BIOCHEMICAL AND BIOINFORMATICS APPROACH TO THE STUDY OF LIPIDS AND THEIR BIOSYNTHETIC PATHWAYS IN *CHROMERA VELIA* AND *VITRELLA BRASSICAFORMIS*

by

Manoj Khadka

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Molecular Bioscience

Middle Tennessee State University

August 2014

Dissertation committee:

Dr. Jeffrey D. Leblond, Chair

Dr. Anthony L. Farone

Dr. J. Brian Robertson

Dr. Jeffrey L. Walck

Dr. Rebecca L. Seipelt-Thiemann

This dissertation work is dedicated to my father Hom Bahadur Khadka and mother Parbati Khadka without whose inspirations and love I would have never accomplished this journey. I also dedicate this work to my beloved wife Sangita Shrestha and my daughter Swarnika Khadka for their inspiration and support. And the dedication also goes to Anoj Khadka, my younger brother who always inspired me to put effort on my doctoral research and who always supported me in every steps of my life.

ACKNOWLEDGMENTS

I feel how difficult it would be to accomplish this dissertation work without the contributions of all the following people that I got the opportunity to work with.

I would like to acknowledge my deepest appreciation and respect to my committee chair, Dr. Jeffrey D. Leblond, for his inspiration, effort, and continuous guidance throughout my doctoral research project and dissertation work. I also would like to thank him for grooming me throughout the research work to overcome the challenges of research and to grow up as a professional, scholar, and a research scientist. I will always be missing your guidance, inspiration and perseverance during my career as a research scientist.

My deepest appreciation goes to my committee members, Dr. Anthony L. Farone, Dr. J. Brian Robertson, Dr. Jeffrey L. Walck and Dr. Rebecca L. Seipelt-Thiemann, for guiding me to accomplish my doctoral research with their expertise. Your suggestions, support, and encouragement have helped me to step up to this level. All of your guidance has made me to realize my potential and the challenges of being research scholar. I heartily appreciate Dr. A. Bruce Cahoon and Dr. Mohamed Salem for their guidance, constructive comments, and expertise support during my doctoral research. I also like to thank Dr. Caroll Hyrum from MTSU's department of Computational Science and Mathematics for providing me computational expertise and resources. His mentoring of different

computational skills helped me to accomplish the bioinformatics related project during my research work.

I like to acknowledge my colleagues, Jeremy L. Dahmen and Vernon Joshua Dodson, for suggesting and supporting during the research period. Your time for invaluable discussions and troubleshooting the research problems in the lab will always be remembered. I am thankful to all the members and friends from the Ph. D. (molecular biosciences) program for their support and suggestions.

I am indebted to all the members of the MTSU's biology and computational sciences and mathematics departments, and the influential people for their help and suggestions. Your kindness and friendliness have added an inspiration to work in a happy and relaxed environment during tedious lab days.

Lastly I like to thank my family, my wife Sangita Shrestha and my daughter Swarnika Khadka, for always being with me sharing the happiness, grief, and supporting me during adverse situation. You are my strength and inspiration to overcome any challenges during my career as a research scientist and scholar.

ABSTRACT

Vitrella brassicaformis and *Chromera velia* are marine microalgae that photosynthetic plastids of red algal origin, are closely related to non-photosynthetic apicomplexan parasites and are classified as the phylum Chromerida and familes Vitrellaceae and Chromeraceae, respectively. To dates, only one study has been performed on the galactolipid of *Chromera velia* and few other studies have been based on the ultrastructural and phylogenetic studies in Chromerida. We used a combination of mass spectrometry techniques and RNA-seq data to elucidate the sterol and galactolipid composition of the Chromerida, and to identify their biosynthetic genes, respectively. The identified genes were further utilized to perform the gene phylogenies of the Chromerida with red algae, dinoflagellates and apicomplexans.

Sterols are amphipathic, ringed lipids that are synthesized mostly by eukaryotes and few prokaryotes via mevalonate and/or non-mevalonate pathways. Some of the major functions of sterols include regulating membrane dynamics, hormonal activity, and acting a secondary messengers in developmental signaling. C *velia* produces two primary sterols 24-ethylcholesta-5,22-dien-3β-ol, and 24-ethylcholest-5-en-3β-ol, not common to dinoflagellates, but commonly found in other groups of algae and plants. Three sterol biosynthetic genes- sterol-24C-methyltransferase (*smt1*), farnesyl diphosphate farnesyl transferase (*fdft1*), and isopentenyl diphosphate isomerase (*idi*1) were

identified after expressed sequence tag (EST) analysis in *C. velia*. *V. brassicaformis* produces two sterols 24-ethylcholest-5-en-3β-ol and an unknown $C_{26:2}$ sterol along with several biosynthetic genes related to the non-mevalonate pathway for sterol biosynthesis.

The galactolipids, MGDG and DGDG, are the major plastid structural lipids in all photosynthetic organisms that play a role in the organization of thylakoid membrane providing stability to the photosystem I and II complexes in chloroplast. *V. brassicaformis* produces a diversity of galactolipids that primarily contained C_{20}/C_{14}, C_{20}/C_{16}, and C_{20}/C_{18} (*sn*-1/*sn*-2 regiochemistry) fatty acid components attached to the glycerol moieties of monogalactosyldiacylglycerol (MGDG) and digalactosyldiacyl glycerol(DGDG). These forms of MGDG and DGDG are different and structurally more diverse than the previously identified C_{20}/C_{20} forms that compose nearly the entirety of *C. velia*'s MGDG and DGDG.

The sterol and galactolipid diversity between these chromerids indicates possible differences in their biosynthesis. I also also present hypotheses for how other aspects of the galactolipid biosynthesis pathways, namely fatty acid elongation and desaturation, in the Chromerida must differ prior to the addition of galactose to MGDG and DGDG.

TABLE OF CONTENTS **PAGE**

LIST OF TABLES

LIST OF FIGURES

CHAPTERS **PAGE**

LIST OF APPENDICES

CHAPTERS **PAGE**

INTRODUCTION

CHROMERIDA: A NEWLY DESCRIBED PHYLUM

The Chromerida is a newly described phylum under the group Alveolata that comprises two recently discovered marine algae, *Chromera velia* and *Vitrella brassicaformis* (Moore et al. 2008; Oborník et al. 2012). The major characteristic that defines the Chromerida is a presence of single secondary plastid bounded by four membranes which are pigmented by chlorophyll *a*, violaxanthin and β-carotene. The presence of of a micropore, single mitochondrion with tubular cristae, and intracellularly assembled cilia are the other major features that define the Chromerida (Moore et al. 2008; Adl et al. 2012). The Chromerida are the closest known photosynthetic relatives of non-photosynthetic apicomplexan parasites, which evolved from a photosynthetic organism and currently retain a non-photosynthetic plastid termed an apicoplast (Janouškovec et al. 2010; Moore et al. 2008; McFadden et al. 1996).

BIOLOGY OF CHROMERIDA

C. velia, a photosynthetic alveolate was recently discovered from Sydney Harbor, Australia, and is free-living or associated with the stony coral, *Plesiastrea versipora* (Moore et al. 2008). *V. brassicaformis* was isolated from the Australian stony coral *Leptastrea purpurea* at One Tree Island, Great Barrier Reef (Oborník

et al. 2012). Both the species being classified under phylum Chromerida possess similar as well as different morphological and ultrastructural features. However, due to some divergence in morphological and ultrastructural characters, *C. velia* and *V. brassicaformis* have been classified under different families, the Chromeraceae and Vitrellaceae, respectively (Oborník et al. 2012).

C. velia differs from *V. brassicaformis* due to the presence of finger-like projections on the shorter flagellum, a pseudoconoid, a chromerosome, four-celled sporangia, and a non-canonical UGA codon encoding for tryptophan in the plastid. The presence of sporangia with dozens of cells, a multiple-laminated thick cell wall, pyrenoid, and highly compacted circular plastid genome are the distinguishing features of *V. brassicaformis* (Janouškovec et al. 2010; Oborník et al. 2011; Oborník et al. 2012).

V. brassicaformis and *C. velia* are differentiated from dinoflagellates, a closely related group of algae, in that they possess a photosynthetic secondary plastid that contains chlorophyll *a*, violaxanthin, and β-carotene but no chlorophyll *c* as in peridinin-containing dinoflagellates, the most abundant group of photosynthetic dinoflagellates (Moore et al. 2008; Adl et al. 2012). Chromerids, apicomplexans, and dinoflagellates share features such as tubular mitochondrial cristae, cortical alveoli, subpellicular microtubules, and micronemes. Furthermore, heterodynamic flagella, a tapered terminal portion of the long flagellum, and bacterial-derived Rubisco are common characteristics of chromerids and

dinoflagellates but not apicomplexans (Moore et al. 2008; Oborník et al. 2011; Oborník et al. 2012).

A phylogenetic analysis of plastid genes and nuclear genes showed that *C. velia* and *V. brassicaformis* represent two different photosynthetic lineages stemming from a red algal plastid ancestor that are closely related to apicomplexans (Janouškovec et al. 2010). *C.velia* and *V. brassicaformis* share 18 sets of plastid genes with heterokonts, hacrobians and red algae which are absent in green algae and plant plastids. The arrangement of genes in a plastidial ribosomal superoperon formed by fusion of the S10+*spc*+alpha operon and *str* operon in red algae, *C. velia* and *V. brassicaformis* and aforementioned characters further reinforce a red algae as an ancestor of *C. velia* and *V. brassicaformis* (Janouskovec et al. 2010).

SIGNIFICANCE OF GALACTOLIPIDS

The galactolipids, monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG), are generally the major plastid structural lipids in all photosynthetic organisms (Hölzl et al. 2007). These galactolipids play an important role in organization of the thylakoid membrane, growth and maintenance of photosynthetic ability, and stability of photosystem I and II complexes in the chloroplast (Guo et al.2005; Ivanov et al. 2006; Hölzl et al. 2007; Kobayashi et al. 2007). It has been reported previously that radiolabeled

UDP-galactose has been incorporated into the MGDG and DGDG of apicomplexan parasites, *Plasmodium falciparum* and *Toxoplasma gondii*, possibly as part of a non-photosynthetic, relic plastid called an apicoplast (Maréchal et al. 2002). A more recent study by Botté et al. (2013) showed the presence of atypical lipids like sphingomyelin and ceramides, but no galactolipids, in the apicoplast of *P. falciparum*. Thus, the composition of the apicoplast membrane appears to be an evolving area of study and the study. The galactolipid composition in Chromerida should provide important information about the possible galactolipid composition of apicomplexans and their biosynthetic genes should reveal how closely the Chromerida is related with other groups of algae.

SIGNIFICANCE OF STEROLS

Sterols are amphipathic ringed lipids that are commonly synthesized by eukaryotes and some prokaryotes (Pearson et al. 2003) via one of two pathways, a classical mevalonate pathway and a non-mevalonate pathway (Eisenreich et al. 2004, Hunter 2007, Lichtenthaler 1999, Rodriguez-Concepcion et al. 2002, Rohmer et al. 1999, Schwender et al. 1996, Takagi et al. 2000, Wanke et al. 2001). The two pathways lead to a common sterol biosynthetic intermediate, isopentenyl diphosphate, utilizing different precursors. Six enzymatic steps lead to formation of isopentenyl diphosphate via the mevalonate pathway.

Acetoacetyl-CoA thiolase first converts acetyl-CoA into acetoacetyl CoA which then is converted into 3-hydroxy-3-methylglutaryl CoA (HMG CoA) by 3-hydroxy-3-methylglutaryl CoA synthase. HMG-CoA is converted to mevalonate and then to phosphomevalonate by the enzymes HMG-CoA reductase and phosphomevalonate kinase, respectively. Phosphomevalonate is converted into diphosphomevalonate by the enzyme phosphomevalonate kinase and diphosphomevalonate is ultimately converted to isopentenyl diphosphate by an enzyme mevalonate diphosphate decarboxylase (Nes 2011). The non-mevalonate pathway (Rohmer et al. 1996) utilizes glyceraldehyde 3-phosphate and pyruvate as precursors to synthesize to synthesize 1-deoxy-D-xylulose-5-phopsphate (DOXP) by the enzyme DOXP synthase. DOXP is further converted to 2-C-methyl-D-erythritol-4-phosphate (MEP) and then to cytidine 5'-diphosphate-2-C-methyl-D-erythritol (CDP-ME) by two enzymes DOXP-reductoisomerase and CDP-ME-synthase respectively. CDP-ME is further converted to CDP-methyl-D-eryhritol-2-phosphate (CDP-ME2P), 2-C-methyl-D-erythritol-2,4-cyclo-diphosphate (MEcPP), 4-hydroxy-3-methyl-2-(E)-butenyl-diphosphate (HMBPP), and then to isopentenyl diphosphate by four enzymes CDP-ME-kinase, MEcPP-synthase, HMBPP-synthase, and HMBPP-reductase respectively (Lichtenthaler 2010). The two molecules of isopentenyl diphosphate are condensed with dimethylallyl pyrophosphate (DMAPP) (Eisenreich et al. 1998, Kuzuyama 2002) to form a condensed product, squalene.

Several enzymes act on squalene to convert it into different types of sterols (Grunwald 1975, Benveniste 2004, Goodwin 1979, Benveniste 1986, Nes 2011). After the synthesis of squalene, the sterol biosynthesis pathway bifurcates into two biosynthetic branches to produce lanosterol and cycloartenol. Lanosterol ultimately produces ergosterol in fungi and cholesterol (mainly in vertebrates). The cycloartenol branch is commonly found in plants, and leads to formation of various sterols like campesterol, sitosterol, and stigmasterol (Desmond et al. 2009).

Among apicomplexans *Toxoplasma gondii* is a feline parasite that is unable to synthesize sterols *de novo*. However, the exogenous cholesterol is salvaged into the parasite through low density lipoprotein (Coppens et al. 2000). *Plasmodium* species though utilize cholesterol synthesized by host hepatic cells (Labaied et. al. 2011) posses endoplasmic reticulum CoA: cholesterol acyltransferases that convert cholesterol to cholesteryl esters (Nishikawa et. al., 2005).

This dissertation work consists of four chapters and each chapter represents the published articles in several journals like Journal of Eukaryotic Microbiology, Algological Studies, and European Journal of Phycology. My work basically focuses on the lipid biochemistry of Chromerida. I worked on two different lipid groups in Chromerida to profile the lipid composition and study their biosynthetic pathway. I conducted the comparative studies of different lipid composition and biosynthetic genes between *C. velia* and *V. brassicaformis* and other closely

related groups of algae. In the first chapter of this dissertation, I studied the sterol composition and biosynthetic genes of *C. velia*. The sterol composition and biosynthetic genes were compared with dinoflagellates and other closely related groups. After the discovery of *V. brassicaformis,*I performed comparative study of the galactolipid composition and biosynthetic genes in Chromerida. The third chapter is about the sterol composition and biosynthetic genes of *V. brassicaformis*. This study is important to make comparison of sterol composition and biosynthetic genes *C. velia* and *V. brassicaformis*. The fourth chapter of my dissertation deals with the biochemical and genomic comparison of galactolipid between dinoflagellates and *C. velia*. I also hypothesized how fatty acids are incorporated in galactolipid in *Lingulodinium polyedrum* and *C. velia*.

The diversity of sterol and galactolipid composition has been described in *C. velia* and *V. brassicaformis*. The biosynthetic genes have been discovered in those species utilizing transcriptome data. The study showed that *C. velia* produces a greater diversity of sterol than *V. brassicaformis*. The major sterol produced by *C. velia* and *V. brassicaformis* are 24-ethylcholesta-5,22E-dien-3β-ol (approx. 86%) and 24-ethylcholest-5-en-3β-ol (approx. 83%) respectively. *V. brassicaformis* produces diversity of galactolipid than *C. velia* when grown at 20℃. The galactolipid composition of *C. velia* is rich in C_{20} and C_{18} fatty acids whereas *V. brassicaformis* produces galactolipids that have C_{20}, C_{18}, C_{16}, and C_{14} fatty acids attached to their *sn*-1 and *sn*-2 carbon. Both these lipids are

absent in apicomplexan parasites. After the study of sterol and galactolipid composition of Chromerida, I conclude that similar sterol and galactolipid would have synthesized by apicomplexan parasites when they were photoautotrophic in the past. The phylogenetic studies using sterol and galactolipid biosynthetic genes suggest the red algal origin of Chromerida.

