Close relationship between coral-associated *Chromera* strains despite major differences within the Symbiodiniaceae

Amin R. Mohamed[1, 2, 3, 4,5*], Cheong Xin Chan[6,7], Mark A. Ragan[6], Jia Zhang[3], Ira Cooke[3], Eldon E. Ball[2,8], David J. Miller[2,3*]

Affiliations:

[1]CSIRO Agriculture and Food, Queensland Bioscience Precinct, St Lucia Brisbane QLD 4067, Australia

[2]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville 4811, Queensland, Australia

[3]Molecular and Cell Biology, James Cook University, Townsville 4811, Queensland, Australia

[4]AIMS@JCU, Australian Institute of Marine Science, Department of Molecular and Cell Biology, James Cook University, Townsville 4811, Queensland, Australia

[5]Zoology Department, Faculty of Science, Benha University, Benha 13518, Egypt

[6]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

[7]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia

[8]Division of Ecology and Evolution, Research School of Biology, Australian National University, Acton, ACT 2601, Australia

***correspondence:** Dr David Miller Email: david.miller@jcu.edu.au Tel: +61747814473 Fax: +61747816078

Keywords: *Chromera*, RNA-Seq, transcriptome, symbiosis, Symbiodiniaceae

Running title: *Chromera* comparative transcriptomics

Summary

Reef-building corals live in a mutualistic relationship with photosynthetic algae (family Symbiodiniaceae) that usually provide most of the energy required by the coral host. This relationship is sensitive to temperature stress; as little as a 1°C increase often leading to collapse of the association. This sensitivity has led to interest in the potential of more stress tolerant algae to supplement or substitute for the normal Symbiodiniaceae mutualists. In this respect, the apicomplexan-like microalga *Chromera* is of particular interest due to its greater temperature tolerance. We generated a *de novo* transcriptome for a *Chromera* strain isolated from a GBR coral (*"GBR Chromera"*) and compared to those of the reference strain of *Chromera* ("Sydney *Chromera*"), and to those of Symbiodiniaceae (*Fugacium, Cladocopium and Breviolum*), as well as the apicomplexan parasite, *Plasmodium falciparum*. By contrast with the Symbiodiniaceae, the two *Chromera* strains had a high level of sequence similarity evident by very low levels of divergence in orthologous genes. Although KEGG categories provide few criteria by which true coral mutualists might be identified, they do supply a molecular rationalization for the ubiquitous association of *Cladocopium* strains with Indo-Pacific reef corals. The presence of HSP20 genes may underlie the higher thermal tolerance of *Chromera*.

Introduction

The ecological success of reef-building corals is generally attributed to their ability to establish mutualistic relationships with specific photosynthetic algae – members of the dinoflagellate family Symbiodiniaceae (formerly, the genus "*Symbiodinium*"). However, as is now widely appreciated, this relationship breaks down under

environmental stress, and coral reefs globally are under threat as a consequence of the increasing frequency of weather events that exceed the thermal thresholds of corals (Hughes *et al*., 2017). In addition to Symbiodiniaceae, a wide range of uncharacterized eukaryotes are associated with corals, including apicomplexan-related lineages (ARLs) (Clerissi *et al*., 2018) that can sometimes occur in high abundance (Kwong *et al*., 2019). Two of these ARLs isolated in association with corals, *Chromera velia* (Moore *et al*., 2008) and *Vitrella brassicaformis* (Obornik *et al.*, 2012), constitute the newly defined phylum Chromerida. As the closest free-living relatives of the parasitic Apicomplexa, these photoautotrophic alveolates are of considerable scientific interest (Moore *et al.*, 2008). Whole genome sequencing (Woo *et al*., 2015) has recently reaffirmed the close relatedness of *Chromera* and *Vitrella*.

The nature of the relationship between corals and *Chromera* has been a subject of debate. Given its photosynthetic ability (Moore *et al.*, 2008) and its ability to colonize coral larvae (Cumbo *et al*., 2013), it was initially thought that *Chromera* might be an alternative coral mutualist, potentially bringing the benefit of higher thermal tolerance than most Symbiodiniaceae (Visser *et al*., 2012; Chakravarti *et al*., 2019). However, several lines of evidence now imply otherwise. It has recently been shown that the transcriptomic response of the coral host post *Chromera* uptake (Mohamed *et al.*, 2018) differed markedly from that of the same coral to a mutualistic strain of Symbiodiniaceae (Mohamed *et al.,* 2016), and resembled the response to incompatible (“incompetent”) Symbiodiniaceae strains (Voolstra *et al*., 2009). The apparently hostile responses of coral larvae to *Chromera* during infection suggested that *Chromera* is more likely to be a parasite or a commensal of corals rather than a

mutualist (Mohamed *et al*., 2018). Other lines of evidence support this suggestion (Barott *et al*., 2011; Janouškovec *et al*., 2012, 2013), including a recent meta-analysis which implies that *Chromera* is near exclusively associated with coral biogenous sediments (Mathur *et al*., 2018).

The present work sought to address two specific issues, in both cases making use of a *de novo* transcriptome assembly generated for a strain of *Chromera* isolated from corals on the Great Barrier Reef (GBR). Whilst *Chromera* was originally isolated from a Sydney Harbor coral, it is known to have a wide distribution (Janouškovec *et al*., 2012; Visser *et al*., 2012), and the diversity within this monospecific genus has not been systematically explored. Given the metabolic diversity that is now known to exist (LaJeunesse *et al*., 2018) within what was previously known as "*Symbiodinium*", the extent to which conclusions about the coral-*Chromera* interaction based on the GBR isolate are generalizable is unknown.

The first goal of the present study was therefore to estimate the degree of divergence between the GBR strain of *Chromera* and that originally isolated from Sydney harbor. Given the evidence that *Chromera* is unlikely to be a coral mutualist, the second goal was to investigate the repertoires of genes that are thought to play roles in symbiosis and environmental stress tolerance in *Chromera* and compare these with those of three members of the Symbiodiniaceae. *Cladocopium goreaui* (formerly Clade C1 *Symbiodinium*) was isolated from a colony of *Acropora tenuis* on the GBR (Howells *et al.,* 2012), and is mutualistic with many Indo-Pacific corals, particularly *Acropora* species. *Breviolum minutum* (formerly Clade B *Symbiodinium*) was isolated from the

Caribbean coral *Orbicella faveolata*, and is a mutualist of Caribbean corals. *Fugacium kawagutii* (formerly Clade F *Symbiodinium*) was originally isolated in association with the Hawaiian reef-building coral *Montipora verrucosa* and the initial whole-genome analyses followed the assumption that *Fugacium* is a coral mutualist (Lin et al. 2015). However, *Fugacium* failed to infect juvenile corals (Yuyama *et al.,* 2016), and the consensus now is that *Fugacium* is probably a surface associate of corals rather than an endosymbiont (Liu *el al.,* 2018; LaJeunesse *et al.,* 2018; González-Pech *et al*., 2019). Thus, the expectation was that, with respect to metabolic repertoire, the *Chromera* strains would resemble *Fugacium* rather than the known coral mutualists, *Cladocopium* and *Breviolum*. Whilst the results suggest that small HSPs may account for the tolerance of *Chromera* to elevated temperatures, they were inconclusive with respect to the nature of relationships between corals and *Chromera* or *Fugacium*. The comparative analyses do, however, provide a molecular rationalization for the near ubiquitous association of *Cladocopium* with Indo-Pacific corals in general and with *Acropora* spp in particular.

Experimental procedures

Chromera culture and culturing conditions

A culture of *Chromera* (Mdig3 strain) from the University of Technology Sydney (Cumbo *et al.,* 2013) was used in this study and referred to as "GBR *Chromera"*. The identity of the culture was confirmed both by microscopy and by using *Chromera*-specific PCR primers (Supplementary information Figure 1, Table 1). This *Chromera* strain was originally isolated from the stony coral *Montipora digitata* (Acroporidae)

from Nelly Bay, Magnetic Island on the inner central part of the Great Barrier Reef. Cultures were maintained at 25 °C in Guillard's f/2 medium on a 12 h/12 h day and night regime. Note that *Chromera* was subjected to a variety of treatments prior to RNA extraction in order to ensure that the transcriptome assembly captured as many genes as possible. Culture conditions included control, dark stress, cold shock, heat shock, motile and mixotrophic (for details see Supplementary Methods). In all cases, exponentially growing cultures were separated and subjected to the treatment condition and harvested at the end of the experimental treatment. During culturing no antibiotics were used to exclude any potential contribution of the antibiotic treatment to the mRNA expression in the cultures.

RNA isolation and high-throughput 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 20s 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 a NanoDrop ND-1000 spectrometer, Qubit® 2.0 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 an 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 (Table 1).

Processing of Illumina data

The raw Illumina reads were filtered and adapters were clipped using TRIMMOMATIC (v0.32) (Bolger *et al.*, 2014). Reads were filtered based on quality and size as follows; both universal and indexed Illumina adapters were clipped, quality trimming was also performed by removing leading and trailing bases with Phred quality score < 25 and 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 (i.e. the chances that a base is called incorrectly is 1 in 100) and reads of < 50 bp were also excluded.

De novo assembly and annotation of transcriptome

The trimmed/filtered Illumina reads were used for *de novo* transcriptome assembly using Trinity (r20140717 version). 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. Minimum contig length of 500 and read normalization were specified. Trinity collects transcripts with shared sequence identity into clusters that are loosely related to genes. The longest isoform per cluster was selected using a custom Perl script from the assembled Trinity output "assembled transcriptome" for the purpose of

annotation. *Chromera* contigs were annotated by similarity search using batch BLASTX conducted locally against the Swiss-Prot protein database downloaded in September 2014 (E-value cut off 10^{-3} and maximum 20 hits). Raw BLASTX outputs were imported to Blast2GO suite (version 2.6.5) (http://blast2go.com/b2ghome) for functional annotation and Gene Ontology (GO) assignment. KEGG analysis was also performed using the KEGG Automatic Annotation Server (KAAS) (Moriya *et al.*, 2007) (http://www.genome.jp/kaas-bin/kaas_main) in order to obtain an overview of the associated metabolic pathways. The bi-directional best hit (BBH) method was used to obtain KEGG orthology (KO) assignments.

In order to validate the accuracy of the *de novo* assembly, reads were mapped back to the *de novo*-assembled transcriptome using the BOWTIE aligner version 0.12.7 (Langmead and Salzberg, 2012) with default mapping parameters. The percent of the mapped reads as proper pairs was used to assess the assembly quality. Moreover, BLASTN (E-value of ≤ 10^{-10}) was performed against bacterial genomes downloaded from the GenBank, NCBI to determine the percentages of putative bacterial transcripts in the dataset. The completeness of the transcriptome assembly was assessed using BUSCO v3.1.0 making use of the function *run_BUSCO.py* (--mode transcriptome) with the Eukaryota_odb9, alveolate_stramenophile and protists_ensembl data (retrieved 23 October 2019).

