO:0070035~purine NTP-dependent helicase activity	9	4.335437272	0.030343231
GO:0031402~sodium ion binding	4	18.14460784	0.030603751
GO:0046872~metal ion binding	85	1.364859882	0.029579362
GO:0070279~vitamin B6 binding	9	4.116843796	0.036191912
GO:0030170~pyridoxal phosphate binding	9	4.116843796	0.036191912
GO:0016791~phosphatase activity	15	2.650997899	0.037078737
GO:0005351~sugar:hydrogen symporter activity	8	4.39869281	0.048077647
GO:0015295~solute:hydrogen symporter activity	8	4.39869281	0.048077647
GO:0005402~cation:sugar symporter activity	8	4.39869281	0.048077647

Table S4.2 Significant functional annotation clusters (FACs) for the set of genes shared between GBR and Sydney *Chromera* using DAVID high classification stringency and ease score ≤ 0.05

Annotation Cluster 1	Enrichment Score: 7.710068650768483			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	plastid	42	4.43329157	6.71E-13
SP_PIR_KEYWORDS	chloroplast	42	4.291198891	1.01E-12
SP_PIR_KEYWORDS	transit peptide	44	3.585401352	3.79E-11
GOTERM_CC_FAT	GO:0044434~chloroplast part	46	2.711821332	2.31E-07
GOTERM_CC_FAT	GO:0044435~plastid part	46	2.629143853	3.05E-07
UP_SEQ_FEATURE	transit peptide:Chloroplast	34	2.294400899	0.004622918
GOTERM_CC_FAT	GO:0009507~chloroplast	86	1.515260089	6.89E-04
GOTERM_CC_FAT	GO:0009536~plastid	87	1.501365763	6.40E-04
GOTERM_CC_FAT	GO:0009570~chloroplast stroma	17	2.276413994	0.047672177
GOTERM_CC_FAT	GO:0009532~plastid stroma	17	2.16310832	0.066722233
Annotation Cluster 2	Enrichment Score: 7.242385737738843			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	transport	46	3.375607652	6.76E-11
SP_PIR_KEYWORDS	membrane	70	2.383999384	3.66E-10
SP_PIR_KEYWORDS	transmembrane	63	2.544704867	3.39E-10
GOTERM_CC_FAT	GO:0016021~integral to membrane	66	1.692610067	5.61E-04
GOTERM_CC_FAT	GO:0031224~intrinsic to membrane	70	1.481135118	0.007132372
UP_SEQ_FEATURE	transmembrane region	53	1.320578255	0.612358665
Annotation Cluster 3	Enrichment Score: 5.907669590578661			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_CC_FAT	GO:0044434~chloroplast part	46	2.711821332	2.31E-07
GOTERM_CC_FAT	GO:0044435~plastid part	46	2.629143853	3.05E-07
GOTERM_CC_FAT	GO:0009941~chloroplast envelope	23	2.939860853	4.74E-04
GOTERM_CC_FAT	GO:0009526~plastid envelope	23	2.805940945	6.99E-04
GOTERM_CC_FAT	GO:0031967~organelle envelope	30	2.264730859	0.001246046
GOTERM_CC_FAT	GO:0031975~envelope	30	2.246637137	0.001287624
Annotation Cluster 4	Enrichment Score: 5.388316282649872			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	atp-binding	61	2.832934333	2.73E-11
SP_PIR_KEYWORDS	nucleotide-binding	62	2.553493028	4.00E-10
GOTERM_MF_FAT	GO:0000166~nucleotide binding	93	1.518855562	0.001556927
GOTERM_MF_FAT	GO:0001883~purine nucleoside binding	75	1.614927518	0.001437774
GOTERM_MF_FAT	GO:0030554~adenyl nucleotide binding	75	1.614927518	0.001437774
GOTERM_MF_FAT	GO:0001882~nucleoside binding	75	1.609833109	0.001340161
GOTERM_MF_FAT	GO:0005524~ATP binding	69	1.605099925	0.002102995
GOTERM_MF_FAT	GO:0032559~adenyl ribonucleotide binding	69	1.586120703	0.002592514
GOTERM_MF_FAT	GO:0017076~purine nucleotide binding	77	1.504992607	0.004303067
GOTERM_MF_FAT	GO:0032553~ribonucleotide binding	71	1.474552259	0.015587752

GOTERM_MF_FAT	GO:0032555~purine ribonucleotide binding	71	1.474552259	0.015587752
UP_SEQ_FEATURE	nucleotide phosphate-binding region:ATP	33	1.797667731	0.139723473
Annotation Cluster 5	Enrichment Score: 4.713579513491883			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_BP_FAT	GO:0006259~DNA metabolic process	22	3.437008929	0.001728089
GOTERM_BP_FAT	GO:0006974~response to DNA damage stimulus	15	4.339657738	0.003070993
GOTERM_BP_FAT	GO:0006281~DNA repair	14	4.229368094	0.005343136
GOTERM_BP_FAT	GO:0033554~cellular response to stress	20	2.567121479	0.03643285
Annotation Cluster 6	Enrichment Score: 3.7613006615383413			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_BP_FAT	GO:0051188~cofactor biosynthetic process	14	4.814563679	0.003319904
GOTERM_BP_FAT	GO:0051186~cofactor metabolic process	16	3.135752688	0.029817288
GOTERM_BP_FAT	GO:0009108~coenzyme biosynthetic process	9	5.46796875	0.031873711
GOTERM_BP_FAT	GO:0006732~coenzyme metabolic process	11	3.084495192	0.205501693
Annotation Cluster 7	Enrichment Score: 3.3027217839102843			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	iron	21	2.652246541	0.001601039
GOTERM_MF_FAT	GO:0005506~iron ion binding	27	2.027190669	0.026085277
UP_SEQ_FEATURE	metal ion-binding site:Iron	10	3.935421004	0.14560993
Annotation Cluster 8	Enrichment Score: 2.948755516830831			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	pyridoxal phosphate	9	8.538610039	1.26E-04
GOTERM_MF_FAT	GO:0048037~cofactor binding	25	2.592086835	0.001845515
INTERPRO	IPR015421:Pyridoxal phosphate-dependent transferase, major region, subdomain 1	9	6.355536771	0.04988529
GOTERM_MF_FAT	GO:0030170~pyridoxal phosphate binding	9	4.116843796	0.036191912
GOTERM_MF_FAT	GO:0070279~vitamin B6 binding	9	4.116843796	0.036191912
GOTERM_MF_FAT	GO:0019842~vitamin binding	10	3.259510391	0.065702414
SP_PIR_KEYWORDS	Aminotransferase	4	7.58987559	0.074009339
GOTERM_MF_FAT	GO:0008483~transaminase activity	5	5.040168845	0.216883552
GOTERM_MF_FAT	GO:0016769~transferase activity, transferring nitrogenous groups	5	3.944479966	0.37323051
Annotation Cluster 9	Enrichment Score: 2.7677576422768615			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	metal-binding	57	2.273543814	2.01E-07
GOTERM_MF_FAT	GO:0046872~metal ion binding	85	1.364859882	0.029579362
GOTERM_MF_FAT	GO:0043169~cation binding	86	1.313498548	0.068123516
GOTERM_MF_FAT	GO:0043167~ion binding	86	1.310556782	0.065012176
SP_PIR_KEYWORDS	zinc	25	1.772546568	0.047051341
GOTERM_MF_FAT	GO:0046914~transition metal ion binding	62	1.224563519	0.456046163
GOTERM_MF_FAT	GO:0008270~zinc ion binding	31	1.020222811	0.985112221