OBJECTIVES

GENERAL OBJECTIVE

- To compare the sterol and galactolipid composition and their biosynthetic genes in the Chromerida utilizing biochemical and bioinformatics approaches.

SPECIFIC OBJECTIVES

- To make comparisons of the sterol and galactolipid composition in *C. velia* and *V. brassicaformis* to assess potential chemotaxonomic relationships.
- To determine if specific sterols or galactolipids can be utilized as distinguishing biomarkers of either species.
- To analyze the transcriptome of *C. velia* and *V. brassicaformis* to discover sterol and galactolipid biosynthetic genes.
- To make phylogenetic inference of Chromerida with apicomplexans and other groups of algae utilizing the gene sequences that are responsible for sterol and galactolipid biosynthesis.

CHAPTER I

STEROL COMPOSITION AND BIOSYNTHETIC GENES OF THE RECENTLY DISCOVERED PHOTOSYNTHETIC ALVEOLATE, *CHROMERA VELIA* (CHROMERIDA), A CLOSE RELATIVE OF APICOMPLEXANS

INTRODUCTION

Chromera velia is a recently discovered marine alveolate that is, based on nuclear-encoded large subunit (LSU) and small subunit (SSU) rDNA genes (Moore et al. 2008) and a concatenated nuclear gene dataset (Janouškovec et al. 2010), the closest free-living, photoautotrophic relative of non-photosynthetic apicomplexans and colpodellids. It is also more distantly related to perkinsids, which are non-photosynthetic parasites of shellfish, and to photosynthetic and non-photosynthetic dinoflagellates (Moore et al. 2008). As put forth by Oborník et al. (2011), "*C. velia* is, at the ultrastructural level, a unique mixture of characters so far described in apicomplexans, colpodellids, dinoflagellates and related groups." In addition, both the photosynthetic plastid of *C. velia* and the non-photosynthetic apicoplast or relict plastid of apicomplexan parasites have the same red algal ancestry as the photosynthetic plastids of peridinin-containing dinoflagellates (Janouškovec et al. 2010; Moore et al. 2008).

Sterols, ringed lipids required by eukaryotes to maintain membrane integrity, can be either synthesized *de novo* as in photoautotrophs or obtained from another

organism, such as the parasitic host of a heterotrophic apicomplexan parasite (Nishikawa et al. 2005; Sehgal et al. 2005). For those organisms, particularly microalgae and other protists, that synthesize sterols *de novo*, there is great potential for using them to derive chemotaxonomic relationships. For example, clustering dinoflagellates based on their sterol composition forms groups that generally support their phylogenetic organization as based on SSU rRNA gene sequences (Leblond et al. 2010). Sterols also have the potential to serve as biomarkers for particular algae, and this is particularly true for dinoflagellates. For example, the genera *Karenia*, *Karlodinium*, and *Takayama* produce gymnodinosterol (i.e. (24*R*)-4α-methyl-5α-ergosta-8(14),22-dien-3β-ol) and its 27-*nor* isomer brevesterol as their predominant sterols (Mooney et al. 2007; Giner et al. 2003; Leblond and Chapman 2002). These two sterols are not found as major sterols of other dinoflagellate genera, and are hence considered biomarkers for the Kareniaceae.

To date, despite its evolutionary importance, the sterol composition of *C. velia* is unknown. Because *C. velia* is a photoautotrophic alveolate, which does not depend on another organism (i.e. a host) to provide it with sterols, and because it has a close evolutionary relationship to apicomplexan parasites while being more distantly related to colpodellids, perkinsids and dinoflagellates, we were therefore interested to determine its sterol composition with the following questions in mind: 1) Does *C. velia* produce sterols that link it chemotaxonomically to dinoflagellates, and if not,

then 2) Does *C. velia* produce a sterol profile that may serve as a potential set of lipid biomarkers? and 3) Can genes responsible for sterol biosynthesis be identified in *C. velia* and can these be utilized as a means of discerning information on the steps involved in its sterol biosynthesis? The first description of the sterols of *C. velia* is presented herein.

MATERIALS AND METHODS

Culture and lipid extraction

Chromera velia CCMP 2878 was acquired from the Provasoli-Guillard National Center for Marine Algae and Microbiota (West Boothbay Harbor, ME, USA), and was grown autotrophically in L1 medium (Guillard and Hargraves 1993) in duplicate at room temperature under a 14/10 h light/dark cycle at an irradiance of approximately 50 µmol photons$\cdot m^{-2} \cdot s^{-1}$. Cells were harvested during the exponential phase of growth and lipids extracted according to the techniques described by Leblond and Chapman (2000).

Processing and analysis of sterols

Saponification of sterols was performed according to the techniques specified by Leblond and Chapman (2002). After saponification, trimethylsilyl ether (TMS) derivatives of sterols were formed with 0.5 ml *N*,*O-bis*(trimethylsilyl)-trifluoroacetamide containing 1% (v/v) trimethylchlorosilane at 80° C for 0.5 h. The

reagent was evaporated under a stream of nitrogen and the derivatives redissolved in 20 µl of 1:1 hexane/ methyl-*tert*-butyl ether.

Derivatives were analyzed via gas chromatography/mass spectrometry (GC/MS) on a Finnigan GCQ (Thermo Finnigan, Waltham, MA, USA) using an RTX-5MS capillary column (i.e. 30 m length, 0.25 mm inner diameter, 0.25 µm film thickness, Restek, Bellefonte, PA, USA) with the following GC/MS conditions: splitless injection with injector set at 280 ℃, column held at 50 ℃ f or 1 min, increased to 170 ℃ at 10 ℃ per min, increased to 300 ℃ at 5 ℃ per min, an d held at 300 ℃ for 20 min. The transfer line was set at 275 ℃, and helium was del ivered at a constant velocity of 40 cm per sec. Relative retention times (RRT) to cholesterol were calculated according to the methodology of Jones et al. (1994).

Identification of genes involved in sterol biosynthesis

Chromera velia EST sequences (2,856) were retrieved from the National Center for Biotechnology Information (Šlapeta and Carter 2010) and used in EGassembler (Masoudi-Nejad et al. 2006) to merge overlapping EST sequences into contigs. BLASTx (Altshul et al. 1990) was employed on the resulting merged and unmerged sequences to identify homologous sequences with possible enzymatic function in the sterol biosynthesis pathway. Accession numbers, species, amino acid sequences, E-values, query coverage values, and per cent identity values were recorded. The amino acid sequences encoded by the *C. velia* gene and the top nine

matches for each species were aligned using CLUSTALW via the Protein Information Resource (Larkin et al. 2007). Amino acid sequences of *C. velia* encoded genes *smt1* (sterol-24C-methyltransferase) and *fdft1* (farnesyl diphosphate farnesyl transferase, squalene synthase) were trimmed at the amino terminal ends to enable reproduction of alignments consistent with BLASTx pairwise alignments.

RESULTS

Sterols produced by *Chromera velia*

Examination of TMS-derivatives of sterols via GC/MS from the total lipids of *C. velia* revealed the presence of five compounds: 1) the C_{28} sterol, 24-methylcholesta-5,22E-dien-3β-ol (*m/z* 470 as its TMS ether derivative, 43.37 min retention time, 1.11 retention time relative to cholesterol, approx. 2% of total sterols); 2) and 3) the C_{29} sterols, 24-ethylcholesta-5,22E-dien-3β-ol (*m/z* 484, 45.77 min, 1.49, approx. 86%) and 24-ethylcholest-5-en-3β-ol (*m/z* 486, 46.63 min, 1.62, approx. 5%); 4) an unknown $C_{29:2}$ sterol (*m/z* 484, 47.00 min, 1.68, 2%); and 5) an unknown $C_{30:2}$ sterol (*m/z* 498, 48.61 min, 1.93, approx. 4%).

The mass spectra for the three known sterols were indistinguishable from the mass spectra of authentic standards (Fig. 1). Key fragments in the mass spectrum of 24-methylcholesta-5,22-dien-3β-ol were M^+ - H - side chain (*m/z* 255), M^+ - TMS-O - CH_3 (*m/z* 365), M^+ - TMS-O - H (*m/z* 380), and M^+ - CH_3 (*m/z* 455). Key fragments in the mass spectrum of 24-ethylcholesta-5,22E-dien-3β-ol were M^+ - H -

side chain (*m/z* 255), M^+ - TMS-O - CH_3 (*m/z* 379), M^+ - TMS-O - H (*m/z* 394), and M^+ - CH_3 (*m/z* 469). Key fragments in the mass spectrum of 24-ethylcholest-5-en-3β-ol were M^+ - H - side chain (*m/z* 255), M^+ - TMS-O - CH_3 (*m/z* 381), M^+ - TMS-O - H (*m/z* 396), and M^+ - CH_3 (*m/z* 471). The mass spectra for the two unknown sterols are also shown in Fig. 1.

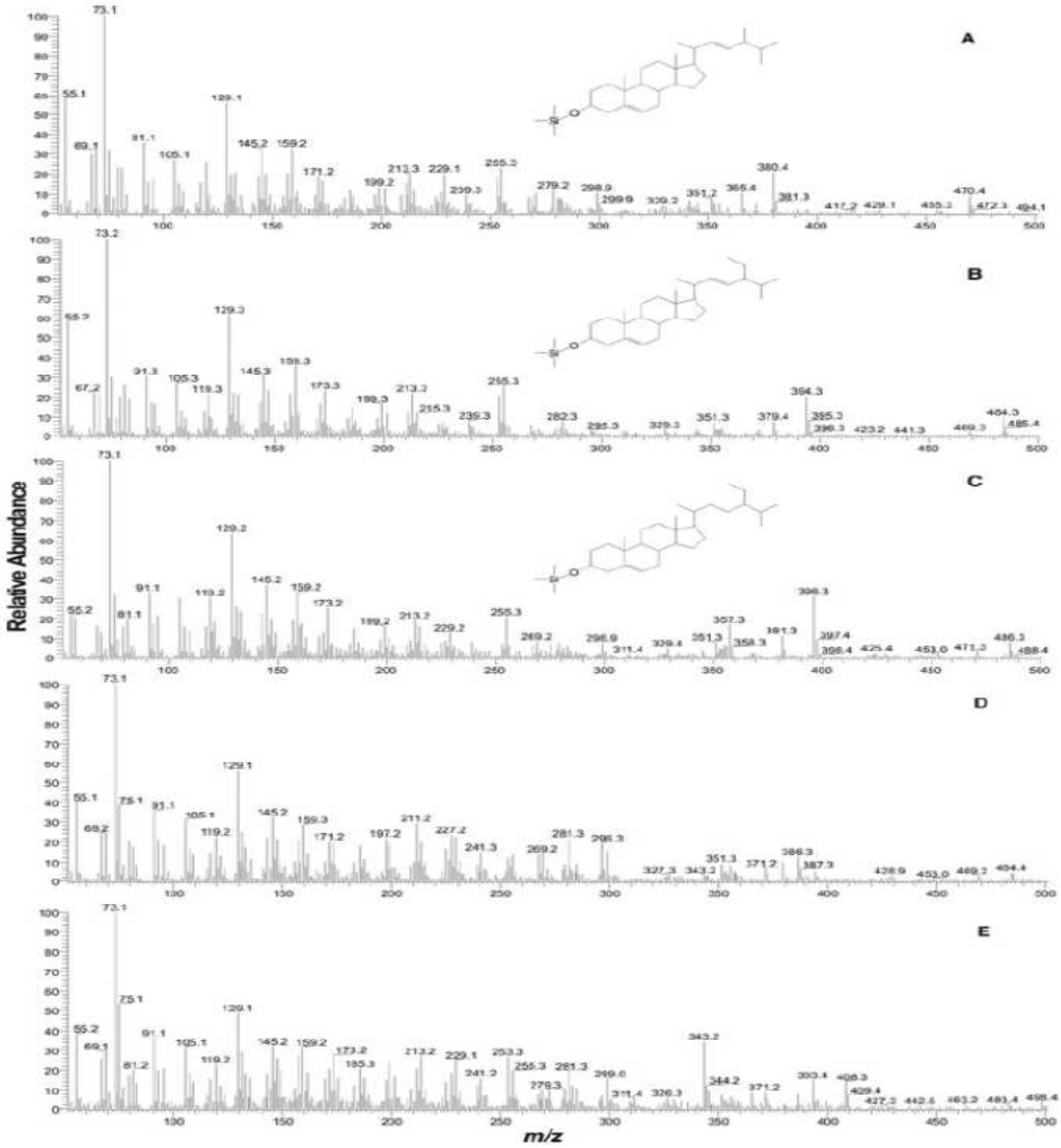


Figure 1. Mass spectra of trimethylsilyl ether (TMS) derivatives of sterols of Chromera velia. A. 24-methylcholesta-5,22E-dien-3β-ol, B. 24-ethylcholesta-5,22E-dien-3β-ol, C. 24-ethylcholest-5-en-3β-ol, D. unknown $C_{29:2}$ sterol, and E. unknown $C_{30:2}$ sterol. Structures of TMS derivatives of the three identified sterols are shown without representation of side chain stereochemistry.

Sterol biosynthesis genes identified in *Chromera velia*

Three sequences produced results with low E-values and showed functionality for the sterol biosynthesis pathway. The only merged EST sequence, which was composed of ESTs 309958118 and 309956050, showed similarity to *smt1* in 172 different species, including animals, plants, bacteria, and fungi (see Table 1 for the ten most similar). Among the potential homologs, the most similar putative homolog with reported *smt1* activity was found in tobacco (*Nicotiana tabacum*; AAC34951.1) with an E-value of 3.00E-43, 57% coverage of the query, and 47% identity. Similarity and identity of the *C. velia* sequence with this putative homolog, along with eight other species, can be observed in the alignment of sequences (Appendix 1). In addition to these putative homologs, ten possible homologs were found in algal species (Table 2).

A homolog search using *C. velia* EST 30955906 showed similarity to proteins encoding *fdft1* activity in 216 different species, including animals, plants, bacteria, and fungi (see Table 1 for the ten most similar). The putative homolog with the most similarity and recognized FDFT1 activity was from zebrafish (*Danio rerio*; NP_001074029.1) with an E-value of 2.00E-28, 73% coverage of the query, and 40% identity (Table 1, Appendix 2). Residues of similarity and identity of the *C. velia* sequence with this putative homolog, along with eight other species, is highlighted by alignment (Appendix 2). Of the more than 200 putative homologs, eleven algae species were observed to have potential homologs (Table 2).

The final *Chromera velia* EST, 309955595, whose homology search results indicated a potential function in the sterol biosynthesis pathway, showed similarity to isopentenyl diphosphate Δ-isomerase (*idi1*) in 356 different species, including animals, plants, bacteria, and fungi (see Table 1 for the ten most similar). The most similar putative homolog with reported function to the protein encoded by this *Chromera* sequence was found in net blotch fungus (*Pyrenophora tritici-repentis*; XP_001939508.1) with an E-value of 4.00E-18, 32% coverage of the query, and 56% identity. Similarity and identity of the *C. velia* sequence with this putative homolog, along with eight other species, is shown by alignment (Appendix 3). Among the numerous potential homologs identified, putative homologs were found in eleven algal species (Table 2).

For these three genes potentially involved in sterol biosynthesis in *C. velia*, putative homologs were identified in the perkinsid, *Perkinsus marinus*, while for *fdft1* putative homologs were identified in a small number of apicomplexans (Table 2).

Table 1. The ten most similar probable homologs of proteins encoded by the putative sterol biosynthesis genes of *Chromera velia*.

Gene	E-value	Type	Organism	Accession Number
smt1	5.00E-44	Plant	*Populus trichocarpa*	XP_002313315.1
	3.00E-43	Plant	*Nicotiana tabacum*	AAC34951.1
	6.00E-43	Plant	*Glycine max*	ACU21534.1
	1.00E-42	Plant	*Ricinus communis*	XP_002517794.1
	2.00E-42	Plant	*Vitis vinifera*	XP_002276734.1
	2.00E-42	Plant	*Medicago truncatula*	ACJ84224.1
	2.00E-42	Plant	*Arabidopsis thaliana*	AAM53553.1
	3.00E-42	Plant	*Dioscorea zingiberensis*	CBX33151.1
	2.00E-41	Plant	*Polysphondylium pallidum*	EFA82153.1
	2.00E-41	Plant	*Arabidopsis lyrata*	XP_002873624.1
fdft1	2.00E-31	Animal	*Nematostella vectensis*	XP_001641349.1
	2.00E-28	Animal	*Danio rerio*	NP_001074029.1
	3.00E-28	Plant	*Catharanthus roseus*	AAQ08894.1
	3.00E-28	Plant	*Pisum sativum*	O24304.1
	4.00E-28	Animal	*Anopheles gambiae*	XP_320743.4
	6.00E-28	Other	*Polysphondylium pallidum*	EFA82811.1

Table 1. Continued: The ten most similar probable homologs of proteins encoded by the putative sterol bisoynthesis genes of *Chromera velia*.