Phylogenomic analyses

A multi-gene phylogenetic analysis was performed to assess the relative phylogenetic distance between *Chromera* strains and infer their evolutionary

relationship to other Apicomplexans and Symbiodiniaceae. The protein sequences of GBR *Chromera* were predicted by Transdecoder (Haas *et al.*, 2013). Transcript nucleotide (CDS) and protein sequences for *Chromera* CCMP2878 strain "Sydney *Chromera*" were downloaded from CryptoDB (release-37; http://cryptodb.org/cryptodb/), and those for *Plasmodium* falciparum and *Toxoplasma gondii* from PlasmoDB and ToxoDB respectively (release-37; http://plasmodb.org/plasmo/ and http://toxodb.org/toxo/). Corresponding data for Symbiodiniaceae species were based on gene models from their respective genome sequencing projects; specifically, *Fugacium kawagutii* (Lin et al 2015) was obtained from the Symka Genome Database (http://web.malab.cn/symka_new/download.jsp), *Cladocopium goreaui* (Liu et al 2018) from ReFuGe 2020 site (http://refuge2020.reefgenomics.org/) and *Breviolum minutum* (Shoguchi *et al.,* 2013) from the OIST Marine Genomics online resource (https://marinegenomics.oist.jp/symb/viewer/info?project_id=21).

The longest transcript for each gene was extracted for all seven species and used to infer orthologous clusters with OrthoFinder (Emms and Kelly, 2015). A total of 693 orthogroups were found to have representative genes in all species, and of these 110 consist of single-copy genes in each species; they represent strictly orthologous gene sets. Amino acid sequences for these 110 orthologous sets were aligned using MAFFT (Katoh and Standley, 2013) and converted into corresponding codon alignment by pal2nal (Suyama *et al.,* 2006). Poorly aligned regions were removed using trimAl (Capella-Gutierrez *et al.,* 2009). Finally, IQ-TREE (Nguyen *et al*., 2015) was used to perform a partitioned phylogenetic analysis allowing independent

estimation of evolutionary model for each protein set. To summarise the effect of phylogenetic relatedness on overlap between gene repertoires, a Venn diagram was generated based on orthogroups present in four species (GBR- *Chromera*,, Sydney *Chromera*, *Plasmodium falciparum* and *Fugacium*). The Venn diagram was plotted using the R package VennDiagram (Chen and Boutros, 2011) based on orthogroup information generated with OrthoFinder (see above).

Comparative transcriptomics

Transcriptomes of Sydney *Chromera*, *Fugacium,* and the known mutualistic Symbiodiniaceae (*Cladocopium* and *Breviolum)* were also mapped against the KEGG database as previously performed for the GBR strain (using the bi-directional best hit (BBH) method). Genes mapped to different KEGG categories were calculated and compared. To better understand relationships between corals and both *Chromera* and *Fugacium*, repertoires of genes in categories that are important for symbiosis, such as ABC transporters, as well as those involved in processes such as nitrogen metabolism and stress tolerance, were compared. The overlap amongst these genes was plotted using the R package UpSetR https://github.com/hms-dbmi/UpSetR/ (Conway *et al.,* 2017).

Results and discussion

GBR Chromera transcriptome assembly and annotation

After confirming the identity of *Chromera* cultures using novel *Chromera*-specific PCR primers, a transcriptome assembly was generated from 166 million paired-end

Illumina reads (~33 million per library; Table 2). The number of putative genes (39 457, based on the longest transcript isoform per Trinity gene cluster) identified in the GBR *Chromera* isolate (Table 2) is comparable to those predicted for various Symbiodiniaceae isolates (30 000-49 000) based on transcriptome and genome data (Bayer *et al*., 2012;Shoguchi *et al.*, 2013;Rosic et al., 2015;Aranda *et al*., 2016; Liu *et al*. 2018 and Shoguchi *et al.,* 2018), but is higher than the number predicted from the *Chromera* genome of the reference strain (26 112 excluding TEs; Woo *et al*., 2015). As only 19.4% of the GBR *Chromera* genes had significant BLASTX hits against the Swiss-Prot protein database, the majority of *Chromera* genes code for unknown functions. This level of novelty is to be expected for organisms such as chromerids that are evolutionarily distant from well-characterized species, and has previously also been observed with Symbiodiniaceae and other dinoflagellates (Lin *et al.*, 2010;Bayer *et al.*, 2012; Stephens *et al.,* 2018). To assess the quality of the *de novo* assembled transcriptome, reads were mapped to the assembly and an average of 84% of the paired Illumina reads were mapped successfully (Table 1). The assembled transcriptome was judged to be relatively comprehensive on the basis of high percentages of reads mapping.

5225 of Swiss-Prot annotated genes (68.35%) were assigned to 38271 GO terms. Biological process (GO-BP) accounted for the majority of GO terms (22 205, 58.02%), followed by cellular component (GO-CC; 11 149, 29.1%) and molecular function (GO-MF; 4 917, 12.8%). Functions involved in *cellular process* and *metabolic process* (16% and 14%, respectively) were highly represented amongst GO-BP. In GO-MF, the most represented terms were *catalytic activity* (46%) followed

by *binding* (36%). In GO-CC, the terms *cell* (39%) and *organelle* (33%) were highly represented (Supporting information Figure 2). 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) (Supporting information Figure 3). Moreover, signal transduction and infectious diseases were the most highly represented pathways (Supporting information Figure 4, Table 2).

Several different approaches were used in order to assess the completeness of the assembled transcriptome. The KEGG annotation was searched for essential protein complexes/ pathways and the majority of genes for the pathways were found. Searched complexes included core cellular/ molecular protein complexes and pathways such as Ribosome biogenesis in eukaryotes, Ribosome, RNA polymerase, Spliceosome and Proteasome (Supporting information Table 3 and Figures 5-9). The use of BUSCO analysis resulted in a moderately high recovery (~60%) of conserved eukaryotic genes and relatively low recovery (~26-46 %) of conserved alveolate and protist genes, respectively, using default settings (Supporting information Table 4). In addition, 0.1% of the assembled *Chromera* sequences (contigs) had BLASTN hits to bacterial databases (E-value ≤ 10^{-10}) indicating very low bacterial contamination.

Sequence and functional similarities of the two Chromera strains

Phylogenomic analyses based on 110 orthologous single copy genes revealed a close relationship (branch length 0.07) between the GBR and Sydney strains of *Chromera* compared with divergences between different species within the Symbiodiniaceae (branch lengths 1.23 to 2.27) (Fig.1). Pairwise nucleotide similarity between the Sydney and GBR *Chromera* isolates was 96.2%, whereas the

corresponding figures for the genera of Symbiodiniaceae were 72-78%. A close relationship between the *Chromera* isolates was further supported by the presence of a relatively large number (14645) of shared orthologous genes and an average nucleotide identity of 99.12% based on alignments between these one-to-one orthologs. The overall distributions of the six main KEGG categories were similar in the two *Chromera* strains, one third of KEGG-annotated genes being assigned to *metabolism* (Supporting information Figure 10). This might reflect broadly similar functions and lifestyles. The only attempt to understand the nature of the coral-*Chromera* association used the Sydney strain (Mohamed *et al*., 2018). Given the high level of similarity between the GBR and Sydney harbour strains, the responses of corals to them are unlikely to differ significantly.

Focus on genes implicated in the symbiotic lifestyle

To gain additional perspectives on the coral-*Chromera* interaction, the assembled transcriptomes for the *Chromera* strains were compared to those for *Fugacium* (which is assumed to be non-mutualistic) and to those of *Cladocopium* and *Breviolum,* which are mutualistic Symbiodiniacae strains, focusing on categories of genes likely involved in symbiosis, particularly nitrogen metabolism, transport and stress.

Metabolic exchanges in coral-algal symbioses

Nutrient exchange between the symbiotic partners is of major importance in coral-*Symbiodinium* mutualisms (Meyer and Weis 2012, Davy *et al.,* 2012, Lin *et al.,* 2015, Aranda *et al.,* 2016). However, the nature of the translocated material(s) and

mechanisms underlying exchanges between the coral host and algal symbionts are still unclear. To assess the potential for nutrient exchange between *Chromera* and coral hosts, the representation of the KEGG pathways *Nitrogen metabolism* and *ATP-binding cassette (ABC) transporters* were investigated in the two *Chromera* strains and compared to those in the three members of the Symbiodiniaceae (*Fugacium, Cladocopium and Breviolum*).

The *Nitrogen metabolism* category was investigated on the basis that nitrogen cycling or conservation appears to be critical to the coral-*Symbiodinium* mutualism. Nitrogen is thought to be a growth limiting factor in nutrient-poor tropical waters, and many marine microbes have the ability to assimilate inorganic nitrogen (Pernice *et al.,* 2012), for a recent review see Radecker *at al*. (2015). The results of a survey of genes captured under the KEGG pathway identifier KO00910 (Nitrogen metabolism) are presented in Supporting information Table 4. Note that this KEGG pathway does not include ammonium transporters, although it has previously been reported that the genomes of Symbiodiniaceae encode multiple ammonium transport proteins (Aranda *et al.,* 2016).

In ocean waters, nitrate is generally the most abundant form of available nitrogen, concentrations often being at least an order of magnitude higher than those of ammonium and nitrite. However, *in hospite*, ammonium is likely to be the dominant nitrogen source available to intracellular symbionts. Thus, facultative symbionts (as most Symbiodiniaceae are thought to be) must not only be nutritionally versatile, but also able to regulate genes involved in the transport and assimilation of different N-

sources. As might be expected, all of the algae surveyed encode nitrate /nitrite transporters as well as enzymes required for assimilation of nitrogen in these forms. Although nitrate/nitrite transporters (NRTs) were present in all of the algae surveyed, some differences were apparent. Whereas the NRTs of *Breviolum*, *Fugacium* and *Chromera* were the MFS-type (K02575), this type was not detected in *Cladocopium*; rather, in this organism, components (K15577 and K15579) of a distinct ABC-type NRT were found. This difference may be significant; in cyanobacteria, the MFS-type has high affinity for both nitrate and nitrite, whereas the ABC-type has a much higher affinity for nitrate (Maeda and Omata, 2009).