Annotation Cluster 10	Enrichment Score: 2.6713710417606684			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	Acyltransferase	11	6.306695184	1.33E-04
INTERPRO	IPR001594:Zinc finger, DHHC-type	4	10.71025641	0.27430841
UP_SEQ_FEATURE	zinc finger region:DHHC-type	4	7.018167457	0.604522437
UP_SEQ_FEATURE	active site:S-palmitoyl cysteine intermediate	4	7.018167457	0.604522437
Annotation Cluster 11	Enrichment Score: 2.5873821833661625			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_MF_FAT	GO:0042623~ATPase activity, coupled	25	3.910475828	1.31E-05
SP_PIR_KEYWORDS	magnesium	19	4.822229873	1.74E-06
GOTERM_MF_FAT	GO:0016887~ATPase activity	26	3.030576899	4.09E-04
GOTERM_MF_FAT	GO:0000287~magnesium ion binding	20	3.240108543	0.001767239
GOTERM_MF_FAT	GO:0015405~P-P-bond-hydrolysis-driven transmembrane transporter activity	16	3.674857285	0.00226523
GOTERM_MF_FAT	GO:0015399~primary active transmembrane transporter activity	16	3.659416708	0.002086886
GOTERM_MF_FAT	GO:0042626~ATPase activity, coupled to transmembrane movement of substances	14	4.119316375	0.00203178
GOTERM_MF_FAT	GO:0043492~ATPase activity, coupled to movement of substances	14	4.119316375	0.00203178
GOTERM_MF_FAT	GO:0016820~hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	14	3.989913767	0.00212288
INTERPRO	IPR005834:Haloacid dehalogenase-like hydrolase	8	7.560180995	0.02577847
INTERPRO	IPR018303:ATPase, P-type phosphorylation site	6	8.203600655	0.144642582
INTERPRO	IPR008250:ATPase, P-type, ATPase-associated region	6	8.203600655	0.144642582
INTERPRO	IPR001757:ATPase, P-type, K/Mg/Cd/Cu/Zn/Na/Ca/Na/H-transporter	6	7.868759812	0.132419275
GOTERM_MF_FAT	GO:0015662~ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	6	5.938235294	0.065855166
UP_SEQ_FEATURE	active site:4-aspartylphosphate intermediate	6	5.492478879	0.308195696
GOTERM_BP_FAT	GO:0009259~ribonucleotide metabolic process	9	3.465613996	0.268934139
UP_SEQ_FEATURE	metal ion-binding site:Magnesium	8	3.583745084	0.38398223
GOTERM_BP_FAT	GO:0009165~nucleotide biosynthetic process	9	2.894806985	0.359926413
GOTERM_BP_FAT	GO:0006163~purine nucleotide metabolic process	8	3.124553571	0.362812347
GOTERM_BP_FAT	GO:0009260~ribonucleotide biosynthetic process	8	3.124553571	0.362812347
GOTERM_MF_FAT	GO:0042625~ATPase activity, coupled to transmembrane movement of ions	6	4.032135076	0.216013036
GOTERM_BP_FAT	GO:0034404~nucleobase, nucleoside and nucleotide biosynthetic process	9	2.689164959	0.39747466
GOTERM_BP_FAT	GO:0034654~nucleobase, nucleoside, nucleotide and nucleic acid biosynthetic process	9	2.689164959	0.39747466
GOTERM_BP_FAT	GO:0044271~nitrogen compound biosynthetic process	17	1.8370646	0.420180445
GOTERM_BP_FAT	GO:0009150~purine ribonucleotide metabolic process	7	2.921815363	0.477736901
GOTERM_BP_FAT	GO:0009152~purine ribonucleotide biosynthetic process	7	2.921815363	0.477736901

GOTERM_BP_FAT	GO:0006164~purine nucleotide biosynthetic process	7	2.793852646	0.497605808
GOTERM_BP_FAT	GO:0006754~ATP biosynthetic process	6	3.009891055	0.552611668
GOTERM_BP_FAT	GO:0046034~ATP metabolic process	6	3.009891055	0.552611668
GOTERM_BP_FAT	GO:0009144~purine nucleoside triphosphate metabolic process	6	2.877878289	0.599556076
GOTERM_BP_FAT	GO:0009201~ribonucleoside triphosphate biosynthetic process	6	2.877878289	0.599556076
GOTERM_BP_FAT	GO:0009199~ribonucleoside triphosphate metabolic process	6	2.877878289	0.599556076
GOTERM_BP_FAT	GO:0009206~purine ribonucleoside triphosphate biosynthetic process	6	2.877878289	0.599556076
GOTERM_BP_FAT	GO:0009205~purine ribonucleoside triphosphate metabolic process	6	2.877878289	0.599556076
GOTERM_BP_FAT	GO:0009145~purine nucleoside triphosphate biosynthetic process	6	2.877878289	0.599556076
GOTERM_BP_FAT	GO:0009142~nucleoside triphosphate biosynthetic process	6	2.852853261	0.59950973
GOTERM_BP_FAT	GO:0009141~nucleoside triphosphate metabolic process	6	2.828259698	0.583657534
INTERPRO	IPR004014:ATPase, P-type cation-transporter, N-terminal	3	7.414792899	0.629081881
UP_SEQ_FEATURE	topological domain:Extracellular	12	1.177874258	0.997778217
Annotation Cluster 12	Enrichment Score: 2.538644284957227			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_MF_FAT	GO:0042626~ATPase activity, coupled to transmembrane movement of substances	14	4.119316375	0.00203178
GOTERM_MF_FAT	GO:0043492~ATPase activity, coupled to movement of substances	14	4.119316375	0.00203178
GOTERM_MF_FAT	GO:0016820~hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	14	3.989913767	0.00212288
INTERPRO	IPR017871:ABC transporter, conserved site	8	4.145905707	0.21595515
UP_SEQ_FEATURE	domain:ABC transmembrane type-2	5	7.519465132	0.330353937
INTERPRO	IPR013525:ABC-2 type transporter	5	7.65018315	0.233836654
INTERPRO	IPR003439:ABC transporter-like	7	3.911571906	0.290193883
UP_SEQ_FEATURE	domain:ABC transporter	6	4.282271668	0.505490656
PIR_SUPERFAMILY	PIRSF002790:Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1	3	13.00303951	0.961078223
SMART	SM00382:AAA	9	1.879010804	0.597646513
INTERPRO	IPR003593:ATPase, AAA+ type, core	9	1.774091553	0.836501002
Annotation Cluster 13	Enrichment Score: 2.2251935356160164			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	thylakoid	10	4.859371567	0.002192757
GOTERM_CC_FAT	GO:0009579~thylakoid	19	2.176324868	0.043782152
GOTERM_CC_FAT	GO:0044436~thylakoid part	15	2.383085438	0.055906218
GOTERM_CC_FAT	GO:0042651~thylakoid membrane	13	2.397149548	0.093376228
GOTERM_CC_FAT	GO:0009534~chloroplast thylakoid	14	2.269081927	0.095976658
GOTERM_CC_FAT	GO:0031976~plastid thylakoid	14	2.269081927	0.095976658
GOTERM_CC_FAT	GO:0031984~organelle subcompartment	14	2.256078592	0.095343938
GOTERM_CC_FAT	GO:0055035~plastid thylakoid membrane	12	2.327206193	0.124504595
GOTERM_CC_FAT	GO:0009535~chloroplast thylakoid	12	2.327206193	0.124504595