Gene	E-value	Type	Organism	Accession Number
	6.00E-28	Animal	*Gallus gallus*	XP_424881.2
	8.00E-28	Animal	*Taeniopygia guttata*	XP_002188148.1
	8.00E-28	Animal	*Hydra magnipapillata*	XP_002153991.1
	8.00E-28	Plant	*Solanum lycopersicum*	P93227.1
idi1	2.00E-18	Fungi	*Penicillium chrysogenum*	XP_002569145.1
	4.00E-18	Fungi	*Pyrenophora triticirepentis*	XP_001939508.1
	4.00E-18	Fungi	*Phaeosphaeria nodorum*	XP_001799259.1
	6.00E-18	Fungi	*Pyrenophora teres*	XP_003303424.1
	1.00E-17	Fungi	*Aspergillus niger*	XP_001392874.1
	2.00E-17	Fungi	*Grosmannia clavigera*	EFW98479.1
	2.00E-17	Fungi	*Paracoccidioides brasiliensis*	XP_002794367.1
	2.00E-17	Fungi	*Aspergillus flavus*	XP_002376785.1
	2.00E-17	Fungi	*Aspergillus oryzae*	XP_001821042.1
	2.00E-17	Fungi	*Aspergillus terreus*	XP_001214230.1

Table 2. Possible homologs in algae, apicomplexans, and perkinsids of proteins encoded by putative sterol biosynthesis genes of *Chromera velia*.

Gene	E-value	Type	Organism	Accession Number
smt1	1.00E-40	Algae	*Aureococcus anophagefferens*	EGB07255.1
	7.00E-39	Algae	*Ectocarpus siliculosus*	CBN76684.1
	9.00E-38	Algae	*Micromonas pusilla*	XP_003058910.1
	2.00E-35	Algae	*Micromonas sp.*	XP_002501753.1
	6.00E-35	Algae	*Ostreococcus lucimarinus*	XP_001419368.1
	1.00E-31	Algae	*Ostreococcus tauri*	XP_003078287.1
	7.00E-30	Algae	*Thalassiosira pseudonana*	XP_002287743.1
	2.00E-28	Algae	*Chlamydomonas reinhardtii*	XP_001690775.1
	4.00E-27	Algae	*Volvox carteri*	XP_002948023.1
	4.00E-25	Algae	*Phaeodactylum tricornutum*	XP_002178531.1
	9.00E-23	Perkinsid	*Perkinsus marinus*	XP_002776821.1
fdft1	7.00E-24	Algae	*Thalassiosira pseudonana*	XP_002289099.1
	1.00E-23	Algae	*Aureococcus anophagefferens*	EGB03382.1
	4.00E-22	Perkinsid	*Perkinsus marinus*	XP_002783518.1
	1.00E-20	Algae	*Phaeodactylum tricornutum*	XP_002180048.1
	2.00E-15	Algae	*Chlorella variabilis*	EFN51976.1
	3.00E-15	Algae	*Micromonas sp.*	XP_002504377.1
	5.00E-14	Apicomplexan	*Cryptosporidium muris*	XP_002140592.1

Table 2. Continued: Possible homologs in algae, apicomplexans, and perkinsids of proteins encoded by putative sterol biosynthesis genes of *Chromera velia*.

Gene	E-value	Type	Organism	Accession Number
	1.00E-13	Apicomplexan	*Cryptosporidium hominis*	XP_668348.1
	1.00E-13	Algae	*Volvox carteri*	XP_002956696.1
	4.00E-13	Algae	*Micromonas pusilla*	XP_003064812.1
	8.00E-12	Algae	*Chlamydomonas reinhardtii*	XP_001689926.1
	2.00E-11	Algae	*Ostreococcus tauri*	XP_003081677.1
	9.00E-11	Algae	*Ostreococcus lucimarinus*	XP_001420167.1
	3.00E-10	Apicomplexan	*Babesia bovis*	XP_001611388.1
	1.00E-08	Apicomplexan	*Theileria annulata*	XP_955405.1
	4.00E-07	Algae	*Ectocarpus siliculosus*	CBN74500.1
	4.00E-05	Cyanobacteria	*Anabaena variabilis*	YP_324009.1
idi1	2.00E-14	Algae	*Ectocarpus siliculosus*	CBN79702.1
	3.00E-12	Algae	*Haematococcus pluvialis*	AAC32209.1
	2.00E-11	Algae	*Chlorella variabilis*	EFN53784.1
	2.00E-10	Algae	*Guillardia theta*	ACI45949.1
	2.00E-10	Algae	*Micromonas sp.*	XP_002505506.1
	3.00E-10	Algae	*Micromonas pusilla*	XP_003063615.1
	1.00E-09	Algae	*Chlamydomonas reinhardtii*	XP_001701418.1

Table 2. Continued: Possible homologs in algae, apicomplexans, and perkinsids of proteins encoded by putative sterol biosynthesis genes of *Chromera velia*.

Gene	E-value	Type	Organism	Accession Number
	6.00E-09	Algae	*Volvox carteri*	XP_002958060.1
	1.00E-08	Algae	*Ostreococcus lucimarinus*	XP_001422616.1
	1.00E-08	Algae	*Thalassiosira pseudonana*	XP_002286536.1
	3.00E-07	Algae	*Aureococcus anophagefferens*	EGB07700.1
	4.00E-06	Perkinsid	*Perkinsus marinus*	XP_002774271.1

DISCUSSION

The most pertinent comparison of the sterol composition of *C. velia* is to the sterol compositions of apicomplexans, the most closely related taxonomic group. However, to our knowledge, for those apicomplexans for which sterol compositions have been determined, they appear to take up cholesterol from their parasitic hosts and have no ability to synthesize sterols *de novo*; see example studies by Nishikawa et al. (2005) and Sehgal et al. (2005). It may be hypothesized that knowledge of the sterol composition of *C. velia*, as the closest photosynthetic relative to apicomplexans, may provide insight into the sterol biosynthetic potential of apicomplexans before they lost the capability to do so.

Similarly, the perkinsids, a group of shellfish parasites phylogenetically intermediate between apicomplexans and dinoflagellates (Moore et al. 2008), have no ability to synthesize sterols *de novo* (Lund *et al*. 2007). To our knowledge, the ability of colpodellids to biosynthesize sterols is unknown. A comparison to the sterols produced by dinoflagellates as the next closest group with the ability to produce sterols related to *C. velia* reveals that they do not possess the sterols observed in *C. velia* [see summary presented by Leblond et al. (2010)]. Rather, the identified sterols of *C. velia* have been observed in other, seemingly unrelated classes of photoautotrophic microalgae, such as chlorarachniophytes (Leblond et al. 2005), glaucocystophytes (Leblond et al. 2010), and plants as reviewed by Volkman (2003), thus implying that *C. velia* shares a sterol biosynthesis pathway(s) ancestral to more than one group of algae. With regard to the search for biomarkers in *C. velia*, because the three identified sterols of *C. velia* are indeed found in other groups of algae, they are of no biomarker utility to this particular taxon. Furthermore, the two unidentified sterols are of too little abundance (less than 5% each of total sterols) to be useful biomarkers.

Regarding the biosynthesis of 24-methylcholesta-522E-dien-3β-ol, 24-ethylcholesta-5,22E-dien-3β-ol, and 24-ethylcholest-5-en-3β-ol, much detail on the enzymatic steps and underlying genes has been revealed in plants since these are common plant sterols [searchable, annotated reference pathways for sterol biosynthesis in plants and fungi, with particular reference to the use of mevalonate

as a precursor (described below), can be found, respectively, at www.plantcyc.org and www.yeastgenome.org]. As summarized in Desmond and Gribaldo (2009), the ring structure of any sterol is formed when squalene is monooxygenated to form squalene-2,3-epoxide (2,3-oxidosqualene), which is then subsequently cyclized. In plants, this cyclization forms cycloartenol (4,4,14α-trimethyl-5α-9,19-cyclocholesta-24-en-3β-ol), whereas in fungi and vertebrates this forms lanosterol (4,4,14α-trimethyl-5α-cholesta-8,24-dien-3β-ol). In plants, approximately fifteen enzymatic steps are used to transform cycloartenol to campesterol (24*R*-methylcholesta-5,22E-dien-3β-ol), sitosterol (24*R*-ethylcholest-5-en-3β-ol), and stigmasterol (24*R*-ethylcholesta-5,22E-dien-3β-ol) as common end products; the product of the *SMT1* gene is used to methylate the side chain, typically at C24. In fungi and vertebrates, a similar number of biosynthetic steps is used to transform lanosterol to ergosterol (24-methylcholesta-5,7,22E-trien-3β-ol) and cholesterol (cholest-5-en-3β-ol), respectively.

Within a given sterol-producing organism, the steps leading to the formation of squalene may occur according to one of two well-characterized routes, the mevalonate (MVA) and 2-C-methyl-D-erythrol 4-phosphate (MEP) pathways [summarized by Lichtenthaler (2003), Rohmer (1999), Rohmer (2003), and Volkman (2005)]. Both pathways ultimately yield isopentenyl diphosphate (IPP), which is combined with geranyl diphosphate to form farnesyl diphosphate. The product of the *fdft1* gene then subsequently combines two farnesyl diphosphates to produce

squalene. In the MVA pathway, the product of the *idi1* gene is able to perform a reversible isomerization of dimethylallyl diphosphate to IPP; this is one of two routes to produce IPP within the MVA pathway, with the other involving mevalonate itself as an intermediate. The MVA pathway is considered to be located within the cytosol and is hence found in fungi and vertebrates, whereas the MEP pathway is considered to be located within the chloroplast and may be used in the biosynthesis of other isoprenoids [reviewed by Grauvogel and Petersen (2007)]. The study of the utilization of the MVA or MEP pathway in sterol biosynthesis in photosynthetic organisms is ongoing with few taxa having been examined so far. However, to date, the limited number of plants and algae that have been examined have generally exhibited the MVA pathway as being involved in sterol biosynthesis, although the MEP pathway may be involved in biosynthesis of other isoprenoids (Volkman 2005), with the Chlorophyceae being a notable exception in that they possess only the MEP pathway (Grauvogel and Petersen 2007; Disch et al. 1998). It should also be noted that the *idi1* gene is found in both fungi and plants, and that the *ERG9* and *ERG6* genes are fungal homologs of plant *fdft1* and *smt1*. Although the apicomplexans and perkinsids have not been observed to produce sterols *de novo*, they have been observed to possess genes indicative of the MEP pathway, presumably to be utilized in the biosynthesis of other isoprenoids (Matsuzaki et al. 2008; Sonda and Hehl 2006). MEP pathway genes have been observed in the non-photosynthetic dinoflagellates, *Crypthecodinium cohnii* and *Oxyrrhis marina* (Slamovits and Keeling

2008; Sanchez-Puerta et al. 2007;); however, whether they are actually used in the biosynthesis of sterols is unknown. The role of the MVA and/or MEP pathway(s) in sterol biosynthesis in dinoflagellates is unstudied, although Mydlarz et al. (2003) did observe in a study on pseudopterosin biosynthesis in the photosynthetic dinoflagellate, *Symbiodinium* sp., that mevastatin sodium, an inhibitor of hydroxymethylglutaryl-CoA (HMG-CoA) reductase, was able to inhibit biosynthesis of mevalonate. Of the approximately twenty genes that are known to be needed in plants (see Benveniste 2004; Fujioka and Yokota 2003) to produce 24-ethylcholesta-5,22E-dien-3β-ol, the most abundant sterol in *C. velia*, we were able to find ESTs for only *SMT1*, *FDFT1*, and *IDI1*. Given that 24-methylcholesta-522E-dien-3β-ol, 24-ethylcholesta-5,22E-dien-3β-ol, and 24-ethylcholest-5-en-3β-ol are common plant sterols, it is not surprising that homologs for the putative *C. velia smt1* were found in plants, although possible homologs were observed in a large number of species, including a number of algae (data not shown). Of note is a putative *smt1* homolog in *Perkinsus marinus*; since perkinsids are not known to synthesize sterols *de novo*, this may be an antiquated gene from a nonfunctional pathway.

Homologs for the putative *fdft1* and *IDI1* of *C. velia* were found in an even larger and more varied number of species, including algae. Of note are putative *fdft1* homologs from *P. marinus* and four members of the apicomplexa, *Cryptosporidium muris*, *Cryptosporidium hominis , Babesia bovis*, and *Theileria annulata*. Similarly, there was a putative homolog of *idi1* in *P. marinus*. Again, these organisms are not

known to synthesize sterols *de novo*, so these may be ancient genes from a nonfunctional pathway. Since ESTs, such as those used in these analyses, are by the nature of the technique incomplete, cloning and sequencing of these full length genes/transcripts would not only be valuable for biochemical studies, but also for reliable phylogenetic analyses.

There was no apparent similarity of these three EST sequences to corresponding sterol biosynthesis genes from the red algae, the presumed ancestor of *C. velia*'s plastid; therefore, it is still an open question as to whether the its plastid ancestor has contributed genetically in any way to its sterol composition. However, a member of the rhodophyceae, *Rhodosorus* sp., has been observed to produce 24-methylcholesta-522E-dien-3β-ol as its major sterol (Dunstan et al. 2005). Thus, the biochemical potential exists for a red algal contribution to the sterol biochemistry of *C. velia*.

At the current time, we cannot say with certainty whether the MVA and/or the MEP pathway is used in *C. velia* to produce the IPP needed to biosynthesize sterols, especially since there is the possibility of crosstalk between the two pathways (Lichtenthaler 2004); however, the presence of the *idi1* gene may indicate utilization of the MVA pathway. In order to resolve this question, future studies should involve feeding *C. velia* with [6-^{13}C]mevalonate per Ohyama et al. (2009) and [1-^{13}C]glucose, an MEP pathway intermediate, per Disch et al. (1998) in order to assess whether either is incorporated into its sterols.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Benveniste P. 2004. Biosynthesis and accumulation of sterols. Annu. Rev. Plant Biol. 55: 429-457.

Desmond E, Gribaldo S. 2009. Phylogenomics of sterol synthesis: Insights into the origin, evolution, and diversity of a key eukaryotic feature. Genome Biol. Evol. 1: 364-381.

Disch A, Schwender J, Müller C, Lichtenthaler HK, Rohmer M. 1998. Distribution of the mevalonate and glyceraldehyde phosphate/pyruvate pathways for isoprenoid biosynthesis in unicellular algae and the cyanobacterium *Synechocystis* PCC 6714. Biochem. J. 333: 381-388.

Dunstan GA, Brown MR, Volkman JK. 2005. Cryptophyceae and rhodophyceae: Chemotaxonomy, phylogeny, and application. Phytochemistry. 66: 2557-2570.

Fujioka S, Yokota T. 2003. Biosynthesis and metabolism of brassinosteroids. Annu. Rev. Plant Biol. 54: 137-164.

Giner JL, Faraldos JA, Boyer G. L. 2003. Novel sterols of the toxic dinoflagellate *Karenia brevis* (Dinophyceae): A defensive function for unusual marine sterols? J. Phycol. 39: 315-319.

Grauvogel C, Petersen J. 2007. Isoprenoid biosynthesis authenticates the classification of the green alga *Mesostigma viride* as an ancient streptophyte. Gene. 396: 125-133.

Guillard RR, Hargraves PE. 1993. *Stichochrysis immobilis* is a diatom, not a chrysophyte. Phycologia. 32: 234-236.

Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ. 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc. Natl. Acad. Sci. USA. 107: 10949-10954.

Jones GJ, Nichols PD, Shaw PM. 1994. Analysis of microbial sterols and hopanoids. In: Goodfellow M, O'Donnel, AG. editors. Chemical Methods in Prokaryotic Systematics. John Wiley & Sons, New York. p. 163-195.

Larkin MA et al. 2007. ClustalW and ClustalX version 2.0. Bioinformatics. 23: 2947-2948.