ATP-binding cassette (ABC) transporters

The ABC class includes the largest number of transporters involved in either or both uptake and export of a wide range of substrates, including inorganic ions, carbohydrates and lipids. ABC transporters have been implicated in translocation of nutrients and metabolites in cnidarian symbioses (Davy *et al.,* 2012; Mathews *et al.,* 2017; Mohamed *et al.,* 2019), and hence were an obvious focus for comparative analyses.

As with the nitrogen metabolism category, surveying ABC transporter complements (Fig. 2) was largely unsuccessful in providing general molecular criteria by which known mutualists (*Breviolum* and *Cladocopium*) can be distinguished from commensals or parasites. One potentially significant difference, however, is that both *Breviolum* and *Cladocopium* encode K10111 members, which are nominal transporters of a variety of different sugars and which may be involved in

carbohydrate translocation *in hospite*, whereas proteins of this type were not detected in either *Fugacium* or *Chromera* (Supporting information Table 5). Unfortunately, the survey provides few other grounds for speculation about photosynthate translocation. Whilst a number of candidates for roles in sterol or lipid translocation were detected, these were generally not restricted to the coral mutualists. For example, components of the mla/lin type transport system for phospholipids/sterols/gamma-HCH (K02065 and K02066) were detected in *Fugacium* as well as *Cladocopium*, and the *Chromera* ABC repertoire does include possible sterol transporters (K05683, K05681, K08712).

Despite these analyses being uninformative on the issue of general characteristics of mutualists, they do provide grounds for speculation about the near-ubiquitous association of *Cladocopium* with Indo-Pacific corals (LaJeunesse *et al.,* 2018) and particularly with *Acropora* spp.. The ABC repertoire of *Cladocopium* by far exceeded those of all of the other organisms surveyed (Fig. 2), and included many genes identified only in this organism, amongst which were transporters for sugars, lipids and amino acids as well as inorganic nutrients. A range of candidate amino acid transporters were restricted to *Cladocopium*, one of which (HisP) is of particular interest in the context of coral symbioses, as it encodes a histidine transport ATP-binding protein (Supporting information Table 6). The *Cladocopium* strain included here was originally isolated from *Acropora tenuis* (Liu *et al.*, 2018), a member of the complex coral superfamily, and whereas robust corals (members of the other coral superfamily) have a complete histidine biosynthetic pathway, complex corals (like all other animals) must acquire this either from the diet or from their resident

endosymbionts (Ying *et al.,* 2018). Also intriguing in the context of mutualism is the detection only in *Cladocopium* of three components required for cystine transport (K02424, K10009, K10010); the interest in these stems from the fact that, amongst corals, members of the genus *Acropora* lack one of the enzymes required for cysteine biosynthesis (Shinzato *et al.,* 2011). In this respect, *Acropora* appears to be unique; all other corals surveyed had a compete cysteine pathway. The presence of an efficient cystine transport system may therefore underlie the near ubiquitous association of *Cladocopium* C1 with *Acropora* spp.

Fugacium – an evolutionary recidivist?

The ABC transporters identified in *Fugacium* include several candidates for roles in nutrient exchange – for example, K05641 encodes an ABCA1 protein known as the cholesterol efflux regulatory protein (CERP) that mediates efflux of cellular cholesterol and phospholipids (reviewed by Zhao and Lappalainen, 2012). Note that an ABCA1 gene was up-regulated in adult *Acropora millepora* colonies in the light, i.e. when transport of photosynthate from alga to host occurs (Bertucci *et al.,*2015). K02065, which was present in all three Symbiodiniaceae, encodes a phospholipid/cholesterol/gamma-HCH transport system ATP-binding protein, again with a potential function in lipid/sterol translocation. Indeed, some ABC family A transporters (ABCA7 and ABCA3) were upregulated in *Cladocopium* during establishment of symbiosis with coral larvae (Mohamed *et al.,* 2019). Thus, although *Fugacium* is now considered not to be a coral mutualist, the comparative analyses presented here do not rule this out. Whilst the nature of the relationship between *Fugacium* and corals is still unclear, the apparent contradictions could, however, be

rationalized by consideration of the phylogenetic position of *Fugacium*. The ability to become endosymbiotic is considered to be a defining characteristic of the family Symbiodiniaceae (LaJeunesse *et al.,* 2018); as a member of a derived clade within the Symbiodiniaceae, *Fugacium* may have undergone (or be undergoing) secondary loss of symbiotic potential but its genome still bear vestiges of this ancestral condition.

The small heat shock protein HSP20 might explain stress tolerance in Chromera

Molecular chaperones are important for refolding damaged proteins (Vierling, 1991), thus their involvement in stress tolerance is inevitable. Amongst these chaperones we looked at heat shock proteins in both *Chromera* and two symbiotic members of Symbiodiniaceae, *Cladocopium* and *Breviolum*. Despite genes coding for chaperones in the KEGG Orthology 03110 *Chaperones and folding catalysts* (including HSP90/70) being present in similar numbers in all of these genera, gene(s) coding for two HSP20 isoforms were found in both *Chromera* strains to the exclusion of the three Symbiodiniaceae algae (Fig. 3). Consistent with a potential role in stress tolerance the HSP20 was reported recently to be correlated with stress tolerance in reef-building corals where coral species containing more HSP20 genes were more stress tolerant (robust corals) than those with smaller numbers of HSP20 genes (complex corals) (Ying *et al.,* 2018).

Conclusions

This paper describes *de novo* assembly and annotation of a transcriptome for a *Chromera* strain isolated from a GBR coral, which was generated in order to

understand diversity within the species and to provide a better understanding of the metabolic capabilities and lifestyle of this photosynthetic apicomplexan alga. The GBR strain has a high degree of similarity with the Sydney (culture collection) strain, hence infection studies based on the former (e.g. Mohamed *et al.,* 2018) can be generalized to the latter. Comparisons between *Chromera*, the Symbiodiniaceae algae *Fugacium, Cladocopium* and *Breviolum* and the apicomplexan parasite *P. falciparum* based on specific categories of genes were inconclusive with respect to common molecular characteristics of mutualists, but suggest that HSP20 genes may underlie the higher thermal tolerance of *Chromera* compared to Symbiodiniaceae. The presence of specific genes implicated in mutualism suggest that *Fugacium* may have secondarily lost the ability to establish symbioses, and the comparative analyses provide a molecular rationale for the near-ubiquitous association of *Cladocopium* with Indo-Pacific corals.

Acknowledgments

The research was supported by the Australian Research Council through Grant CE140100020 to DM via the ARC Centre of Excellence for Coral Reef Studies at James Cook University. AM was supported by PhD scholarships provided by the Cultural Affairs and Mission sectors of the Egyptian Ministry of Higher Education and Scientific Research and James Cook University Postgraduate Research Scholarship (JCUPRS) and through the AIMS@JCU joint- scheme. The authors gratefully thank Dr Vivian Cumbo at Macquarie University for supplying *Chromera* cultures.

Conflict of Interest

The authors declare that they have no conflict of interest.

DATA AVAILABILITY. Data reported in this study have been submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE139820 including raw illumina data, transcriptome assembly and annotations.

Author contributions

AM and DM conceived and designed the study. AM conducted the experiments and generated the RNA-Seq data. AM performed bioinformatics required for transcriptome assembly, annotation and comparative analyses with guidance from CXC and MR. JZ and IC performed the phylogenomic analysis. AM, EB and DM interpreted that data and wrote the manuscript. All authors read the article and approved the final version.

Table legends

Table 1. Illumina sequencing and mapping statistics: number of reads and bases of raw and processed data after quality control for each library and percentages of reads successfully mapped to the GBR *Chromera* transcriptome.

Table 2. Summary statistics of the GBR *Chromera* transcriptome assembly using Trinity and annotation carried out based on the Swiss-Prot (SP), Gene Ontology (GO) and KEGG databases

Figure legends

Figure 1 (A) Unrooted phylogenetic tree based on maximum-likelihood phylogenetic analysis of 110 orthologous nucleotide sequences from Sydney *Chromera*, GBR *Chromera*, *P. falciparum*, *F. kawagutii*, *B. minutum* and C. *goreaui*. All nodes in the tree were fully supported (100% of bootstraps) based on 1000 bootstraps by ultrafast bootstrap in IQtree. The sizes of filled circles at some nodes in the tree indicate numbers of shared othologs. (B) Heat map summarizing Hamming distances between taxa based on nucleotide similarity in the 110 orthologous sequences used for phylogeny reconstruction.

Figure 2 UpsetR plot (which allows comparative visualisation of the presence/absence of genes of a pathway in different taxa) illustrating the intersection between genes of two KEGG pathways: (A) Nitrogen metabolism (KO00910) and (B) ABC transporters (KO02010) in the two *Chromera* strains, *Fugacium*, mutualistic Symbiodinianceae (*Cladocopium* and *Breviolum*) and the parasite *P. falciparum*. Each plot is generated from a binary matrix with unique KEGG identifiers for each pathway and each taxon in a column. The number of genes per taxon is represented by the length of the orange bars. The black dots represent presence/absence. The purple columns show the number of genes shared amongst

taxa or exclusive to one of them. So, reading the upper plot from the left there are 8 different genes (each gene is represented once) that are exclusively found in *Cladocopium* and not the others, second column shows 3 genes that are found in the 3 symbiodiniaceae and the 2 *Chromera*, but not present in *P. falciparum*, the third

column shows 3 genes that are found in all taxa considered, etc. The intersection size reflects the number of unique or shared genes.

Figure 3 Comparison of the repertoires of genes encoding chaperones and folding catalysts (KO03110) of *Chromera* strains and Symbiodiniaceae. (A) A five-way Venn diagram of the gene overlap amongst the two *Chromera* strains, mutualistic Symbiodiniaceae (*Cladocopium* and *Breviolum*) and *Fugacium*. (B) A bar graph illustrating numbers of HSP20, HSP70 and HSP90 genes identified in the five algal datasets.

References

Aranda, M., Li, Y., Liew, Y.J., Baumgarten, S., Simakov, O., Wilson, M.C. et al. (2016). Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. Sci Rep 6: 39734.

Barott, K.L., Rodriguez-Brito, B., Janouškovec, J., Marhaver, K., Smith, J.E., Keeling, P., et al. (2011) Microbial diversity associated with four functional groups of benthic reef algae and the reef-building coral Montastraea annularis. Environ Microbiol 13:1192–204.

Bayer, T., Aranda, M., Sunagawa, S., Yum, L.K., DeSalvo, M.K., Lindquist, E. et al. (2012). Symbiodinium Transcriptomes: Genome Insights into the Dinoflagellate Symbionts of Reef-Building Corals. Plos One 7.