	membrane			
GOTERM_CC_FAT	GO:0034357~photosynthetic membrane	13	2.21554731	0.121471303
Annotation Cluster 14	Enrichment Score: 2.108019531671565			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	nadp	9	4.84623813	0.004243275
INTERPRO	IPR017927:Ferredoxin reductase-type FAD-binding domain	4	8.568205128	0.294372087
GOTERM_MF_FAT	GO:0016651~oxidoreductase activity, acting on NADH or NADPH	5	3.128380663	0.50972309
Annotation Cluster 15	Enrichment Score: 1.9936062416867306			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	helicase	9	5.603462838	0.002106821
GOTERM_MF_FAT	GO:0008026~ATP-dependent helicase activity	9	4.335437272	0.030343231
GOTERM_MF_FAT	GO:0070035~purine NTP-dependent helicase activity	9	4.335437272	0.030343231
SMART	SM00490:HELICc	8	3.889254109	0.101501178
SMART	SM00487:DEXDc	8	3.755141898	0.092701727
INTERPRO	IPR014021:Helicase, superfamily 1 and 2, ATP-binding	8	3.725306577	0.294976826
INTERPRO	IPR001650:DNA/RNA helicase, C-terminal	8	3.672087912	0.269952743
UP_SEQ_FEATURE	domain:Helicase ATP-binding	7	4.03784977	0.37368093
INTERPRO	IPR014001:DEAD-like helicase, N-terminal	8	3.545464191	0.277506683
INTERPRO	IPR011545:DNA/RNA helicase, DEAD/DEAH box type, N-terminal	6	4.536108597	0.293571547
GOTERM_MF_FAT	GO:0004386~helicase activity	9	2.799453782	0.208910062
UP_SEQ_FEATURE	domain:Helicase C-terminal	6	3.558507443	0.611729477
PIR_SUPERFAMILY	PIRSF001321:ATP-dependent RNA helicase	3	9.58118701	0.945909138
INTERPRO	IPR014014:RNA helicase, DEAD-box type, Q motif	4	4.673566434	0.610911245
INTERPRO	IPR000629:RNA helicase, ATP-dependent, DEAD-box, conserved site	4	4.509581646	0.633449836
UP_SEQ_FEATURE	short sequence motif:Q motif	4	3.007786053	0.928968162
UP_SEQ_FEATURE	short sequence motif:DEAD box	4	2.955017876	0.924419557
Annotation Cluster 16	Enrichment Score: 1.9892114346404128			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_MF_FAT	GO:0051119~sugar transmembrane transporter activity	9	4.494535888	0.025453114
GOTERM_MF_FAT	GO:0005402~cation:sugar symporter activity	8	4.39869281	0.048077647
GOTERM_MF_FAT	GO:0015295~solute:hydrogen symporter activity	8	4.39869281	0.048077647
GOTERM_MF_FAT	GO:0005351~sugar:hydrogen symporter activity	8	4.39869281	0.048077647
GOTERM_MF_FAT	GO:0015294~solute:cation symporter activity	8	3.45611578	0.128422375
SP_PIR_KEYWORDS	sugar transport	5	6.13028413	0.04857424
INTERPRO	IPR003663:Sugar/inositol transporter	5	6.062409289	0.276681949
INTERPRO	IPR005829:Sugar transporter, conserved site	6	4.381468531	0.29038037
SP_PIR_KEYWORDS	Symport	5	5.61223195	0.062223085

INTERPRO	IPR005828:General substrate transporter	5	5.1001221	0.351816548
GOTERM_BP_FAT	GO:0008643~carbohydrate transport	6	3.952748494	0.392542726
GOTERM_MF_FAT	GO:0015293~symporter activity	8	2.982675262	0.22263299
PIR_SUPERFAMILY	PIRSF005322:glucose transport protein	3	4.667757774	0.999254044
GOTERM_BP_FAT	GO:0055085~transmembrane transport	7	1.788587909	0.876603787
Annotation Cluster 17	Enrichment Score: 1.9663575102020243			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_BP_FAT	GO:0006511~ubiquitin-dependent protein catabolic process	14	2.743783602	0.180783881
GOTERM_BP_FAT	GO:0044265~cellular macromolecule catabolic process	21	1.949530454	0.276481864
GOTERM_BP_FAT	GO:0019941~modification-dependent protein catabolic process	19	1.888934659	0.365121283
GOTERM_BP_FAT	GO:0043632~modification-dependent macromolecule catabolic process	19	1.888934659	0.365121283
GOTERM_BP_FAT	GO:0051603~proteolysis involved in cellular protein catabolic process	19	1.87191723	0.355984074
GOTERM_BP_FAT	GO:0009057~macromolecule catabolic process	22	1.763860887	0.346478764
SP_PIR_KEYWORDS	ubl conjugation pathway	12	2.355478631	0.068038073
GOTERM_BP_FAT	GO:0044257~cellular protein catabolic process	19	1.851896725	0.358931473
GOTERM_BP_FAT	GO:0030163~protein catabolic process	19	1.803670247	0.396196916
GOTERM_BP_FAT	GO:0006508~proteolysis	29	1.530609013	0.42742406
Annotation Cluster 18	Enrichment Score: 1.7891297762197387			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_BP_FAT	GO:0016114~terpenoid biosynthetic process	7	4.907151442	0.231384091
GOTERM_BP_FAT	GO:0006721~terpenoid metabolic process	7	3.905691964	0.330220678
GOTERM_BP_FAT	GO:0016109~tetraterpenoid biosynthetic process	4	8.74875	0.341447862
GOTERM_BP_FAT	GO:0016117~carotenoid biosynthetic process	4	8.74875	0.341447862
GOTERM_BP_FAT	GO:0019748~secondary metabolic process	16	2.068262411	0.345782857
GOTERM_BP_FAT	GO:0008299~isoprenoid biosynthetic process	7	3.387237279	0.393471576
GOTERM_BP_FAT	GO:0016108~tetraterpenoid metabolic process	4	6.834960938	0.406060737
GOTERM_BP_FAT	GO:0016116~carotenoid metabolic process	4	6.834960938	0.406060737
GOTERM_BP_FAT	GO:0006720~isoprenoid metabolic process	7	2.856401586	0.484214343
GOTERM_BP_FAT	GO:0042440~pigment metabolic process	6	3.248298267	0.484239672
GOTERM_BP_FAT	GO:0046148~pigment biosynthetic process	5	3.142510776	0.634170818
Annotation Cluster 19	Enrichment Score: 1.7553011091663233			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	phosphotransferase	11	6.641141141	9.57E-05
SP_PIR_KEYWORDS	kinase	30	2.464753413	1.73E-04
SP_PIR_KEYWORDS	ATP	8	6.071900472	0.003100202
SP_PIR_KEYWORDS	serine/threonine-specific protein kinase	5	11.38481338	0.006658521
UP_SEQ_FEATURE	nucleotide phosphate-binding region:ATP	33	1.797667731	0.139723473