Leblond JD, Chapman PJ. 2000. Lipid class distribution of highly unsaturated long chain fatty acids in marine dinoflagellates. J. Phycol. 36: 1103-1108.

Leblond JD, Chapman PJ. 2002. A survey of the sterol composition of the marine dinoflagellates *Karenia brevis*, *Karenia mikimotoi*, and *Karlodinium micrum*: Distribution of sterols within other members of the class Dinophyceae. J. Phycol. 38: 670-682.

Leblond JD, Timofte HI, Roche SA, Porter NM. 2011. Sterols of glaucocystophytes. Phycolog. Res. 59(2): 129-134.

Leblond JD, Lasiter AD, Li C, Logares R, Rengefors K, Evens TJ. 2010. A data mining approach to dinoflagellate clustering according to sterol composition: Correlations with evolutionary history. Int. J. Data Mining Bioinformatics. 4: 431-451.

Leblond JD, Dahmen JL, Seipelt RL, Elrod-Erickson MJ, Kincaid R, Howard JC, Evenn TJ, Chapman PJ. 2005. Lipid composition of chlorarachniophytes (Chlorarachniophyceae) from the genera *Bigelowiella*, *Gymnochlora*, and *Lotharella*. J. Phycol. 41: 311-321.

Lichtenthaler HK. 2004. Evolution of carotenoid and isoprenoid biosynthesis in photosynthetic and non-photosynthetic organisms. Proceedings of the 16th International Plant Lipid Symposium. Available from: http://www.mete.mtesz.hu/pls/proceedings/eloadasok-pdf/16pls-proceeding-20050129.pdf

Lund ED, Chu FL, Soudant P, Harvey E. 2007. *Perkinsus marinus*, a protozoan parasite of the eastern oyster, has a requirement for dietary sterols. Comp. Biochem. Physiol. A. 146: 141-147.

Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, Kanehisa M, Endo T, Goto S. 2006. EGassembler: Online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. Nucleic Acids Res. 34: W459-462.

Matsuzaki M, Kuroiwa H, Kuroiwa T, Kita K, Nozaki H. 2008. A cryptic algal group unveiled: A plastid biosynthesis pathway in the oyster parasite *Perkinsus marinus*. Mol. Biol. Evol. 25: 1167-1179.

Mooney BD, Nichols PD, de Salas MF, Hallegraeff GM. 2007. Lipid, fatty acid, and sterol composition of eight species of Kareniaceae (Dinophyta): Chemotaxonomy and putative lipid phycotoxins. J. Phycol. 43: 101-111.

Moore RB, Oborník M, Janouškovec J, Chrudimský T, Vancová M, Green DH, Wright SW, Davies NW, Bolch CJS, Heimann K, Šlapeta J, Hoegh-Guldberg O, Logsdon JM, Carter DA. 2008. A photosynthetic alveolate closely related to apicomplexan parasites. Nature. 451: 959-963.

Mydlarz LD, Jacobs RS, Boehnlein J, Kerr RG. 2003. Pseudopterin biosynthesis in *Symbiodinium* sp., the dinoflagellate symbiont of *Pseudopterogorgia elisabethae*. Chem. Biol. 10: 1051-1056.

Nishikawa Y, Quittnat F, Stedman TT, Voelker DR, Choi JY, Zahn, M, Yang M, Pypaert M, Joiner KA, Coppens I. 2005. Host cell lipids control cholesteryl ester synthesis and storage in intracellular *Toxoplasma*. Cell. Microbiol. 7: 849-867.

Oborník M, Vancová M, Lai DH, Janouškovec J, Keeling PJ, Lukeš J. 2011. Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of apicomplexa, *Chromera velia*. Protist.162: 115-130.

Ohyama K, Suzuki M, Kikuchi J, Saito K, Muranaka T. 2009. Dual biosynthetic pathways to phytosterols via cycloartenol and lanosterol in *Arabidopsis*. Proc. Natl. Acad. Sci. USA. 106: 725-730.

Rohmer M. 1999. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. Nat. Prod. Rep. 16: 565-574.

Rohmer M. 2003. Mevalonate-independent methylerythritol phosphate pathway for isoprenoid biosynthesis. Elucidation and distribution. Pure Appl. Chem. 75: 375-387.

Sanchez-Puerta MV, Kippmeier JC, Apt KE, Delwiche CF. 2007. Plastid genes in a non-photosynthetic dinoflagellate. Protist. 158: 105-117.

Sehgal A, Bettiol S, Pypaert M, Wenk MR, Kaasch A, Blader IJ, Joiner KA, Coppens I. 2005. Peculiarities of host cholesterol transport to the unique intracellular vacuole containing *Toxoplasma*. Traffic. 6: 1125-1141.

Slamovits CH, Keeling PJ. 2008. Plastid-derived genes in the nonphotosynthetic alveolate *Oxyrrhis marina*. Mol. Biol. Evol. 25: 1297-1306.

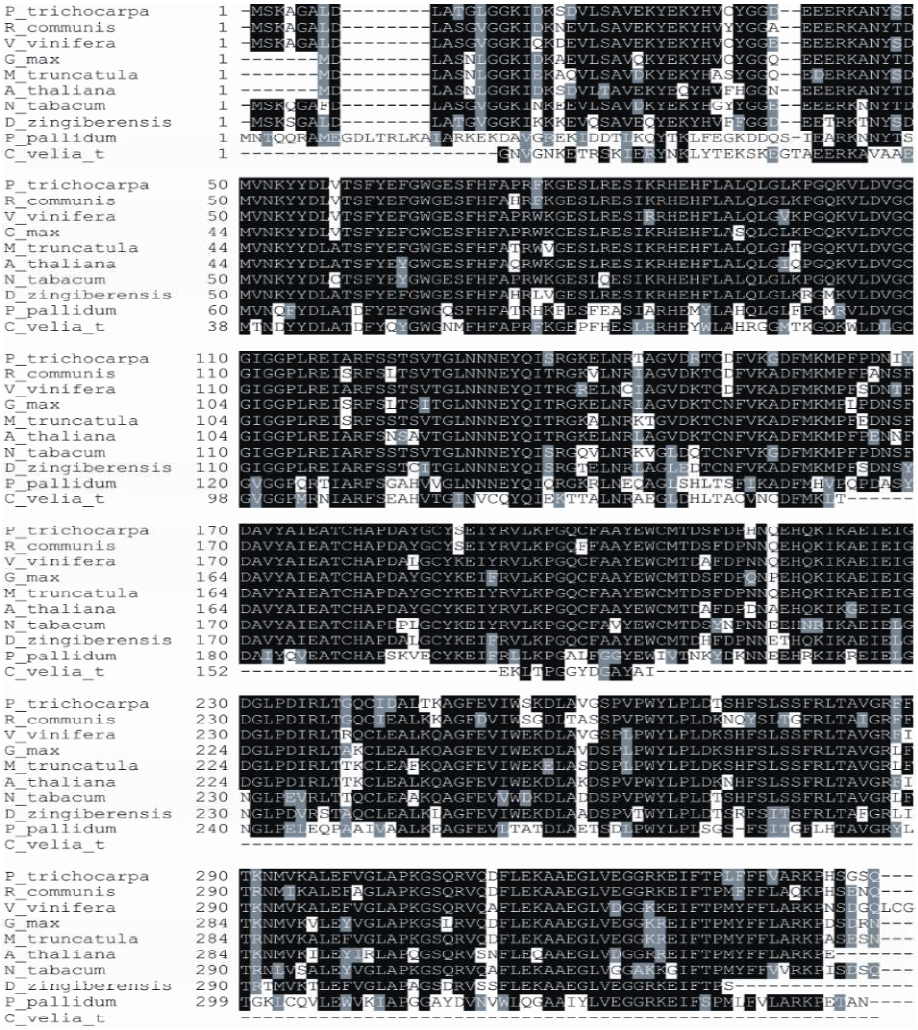
Šlapeta J, Carter D. 2010. Expression sequence tags of *Chromera velia*. (unpubl. Data via the National Center for Biotechnology Information). Available from: http://www.ncbi.nlm.nih.gov/nucest/?term=chromera%20velia.

Sonda S, Hehl AB. 2006. Lipid biology of apicomplexa: perspectives for new drug targets, particularly for *Toxoplasma gondii*. Trends Parasitol. 22: 41-47.

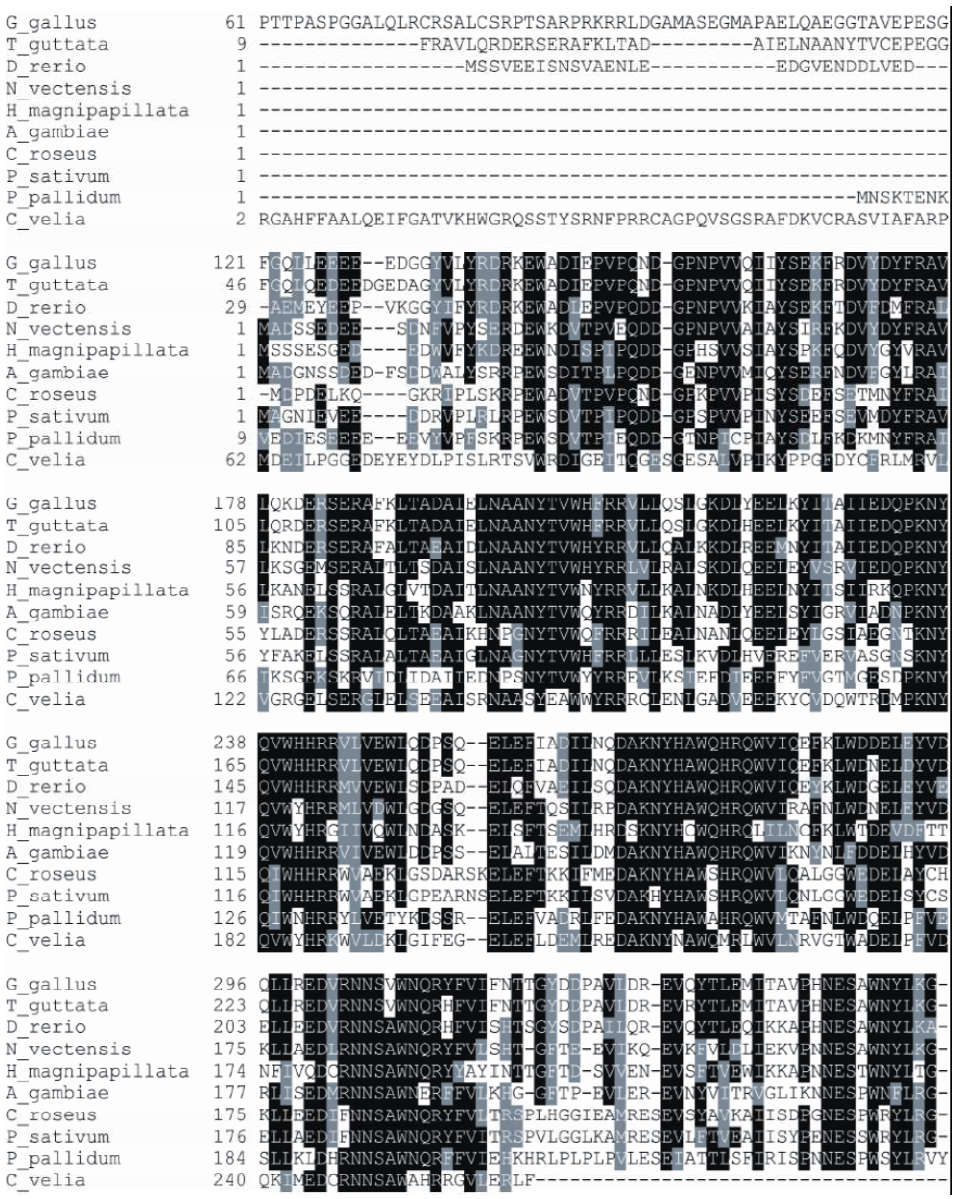
Volkman JK. 2003. Sterols in microorganisms. Appl. Microbiol. Biotech. 60: 495-506.

Volkman JK. 2005. Sterols and other triterpenoids: Source specificity and evolution of biosynthetic pathways. Org. Geochem. 36: 139-159.

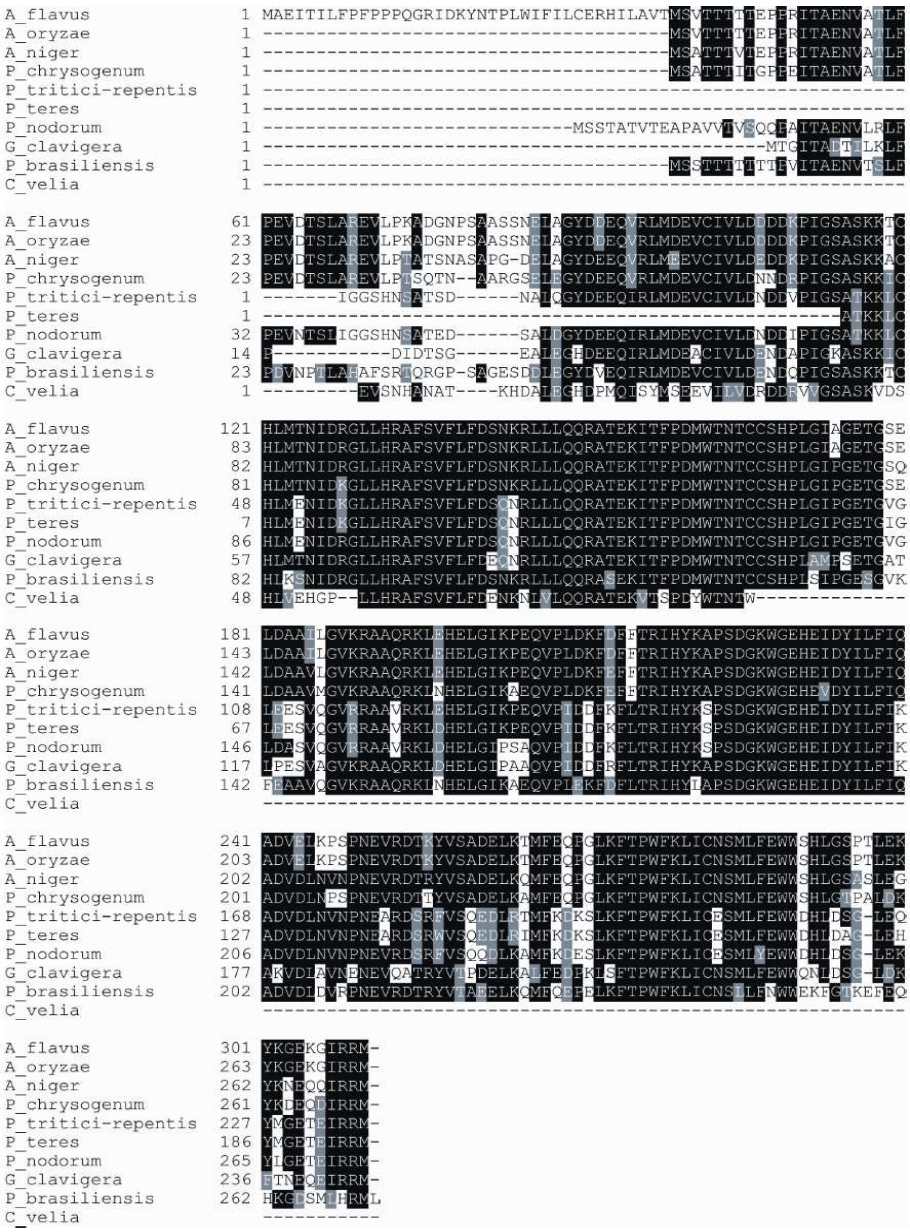
APPENDICES



Appendix A: Comparison of putative sterol-24C-methyltransferase homologs most similar to *C. velia.* Full species names and sequence identifiers are found in Table 1. *C. velia* EST 309956050 was translated *in silico* and trimmed at the amino terminal end. The remaining sequences were retrieved from NCBI. Black boxes indicate identity, while gray boxes denote amino acid similarity.



Appendix B: Comparison of homologs most similar to putative *C. velia* farnesyl diphosphate farnesyl transferase. Full species names and sequence identifiers are found in Table 1. *C. velia* EST 309955906 was translated *in silico* and trimmed at the amino terminal end. The remaining sequences were retrieved from NCBI. Black boxes indicate identity, while gray boxes denote amino acid similarity.