Bertucci, A., Foret, S., Ball, E.E., and Miller, D.J. (2015) Transcriptomic differences between day and night in Acropora millepora provide new insights into metabolite exchange and light - enhanced calcification in corals. Mol Ecol, 24: 4489–4504.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972–1973. doi: 10.1093/bioinformatics/btp348

Chakravarti, L.J., Negri, A.P. and van Oppen, M. (2019) Thermal and Herbicide Tolerances of Chromerid Algae and Their Ability to Form a Symbiosis with Corals. Front Microbiol 10: 173. https://doi.org/10.3389/fmicb.2019.00173

Chen, H., and Boutros, P.C. (2011) VennDiagram: a package for the generation of highly‐customizable Venn and Euler diagrams in R. BMC Bioinformatics 12: 35.

Clerissi, C., Brunet, S., Vidal-Dupiol, J., Adjeroud, M., Lepage, P., Guillou, L., et al. (2018). Protists within corals: the hidden diversity. Front Microbiol 9: 2043. doi: 10.3389/fmicb.2018.02043

Conway, J.R., Lex, A., and Gehlenborg, N. (2017) Upsetr: An r package for the visualization of intersecting sets and their properties. Bioinformatics 33: 2938–40.

Cumbo, V.R., Baird, A.H., Moore, R.B., Negri, A.P., Neilan, B.A., Salih, A. et al. (2013). Chromera velia is Endosymbiotic in Larvae of the Reef Corals Acropora digitifera and A. tenuis. Protist 164: 237-244.

Davy, S.K., Allemand, D., and Weis, V.M. (2012) Cell biology of cnidarian‐dinoflagellate symbiosis. Microbiol and Mol Biol Rev 76: 229–261.

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16: 157.

Gonzalez-Pech, R.A., Bhattacharya, D., Ragan, M.A., and Chan, C.X. (2019) Genome evolution of coral reef symbionts as intracellular residents. Trends Ecol Evol https://doi.org/10.1016/j.tree.2019.04.010

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8: 1494-1512.

Howells, E.J., Beltran, V.H., Larsen, N.W., Bay, L.K., Willis, B.L., and Van Oppen, M.J. (2012) Coral thermal tolerance shaped by local adaptation of photosymbionts. Nature Climate Change 2: 116-120

Hughes, T. P., Kerry, J. T., Álvarez-Noriega, M., Álvarez-Romero, J. G., Anderson, K. D., Baird, A. H. et al. (2017). Global warming and recurrent mass bleaching of corals. Nature 543: 373–377. doi: 10.1038/nature21707

Janouškovec, J., Horak, A., Barott, K.L., Rohwer, F.L., and Keeling, P.J. (2012) Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. Curr Biol. 22: R518–R519.

Janouškovec, J., Horák, A., Barott, K. L., Rohwer, F. L., and Keeling, P. J. (2013). Environmental distribution of coral-associated relatives of apicomplexan parasites. ISME J. 7: 444–447. doi: 10.1038/ismej.2012.129

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol 30: 772–780. doi: 10.1093/molbev/mst010

Kwong, W. K., Campo, J., del Mathur, V., Vermeij, M. J. A., and Keeling, P. J. (2019). A widespread coral-infecting apicomplexan contains a plastid encoding chlorophyll biosynthesis. Nature. 568:103-107. doi: 10.1038/s41586-019-1072-z.

LaJeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R., and Santos, S.R. (2018). Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. Curr Biol, 28: 2570-2580.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods 9*: 357.

Lin, S., Zhang, H., Zhuang, Y., Tran, B., and Gill, J. (2010). Spliced leader-based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates. Proc Nat Acad Sci USA 107: 3-20038.

Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W. et al. (2015). The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. Science 350: 691-694.

Liu, H., Stephens, T.G., González-Pech, R., Beltran, V.H., Lapeyre, B., Bongaerts, P. et al. (2018) Symbiodinium genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. Comm Biol 1: 95.

Maeda, S-I., and Omata, T. (2009) Nitrite transport activity of the ABC-type cyanate transporter of the cyanobacterium Synechococcus elongatus. J Bacteriol 191: 3265-3272.

Mathur, V., del Campo, J., Kolisko, M., Keeling, P.J. (2018). Global diversity and distribution of close relatives of apicomplexan parasites. Environ Microbiol doi:10.1111/1462-2920.14134ms

Matthews, J.L., Crowder, C.M., Oakley, C.A., Lutz, A., Roessner, U., Meyer, E., Grossman, A.R. et al. (2017) Optimal nutrient exchange and immune responses operate in partner specificity in the cnidarian-dinoflagellate symbiosis. Proc Nat Acad Sci USA 114:13194-13199.

Meyer, E., and Weis, V.M. (2012) Study of cnidarian‐algal symbiosis in the “omics” age. Biol Bull 223:44–65.

Mohamed, A.R., Cumbo, V., Harii, S., Shinzato, C., Chan, C.X., Ragan, M.A. et al. (2016) The transcriptomic response of the coral Acropora digitifera to a competent Symbiodinium strain: the symbiosome as an arrested early phagosome. Mol Ecol 25:3127–41.

Mohamed, A.R., Cumbo, V.R., Harii, S., Shinzato, C., Chan, C.X., Ragan, M.A. (2018). Deciphering the nature of the coral-Chromera association. Isme J 12: 776-790.

Mohamed, A.R., Andrade, N., Moya, A., Chan, C.X., Negri, A.P., Bourne, D.G. et al. (2019). Transcriptomic insights into the establishment of coral-algal symbioses from the symbiont perspective. BioRxiv, 652131. https://doi.org/10.1101/652131.

Moore, R.B., Obornik, M., Janouskovec, J., Chrudimsky, T., Vancova, M., Green, D.H. 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 research 35: W182-185.

Nguyen, L.T., Schmidt, H. A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. doi: 10.1093/molbev/msu300

Oborník M, Vancová M, Lai DH, Janous□kovec J, Keeling PJ, Lukes□ J. Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of apicomplexa. Chromera velia Protist. 2011;162:115–30.

Pernice, M., Meibom, A., Van Den Heuvel, A., Kopp, C., Domart-Coulon, I., Hoegh-Guldberg, O., et al. (2012). A single-cell view of ammonium assimilation in coral-dinoflagellate symbiosis. ISME J. 6:1314–1324. doi: 10.1038/ismej.2011.196

Rädecker, N., Pogoreutz, C., Voolstra, C.R., Wiedenmann, J., and Wild, C. (2015). Nitrogen cycling in corals: the key to understanding holobiont functioning? Trends Microbiol 23:490–497. doi: 10.1016/j.tim.2015.03.008

Rosic, N., Ling, E.Y., Chan, C.K., Lee, H.C., Kaniewska, P., Edwards, D. et al. (2015) Unfolding the secrets of coral-algal symbiosis. ISME J 9:844-856.

Shinzato, C., Shoguchi, E., Kawashima, T., Hamada, M., Hisata, K., Tanaka, M., 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., Gyoja, F., Mungpakdee, S., Koyanagi, R. et al. (2013). Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. Curr Biol 23:1399-1408.

Shoguchi, E., Beedessee, G., Tada, I., Histata, K., Kawashima, T., Takeuchi T. et al (2018) Two divergent Symbiodinium genomes reveal conservation of a gene cluster for sunscreen biosynthesis and recently lost genes. BMC Genomics 19:458.

Stephens, T.G., Ragan, M.A., Bhattacharya, D., Chan, C.X. (2018) Core genes in diverse dinoflagellate lineages include a wealth of conserved dark genes with unknown functions. Sci Reports 8: 17175.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34: W609–W612.

Vierling, E. (1991) The roles of heat shock proteins in plants. Ann Rev Plant Biol 42:579-620.

Visser, P., Bintoudi, E., Boschker, E., Frade, P., van Bleijswijk, J., and Matthijs, H. (2012). Newly discovered coral endosymbiont Chromera is more thermotolerant than Symbiodinium.

in 12th International Coral Reef Symposium Abstracts (Cairns), 198. Available online at: www.icrs2012.com

Voolstra, C.R., Schwarz, J.A, Schnetzer, J., Sunagawa, S., Desalvo, M.K., Szmant, A.M, et al. (2009) The host transcriptome remains unaltered during the establishment of coral–algal symbioses. Mol Ecol. 18:1823–33.

Woo, Y.H., Ansari, H., Otto, T.D, Klinger, C.M., Kolisko, M., Michalek, J. et al. (2015). Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. eLife 4: e06974.

Ying, H., Cooke, I., Sprungala, S., Wang, W., Hayward, D.C., Tang, Y., et al. (2018) Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineage. Genome Biol. 19:175

Zhao, H., and Lappalainen, P. (2012). A simple guide to biochemical approaches for analyzing protein-lipid interactions. Mol Biol Cell 23, 2823–2830. doi: 10.1091/mbc.E11-07-0645

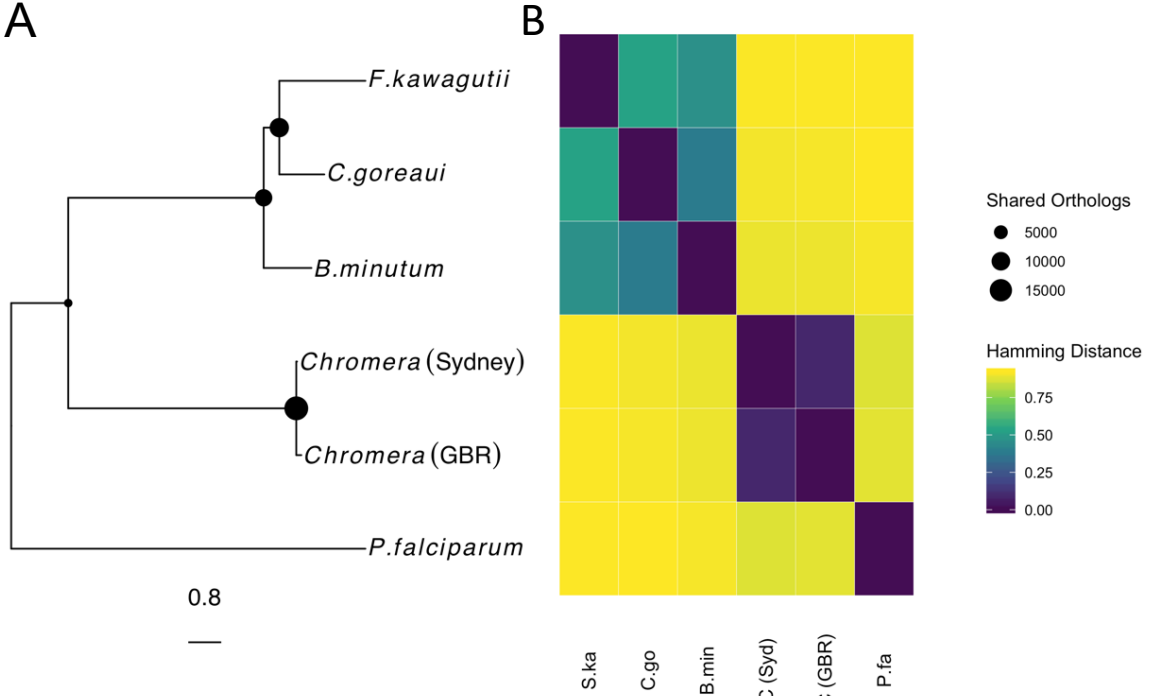
Table 1. Illumina sequencing and mapping statistics: number of reads and bases of raw and processed data after quality control (adapter removal, trimming and filtering low-quality bases) for each library and percentages of reads successfully mapped onto the GBR *Chromera* transcriptome. Note that *Chromera* was subjected to a variety of treatments prior to RNA extraction in order to ensure that the transcriptome assembly captured as many genes as possible.