SMART	SM00220:S_TKc	12	3.190403761	0.087338355
INTERPRO	IPR002290:Serine/threonine protein kinase	12	3.012259615	0.204161236
SP_PIR_KEYWORDS	serine/threonine-protein kinase	18	2.046343062	0.045368465
UP_SEQ_FEATURE	binding site:ATP	18	1.676907268	0.713024151
UP_SEQ_FEATURE	domain:Protein kinase	17	1.676470915	0.729009732
UP_SEQ_FEATURE	active site:Proton acceptor	20	1.542454386	0.793129953
INTERPRO	IPR017441:Protein kinase, ATP binding site	17	1.527896719	0.751166137
INTERPRO	IPR008271:Serine/threonine protein kinase, active site	17	1.418761239	0.840901948
GOTERM_BP_FAT	GO:0006796~phosphate metabolic process	29	1.245648812	0.856842993
GOTERM_BP_FAT	GO:0006793~phosphorus metabolic process	29	1.244671066	0.855770531
GOTERM_MF_FAT	GO:0004674~protein serine/threonine kinase activity	21	1.172420814	0.90693046
GOTERM_MF_FAT	GO:0004672~protein kinase activity	23	1.127908055	0.929947456
INTERPRO	IPR017442:Serine/threonine protein kinase-related	15	1.175515947	0.994293335
GOTERM_BP_FAT	GO:0016310~phosphorylation	24	1.114017402	0.953515933
INTERPRO	IPR000719:Protein kinase, core	17	1.126233148	0.995812966
GOTERM_BP_FAT	GO:0006468~protein amino acid phosphorylation	21	1.078191021	0.972692964
Annotation Cluster 20	Enrichment Score: 1.727969541989424			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_BP_FAT	GO:0031163~metallo-sulfur cluster assembly	4	10.41517857	0.274733294
GOTERM_BP_FAT	GO:0016226~iron-sulfur cluster assembly	4	10.41517857	0.274733294
SP_PIR_KEYWORDS	iron-sulfur	5	6.324896325	0.045968799
GOTERM_MF_FAT	GO:0051540~metal cluster binding	5	2.990869425	0.532029627
GOTERM_MF_FAT	GO:0051536~iron-sulfur cluster binding	5	2.990869425	0.532029627
Annotation Cluster 21	Enrichment Score: 1.7179651528418045			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	Protease	14	3.350485621	0.002869537
GOTERM_MF_FAT	GO:0008237~metallopeptidase activity	7	4.329963235	0.08798486
GOTERM_BP_FAT	GO:0006508~proteolysis	29	1.530609013	0.42742406
GOTERM_MF_FAT	GO:0070011~peptidase activity, acting on L-amino acid peptides	16	1.5864138	0.518264348
GOTERM_MF_FAT	GO:0004175~endopeptidase activity	10	1.750283715	0.62841503
GOTERM_MF_FAT	GO:0008233~peptidase activity	16	1.468703502	0.637245191
Annotation Cluster 22	Enrichment Score: 1.7107047512125682			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_BP_FAT	GO:0042742~defense response to bacterium	11	3.268894361	0.179121657
GOTERM_BP_FAT	GO:0009617~response to bacterium	12	2.689164959	0.273181808
GOTERM_BP_FAT	GO:0006952~defense response	18	0.955567354	0.996932839
Annotation Cluster 23	Enrichment Score: 1.5951967712319624			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_BP_FAT	GO:0042325~regulation of	4	6.627840909	0.423925988

	phosphorylation			
GOTERM_BP_FAT	GO:0019220~regulation of phosphate metabolic process	4	6.075520833	0.438571748
GOTERM_BP_FAT	GO:0051174~regulation of phosphorus metabolic process	4	6.075520833	0.438571748
Annotation Cluster 24	Enrichment Score: 1.5402540306647332			
Category	Term	Count	Fold Enrichment	Benjamini
INTERPRO	IPR006204:GHMP kinase	3	16.06538462	0.323772219
INTERPRO	IPR013750:GHMP kinase, C-terminal	3	16.06538462	0.323772219
INTERPRO	IPR014721:Ribosomal protein S5 domain 2-type fold	3	5.073279352	0.809402394
Annotation Cluster 25	Enrichment Score: 1.5043410224459075			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	FAD	7	5.810998498	0.00917376
INTERPRO	IPR017927:Ferredoxin reductase-type FAD-binding domain	4	8.568205128	0.294372087
UP_SEQ_FEATURE	domain:FAD-binding FR-type	3	7.431000836	0.772341088
GOTERM_MF_FAT	GO:0050660~FAD binding	7	2.323394907	0.517686806
GOTERM_MF_FAT	GO:0009055~electron carrier activity	14	1.161697453	0.95062994

Table S4.3 Swiss-Prot IDs and description of some genes that have been found to be shared in the two *Chromera* strains

ID	Gene Name	Species
HSP genes		
Q6F2Y7	Heat shock protein 101	Oryza sativa
Q4UJB1	Small heat shock protein C4	Rickettsia felis
Q8BM72	heat shock protein 70 family, member 13	Mus musculus
Q4UJB1	kda class i heat shock protein	Medicago sativa
Antioxidant genes		
Q5X8J8	Catalase-peroxidase 2	Legionella pneumophila
Q06652	gpx4_citsiprobable phospholipid hydroperoxide glutathione peroxidase	Citrus sinensis
P00449	sodm_geosesuperoxide dismutase	
Q13162	prdx4_humanperoxiredoxin-4	Homo sapiens
O94561	Peroxiredoxin C1773.02c	Schizosaccharomyces pombe
Q6QPJ6	Peroxiredoxin Q, chloroplastic	Populus trichocarpa
Q9LU86	Peroxiredoxin Q, chloroplastic	Arabidopsis thaliana
Q949U7	Peroxiredoxin-2E, chloroplastic	Arabidopsis thaliana
Q69TY4	Peroxiredoxin-2E-1, chloroplastic	Oryza sativa
Q7F8S5	Peroxiredoxin-2E-2, chloroplastic	Oryza sativa
O22229	Thioredoxin reductase	Arabidopsis thaliana
O14463	Thioredoxin-1	Schizosaccharomyces pombe
Q39027	Mitogen-activated protein kinase 7	Arabidopsis thaliana
Q9FIJ0	Respiratory burst oxidase homolog protein D	Arabidopsis thaliana
Q539E5	Putative ascorbate peroxidase	Hydra viridissima
Q9DCM2	glutathione S-transferase kappa 1	Mus musculus
P30711	glutathione S-transferase theta 1	Homo sapiens
Q9WVL0	glutathione transferase zeta 1 (maleylacetoacetate isomerase)	Mus musculus
Q3T100	microsomal glutathione S-transferase 3	Bos taurus
P46428	Glutathione S-transferase	Anopheles gambiae
Q9BEA9	Glutathione S-transferase Mu 3	Macaca fuscata
Q10075	Glutathione gamma-glutamylcysteinyltransferase	Schizosaccharomyces pombe
Q873E8	Glutathione reductase	Neurospora crassa
P46436	gst1_ascuglutathione s-transferase 1	ascaris suum
Q9LZ06	gstl3_arathglutathione s-transferase l3	arabidopsis thaliana
Q9FRL8	dhar2_arathglutathione s-transferase dhar2	arabidopsis thaliana
Regulation of incoming light		
Q9FN03	uvr8_arathultraviolet-b receptor uvr8	os=arabidopsis thaliana
Q2NI00	dhqs_metst3-dehydroquinase synthase	methanosphaera stadtmanae
photosynthetic genes		
Q39709	Fucoxanthin-chlorophyll a-c binding protein, chloroplastic	Isochrysis galbana
Q38833	Chlorophyll synthase, chloroplastic	Arabidopsis thaliana

O78502	Photosystem I reaction center subunit II	Guillardia theta
Q9AW48	Photosystem II stability/assembly factor HCF136, chloroplastic	Guillardia theta
Q37D32	Ribulose biphosphate carboxylase	Rhodopseudomonas palustris
Q43088	Ribulose-1,5 biphosphate carboxylase/oxygenase large subunit N-methyltransferase, chloroplastic	Pisum sativum
P44756	Ribulose-phosphate 3-epimerase	Haemophilus influenzae
P46969	Ribulose-phosphate 3-epimerase	Saccharomyces cerevisiae
Cytochrome		
Q50EK3	Cytochrome P450 704C1	Pinus taeda
P48422	Cytochrome P450 86A1	Arabidopsis thaliana
O23365	Cytochrome P450 97B3	Arabidopsis thaliana
Q9VS79	Cytochrome P450-4d8	Drosophila melanogaster
P19967	Cytochrome b5-related	Drosophila melanogaster
B2XTQ5	Cytochrome c biogenesis protein ccsA	Heterosigma akashiwo
Q02212	Cytochrome c oxidase subunit 2	Phytophthora megasperma
P00110	Cytochrome c6	Bumilleriopsis filiformis
Q93VA3	Cytochrome c6, chloroplastic	Arabidopsis thaliana
P51869	cytochrome P450, family 4, subfamily f, polypeptide 4	Rattus norvegicus
Q02212	Cytochrome c oxidase subunit 2	Phytophthora megasperma
Calcium-dependent protein kinase genes		
O49717	Calcium-dependent protein kinase 15	Arabidopsis thaliana
Q9SSF8	Calcium-dependent protein kinase 30	Arabidopsis thaliana
P62345	Calcium-dependent protein kinase 4	Plasmodium berghei
Q5ZK10	calcium/calmodulin-dependent protein kinase (CaM kinase) II delta	Gallus gallus
Q16566	calcium/calmodulin-dependent protein kinase IV	Homo sapiens
House keeping genes		
A8CEP3	Calmodulin	Saccharina japonica
P27166	Calmodulin	Stylonychia lemnae
P69005	actin	Strongylocentrotus purpuratus
P20365	Tubulin beta chain	Moneuplotes crassus
P51469	glyceraldehyde-3-phosphate dehydrogenase	Xenopus laevis
O02367	calmodulin homologue	Ciona intestinalis
Q9I8D1	myosin VI	Gallus gallus
Q8MJU9	myosin, heavy chain 7, cardiac muscle, beta	Equus caballus
Nutrient and metabolite transport		
Q8CF82	ATP-binding cassette, sub-family A (ABC1), member 5	Rattus norvegicus
Q9NRK6	ATP-binding cassette, sub-family B (MDR/TAP), member 10	Homo sapiens
Q8R4P9	ATP-binding cassette, sub-family C (CFTR/MRP), member 10	Mus musculus
Q8VI47	ATP-binding cassette, sub-family C (CFTR/MRP), member 2	Mus musculus
Q8K268	ATP-binding cassette, sub-family F (GCN20), member 3	Mus musculus