Appendix C: Alignment of putative isopentenyl diphosphate Δ-isomerase homologs most similar to *C. velia.* Full species names and sequence identifiers are found in Table 1. *C. velia* EST 309955595 was translated *in silico* and the remaining sequences were retrieved from NCBI. Black boxes indicate identity, while gray boxes denote amino acid similarity.

CHAPTER II

COMPARATIVE STUDY OF GALACTOLIPID COMPOSITION AND BIOSYNTHETIC GENES FOR GALACTOLIPID SYNTHASES IN *VITRELLA BRASSICAFORMIS* AND *CHROMERA VELIA*, TWO RECENTLY IDENTIFIED CHROMERIDS WITH RED ALGAL-DERIVED PLASTIDS

INTRODUCTION

The discovery of *Chromera velia* as the closest known photosynthetic relative to non-photosynthetic apicomplexan parasites has led to formation of the new phylum Chromerida in the group Alveolata (Adl et al. 2005; Adl et al. 2012; Moore et al. 2008). The discovery of *Vitrella brassicaformis*, initially referred to as unnamed isolate CCMP3155 from the Provasoli-Guillard National Center for Marine Algae and Microbiota (NCMA), added a second member to the phylum Chromerida that is also a close relative of apicomplexan parasites (Janouškovec et al. 2010; Oborník et al. 2012). Taken together, *C. velia* and *V. brassicaformis* provide a unique opportunity to study photosynthetic organisms that may provide physiological and biochemical insights into the evolution of apicomplexans from photosynthetic organisms to non-photosynthetic, obligate parasites.

In order to understand the evolutionary differences between chromerids and apicomplexan parasites, one must assess the physiological and biochemical traits that are either shared or different between these groups of organisms along

with other closely related organisms, such as dinoflagellates. Features such as tubular mitochondrial cristae, cortical alveoli, subpellicular microtubules, and micronemes are common in chromerids, apicomplexans, and dinoflagellates, whereas heterodynamic flagella, a tapered terminal portion of the long flagellum, and bacterial-derived Rubisco are common characteristics of chromerids and dinoflagellates but not apicomplexans (Moore et al. 2008; Oborník et al. 2011; Oborník et al. 2012). Moreover, *V. brassicaformis* and *C. velia* are further differentiated from dinoflagellates in that they possess a photosynthetic secondary plastid that contains chlorophyll *a*, violaxanthin, and β-carotene but no chlorophyll *c* as in peridinin-containing dinoflagellates (Moore et al. 2008; Adl et al. 2012).

Within the Chromerida themselves, features such as the presence of finger-like projections on the shorter flagellum, pseudoconoid, chromerosome, four-celled sporangia, and a non-canonical UGA codon encoding for tryptophan in the plastid are found in *C. velia* but not in *V. brassicaformis*. Furthermore, the presence of sporangia with dozens of cells, a multiple-laminated thick cell wall, pyrenoid, and highly compacted circular plastid genome are the distinguishing features of *V. brassicaformis* from *C. velia* (Janouškovec et al. 2010; Oborník et al. 2011; Oborník et al. 2012).

The presence of 18 different plastid genes in alveolates, that are also found in red algae, hacrobians, and heterokonts, but not in green algae and plants, has

suggested a red algal origin of all alveolate plastids (Janouškovec et al. 2010). Additionally, phylogenetic analysis of 34 conserved genes from the plastid genomes and a ribosomal superoperon of *C. velia* and *V. brassicaformis* indicated a red algal plastid origin and a close relationship to the plastids of heterokonts (Janouškovec et al. 2010).

The galactolipids monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG) are generally the major plastid structural lipids in all photosynthetic organisms (Hölzl et al. 2007). These galactolipids play an important role in organization of the thylakoid membrane, growth and maintenance of photosynthetic ability, and stability of photosystem I and II complexes in the chloroplast (Guo et al.2005; Ivanov et al. 2006; Hölzl et al. 2007; Kobayashi et al. 2007). It has been reported previously that radiolabeled UDP-galactose has been incorporated into the MGDG and DGDG of apicomplexan parasites, *Plasmodium falciparum* and *Toxoplasma gondii*, possibly as part of a non-photosynthetic, relic plastid called an apicoplast (Maréchal et al. 2002). A more recent study by Botté et al. (2013) showed the presence of atypical lipids like sphingomyelin and ceramides, but no galactolipids, in the apicoplast of *P. falciparum*. Thus, the composition of the apicoplast membrane appears to be an evolving area of study.

In this study we have used positive-ion electrospray ionization/mass spectrometry (ESI/MS) and ESI/MS/MS and publicly available transcriptomic

data to achieve the following objectives: 1) To profile the galactolipid composition of *V. brassicaformis* and compare it with published studies of *C. velia*, and 2) To perform phylogenetic analyses of the MGDG and DGDG synthase genes from *V. brassicaformis* and *C. velia* (and other algal species) in order to assess the phylogenetic relationship of the galactolipid biosynthetic pathways in these two organisms to each other and to other algae in general; typically these two terminal genes of the galactolipid biosynthetic pathway are those that are chosen for phylogenetic comparisons because of the transport and corresponding localization of the resulting enzymes within the chloroplast. Utilizing these biochemical and bioinformatics approaches, we have demonstrated that *V. brassicaformis* and *C. velia* differ in their galactolipid content (i.e. associated fatty acids) as they do in other ultrastructural and physiological traits, yet share the same red algal origin of their galactolipid synthases. This indicates that there are differences in the steps of fatty acids biosynthesis pathway that are ultimately incorporated in galactolipids. Thus to explain the difference in fatty acid biosynthesis and galactolipid ultimately, we proposed a hypotheses on fatty acid biosynthesis on the basis of our finding and provided future avenues for further study to explain the mechanism of galactolipid biosynthesis in these two chromerids.

MATERIALS AND METHODS

Cultures and growth conditions

Vitrella brassicaformis was acquired from the Provasoli-Guillard National Center for Marine Algae and Microbiota (West Boothbay Harbor, ME, USA), and was grown autotrophically in L1 medium (Guillard et al. 1993) in triplicate at room temperature under a 14/10 h light/dark cycle at an irradiance of approximately 50 µmol photons•m^{-2}•s^{-1}. Cells were harvested during the exponential phase of growth and lipids extracted according to the techniques described by Leblond and Chapman (2000).

Lipid processing and mass spectrometry

Total lipids were extracted and galactolipids separated from other lipid classes according to the techniques described by Leblond & Chapman (2000). Briefly, the total lipid extracts were separated into five component lipid fractions on columns of activated Unisil silica (1.0 g, 100-200 mesh, activated at 120℃, Clarkson Chromatography, South Williamsport, PA, USA). The following solvent regime was used to separate lipids according to polarity, with fraction 5 eluting the most polar lipids (Leblond& Chapman 2000): 1) 12 ml methylene chloride (sterol esters), 2) 15 ml 5% acetone in methylene chloride with 0.05% acetic acid (free sterols, tri- and diacylglycerols, and free fatty acids), 3) 10 ml 20% acetone in methylene chloride (monoacylglycerols), 4) 45 ml acetone (MGDG and

DGDG), and 5) 15 ml methanol with 0.1% acetic acid (polar lipids, including betaine lipids).

Following the procedure of Welti et al. (2002), MGDG and DGDG were suspended in methanol, chloroform, and 50 mM sodium acetate prior to examination via mass spectrometry. The resulting sodium adducts were examined using positive-ion ESI/MS full scans from 100-2,000 Daltons through direct injection (5 µL sample volume into a methylene chloride carrier solvent at 0.5 mL/min) into a Finnigan DecaXP ion trap mass spectrometer (currently Thermo Scientific, Waltham, MA, USA). The relative abundance of each lipid was determined as part of data processing by calculating its relative percent composition based on relative peak height of raw data in the positive-ion ESI/MS full scan mode. Subsequent positive-ion ESI/MS/MS was performed using collision energy between 37.5 and 48%, and major cleaved fatty acids were identified by the differences between the masses of the original ions and their fragments (Guella et al. 2003; Gray et al. 2009a).

In silico **analyses of transcriptome data**

RNA-seq data with accession number SRX215482 for *V. brassicaformis* were downloaded from the Sequence Read Archive (SRA) database of the National Center for Biotechnology Information (NCBI), (Wheeler et al. 2008). The RNA-seq data were deposited by the Broad Institute (http://www.broadinstitute.org/)

and consisted of a paired-end cDNA library using an Illumina HiSeq 2000 Solexa-131194 sequencer (Illumina Inc., San Diego, CA, USA). The data can be retrieved using the following link: http://www.ncbi.nlm.nih.gov/sra?term=(SRX215482)%20NOT%20cluster_dbgap%5BPROP%5D (Wheeler et al. 2008).

The CLC Genomics Workbench (version 6.0.2) (http://www.clcbio.com/) algorithm was used to *de novo* assemble contigs from the non-redundant short-read data using specific parameters (automatic word/bubble size, minimum contig length 200, similarity = 0.8, length fraction = 0.5, insertion/deletion cost = 3, mismatch cost =2). The short read of the transcriptome was assembled into 50,755 contigs. Potential biochemical functions of the contigs were predicted using the Kyoto Encyclopedia of Genes and Genomes Pathway (KEGG Pathway mapping) automatic annotation server (Moriya et al. 2007). Contigs related to galactolipid biosynthesis were identified and their sequences were retrieved. Two contigs (5672 and 11390) were found to be related to MGDG and DGDG synthase and were chosen for phylogenetic analysis (below).

Phylogenetic analyses

In order to determine the phylogenetic relationships of MGDG and DGDG synthases between *V. brassicaformis* with *C. velia*, other algae, and plants, phylogenetic trees were constructed. The amino acid sequences of MGDG and

DGDG synthases from *V. brassicaformis*, *C. velia*, green algae, red algae, brown algae, diatoms, and plants (sequenceslisted in supplementary data Data 1 and Data 2) were retrieved from NCBI protein database. The sequences were ClustalW aligned with pairwise and multiple alignment option using MEGA 5.2.2 (Tamura et al. 2011). The aligned sequences were converted to Phylip format using Mesquite 2.75 tool (Madison et al. 2011). The phylip formatted aligned sequences were analysed to find the best-fit model of protein evolution using ProtTest 3.3 software (Darriba et al. 2011). The result obtained from ProtTest analysis was used to construct maximum likelihood phylogenetic trees in RAxML 7.2.8 software (Stamatakis 2006) using LG+I+G model with 100 bootstrap replications.

RESULTS

Mass spectrometry analysis of galactolipids

Positive-ion ESI/MS analysis of *V. brassicaformis* galactolipids indicated seven forms of MGDG and four forms of DGDG (Table 1). A C_{20}/C_{16} (*sn*-1/*sn*-2combination) was found in two forms of MGDG and one form of DGDG at *m/z* 799, 801, and 961, respectively (see next paragraph for discussion of structural characterization of example galactolipid species). A C_{20}/C_{18} combination was found in three forms of MGDG and one form of DGDG at *m/z* 823, 825(2), and 985, respectively. Two forms of MGDG and one form of DGDG were identified

with fatty acid combinations of C_{18}/C_{18} and C_{18}/C_{16} at *m/z* 801(2) and 939, respectively. The most abundant galactolipids in terms of relative abundance were 20:5/18:2 MGDG (*m/z* 823) at 31.4±4.4% and 20:5/14:0MGDG (*m/z* 771) at 23.6±4.9%. 20:4/16:0 and 18:2/18:2 were the minor forms of MGDG with relative abundances of 1.3±0.9% and 1.4±0.9%. The major form of DGDG was 20:5/16:0 with a relative abundance of 8.4±4.3% at *m/z* 961 and the other two forms of DGDG were found to be 18:2/16:0 and 20:5/14:0 with relative abundance of 3.5±1.1% and 3.9±2.7% at *m/z* of 939 and 933 respectively.

Structural characterization of galactolipids was performed by positive-ion ESI/MS/MS. As examples, the ESI/MS/MS spectrum of MGDG at *m/z* 771 produced two fragments at *m/z* 543 and 469 that indicated the loss of 14:0 and 20:5 fatty acids, with preferential cleavage of the 20:5 fatty acid from the *sn*-1 position to yield the larger daughter ion (fragment) at *m/z* 469 containing glycerol, galactose, and the 14:0 fatty acid at the *sn*-2 position (Fig. 1). Similarly, preferential cleavage of the 20:5 fatty acid from MGDG at *m/z* 799 to yield the larger *m/z* 497 daughter ion indicated that this form of MGDG was 20:5/16:0 (*sn*-1/*sn*-2; Fig. 2). Similarly, the parent ion of DGDG at *m/z* of 933 showed the presence of fatty acid combinations with regiochemical assignments of 20:5/14:0 because of preferential cleavage of the 20:5 fatty acid to form the larger *m/z* 631 daughter ion (Fig. 3).

Table 1: Average relative percentages of galactolipids as determined via positive-ion ESI/MS and ESI/MS/MS.

Galactolipids[1]	***m/z***[2]	**Average relative percentage**[3]	**Standard deviation**
20:5/14:0 MGDG	771	23.5	4.9
20:5/16:0 MGDG	799	12.4	3.8
20:4/16:0 MGDG	801	1.3	0.9
18:2/18:2 MGDG	801	1.4	0.9
20:5/18:2 MGDG	823	31.4	4.4
20:4/18:2 MGDG	825	6.8	2.1
20:5/18:1 MGDG	825	3.4	1.0
20:5/14:0 DGDG	933	3.9	2.7
18:2/16:0 DGDG	939	3.5	1.1
20:5/16:0 DGDG	961	8.4	4.3
20:5/18:2 DGDG	985	3.9	2.3

[1][M + Na] $^{+}$; represents the number of carbon: number of double bonds (*sn*-1/*sn*-2)
[2]Mass rounded down to nearest odd number.
[3]The average relative percentage was calculated from the three replicates of *Vitrella brassicaformis* cultures.
Note: The average total may not add to exactly 100% due to rounding.

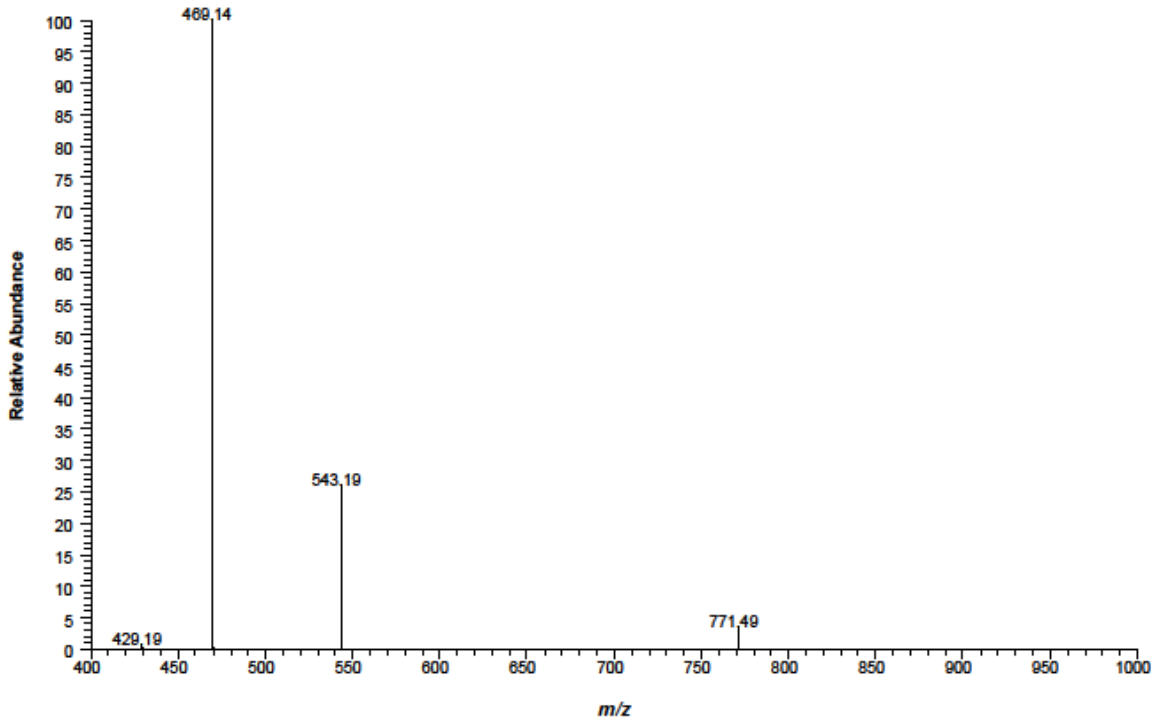


Figure 1: Positive-ion ESI/MS/MS spectra of 20:5/14:0 MGDG (*m/z* 771) from *V. brassicaformis*. The *m/z* 469 and 543 fragments represent the loss of 20:5 and 14:0 fatty acids from *sn*-1 and *sn*-2 positions, respectively.