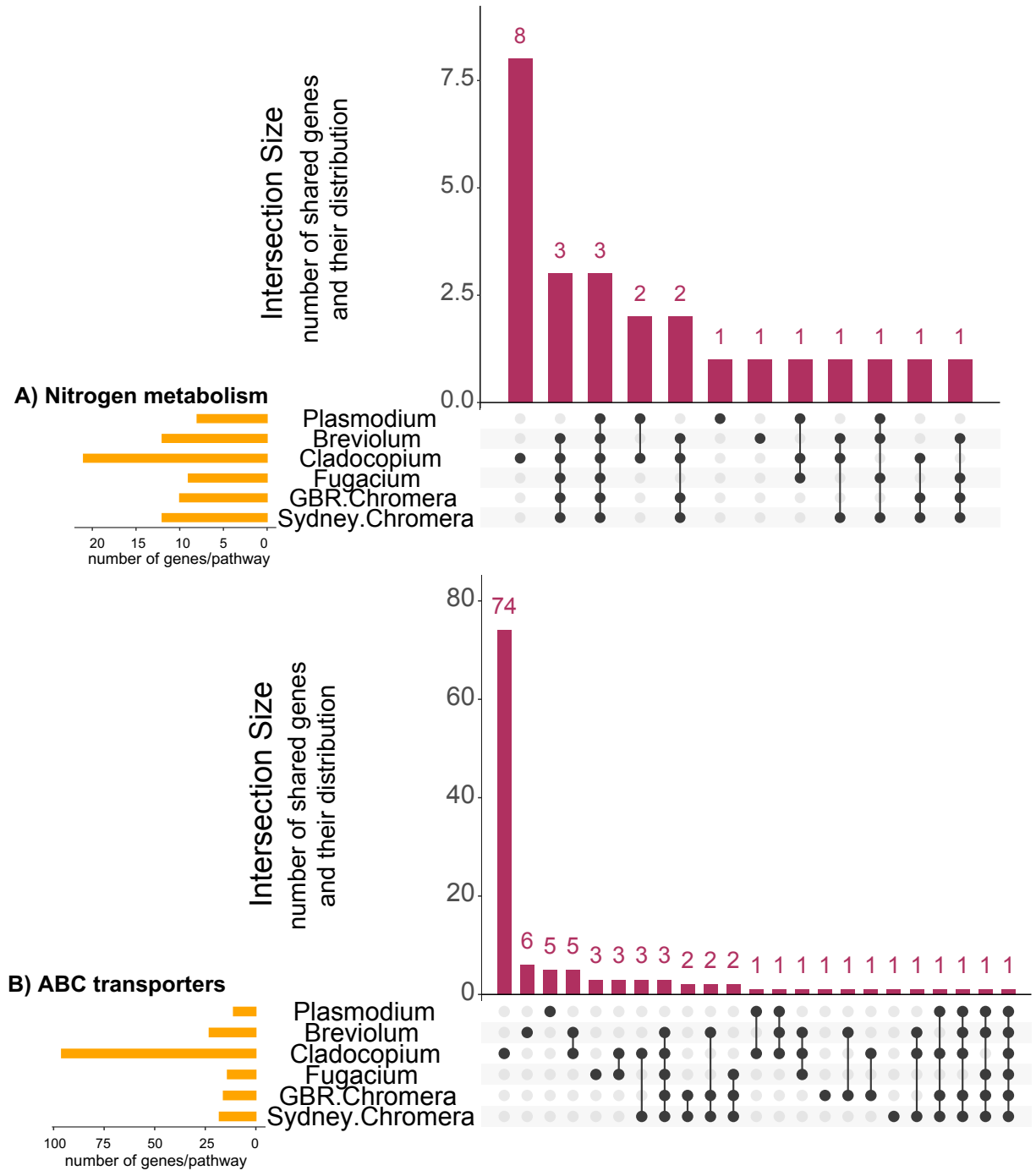
	RNA-Seq library/ culturing condition							
	Control	Cold	Heat	Dark	Motile	Mixotrophic	Total	Average
No. raw paired-end reads (M)	31.35	31.75	31.49	31.8	30.17	32.91	189.47	31.57
No. raw bases (Gb)	6.27	6.35	6.29	6.36	6.03	6.58	37.88	6.31
No. reads passed QC (M)	27.59	28.03	27.46	27.9	26.47	28.93	166.38	27.73
No. bases passed QC (Gb)	5.51	5.6	5.49	5.58	5.29	5.78	33.25	5.54
Mapping rate (%)	85.2	83.9	86.9	83.7	81.3	82.5	-	84.2

M= million; Gb= gigabase

Table 2. Summary statistics for the GBR *Chromera de novo* transcriptome assembly using Trinity and annotation based on Swiss-Prot (SP), Gene Ontology (GO) and KEGG databases

***Trinity outputs*:**	
Total Trinity 'genes'	39 457
Total Trinity transcripts	79 842
GC %	53.42
Statistics based on ALL transcript contigs:	
Contig N50	2 289
Median contig length	1 461
Average contig	1 838.03
Total assembled bases	146 752 195
Statistics based on ONLY LONGEST ISOFORM per 'GENE':	
Contig N50	2 220
Median contig length	1 403
Average contig	1 764.26
Total assembled bases	69 612 282
Annotation statistics based on longest isoform per gene data:	
Swiss-Prot (SP) database	7 644
Gene Ontology (GO) database	5 225
KEGG database	4 220

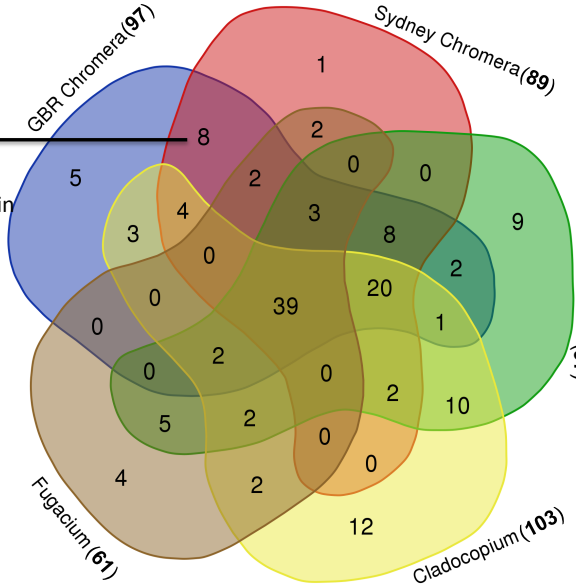

A
F.kawagutii
C.goreaui
B.minutum
Chromera (Sydney)
Chromera (GBR)
P.falciparum
0.8
B
S.ka
C.go
B.min
C (Syd)
C (GBR)
P.fa
Shared Orthologs
5000
10000
15000
Hamming Distance
0.75
0.50
0.25
0.00


A) Nitrogen metabolism
Intersection Size
number of shared genes and their distribution
8
3
3
2
2
1
1
1
1
1
1
1
7.5
5.0
2.5
0.0
Plasmodium
Breviolum
Cladocopium
Fugacium
GBR.Chromera
Sydney.Chromera
20
15
10
5
0
number of genes/pathway
B) ABC transporters
Intersection Size
number of shared genes and their distribution
74
6
5
5
3
3
3
3
2
2
2
1
1
1
1
1
1
1
1
1
1
1
80
60
40
20
0
Plasmodium
Breviolum
Cladocopium
Fugacium
GBR.Chromera
Sydney.Chromera
100
75
50
25
0
number of genes/pathway

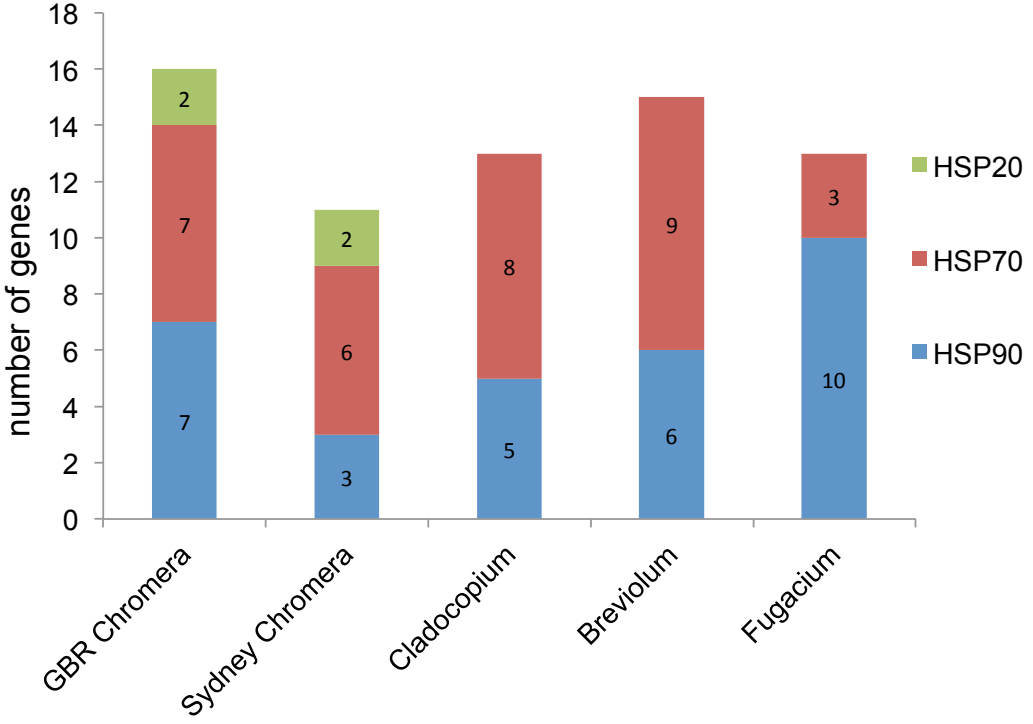
A) Chaperones and folding catalysts

8 molecular chaperones in both *Chromera*

1. **HSP20 family protein**
2. stress 70 protein chaperone microsome-associated protein
3. DnaJ homolog subfamily B member 8
4. molecular chaperone HscB
5. STIP1 homology and U-box containing protein 1
6. major intracellular serine protease (isp)
7. serine protease
8. minor extracellular serine protease Vpr