Q7TMS5	ATP-binding cassette, sub-family G (WHITE), member 2	Mus musculus
P28605	Glutamine synthetase	Synechococcus sp. PCC 7002
Q96QE2	solute carrier family 2 (facilitated glucose transporter), member 13	Homo sapiens
Q9R0M8	solute carrier family 35 (UDP-galactose transporter), member A2	Mus musculus
Q56ZZ7	Plastidic glucose transporter 4	Arabidopsis thaliana
Q0WVE9	Probable plastidic glucose transporter 1	Arabidopsis thaliana
P22152	Nitrate transporter	Emericella nidulans
P54147	Putative ammonium transporter sl10108	Synechocystis sp. PCC 6803
Vesicular trafficking		
A2AWA9	RAB GTPase activating protein 1	Mus musculus
Q5R372	RAB GTPase activating protein 1-like	Homo sapiens
Q5FVJ7	RAB, member RAS oncogene family-like 5	Rattus norvegicus
P46638	RAB11B, member RAS oncogene family	Mus musculus
P35288	RAB23, member RAS oncogene family	Mus musculus
Q99KL7	RAB28, member RAS oncogene family	Mus musculus
P61018	RAB4B, member RAS oncogene family	Homo sapiens
Q1KME6	RAB6A, member RAS oncogene family	Gallus gallus
Q9NVG8	TBC1 domain family, member 13	Homo sapiens
Q9P2M4	TBC1 domain family, member 14	Homo sapiens
Q5BKM3	vacuolar protein sorting 24 (yeast)	Xenopus (Silurana) tropicalis
Q7ZV68	vacuolar protein sorting 29 (yeast)	Danio rerio
Q96JC1	vacuolar protein sorting 39 homolog (S. cerevisiae)	Homo sapiens
O75351	vacuolar protein sorting 4 homolog B (S. cerevisiae)	Homo sapiens
P59015	vacuolar protein sorting protein 18	Danio rerio
Q9URZ5	Vacuolar protein sorting-associated protein 1	Schizosaccharomyces pombe
Q9S9T7	Vacuolar protein sorting-associated protein 28 homolog 2	Arabidopsis thaliana
O82197	Vacuolar protein sorting-associated protein 32 homolog 1	Arabidopsis thaliana
Q7T385	ATPase, H ⁺ transporting, lysosomal, V1 subunit C, isoform 1	Danio rerio
O93428	Cathepsin D	Chionodraco hamatus
Q5RB02	cathepsin C	Pongo abelii
O35186	cathepsin K	Rattus norvegicus
Q9UBR2	cathepsin Z	Homo sapiens
Q9SFB8	Phosphatidylinositol-4-phosphate 5-kinase 6	Arabidopsis thaliana
O13853	Phosphatidylinositol-4-phosphate 5-kinase its3	Schizosaccharomyces pombe
Q8BKX6	SMG1 homolog, phosphatidylinositol 3-kinase-related kinase (C. elegans)	Mus musculus
Ribosome genes		
P46228	30S ribosomal protein S1	Synechococcus elongatus
A8F4Q6	30S ribosomal protein S12	Thermotoga lettingae

B5ZB57	30S ribosomal protein S5	Ureaplasma urealyticum
B1YH85	30S ribosomal protein S9	Exiguobacterium sibiricum
Q9GT45	40S ribosomal protein S26	Anopheles gambiae
Q6BXH8	40S ribosomal protein S6	Debaryomyces hansenii
B2GAC3	50S ribosomal protein L1	Lactobacillus fermentum
P49544	50S ribosomal protein L1, chloroplastic	Odontella sinensis
Q6ANL8	50S ribosomal protein L13	Desulfotalea psychrophila
Q06SI6	50S ribosomal protein L20, chloroplastic	Stigeoclonium helveticum
P49557	50S ribosomal protein L21, chloroplastic	Odontella sinensis
B1YGW1	50S ribosomal protein L24	Exiguobacterium sibiricum
Q68W78	50S ribosomal protein L3	Rickettsia typhi
Q9LUQ6	60S ribosomal protein L19-2	Arabidopsis thaliana
O42706	60S ribosomal protein L21-B; 60S ribosomal protein L21-A	Schizosaccharomyces pombe
P51414	60S ribosomal protein L26-1	Arabidopsis thaliana
Q7KF90	60S ribosomal protein L31	Spodoptera frugiperda
Q6FSN6	60S ribosomal protein L7	Candida glabrata
Q1MTQ9	60S ribosome subunit biogenesis protein nip7	Schizosaccharomyces pombe
P48166	Ribosomal Protein, Large subunit	Caenorhabditis elegans
Q9BL19	Ribosomal Protein, Large subunit	Caenorhabditis elegans
Q88P77	Ribosomal RNA large subunit methyltransferase F	Pseudomonas putida
A1TM24	Ribosomal RNA large subunit methyltransferase N	Acidovorax avenae
Q1DCU1	Ribosomal RNA large subunit methyltransferase N 1	Myxococcus xanthus
P94464	Ribosomal RNA small subunit methyltransferase B	Bacillus subtilis
Q10257	Ribosomal RNA-processing protein 8	Schizosaccharomyces pombe
Q8FL93	Ribosomal large subunit pseudouridine synthase A	Escherichia coli
Q8PQN3	Ribosomal large subunit pseudouridine synthase E	Xanthomonas axonopodis
A8IR43	Ribosome biogenesis protein WDR12 homolog	Chlamydomonas reinhardtii
A8MHH2	Ribosome-recycling factor	Alkaliphilus oremlandii
P41116	ribosomal protein L8	Xenopus laevis
P63326	ribosomal protein S10; similar to 40S ribosomal protein S10; similar to ribosomal protein S10	Rattus norvegicus
P62979	ribosomal protein S27a pseudogene 12; ribosomal protein S27a; ribosomal protein S27a pseudogene 11; ribosomal protein S27a pseudogene 16	Homo sapiens

Table S4.4 GO categories (BP, CC, MF) enriched in the list of 507 shared genes between GBR/ Sydney *Chromera* and *Symbiodinium* with corrected *P*-value ≤ 0.05

Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_BP	GO:0046488~phosphatidylinositol metabolic process	5	42.83353733	0.002030636
	GO:0015994~chlorophyll metabolic process	6	21.70232558	0.001641357
	GO:0006778~porphyrin metabolic process	6	16.00991231	0.004929599
	GO:0033013~tetrapyrrole metabolic process	6	15.75168792	0.004004306
	GO:0051188~cofactor biosynthetic process	8	8.189556823	0.004449178
	GO:0015995~chlorophyll biosynthetic process	5	23.93638851	0.003774651
	GO:0030384~phosphoinositide metabolic process	5	18.92644673	0.008240886
	GO:0006779~porphyrin biosynthetic process	5	17.3156853	0.010221216
	GO:0033014~tetrapyrrole biosynthetic process	5	16.27674419	0.01156223
	GO:0051186~cofactor metabolic process	9	5.250562641	0.012684307
	GO:0042440~pigment metabolic process	6	9.669352982	0.014734596
	GO:0006650~glycerophospholipid metabolic process	5	13.79385101	0.016402478
	GO:0046486~glycerolipid metabolic process	5	12.33086681	0.023139436
	GO:0018130~heterocycle biosynthetic process	6	7.398520085	0.038587166
	GO:0046148~pigment biosynthetic process	5	9.354450682	0.055346314
GOTERM_CC	GO:0044435~plastid part	26	4.667682927	6.24E-09
	GO:0044434~chloroplast part	25	4.629293662	1.05E-08
	GO:0009507~chloroplast	40	2.213707345	5.21E-06
	GO:0009536~plastid	40	2.168196946	7.00E-06
	GO:0009941~chloroplast envelope	14	5.620804196	2.05E-05
	GO:0010287~plastoglobule	7	20.60961538	1.95E-05
	GO:0009526~plastid envelope	14	5.364758885	2.46E-05
	GO:0031967~organelle envelope	16	3.793908105	1.95E-04
	GO:0031975~envelope	16	3.763597255	1.90E-04
	GO:0009534~chloroplast thylakoid	9	4.581800044	0.008202774
	GO:0031976~plastid thylakoid	9	4.581800044	0.008202774
	GO:0031984~organelle subcompartment	9	4.555543311	0.007744066
	GO:0031090~organelle membrane	14	2.913019842	0.007750608
	GO:0044436~thylakoid part	9	4.491199478	0.00719094
	GO:0055035~plastid thylakoid membrane	8	4.873209549	0.009841586
	GO:0009535~chloroplast thylakoid membrane	8	4.873209549	0.009841586
	GO:0042651~thylakoid membrane	8	4.633543506	0.012264567
	GO:0009579~thylakoid	10	3.597838007	0.011945211
	GO:0034357~photosynthetic membrane	8	4.282517483	0.016871533
	GO:0009532~plastid stroma	9	3.597024017	0.020931549
GO:0009570~chloroplast stroma	8	3.364835165	0.05471395	
GOTERM_MF	GO:0016308~1-phosphatidylinositol-4-phosphate 5-kinase activity	5	52.5035461	4.19E-04
	GO:0016307~phosphatidylinositol phosphate kinase activity	5	49.22207447	2.78E-04
	GO:0001727~lipid kinase activity	5	35.79787234	7.24E-04
	GO:0032559~adenyl ribonucleotide binding	33	2.195038456	5.53E-04

GO:0032553~ribonucleotide binding	35	2.103346944	5.00E-04
GO:0032555~purine ribonucleotide binding	35	2.103346944	5.00E-04
GO:0030554~adenyl nucleotide binding	34	2.118418395	5.27E-04
GO:0001883~purine nucleoside binding	34	2.118418395	5.27E-04
GO:0001882~nucleoside binding	34	2.111735687	4.83E-04
GO:0017076~purine nucleotide binding	36	2.036044158	4.40E-04
GO:0005524~ATP binding	32	2.153991635	5.42E-04
GO:0004428~inositol or phosphatidylinositol kinase activity	5	17.12072155	0.004309586
GO:0000166~nucleotide binding	36	1.701285022	0.013637066
GO:0016887~ATPase activity	10	3.372818807	0.047700994

Table S4.5 Significant Kegg pathways enriched in the list of 507 shared genes between GBR/ Sydney *Chromera* and *Symbiodinium* with corrected *P*-value ≤ 0.05

Term	Count	%	PValue
ath00562:Inositol phosphate metabolism	4	0.99009901	0.00219388
ath04070:Phosphatidylinositol signaling system	4	0.99009901	0.00219388

Table S4.6 Significant functional annotation clusters (FACs) for the set of genes shared between GBR/ Sydney *Chromera* and *Symbiodinium* using DAVID high classification stringency and ease score ≤ 0.05

Annotation Cluster 1	Enrichment Score: 12.73314649577858			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	plastid	25	8.290016244	2.38E-13
SP_PIR_KEYWORDS	chloroplast	25	8.024310595	2.45E-13
SP_PIR_KEYWORDS	transit peptide	26	6.655747193	2.65E-12
UP_SEQ_FEATURE	transit peptide:Chloroplast	23	4.425567394	6.20E-07
Annotation Cluster 2	Enrichment Score: 5.61636537757201			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	atp-binding	29	4.231000132	2.38E-09
SP_PIR_KEYWORDS	nucleotide-binding	30	3.881526986	5.99E-09
GOTERM_MF_FAT	GO:0032559~adenyl ribonucleotide binding	33	2.195038456	5.53E-04
GOTERM_MF_FAT	GO:0032553~ribonucleotide binding	35	2.103346944	5.00E-04
GOTERM_MF_FAT	GO:0032555~purine ribonucleotide binding	35	2.103346944	5.00E-04
GOTERM_MF_FAT	GO:0030554~adenyl nucleotide binding	34	2.118418395	5.27E-04
GOTERM_MF_FAT	GO:0001883~purine nucleoside binding	34	2.118418395	5.27E-04
GOTERM_MF_FAT	GO:0001882~nucleoside binding	34	2.111735687	4.83E-04
GOTERM_MF_FAT	GO:0017076~purine nucleotide binding	36	2.036044158	4.40E-04
GOTERM_MF_FAT	GO:0005524~ATP binding	32	2.153991635	5.42E-04
GOTERM_MF_FAT	GO:0000166~nucleotide binding	36	1.701285022	0.013637066
Annotation Cluster 3	Enrichment Score: 5.407386600903847			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_CC_FAT	GO:0009941~chloroplast envelope	14	5.620804196	2.05E-05
GOTERM_CC_FAT	GO:0009526~plastid envelope	14	5.364758885	2.46E-05
GOTERM_CC_FAT	GO:0031967~organelle envelope	16	3.793908105	1.95E-04
GOTERM_CC_FAT	GO:0031975~envelope	16	3.763597255	1.90E-04
Annotation Cluster 4	Enrichment Score: 4.0959796430426945			
Category	Term	Count	Fold Enrichment	Benjamini
INTERPRO	IPR002498:Phosphatidylinositol-4-phosphate 5-kinase, core	5	66.93910256	1.60E-04
GOTERM_MF_FAT	GO:0016308~1-phosphatidylinositol-4-phosphate 5-kinase activity	5	52.5035461	4.19E-04
GOTERM_MF_FAT	GO:0016307~phosphatidylinositol phosphate kinase activity	5	49.22207447	2.78E-04
GOTERM_BP_FAT	GO:0046488~phosphatidylinositol metabolic process	5	42.83353733	0.002030636
PIR_SUPERFAMILY	PIRSF037274:PIP5K_plant_prd	4	105.6296296	2.07E-04
PIR_SUPERFAMILY	PIRSF037274:phosphatidylinositol-4-phosphate 5-kinase, plant type	4	105.6296296	2.07E-04
INTERPRO	IPR017163:Phosphatidylinositol-4-phosphate 5-kinase, plant	4	89.25213675	0.001048128
GOTERM_MF_FAT	GO:0001727~lipid kinase activity	5	35.79787234	7.24E-04
INTERPRO	IPR016034:Phosphatidylinositol-4-phosphate 5-kinase, core, subgroup	4	73.02447552	0.001362512