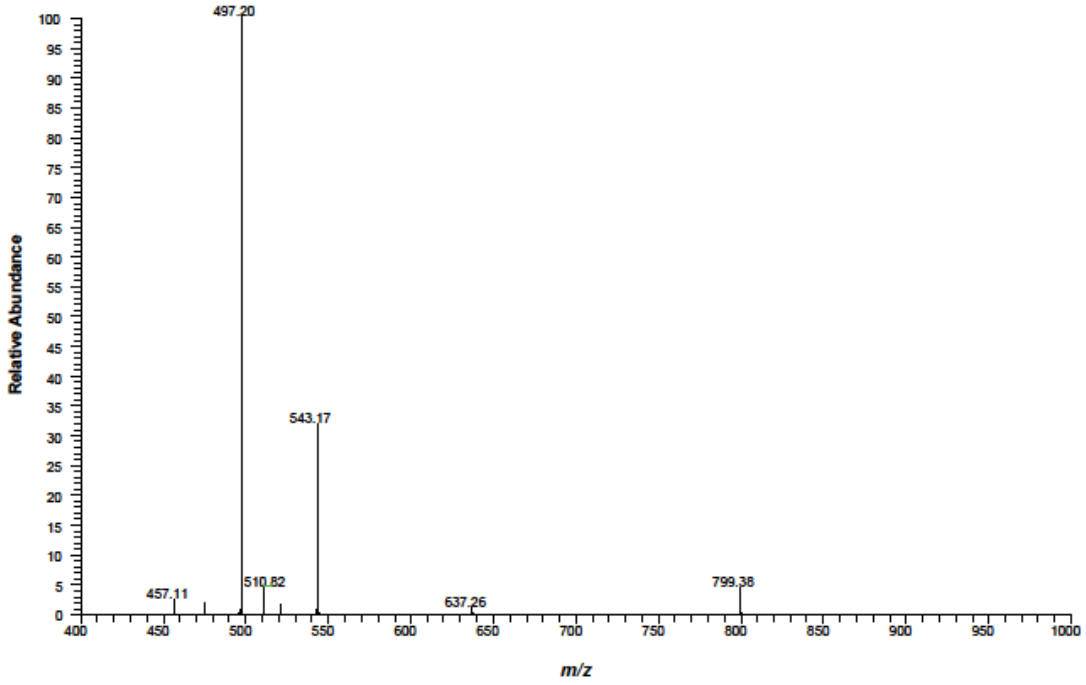


Figure 2. Positive-ion ESI/MS/MS spectra of 20:5/16:0 MGDG (*m/z* 799) from *V. brassicaformis*. The *m/z* 497 and 543 fragments represent the loss of 20:5 and 16:0 fatty acids from *sn*-1 and *sn*-2 positions, respectively

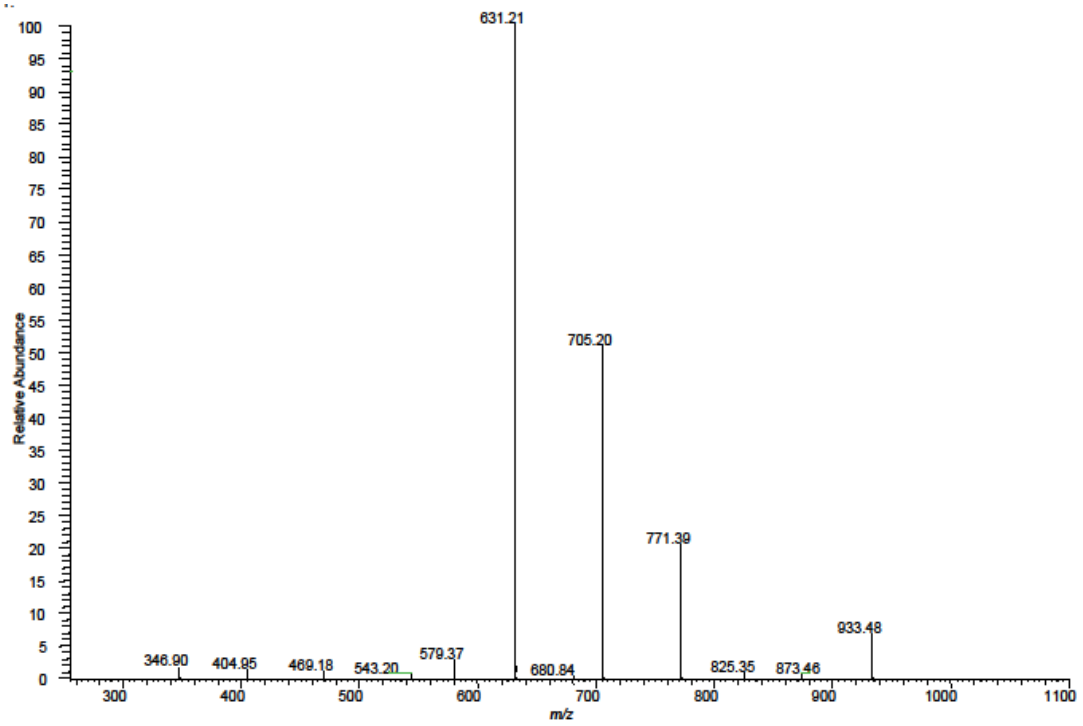


Figure 3. Positive-ion ESI/MS/MS spectra of 20:5/14:0 DGDG (*m/z* 933) from *V. brassicaformis*. The *m/z* 631 and 705 fragments represent the loss of 20:5 and 14:0 fatty acids from *sn*-1 and *sn*-2 positions, respectively. The loss of galactose yields the *m/z* 771 fragment.

In silico analyses of RNA-seq data and phylogenetic analysis

The *de novo* assembly of RNA-seq data for *V. brassicaformis* produced 50,755 contigs. Contig 5672 was found to be associated with MGDG synthase and contig 11390 was found to be associated with DGDG synthase, which have KEGG entry codes of K03715 and K09480 and enzyme commission numbers (EC) 2.4.1.46 and (EC) 2.4.1.241, respectively. These predicted amino acid sequences of the terminal genes of galactolipid synthesis were analyzed through

alignment with sequences from *C. velia* and *A. thaliana* (Figs. 4 and 5). The galactolipid synthases were also used in a phylogenetic analysis to better elucidate the association of *V. brassicaformis*' galactolipid biosynthesis with the first reported chromerid, *C. velia*, and to confirm its plastid ancestry. Both the MGDG and DGDG synthases from *V. brassicaformis* were found to associate with *C. velia* within other algae with a red algal plastid (Figs. 6 and 7), thus strongly indicating a shared red algal origin.

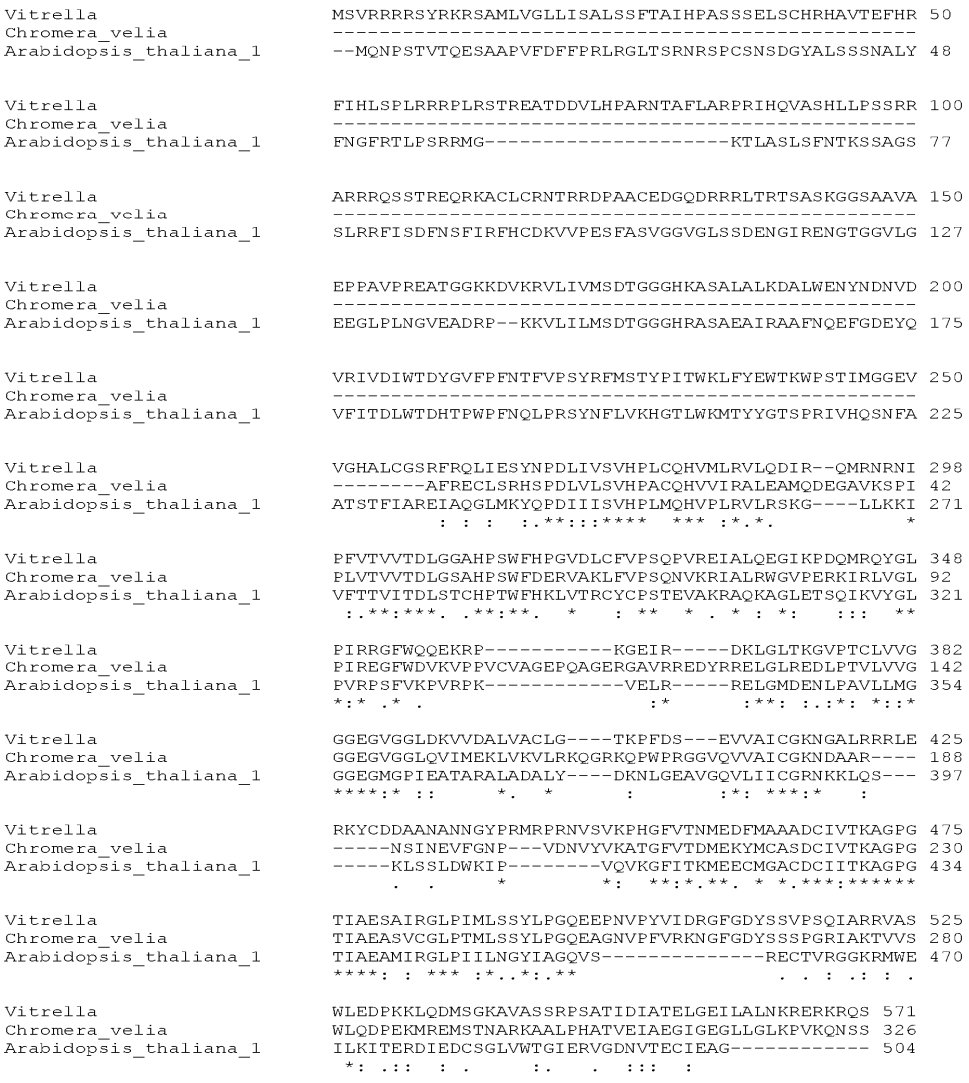


Figure 4. Predicted amino acid sequence of MGDG synthase from *V. brassicaformis* aligned with the MGDG synthases from *C. velia* and *A. thaliana.*

Figure 5. Predicted amino acid sequence of DGDG synthase from *V. brassicaformis* aligned with the DGDG synthases from *C. velia* and *A. thaliana.*

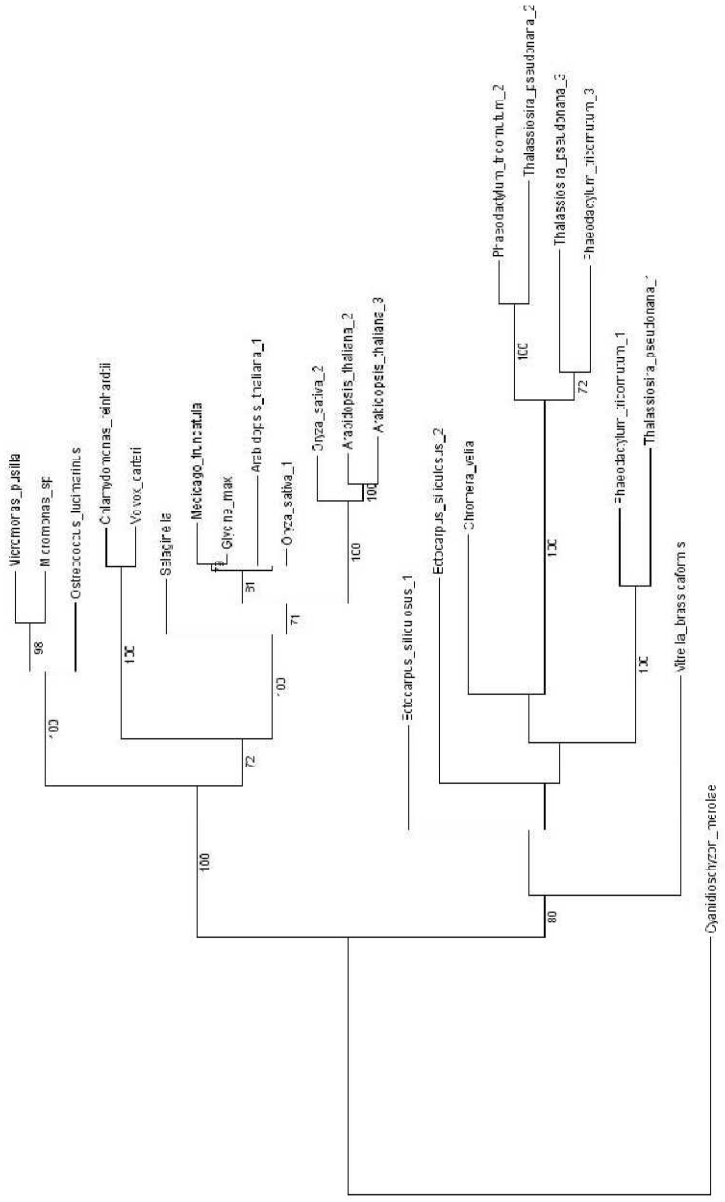


Figure 6. Phylogenetic tree constructed from RAxML using MGDG synthase amino acid predicted sequences to determine plastidial ancestry of the chromerids *V. brassicaformis* and *C. velia* with plants, green algae, red algae, and other chromalveolates. Nodes with >50% bootstrap support values are labeled.

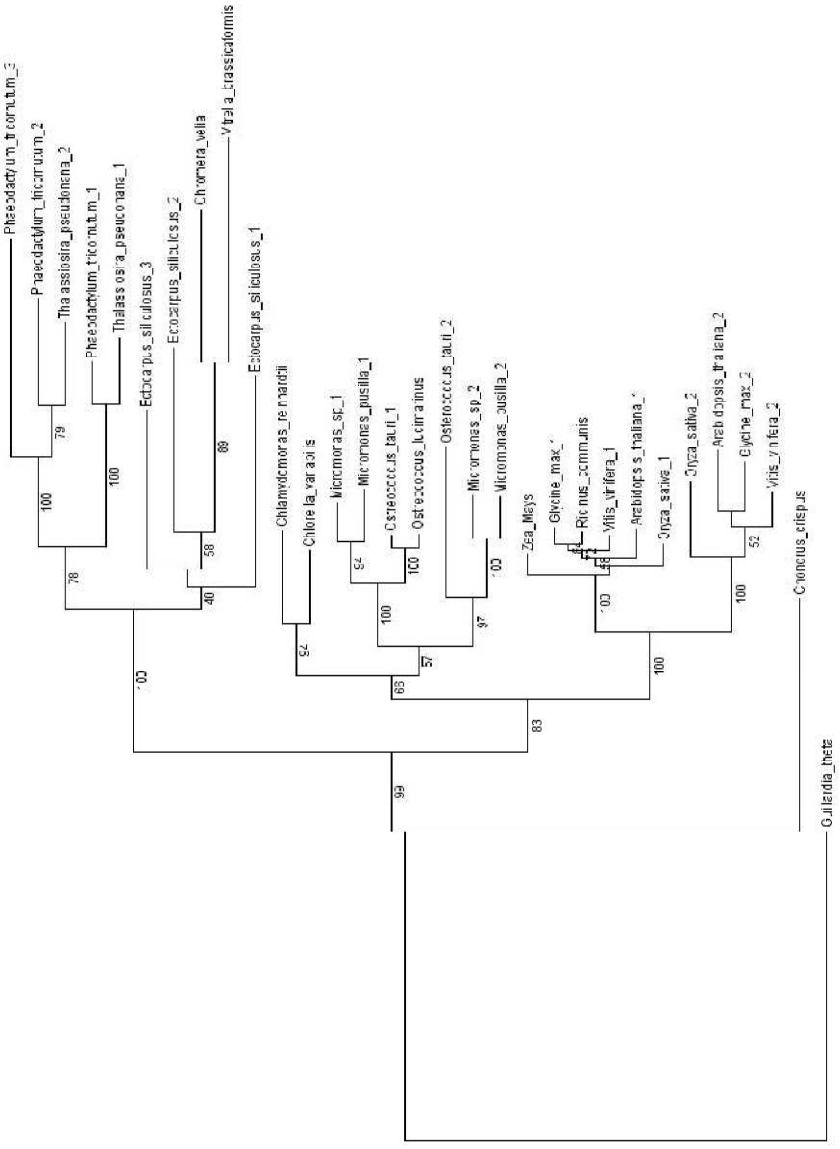


Figure 7. Phylogenetic tree constructed from RAxML using DGDG synthase predicted amino acid sequences determine plastidial ancestry of the chromerids. *V. brassicaformis* and *C. velia* with plants, green algae, red algae, and other chromalveolates. Nodes with >50% bootstrap support are labeled.