B) Heat shock protein genes



Supplementary Information

Close relationship between coral-associated *Chromera* strains despite major differences within the Symbiodiniaceae

Amin R. Mohamed[1, 2, 3, 4,5*], Cheong Xin Chan[6,7], Mark A. Ragan[6], Jia Zhang[3], Ira Cooke[3], Eldon E. Ball[2,8], David J. Miller[2,3*]

Affiliations:

[1]CSIRO Agriculture and Food, Queensland Bioscience Precinct, St Lucia Brisbane QLD 4067, Australia

[2]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville 4811, Queensland, Australia

[3]Molecular and Cell Biology, James Cook University, Townsville 4811, Queensland, Australia

[4]AIMS@JCU, Australian Institute of Marine Science, Department of Molecular and Cell Biology, James Cook University, Townsville 4811, Queensland, Australia

[5]Zoology Department, Faculty of Science, Benha University, Benha 13518, Egypt

[6]Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

[7]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia

*Correspondence: David Miller david.miller@jcu.edu.au

Supplementary Methods

Identity check of *Chromera* cultures

Microscopic check of Chromera culture

The starting *Chromera* culture was checked with an inverted microscope for protist and bacterial 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 resuspended 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 the manufacturer's instructions. DNA was eluted in 100 µl of elution buffer in a 1.5 ml tube. DNA was checked by running onto an agarose gel and a 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 (λ230) and the ratio 260/280 were recorded.

Polymerase chain reaction (PCR) Primer Design

Chromera large subunit ribosomal RNA gene, partial sequence, GenBank: EU106870.1, (Moore *et al.*, 2008) and *C. velia* 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 GenBank and used as templates for designing the PCR primers. The following primers (Table 1) were used to check the identity of the starting culture.

Amplification of Chromera ribosomal genes using PCR

Amplification of *Chromera* ribosomal genes was undertaken using specific primers (above) to obtain a PCR product ranging between 416 to 778 bp in size. PCR reaction was conducted in 50 µl using 1 µl of *Chromera* gDNA (approx. 100ng of DNA) as template. 1 µl of GoTaq® DNA polymerase and 2X GoTaq® reaction buffer and, 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 a UV trans-illuminator.

Chromera culture and culturing conditions

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 a variety of treatments before RNA isolation and preparation of cDNA libraries. 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 cultures were separated and subjected to the treatment condition and harvested at the end of the experimental treatment. During culturing no antibiotics were used to exclude any potential contribution of the antibiotic treatment to the mRNA expression in the cultures.

Illumina data quality check

Illumina raw reads from each paired end file were visualized using FASTQC version 0.11.2 in order to determine the quality of the data. In addition, reads were inspected for adapter contamination by searching for the IIlumina universal and indexed adapters.

Supplementary Figures

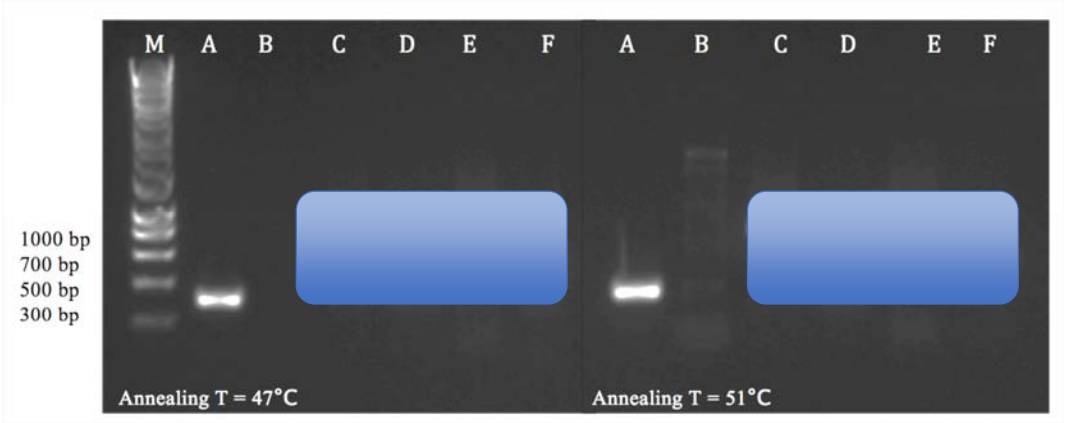


Fig.1. Amplification of *Chromera* ribosomal genes using newly-designed *Chromera*-specific PCR primers. M refers to the marker or DNA ladder. A refers to positive control reaction (*Symbiodinium* gDNA and *Symbiodinium*-specific primers), while B refers to negative control reaction (MQ water as a template).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

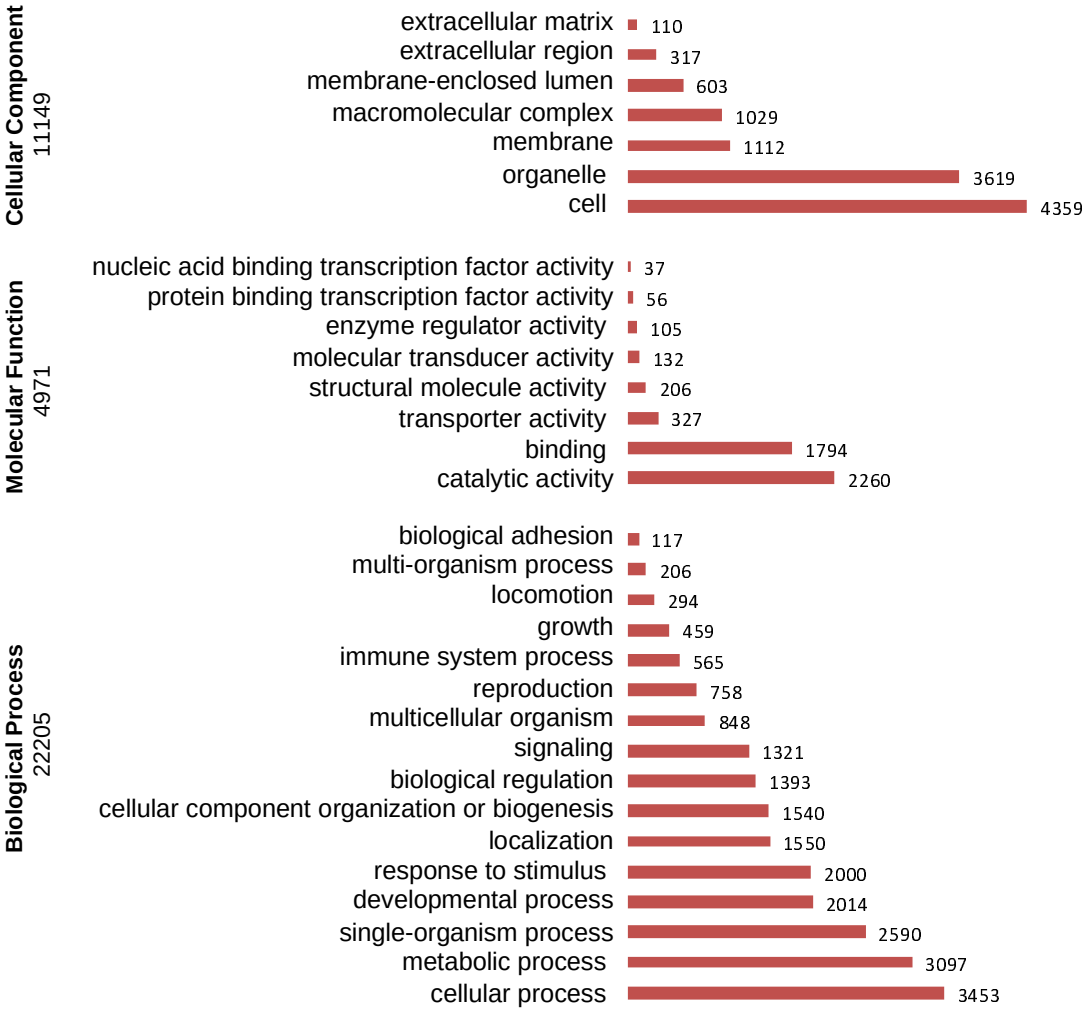


Fig.2. Gene Ontology (GO) assignment (2nd level GO terms) of the GBR *Chromera* transcriptome. Biological processes (A) constituted that majority of GO assignment of contigs (22,205 counts, 58.02%), followed by cellular components (C) (11,149 counts, 29.1%) and molecular function (B) (4,917 counts, 12.8%).

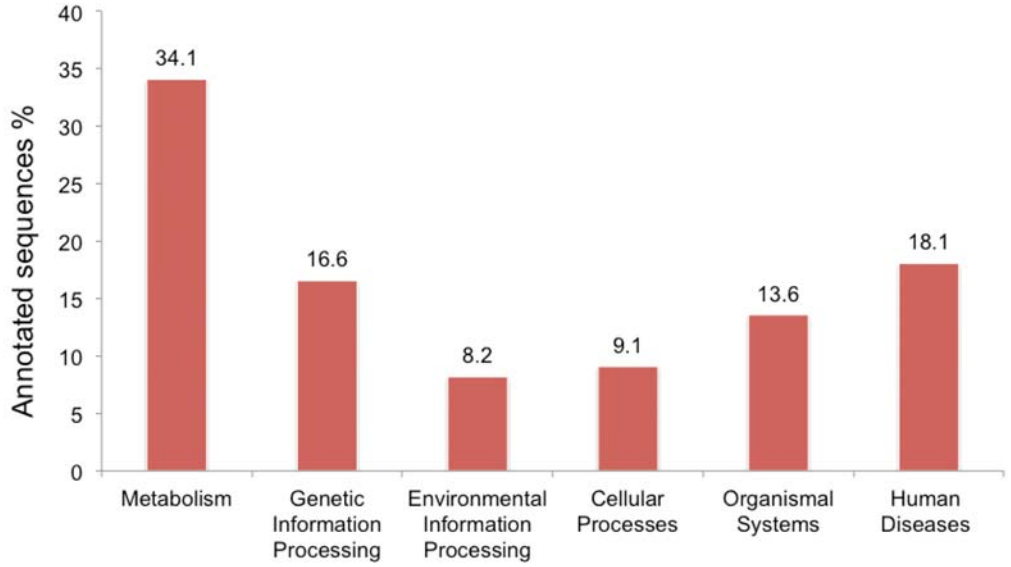


Fig.3. Main KEGG pathway category representation and percentages in the case of the GBR *Chromera* strain. Numbers above the bars give percent of annotated sequences in each category.