SMART	SM00330:PIPKc	4	65.04016913	9.45E-04
UP_SEQ_FEATURE	repeat:MORN 6	4	53.36336336	0.005378734
UP_SEQ_FEATURE	repeat:MORN 7	4	53.36336336	0.005378734
UP_SEQ_FEATURE	repeat:MORN 4	4	53.36336336	0.005378734
UP_SEQ_FEATURE	repeat:MORN 5	4	53.36336336	0.005378734
UP_SEQ_FEATURE	repeat:MORN 2	4	53.36336336	0.005378734
UP_SEQ_FEATURE	repeat:MORN 3	4	53.36336336	0.005378734
UP_SEQ_FEATURE	repeat:MORN 1	4	53.36336336	0.005378734
INTERPRO	IPR003409:MORN motif	4	53.55128205	0.002775813
SMART	SM00698:MORN	4	47.69612403	0.001283059
UP_SEQ_FEATURE	domain:PIPK	4	43.66093366	0.006955354
GOTERM_BP_FAT	GO:0030384~phosphoinositide metabolic process	5	18.92644673	0.008240886
GOTERM_MF_FAT	GO:0004428~inositol or phosphatidylinositol kinase activity	5	17.12072155	0.004309586
GOTERM_BP_FAT	GO:0006650~glycerophospholipid metabolic process	5	13.79385101	0.016402478
GOTERM_BP_FAT	GO:0046486~glycerolipid metabolic process	5	12.33086681	0.023139436
UP_SEQ_FEATURE	repeat:MORN 8	3	60.03378378	0.059305081
KEGG_PATHWAY	ath00562:Inositol phosphate metabolism	4	13.98577525	0.045074635
KEGG_PATHWAY	ath04070:Phosphatidylinositol signaling system	4	13.98577525	0.045074635
UP_SEQ_FEATURE	region of interest:Activation loop	4	12.63869132	0.141413194
GOTERM_BP_FAT	GO:0006644~phospholipid metabolic process	5	6.670796798	0.155749657
GOTERM_BP_FAT	GO:0019637~organophosphate metabolic process	5	6.260286225	0.180896166
KEGG_PATHWAY	ath04144:Endocytosis	4	7.961133603	0.073486419
Annotation Cluster 5	Enrichment Score: 3.86203095542081			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_BP_FAT	GO:0015994~chlorophyll metabolic process	6	21.70232558	0.001641357
GOTERM_BP_FAT	GO:0006778~porphyrin metabolic process	6	16.00991231	0.004929599
GOTERM_BP_FAT	GO:0033013~tetrapyrrole metabolic process	6	15.75168792	0.004004306
GOTERM_BP_FAT	GO:0015995~chlorophyll biosynthetic process	5	23.93638851	0.003774651
GOTERM_BP_FAT	GO:0006779~porphyrin biosynthetic process	5	17.3156853	0.010221216
GOTERM_BP_FAT	GO:0033014~tetrapyrrole biosynthetic process	5	16.27674419	0.01156223
GOTERM_BP_FAT	GO:0042440~pigment metabolic process	6	9.669352982	0.014734596
GOTERM_BP_FAT	GO:0018130~heterocycle biosynthetic process	6	7.398520085	0.038587166
GOTERM_BP_FAT	GO:0046148~pigment biosynthetic process	5	9.354450682	0.055346314
Annotation Cluster 6	Enrichment Score: 2.9612355233374568			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_CC_FAT	GO:0031976~plastid thylakoid	9	4.581800044	0.008202774
GOTERM_CC_FAT	GO:0009534~chloroplast thylakoid	9	4.581800044	0.008202774

GOTERM_CC_FAT	GO:0031984~organelle subcompartment	9	4.555543311	0.007744066
GOTERM_CC_FAT	GO:0044436~thylakoid part	9	4.491199478	0.00719094
GOTERM_CC_FAT	GO:0055035~plastid thylakoid membrane	8	4.873209549	0.009841586
GOTERM_CC_FAT	GO:0009535~chloroplast thylakoid membrane	8	4.873209549	0.009841586
GOTERM_CC_FAT	GO:0042651~thylakoid membrane	8	4.633543506	0.012264567
GOTERM_CC_FAT	GO:0009579~thylakoid	10	3.597838007	0.011945211
GOTERM_CC_FAT	GO:0034357~photosynthetic membrane	8	4.282517483	0.016871533
Annotation Cluster 7	Enrichment Score: 2.1508311688174118			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	Rotamase	4	18.54507338	0.017899257
GOTERM_MF_FAT	GO:0003755~peptidyl-prolyl cis-trans isomerase activity	4	11.05337813	0.089794345
GOTERM_MF_FAT	GO:0016859~cis-trans isomerase activity	4	10.86280264	0.087646861
INTERPRO	IPR001179:Peptidyl-prolyl cis-trans isomerase, FKBP-type	3	24.09807692	0.18995396
GOTERM_MF_FAT	GO:0005528~FK506 binding	3	21.4787234	0.11616141
GOTERM_MF_FAT	GO:0005527~macrolide binding	3	21.4787234	0.11616141
GOTERM_MF_FAT	GO:0008144~drug binding	3	19.68882979	0.128291737
SP_PIR_KEYWORDS	Isomerase	4	5.657818996	0.234337822

Table S4.7 GO categories (BP, CC, MF) enriched in the list of 240 shared genes between GBR/ Sydney *Chromera* and *Plasmodium* with corrected *P*-value ≤ 0.05

Category	Term	Count	Fold Enrichment	Benjamini	
GOTERM_BP_FAT	GO:0044265~cellular macromolecule catabolic process	11	6.376164644	0.001323319	
	GO:0006511~ubiquitin-dependent protein catabolic process	8	9.789666929	0.001868209	
	GO:0009057~macromolecule catabolic process	11	5.506687648	0.001601535	
	GO:0043632~modification-dependent macromolecule catabolic process	10	6.207538803	0.001458685	
	GO:0019941~modification-dependent protein catabolic process	10	6.207538803	0.001458685	
	GO:0051603~proteolysis involved in cellular protein catabolic process	10	6.15161503	0.00125398	
	GO:0044257~cellular protein catabolic process	10	6.085822356	0.001138002	
	GO:0030163~protein catabolic process	10	5.927337398	0.001201529	
	GO:0006508~proteolysis	12	3.954609662	0.004111733	
	GOTERM_CC_FAT	GO:0008540~proteasome regulatory particle, base subcomplex	4	153.1	2.02E-04
		GO:0000502~proteasome complex	5	38.275	3.95E-04
		GO:0022624~proteasome accessory complex	4	70.66153846	7.80E-04
GO:0005838~proteasome regulatory particle		4	70.66153846	7.80E-04	
GO:0043232~intracellular non-membrane-bounded organelle		10	4.01486014	0.010908409	
GO:0043228~non-membrane-bounded organelle		10	4.01486014	0.010908409	
GO:0005829~cytosol		8	5.189830508	0.011715427	
GOTERM_MF_FAT		GO:0000166~nucleotide binding	24	2.733688753	5.38E-05
	GO:0005524~ATP binding	17	2.758075827	0.005219892	
	GO:0032555~purine ribonucleotide binding	18	2.607225663	0.003642088	
	GO:0032553~ribonucleotide binding	18	2.607225663	0.003642088	
	GO:0032559~adenyl ribonucleotide binding	17	2.725463444	0.00302256	
	GO:0016887~ATPase activity	8	6.503486521	0.003427499	
	GO:0017076~purine nucleotide binding	18	2.453694241	0.00394897	
	GO:0001883~purine nucleoside binding	17	2.552965758	0.003801319	
	GO:0030554~adenyl nucleotide binding	17	2.552965758	0.003801319	
	GO:0001882~nucleoside binding	17	2.544912238	0.003454308	