DISCUSSION

V. brassicaformis is a recently identified alga that along with *C. velia* makes up the phylum Chromerida. Together, these two chromerids are the closest known photosynthetic relatives of parasitic apicomplexans, which are postulated because of the presence of apicoplasts to have lost the ability to mediate photosynthesis at some point in their evolution into parasites of animals. The apicoplast is a relic, non-photosynthetic plastid that is essential for the survival of apicomplexans (Fichera et al. 1997; Ramya et al. 2007) because it is involved in various essential metabolic functions, such as fatty acid synthesis, iron sulfur-cluster synthesis, heme synthesis, and isoprenoid synthesis (Lim et al. 2010).

The galactolipid content of *C. velia* was described initially by Botté et al. (2011), where it was found that $C_{40:10}$ and $C_{36:5}$ forms were the only detectable forms of MGDG and DGDG. Regiochemical assignment of the fatty acids was not performed by Botté et al. (2011), but these forms presumably correspond to 20:5/20:5 and 20:5/16:0, respectively, because this agrees with a more recent study by Dahmen et al. (2013), who found 20:5/20:5 MGDG and DGDG as major form of galactolipids in *C. velia* at 20 ℃ and 30 ℃, and 20:5/16:0 DGDG as a minor form at 30 ℃. In addition, Dahmen et al. (2013) found 20:5/18:1 and 20:5/20:4 as minor forms of MGDG along with the minor forms 20:5/16:0, 20:4/16:0, and 20:5/20:4 of DGDG at 30 ℃. Note that these minor forms were typically less than 5% each of the total galactolipids, 20:5/20:4 DGDG had the

highest relative percentage of approximately 8%, and that with the exception of 20:5/18:1 MGDG these minor forms were absent at 20 ℃.

Because of the morphological and ultrastructural differences between *V. brassicaformis* and *C. velia*, we were interested in determining the potential similarity, or difference, in the forms of MGDG and DGDG in these two chromerids. Positive-ion ESI/MS and ESI/MS/MS analyses of galactolipids in *V. brassicaformis* showed distinct differences in MGDG and DGDG composition from *C. velia*. For example, 20:5, the most common fatty acid found in *V. brassicaformis*' galactolipids, was found to occupy the *sn*-1 position in all forms of MGDG and DGDG identified. However, unlike *C. velia*, no MGDG or DGDG with a 20:5/20:5 combination was observed in *V. brassicaformis*.

MGDG and DGDG with a $C_{36:5}$ fatty acid combination, which were likely to be the 20:5/16:0 forms of MGDG and DGDG, described initially in *C. velia* as very minor galactolipids (about 4% relative percentage; Botté et al 2011; Dahmen et al. 2013), were also found in *V. brassicaformis* yet at slightly higher relative percentages of approximately 12 and 8%, respectively (Table 1). Importantly, fatty acid combinations of 20:5/14:0, 20:4/16:0, 18:2/18:2, 20:5/18:2, and 20:4/18:2 within MGDG, and 18:2/16:0 and 20:5/18:2 within DGDG were observed in *V. brassicaformis* (Table 1), which was not reported previously in *C. velia* (Botté et al. 2011; Dahmen et al. 2013). These results indicate that at least under these growth conditions, *V. brassicaformis* has a greater diversity of

galactolipids than *C. velia*; this provides biochemical support for the separation of these two species in different families Vitrellaceae and Chromeraceae (Oborník et al. 2012).

The 20:5 fatty acid found in *V. brassicaformis*' MGDG and DGDG was consistently found at the *sn*-1 position as in 20:5/18:5 MGDG and DGDG. A similar phenomenon was also observed previously in a C_{20}/C_{18} cluster of peridinin-containing dinoflagellates (Gray et al. 2009a, b; Leblond et al. 2010). However, MGDG and DGDG with a 20:5/20:5 combination, as observed in *C. velia*, were rarely found in these same peridinin-containing dinoflagellates, often at a relative abundance of less than 10% of the total galactolipids (Gray et al. 2009a), and were absent in *V. brassicaformis*.

In the biosynthesis of galactolipids, the composition of the *sn*-2 fatty acid has been used traditionally to imply differences in the biosynthetic steps, referred to as either the prokaryotic pathway that exclusively occurs in the plastid, or eukaryotic pathway that begins with phosphatidic acid biosynthesis in the endoplasmic reticulum before continuation of galactolipid biosynthesis in the chloroplast. The presence of C_{14} and C_{16} fatty acids in the *sn*-2 position indicates the use of the prokaryotic pathway for galactolipid fatty acid biosynthesis, , whereas the eukaryotic pathway yields C_{18} and C_{20} fatty acids in the *sn*-2 position (Browse & Sommerville 1991).

In the prokaryotic pathway, a diacylglyerol can be synthesized within the chloroplast membrane through a series of biochemical steps known as the Kornberg-Pricer pathway (Joyard et al. 1998). Two acyltransferases, glycerol-3-phosphate acyltransferase and 1-acylglycerol-3-phosphate acyl-transferase, simultaneously acylate 18:1 and 16:0 fatty acids to the *sn*-1 and *sn*-2 positions of glycerol-3-phosphate to synthesize phosphatidic acid, which is further metabolized into a diacylglcyerol or phosphatidyl glycerol (Frentzen 1993; Joyard et al. 1998). A galactose molecule from uridine diphosphate-galactose (UDP-galactose) is then transferred to the *sn*-3 position of the diacylglyerol by MGDG synthase to synthesize MGDG in the plastid. In the subsequent production of DGDG, a galactolipid:galactolipid galactosyltransferase catalyzes the transfer of a galactose molecule between MGDG molecules leading to the formation of DGDG (Joyard et al. 1998).

In the eukaryotic pathway, an 18:1 fatty acid is transported to the endoplasmic reticulum from the plastid (i.e. the original site of biosynthesis) where further desaturation and elongation occurs. The endoplasmic reticulum 1-acyl-*sn*-glycerol-3-phosphate acyltransferase has strong affinity for unsaturated C_{18} acyl groups, and acylates the *sn*-2 position of lysophosphatidic acid to synthesize phosphatidic acid, which is a precursor for the synthesis of the phosphatidylcholine, lysophosphatidaylcholine and diacylglycerol that are then transported to the plastid for utilization in galactolipid biosynthesis (Andersson et

al. 2004; Mongrand et al. 1997; Williams et al. 2000). The differences in galactolipid composition between *V. brassicaformis* and *C. velia* can be used to hypothesize differences in galactolipid biosynthesis. Both MGDG and DGDG in *V. brassicaformis* have a mixture of C_{14}, C_{16}, and C_{18} fatty acids in the *sn*-2 position, indicating the use of both prokaryotic and eukaryotic pathways. *C. velia*, however, completely lacks C_{14} and C_{16} fatty acids in the *sn*-2 position of its forms of MGDG, while C_{16} fatty acids only comprise approximately 6% of *C. velia* DGDG when grown at 30ºC (Dahmen et al. 2013). Thus, *C. velia*'s abundance in C_{18} and C_{20} fatty acids and lack of C_{14} and C_{16} fatty acids in the *sn*-2 position can be explained by *C. velia* using primarily the eukaryotic pathway in galactolipid synthesis.

RNA-seq is a sequencing-based, high-throughput and quantitative technique to survey the entire transcriptome that is cost-effective and is able to annotate and quantitate gene expression levels at genome-wide scale (Wang et al. 2009). As an approach to elucidate the similarity of key galactolipid biosynthetic genes between these two chromerids, we identified expression of galactolipid synthase genes in *V. brassicaformis* after *de novo* assembly of RNA-seq data, pathway mapping using the KEGG automatic annotation server (Moriya et al. 2007), BLAST analysis using KEGG (Kanehisa & Goto 2000), and NCBI BLAST (Altschul et al. 1997) to confirm the function of the assembled contigs. A comparison of the *V. brassicaformis* MGDG synthase revealed similarities with

MGDG synthases from both *A. thaliana* and *C. velia*. The MGDG synthase cDNA we obtained from *V. brassicaformis* appears to be a complete sequence based on an alignment with *A. thaliana* (Fig. 4). Previous studies have determined amino acid residues important in function; Arg260 and Trp287, which have been proposed to be involved in activation by phosphatidylglycerol and phosphatidic acid were both identified in *V. brassicaformis* (Dubots et al. 2010). There were also major differences in sequence found between *V. brassicaformis* and *C. velia*; an insertion between phenylalanine and arginine identified in *C. velia*, which is not found in other algae or plants was lacking in *V. brassicaformis* (Botté et al. 2011). There was also an insertion found in the MGDG synthase between Ala419 and Arg431 composed of several asparagine residues, which was not found in either *A. thaliana* or *C. velia* (Botté et al. 2011). The DGDG synthase amino acid sequence from *V. brassicaformis* was also found to have essential residues including Val32 along with Trp26 and Trp57, which are predicted to be involved in mediating membrane interaction (Ge et al. 2011). However, the *V. brassicaformis* DGDG synthase sequence was also found to have several additional amino acids at the C-terminus when compared to *A. thaliana* and *C. velia*.

Phylogenetic analysis using contigs 5672 and 11390 of MGDG and DGDG synthases, respectively, from *V. brassicaformis* along with MGDG and DGDG synthase sequences from other photosynthetic organisms showed grouping with

algae of red algal origin, thus confirming their shared red algal origin. The phylogenetic analysis of MGDG synthase (Fig. 6) showed that for this particular enzyme *V. brassicaformis* is not as closely related to *C. velia* as for the same type of analysis of DGDG synthase (Fig. 7), though both of them group displayed a red algal origin. We found the galactolipid composition in the two chromerids, *V. brassicaformis* and *C. velia*, to be widely varied, however, the 20:5 fatty acid is highly abundant in both species.

In summary, these biochemical data along with phylogenetic analysis of MGDG and DGDG synthases were found to be indicative of a shared red algal plastid lineage in *V. brassicaformis* and *C. velia*. However, the differences in galactolipid composition also represent a broader than expected diversity of galactolipid-associated fatty acids within the same phyla that appears to be in line with previously described morphological and ultrastructural differences (see Introduction). The varied galactolipid compositions between *C. velia* and *V. brassicaformis* could be attributed to the type of precursors (i.e. diacylglycerols and/or free fatty acids) available for galactolipid synthesis; future studies should focus on radiolabelling studies to decipher differences in galactolipid synthesis by determining the precursors available for galactolipid synthesis in the plastid (and/or those transported from the endoplasmic reticulum to the plastid.

REFERENCES

Adl SM et al. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. 52: 399-451.

Adl SM et al. 2012. The revised classification of eukaryotes. J. Eukaryot. Microbiol. 59: 429-514.

Adlerstein D, Bigogno C, Kohzin I, Cohen Z. 1997. The effect of growth temperature and culture density on the molecular species composition of the galactolipids in the red microalga *Porphyridium cruentum* (Rhodophyta). J. Phycol. 33: 975-979.

Altschul SF, Madden TL, Schaeffer A, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25: 3889-3402.

Anderson MX, Kjellberg JM, Sandelius AS. 2004. The involvement of cytosolic lipases in converting phosphatidylcholine to substrate for galactolipid synthesis in the chloroplast envelope. Biochim. Biophys. Acta. 30: 46-53.

Botté CY, Yamaryo-Botté Y, Janouškovec J, Rupasinghe T, Keeling PJ, Crellin P, Coppel R, Maréchal lE, McConville MJ, McFadden GI. 2011. Identification of plant-like galactolipids in *Chromera velia*, a photosynthetic relative of malaria parasite. J. Biol. Chem. 286: 29893-29903.

Browse J, Somerville C. 1991. Glycerolipid synthesis: Biochemistry and regulation. Annu. Rev. Plant Physiol. Plant Mol. Biol. 42: 467-506.

Burki F et al. 2009. Large-scale phylogenomic analyses reveal that two enigmatic protest lineages, telonemia and centroheliozoa are related to photosynthetic chromalveolates. Genome Biol. Evol. 1: 231-238.

Burki F, Shalchian-Tabrizi K, Pawlowski J. 2008. Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. Biol. Lett. 4: 366-369.

Cavalier-Smith T. 1993. Kingdom protozoa and its 18 phyla. Microbiol. Rev. 57: 953-994.

Dahmen J L, Khadka M, Dodson J, Leblond JD. 2013. Mono- and digalactosyldiacylglycerol composition of dinoflagellates. VI. Biochemical and genomic comparison of galactolipid biosynthesis between *Chromera velia* (Chromerida), a photosynthetic alveolate with a red algal plastid ancestry, to the dinoflagellate, *Lingulodinium polyedrum*. Eur. J. Phycol. 48: 268-277.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: Fast selection of best-fit models of protein evolution. Bioinformatics. 27: 1164-1165.

Dörmann P, Benning C. 2002. Galactolipids rule in seed plants. Trends in Plant Sci. 7: 112-118.

Dubots E, Audry M, Yamaryo Y, Bastien O, Ohta H, Breton C, Maréchal E, Block MA. 2010. Activation of the Chloroplast Monogalactosyldiacylglycerol

Synthase MGD1 by Phosphatidic Acid and Phosphatidylglycerol. J. Biol. Chem. 285: 6003–6011.

Fichera ME, Roos DS. 1997. A plastid organelle as a drug target in apicomplexan parasites. Nature. 390: 407-409.

Frentzen M. 1993. Acyltransferases and triacylglycerol. Moore ST Jr. editors: Lipid metabolism in plants. pp 195-230. CRC Press, Boca Raton.

Funes S, Davidson E, Reyes-Prieto A, MagallónS, Herion P, King MP, González-Halphen D. 2002. A green algal apicoplast ancestor. Science. 298: 2155.

Gardner MJ et al. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature. 419: 498-511.

Ge C, Georgiev A, Öhman A, Wieslander Å, Kelly AA. 2011. Tryptophan residues promote membrane association for a plant lipid glycosyltransferase involved in phosphate stress. J Biol.Chem. 286: 6669-6684.

Gray CG, Lasiter AD, Leblond JD. 2009b. Mono- and digalactosyldiacylglycerol composition of dinoflagellates. III. Four cold adapted, peridinin-containing taxa and the presence of trigalactosyldiacylglycerol as an additional glycolipid. Eur. J. Phycol. 44: 439-445.

Gray CG, Lasiter AD, Li C, Leblond JD. 2009a. Mono-and digalactosyldiacylglycerol composition of dinoflagellates. I. Peridinin-containing taxa. Eur. J. Phycol. 44: 191-197.

Guella G, Frassanito R, Mancini I. 2003. A new solution for an old problem: the regiochemical distribution of the acyl chains in galactolipids can be established by electrospray ionization tandem mass spectrometry. Rapid Commun. Mass Spectrom. 17: 1982-1984.

Guschina IA, Harwood JL. 2006. Lipids and lipid metabolism in eukaryotic algae. Prog. Lipid Res. 45: 160-186

Hölzl G, Dörmann P. 2007. Structure and function of glycoglycerolipid in plants and bacteria. Prog. Lipid Res. 46: 225-243.

Janero DR, Barrnett R. 1981. Cellular and thylakoid-membrane glycolipids of Chlamydomonas reinhardtii 137^{+}. J. Lipid Res. 22: 1119-1125.

Janouškovec J, Horak A, Oborník M, Lukeš J, Keeling PJ. 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc. Natl. Acad. Sci. 107: 10949-10954.

Jones MR. 2007. Lipid in photosynthetic reaction centers: Structural roles and functional holes. Prog. Lipid Res. 46: 56-87.