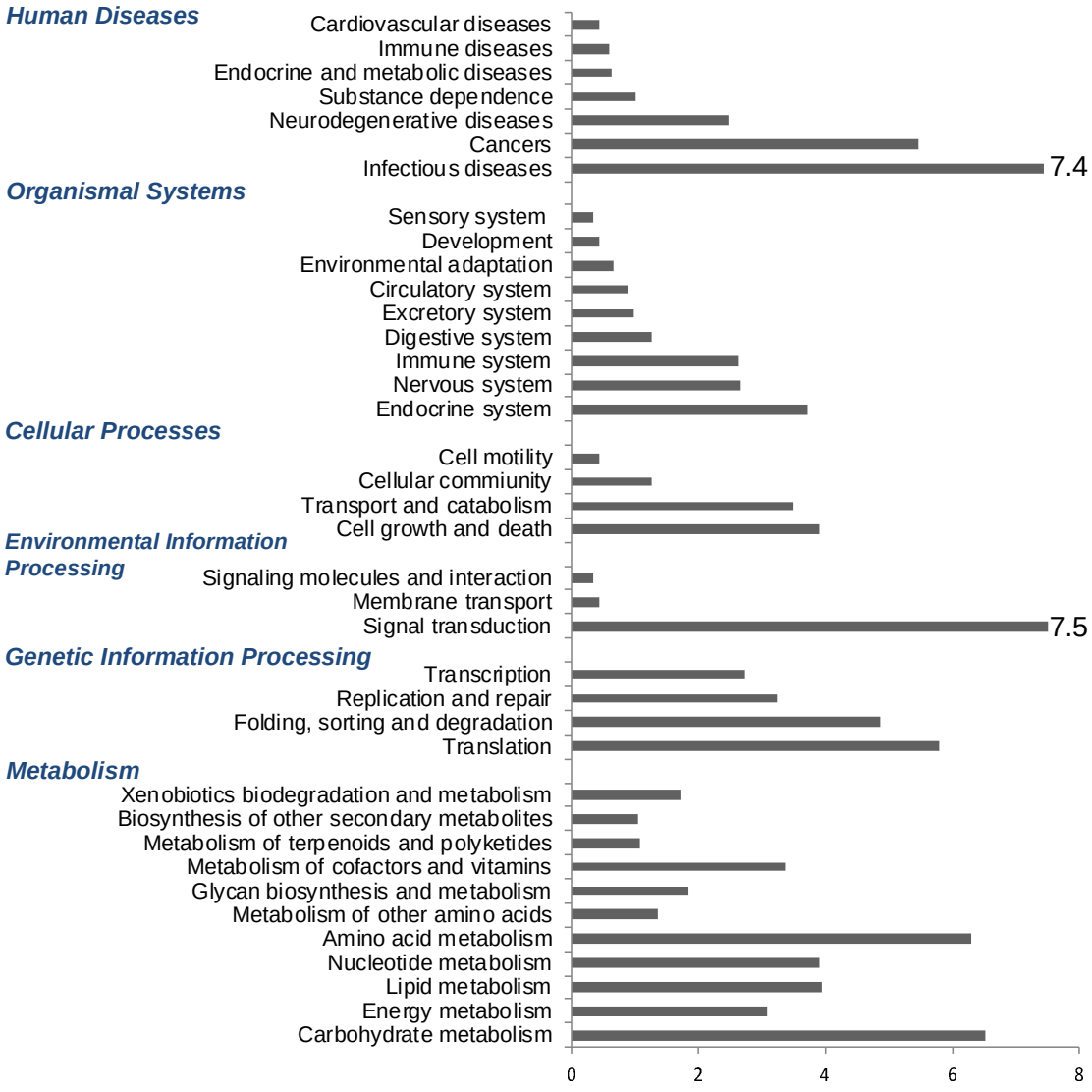


Fig.4. Distribution of KEGG pathways in transcriptome of the GBR *Chromera* strain. The chart shows the percentage of sequences assigned to each category.

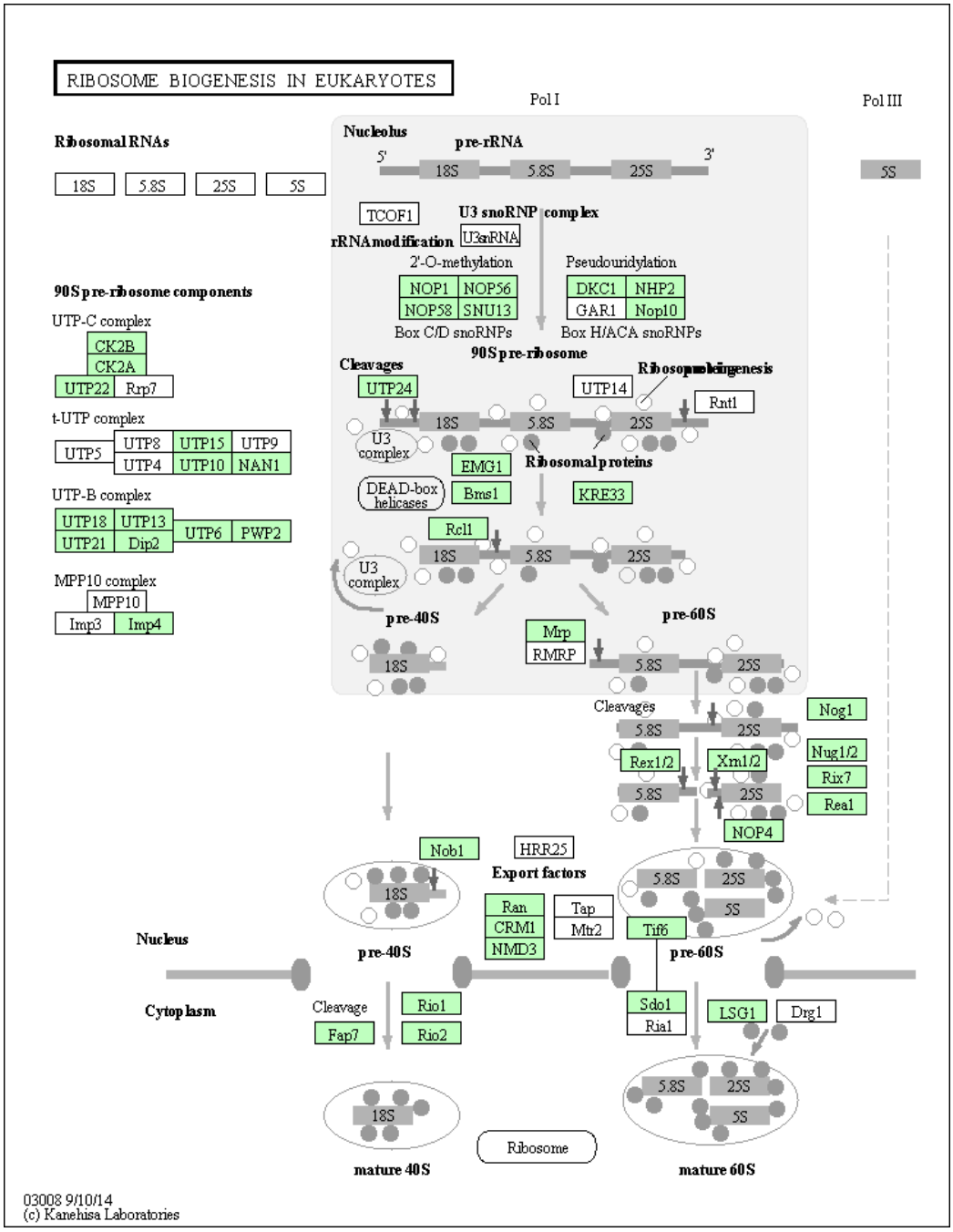


Fig.5. Ribosome biogenesis (eukaryotes) pathway (KO03008) identified in the *Chromera* transcriptome. KEGG pathways analysis shows *Chromera* orthologs involved in ribosome biogenesis (highlighted in green).

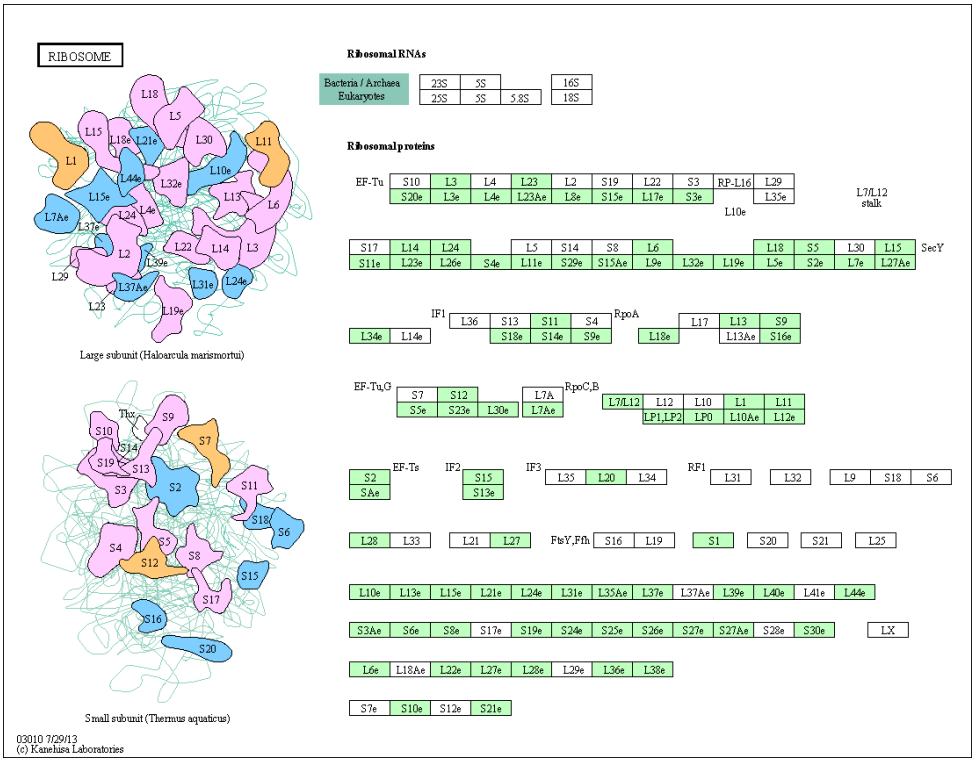


Fig.6. Ribosome pathway (KO03010) identified in the *Chromera* transcriptome. KEGG pathways analysis shows *Chromera* orthologs involved in ribosome (highlighted in green).