Table S4.8 Significant functional annotation clusters (FACs) for the set of genes shared between GBR/ Sydney *Chromera* and *Plasmodium* using DAVID high classification stringency and ease score ≤ 0.05

Annotation Cluster 1	Enrichment Score: 4.889352999775386			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_CC_FAT	GO:0008540~proteasome regulatory particle, base subcomplex	4	153.1	2.02E-04
SP_PIR_KEYWORDS	proteasome	5	51.11325116	6.20E-05
GOTERM_CC_FAT	GO:0000502~proteasome complex	5	38.275	3.95E-04
GOTERM_CC_FAT	GO:0022624~proteasome accessory complex	4	70.66153846	7.80E-04
GOTERM_CC_FAT	GO:0005838~proteasome regulatory particle	4	70.66153846	7.80E-04
KEGG_PATHWAY	ath03050:Proteasome	5	13.24353448	0.005013596
Annotation Cluster 2	Enrichment Score: 4.7279589331460565			
Category	Term	Count	Fold Enrichment	Benjamini
SP_PIR_KEYWORDS	nucleotide-binding	18	5.610570825	2.14E-07
SP_PIR_KEYWORDS	atp-binding	17	5.975127146	1.59E-07
GOTERM_MF_FAT	GO:0005524~ATP binding	17	2.758075827	0.005219892
GOTERM_MF_FAT	GO:0032553~ribonucleotide binding	18	2.607225663	0.003642088
GOTERM_MF_FAT	GO:0032555~purine ribonucleotide binding	18	2.607225663	0.003642088
GOTERM_MF_FAT	GO:0032559~adenyl ribonucleotide binding	17	2.725463444	0.00302256
GOTERM_MF_FAT	GO:0017076~purine nucleotide binding	18	2.453694241	0.00394897
GOTERM_MF_FAT	GO:0030554~adenyl nucleotide binding	17	2.552965758	0.003801319
GOTERM_MF_FAT	GO:0001883~purine nucleoside binding	17	2.552965758	0.003801319
GOTERM_MF_FAT	GO:0001882~nucleoside binding	17	2.544912238	0.003454308
Annotation Cluster 3	Enrichment Score: 4.706804811585773			
Category	Term	Count	Fold Enrichment	Benjamini
GOTERM_BP_FAT	GO:0044265~cellular macromolecule catabolic process	11	6.376164644	0.001323319
GOTERM_BP_FAT	GO:0006511~ubiquitin-dependent protein catabolic process	8	9.789666929	0.001868209
GOTERM_BP_FAT	GO:0009057~macromolecule catabolic process	11	5.506687648	0.001601535
GOTERM_BP_FAT	GO:0019941~modification-dependent protein catabolic process	10	6.207538803	0.001458685
GOTERM_BP_FAT	GO:0043632~modification-dependent macromolecule catabolic process	10	6.207538803	0.001458685
GOTERM_BP_FAT	GO:0051603~proteolysis involved in cellular protein catabolic process	10	6.15161503	0.00125398
GOTERM_BP_FAT	GO:0044257~cellular protein catabolic process	10	6.085822356	0.001138002
GOTERM_BP_FAT	GO:0030163~protein catabolic process	10	5.927337398	0.001201529
GOTERM_BP_FAT	GO:0006508~proteolysis	12	3.954609662	0.004111733
Annotation Cluster 4	Enrichment Score: 2.0706289583480983			
Category	Term	Count	Fold Enrichment	Benjamini

UP_SEQ_FEATURE	active site:Glycyl thioester intermediate	4	20.90588235	0.041382344
SP_PIR_KEYWORDS	ubl conjugation pathway	5	7.42778773	0.027284019
INTERPRO	IPR000608:Ubiquitin-conjugating enzyme, E2	3	30.35610465	0.177169645
INTERPRO	IPR016135:Ubiquitin-conjugating enzyme/RWD-like	3	28.02101968	0.157276217
SMART	SM00212:UBCc	3	20.02864583	0.109446718
SP_PIR_KEYWORDS	ligase	4	8.902381751	0.057094385
GOTERM_MF_FAT	GO:0004842~ubiquitin-protein ligase activity	4	5.730430576	0.305288934
GOTERM_MF_FAT	GO:0019787~small conjugating protein ligase activity	4	5.542204754	0.300490953
GOTERM_MF_FAT	GO:0016881~acid-amino acid ligase activity	4	5.165184022	0.322408182

Table S4.9 KEGG pathways comparison in GBR/Sydney *Chromera*, *Symbiodinium* and *Plasmodium*

	GBR <i>Chromera</i>	Sydney <i>Chromera</i>	<i>Symbiodinium</i>	<i>Plasmodium</i>
Carbohydrate metabolism	6.516587678	6.660277911	8.189795255	4.738832427
Energy metabolism	3.08056872	3.002715221	3.41241469	6.275112267
Lipid metabolism	3.957345972	3.753394027	4.06239844	1.808083195
Nucleotide metabolism	3.909952607	3.210349784	3.087422814	3.10801229
Amino acid metabolism	6.303317536	5.829739658	6.532336692	1.264476483
Metabolism of other amino acids	1.3507109	1.549273279	1.332466688	4.289766013
Glycan biosynthesis and metabolism	1.848341232	1.565245169	1.364965876	0.661782085
Metabolism of cofactors and vitamins	3.36492891	2.954799553	2.924926877	1.087213425
Metabolism of terpenoids and polyketides	1.066350711	1.038172816	0.844978876	0.354526117
Biosynthesis of other secondary metabolites	1.042654028	1.102060374	0	0.319073505
Xenobiotics biodegradation and metabolism	1.72985782	1.884682958	1.559961001	0.76813992
Transcription	2.725118483	2.060373742	2.339941501	2.540770503
Translation	5.805687204	5.126976521	5.979850504	11.9357126
Folding, sorting and degradation	4.857819905	4.26449449	4.419889503	4.242495864
Replication and repair	3.246445498	2.922855774	2.534936627	4.384306311
Membrane transport	0.426540284	0.463184795	0.584985375	0.224533207
Signal transduction	7.511848341	9.151892669	7.99480013	5.908768613
Signalling molecules and interaction	0.331753555	0.335409679	0.227494313	0.011817537
Cell growth and death	3.909952607	3.370068679	3.217419565	2.398960057
Transport and catabolism	3.507109005	3.625618911	3.80240494	2.647128338
Cellular community	1.279620853	1.102060374	1.234969126	0.42543134
Cell motility	0.450236967	0.367353458	0.38999025	0.224533207
Endocrine system	3.720379147	5.414470532	5.102372441	3.947057433
Nervous system	2.677725118	3.178406005	3.704907377	1.264476483
Immune system	2.654028436	2.491614758	2.372440689	1.678090286
Digestive system	1.279620853	1.245807379	1.039974001	0.661782085
Excretory system	0.995260664	0.830538253	0.942476438	0.413613803
Circulatory system	0.876777251	0.814566363	0.7799805	0.803592531
Environmental adaptation	0.663507109	0.686791247	0.714982125	0.791774994
Development	0.450236967	0.638875579	0.487487813	0.259985819
Sensory system	0.355450237	0.654847468	0.714982125	0.862680217
Infectious diseases	7.440758294	7.011659479	6.954826129	11.87662491
Cancers	5.473933649	5.829739658	5.102372441	4.431576459
Neurodegenerative diseases	2.488151659	2.523558537	2.957426064	2.859844009
Substance dependence	1.018957346	1.373582495	1.429964251	5.022453321
Endocrine and metabolic diseases	0.639810427	0.974285258	0.714982125	0.590876861
Immune diseases	0.592417062	0.574988021	0.552486188	4.715197353
Cardiovascular diseases	0.450236967	0.351381568	0.38999025	0.200898133
Drug resistance	0	0.063887558	0	0