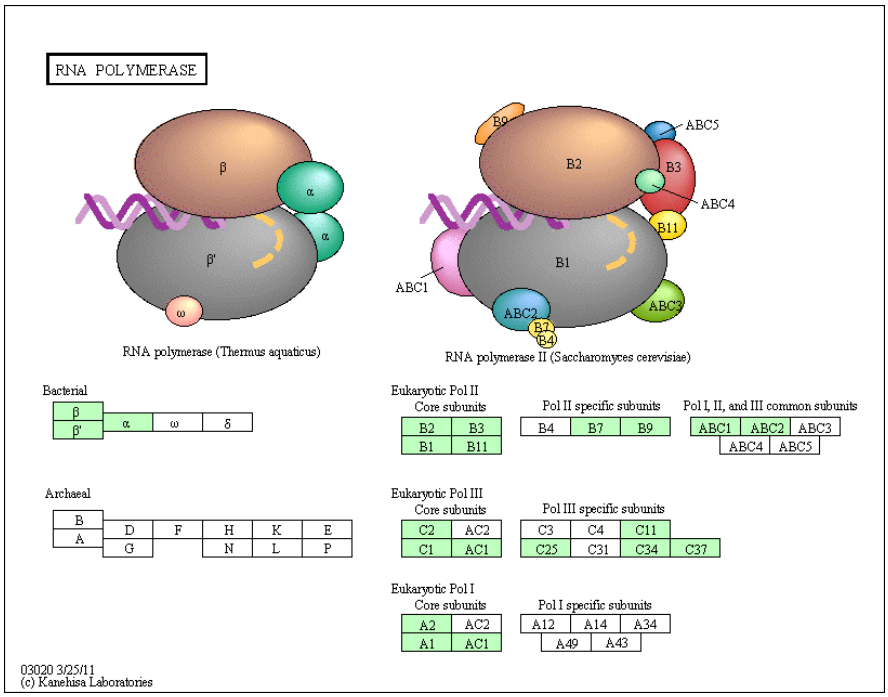


Fig.7. RNA polymerase pathway (KO03020) identified in the *Chromera* transcriptome. KEGG pathways analysis shows *Chromera* orthologs involved in RNA polymerase (highlighted in green).

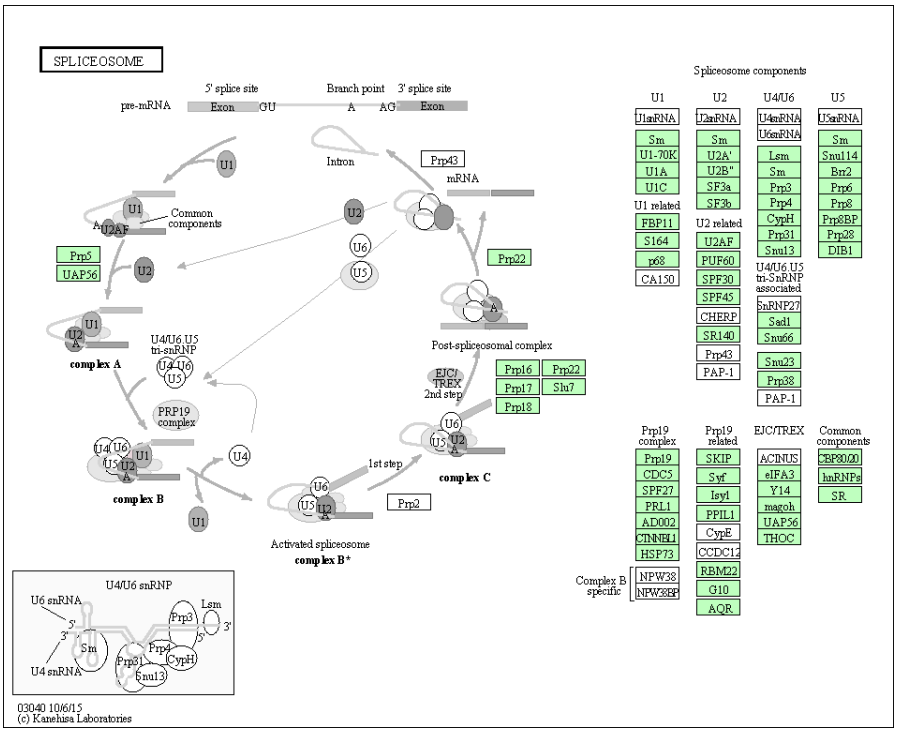


Fig.8. Spliceosome pathway (KO03040) identified in the *Chromera* transcriptome. KEGG pathways analysis shows *Chromera* orthologs involved in spliceosome (highlighted in green).

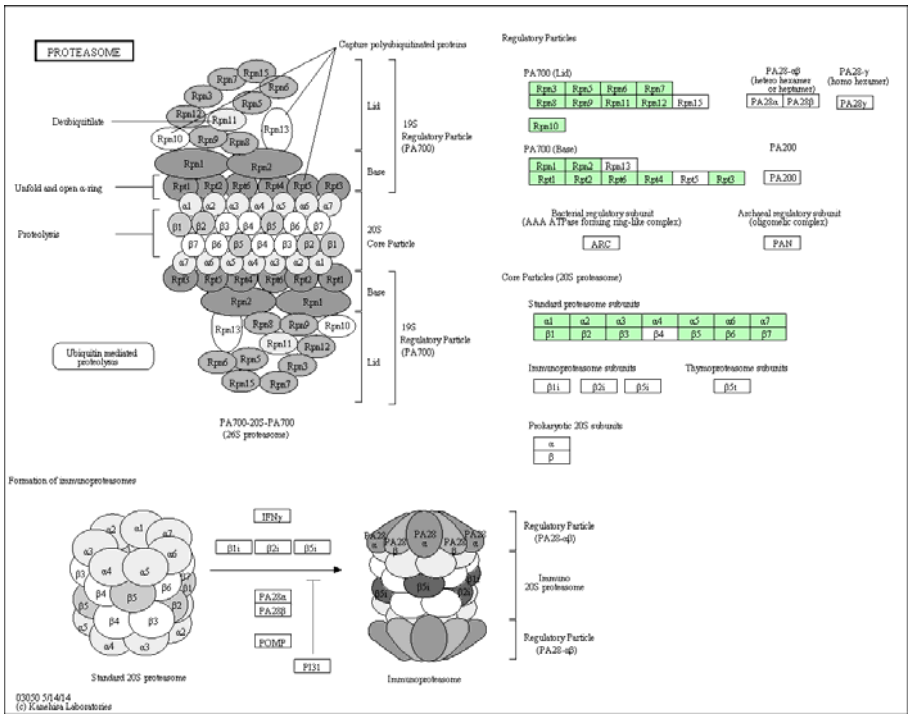


Fig.9. Proteasome pathway (hsa03050) identified in the *Chromera* transcriptome. KEGG pathways analysis shows *Chromera* orthologs involved in proteasome (highlighted in green).

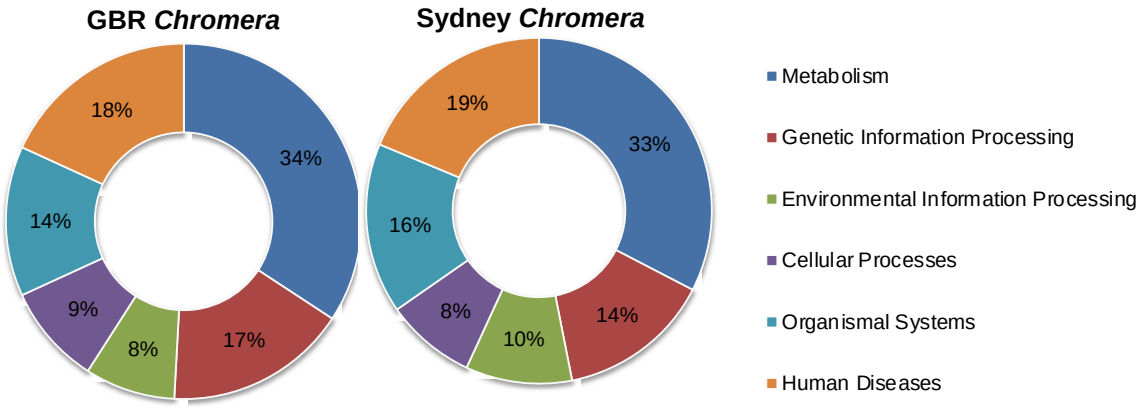


Fig.10. Overall distribution of the main KEGG categories in GBR and Sydney *Chromera*. The doughnut charts show the percentages of the sequences assigned to the six KEGG categories.

Table 1. List of *Chromera velia* specific PCR primers used for verifying the identity of the cultures. Primers were designed using the NCBI Primer-Blast tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/)

Primer Pairs	**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**

Table 2. Summary of KEGG orthology data for the GBR *Chromera* strain

KEGG categories	No. of KO-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

Table 3. 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

Table 4. Recovery of BUSCO genes in *de novo* transcriptome of GBR *Chromera*

BUSCO (Eukaryote, n=303)		**# genes**	**%**
	Single	180	59.4
	Duplicated	9	2.97
	Fragmented	54	17.8
	Missing	60	19.8
BUSCO (Alveolate-Stramenophile, n=234)		**# genes**	**%**
	Single	60	25.6
	Duplicated	1	0.43
	Fragmented	2	0.85
	Missing	171	73.07
BUSCO (Protist, n=215)		**# genes**	**%**
	Single	99	46
	Duplicated	1	0.46
	Fragmented	2	0.93
	Missing	113	52.55

Table 5. Genes mapped to the KEGG pathway Nitrogen metabolism “KO00910” in in GBR-*Chromera*, Sydney *Chromera*, *Cladocopium*, *Breviolum and P. falciparum.*

Provided as excel file

Table 6. Genes mapped to the KEGG pathway ABC transporters “KO02010” in GBR-*Chromera*, Sydney *Chromera*, *Cladocopium*, *Breviolum and P. falciparum.*

Provided as excel file