Joyard J, Maréchal E, Miège C, Block MA, Dorne AJ, Douce R. 1998. Structure, distribution, and biosynthesis of glycerolipids from higher plant chloroplasts. Siegenthaler, P., A., & Murata, N. editors. Lipids in photosynthesis : Structure, function, and genetics. pp. 21-52. Kluwer Academic Publishers. The Netherlands.

Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28: 27-30.

Katoh K, Kumar K, Toh H, Miyata T. 2005. MAFFT version 5: Improvement in accuracy of multiple sequence alignment describes ancestral version of the G-INS-I, L-INS-I and E-INS-I strategies. Nucleic Acids Res. 33: 511-518.

Kobayashi K, Kondo K, Fukuda H, Nishimura M, Ohta H. 2007. Galactolipid synthesis in chloroplast inner envelope is essential for proper thylakoid biogenesis, photosynthesis, and embryogenesis. Proc. Natl. Acad. Sci. 104: 17216-17221.

Köhler S. 1997. A plastid of probable green algal origin in apicomplexan parasites. Science. 275: 1485-1489.

Lau AOT, McElwain TF, Brayton KA, Knowles DP, Roalson EH. 2009. Babesia bovis: A comprehensive phylogenetic analysis of plastid-encoded genes supports green algal origin of apicoplasts. Exp. Parasitol. 123: 236-243.

Leblond JD, Chapman PJ. 2000. Lipid class distribution of highly unsaturated long-chain fatty acids in marine dinoflagellates. J. Phycol. 36: 1103-1108.

Leblond JD, Dahmen JL, Evens TJ. 2010. Mono- and diagalactosyldiacylglycerol composition of dinoflagellates. IV. Temperature-induced modulation of fatty acid regiochemistry as observed by electrospray ionization/mass spectrometry. Eur. J. Phycol. 45: 13-18.

Lim L, McGadden GI. 2010. The evolution, metabolism and function of the apicoplast. Phil. Trans. R. Soc. B. 365: 749-763.

Maddison, WP, Maddison DR. 2011. Mesquite: A modular system for evolutionary analysis. Version 2.75. Available from: http://mesquiteproject.org.

Maréchal E, Azzouz N, de Macedo CS, Block MA, Feagin JE, Schwarz RT, Joyard J. 2002. Synthesis of chloroplast galactolipids in apicomplexan parasites. Eukaryot. Cell. 1: 653-656.

Mongrand S, Bessoule JJ, Cassagne C. 1997. A re-examination in vivo of the phosphatidylcholine- galactolipid metabolic relationship during plant lipid biosynthesis. Biochem. J. 327: 853-858.

Moore RB et al. 2008. A photosynthetic alveolate closely related to apicomplexan parasites. Nature. 451: 959-963.

Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35: W182-W185.

Oborník M, Modrý D, Lukeš M, Černotíková-Stříbrná E, Cihlář J, Tesařová M, Kotabová E, Vancová M, Prášil O, Lukeš J. 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.

Oborník M, Vancová M, Lai D, Janouškovec J, Keeling PJ, Lukeš J. 2011. Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of apicomplexa, *Chromera velia*. Protist. 162: 115-130.

Ramya TN, Mishra S, Armodiya K, Surolia N, Surolia A. 2007. Inhibitors of nonhouse keeping functions of the apicoplast defy delayed death in Plasmodium falciparum. Antimicrob. Agents Chemother. 51: 307-316.

Slabas T. 1997. Galactolipid biosynthesis genes and endosymbiosis. Trends Plant Sci. 2: 161-162.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22: 2688-2690.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolution genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731-2739.

Thompson GA Jr. 1996. Lipids and membrane function in green algae. Biochim. Biophys. Acta. 1302: 17-45.

Welti R, Li W, Li M, Sang Y, Biesiada H, Zhou H, Rajashekar C, Williams T, Wang X. 2002. Profiling membrane lipids in plant stress responses: Role

of phospholipase Dα in freezing-induced lipid changes in Arabidopsis. J. Biol. Chem. 277: 31994-32002.

Wheeler DL et al. 2008. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 36: D13-D21.

Williams JP, Imperial V, Khan MU, Hodson JN. 2000. The role of phosphatidylcholine in fatty acid exchange and desaturation in *Brassica napus* L. leaves. Biochem. J. 349: 127-133.

Woehle C, Dagan T, Martin WF, Gould SB. 2012. Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related *Chromera velia*. Genome Biol. Evol. 3: 1220-1230.

CHAPTER III

STEROL COMPOSITION AND BIOSYNTHETIC GENES OF *VITRELLA BRASSICAFORMIS*, A RECENTLY DISCOVERED CHROMERID: COMPARISON TO *CHROMERA VELIA*

INTRODUCTION

The phylum Chromerida was first described under the group Alveolata after the discovery of *Chromera velia* in 2008 (Adl et al. 2012; Moore et al. 2008). The second discovered species, now called *Vitrella brassicaformis,* was first used as an unnamed chromerid to demonstrate a shared red algal plastid ancestry with *C. velia*, dinoflagellates, heterokonts, and non-photosynthetic apicomplexans which possess a non-photosynthetic relic plastid termed an apicoplast (Janouškovec et al. 2010). Later the formal description of *V. brassicaformis* was provided by studying its morphology, ultrastructure, and life cycle, and it was formally classified under the phylum Chromerida and family Vitrellaceae (Oborník et al. 2012). Phylogenetic analysis of eight concatenated nuclear genes has also shown that *V. brassicaformis* and *C. velia* are both close relatives of apicomplexan parasites (Janouškovec et al. 2010).

Comparatively, *C. velia*, *V. brassicaformis,* and dinoflagellates possess a combination of similar, as well as different, features, thus illustrating that even though they are in the same phylum, there are distinct differences between these

only two chromerids isolated to date. For example, morphological and genetic features, such as tubular mitochondrial cristae, cortical alveoli, subpellicular microtubules, heterodynamic flagella, terminally tapered flagella, micronemes, and a bacterial Rubisco gene, are found in all of them, whereas a pseudoconoid, chromerosome, finger-like projection on shorter flagellum, four-celled sporangia are only found in *C. velia* but not in *V. brassicaformis* (Oborník et al. 2011; Oborník et al. 2012). *V. brassicaformis* also possesses multiple cells containing sporangia, a multiple laminated cell wall, pyrenoid, and compacted plastid genome that are absent in *C. velia* (Oborník et al. 2012).

Sterols are isoprenoid-derived, amphipathic, ringed lipids that are commonly biosynthesized *de novo* in many eukaryotes via either a classical acetate/mevalonate pathway (MVA) and/or a more recently discovered non-mevalonate/methylerythritol pathway (MEP) (Bloch 1965; Eisenreich et al. 2004; Nes 2011). Sterols perform several essential functions, such as regulating membrane dynamics that affect cell proliferation and differentiation, displaying hormonal activity in plants, and acting as an effector molecule by modulating function of membrane bound proteins such as H^+-APTase (Dufourc 2008; Grunwald et al 1971; Hartmann 1998; Nes 2011). Because of their prevalence in eukaryotes, sterols have also been utilized as molecular biomarkers to trace algal productivity and to make chemotaxonomic inference for classifying various algal groups (Rampen et al. 2010; Moldowan et al. 1990). For example,

4α,23,24*R*-trimethyl-5α-cholest-22E-en-3β-ol (dinosterol) is a “molecular fossil” which is indicative of the class Dinophyceae, and has been utilized as a biomarker for determining dinoflagellate blooms in the distant past (Boon et al. 1979; Volkman 1986). As an example, abundant sterols were found in the extract of a sapropel layer of the Black sea sediment, which dates back to 7,000-3,000 yr before present (BP), during which the sea water was brackish due to gradual change from fresh water to marine water, suggesting periodic blooming of the dinoflagellates (Boon et al. 1979). Studies have also been done to correlate the clustering of various algal groups as based on sterol composition with the phylogenetic clustering using ribosomal RNA gene sequences (Leblond et al. 2010; Rampen et al. 2010).

Most eukaryotes synthesize sterols *de novo*, however parasitic apicomplexans scavenge sterols from the host cell in the form of cholesteryl esters (Coppens 2006; Nishikawa et al. 2005). *V. brassicaformis* and *C. velia*, being closely related to apicomplexans (Janouškovec et al. 2010), can be utilized to speculate how the sterols were biosynthesized in apicomplexans before the biosynthetic machinery was lost. A recent study on the sterol composition of *C. velia* showed the presence of 24*R*-methylcholesta-5,22E-dien-3β-ol, 24*R*-ethylcholesta-5,22E-dien-3β-ol and 24-ethylcholest-5-en-3β-ol as major sterols; these three sterols are common in higher plants and seemingly unrelated algae (see Discussion). Referring to the Kyoto Encyclopedia of Gene and Genomes (KEGG) pathway

database, there are 6 and 7 enzymatic steps that leads to the synthesis of isopentenyl diphosphate via mevalonate and non mevalonate pathways, respectively. Out of these and following enzymatic steps that lead to formation of these three sterols, only three sterol biosynthetic genes, namely- *smt1*, *fdft1*, and *idi1*, were discovered in *C. velia* after examination of an expressed sequence tag (EST) database (Leblond et al. 2012). Out of two possible pathways, the MVA and MEP pathways, for sterol biosynthesis, the presence of the *fdft1* gene suggested the presence of the mevalonate pathway in *C. velia*. However, further study of the sterol composition and biosynthetic genes in *V. brassicaformis* and *C. velia* (using the more powerful RNA Seq technique) can be useful to better decipher sterol biosynthesis in these two chromerids, in addition to identifying chemotaxonomic and phylogenetic relationships between themselves, and possibly dinoflagellates and apicomplexans. Therefore, we conducted this study with the following objectives: 1) To determine the sterol composition of *V. brassicaformis* and compare it to the previously determined sterol composition of *C. velia* (and other algal groups, such as dinoflagellates), 2.) To identify via RNA Seq genes involved with sterol biosynthesis in both *V. brassicaformis* and *C. velia* with the goal of identifying usage of the mevalonate and/or non-mevalonate pathway(s) for sterol biosynthesis, and 3.) To make phylogenetic comparisons of these genes with other algae, such as dinoflagellates, and parasitic apicomplexans (which have lost the ability to produce sterols *de novo* but still

retain some biosynthetic genes). To this end, we present this first characterization of the sterols and related biosynthetic genes of *V. brassicaformis*.

MATERIALS AND METHODS

Culture and lipid extraction

Vitrella brassicaformis NCMA 3155 (also known as CCMP 3155) was procured from the Provasoli-Guillard National Center for Marine Algae and Microbiota (West Boothbay Harbor, ME, USA). The cultures were grown autotrophically in triplicate in L1 medium (Guillard and Hargraves 1993) at room temperature under a 14/10h light/dark cycle at an irradiance of 50 µmol photon/m^2/sec and harvested during late exponential phase by filtration onto a precombusted 934-AH Whatman glass fiber filter. Lipids were extracted according to the technique described by Leblond and Chapman (2000).

Processing and analysis of sterols

The total lipid extract was separated into five different fractions via column chromatography using 1 g of 100-200 mesh Unisil silica (Clarkson Chromatography, South Williamsport, PA, USA) activated at 120℃ for 4h. The following solvent regime was used to separate lipids according to polarity (Leblond and Chapman 2000): 1) 12 ml methylene chloride (sterol esters), 2) 15

ml 5% acetone in methylene chloride with 0.05% acetic acid (free sterols, tri-and diacylglycerols, and free fatty acids), 3) 10ml 20% acetone in methylene chloride (monoacylglycerols), 4) 45 ml acetone (galactolipids), and 5) 15 ml methanol with 0.1% acetic acid (polar lipids, including phospholipids and betaine lipids).

Fraction two that contains free sterols was saponified according to the techniques described by Leblond and Chapman (2002). After saponification, sterols were derivatized with 0.5 ml *N,O-bis*(trimethylsilyl)trifluoroacetamide containing 1% trimethylethylchlorosilane at 80℃ fo r 0.5 h. The reagent was evaporated under a stream of nitrogen and the derivatives redissolved in 20 µl of 1:1 hexane/methyl-*tert*-butyl ether. Gas chromatography/mass spectrometry (GC/MS) was used to analyze the derivatized sterols on a using the following conditions: 1 µl injected via splitless injection with injector set at 280 ℃, transfer line set at 275 ℃, helium carrier at 28 cm/sec, 70 eV with a scanning range of 50-600 amu and a cycle time of 1.1 sec using a Schimadzu GC-17A with a QP-5000 MS unit (Shimadzu Scientific Instruments, Columbia, MD, USA) and a DB-5 column (30 m X 0.25 µm film thickness, J&W Scientific Incorporated, Golsom, CA). The GC temperature was 50 ℃ for 1 min, 50-170 ℃ at 15 ℃/min, 170 ℃-300 ℃ at 10 ℃/min with a hold of 11 minRelative r etention times (RRT) to cholest-5-en-3β-ol (cholesterol) were calculated according to the methodology of Jones et al. (1994).

Identification of sterol biosynthetic genes using transcriptome data

The Sequence Read Archive (SRA) database from National Center for Biotechnology and Information (NCBI) (Wheeler et al. 2008) was utilized to retrieve transcriptome data for *V. brassicaformis* NCMA 3155. The data can be accessed using SRA accession number SRX215482 or using the following link: http://www.ncbi.nlm.nih.gov/sra?term=(SRX215482)%20NOT%20cluster_dbgap%5BPROP%5D. The short transcripts were *de novo* assembled to 50,755 contigs using CLC Genomics Workbench (version 6.0.2) and mapped to Kyoto Encyclopedia Genes and Genomes Pathway (KEGG) automatic annotation server (KAAS) (Moriya et al. 2007) for the identification of sterol biosynthetic genes. The KAAS mapping identified 12 genes related to sterol biosynthesis that included the genes related to an isoprenoid biosynthetic pathway. The transcriptome data for *C. velia* was accessed from NCBI-SRA database (Wheeler et al. 2008) using the link http://www.ncbi.nlm.nih.gov/sra/SRX275147 and accession number SRX275147. All the assembled transcripts were further utilized to identify the genes related to sterol and isoprenoid biosynthetic pathway using KAAS. The KAAS annotation showed six genes related to sterol biosynthetic pathway that included genes related to an isoprenoid biosynthetic pathway.

Phylogenetic analyses of sterol biosynthetic genes

The genes sequences for 1-D-deoxy-xylulose phosphate reductoisomerase (*dxr*), 2-C-methyl-D-erythritol 2, 4-cyclodiphospahte synthase (*isp*F), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (*isp*D) and isopentenyl diphophaste isomerase (*idi*) identified from the transcriptome analyses of *V. brassicaformis* and *C. velia*, and were utilized to predict the protein sequence using an online tool "The genescan server at MIT" (Burge et al. 1998). The parameters were set up as organisms: *Arabidopsis*; suboptimal exon cutoff: 1; print option: predicted CDS and peptides. The FASTA-formatted protein sequences for the apicomplexans, dinoflagellates, diatoms, and plants were searched using gene names in NCBI database (Sayers et al. 2009). The assembled transcript along with the NCBI-retrieved protein sequences were utilized for the construction of maximum likelihood trees for the four genes *dxr*, *ispF*, *isp*D and *idi*. All the protein sequences were MUSCLE aligned using MEGA 6.0 (Tamura et al. 2013) and converted to PHYLIP format using Mesquite version 2.75 (Madison et al. 2011). The aligned protein sequences were analysed to find the best model fit using ProtTEST 3.3 software. The best model was chosen and used in RAxML (Stamatakis 2014) for the construction of phylogenetic trees that was installed in Cyberinfrastructure for phylogenetic research (CIPRES) gateway version 3.3 (Miller et al. 2010).

RESULTS

Sterol profile of *Vitrella brassicaformis*

Fraction 2 of the total lipid that contained free sterol was derivatized and analyzed via GC/MS to examine the sterol profile of *V. brassicaformis*. The analysis identified two sterols- 24-ethylcholest-5-en-3β-ol at *m/z* 486, retention time 42.54 minutes, relative retention time 1.13 relative to cholesterol and relative abundance of approximately 83.32±1.92%; and an unknown $C_{26:2}$ sterol at *m/z* 442, retention time 37.65 minutes, relative retention time 0.29 relative to cholesterol and relative abundance of approximately 16.68±1.92% (Figures 1 and 2). The identified sterol was confirmed by comparing to the spectrum of a sitosterol standard procured from TCI America, Montgomeryville, PA, USA, product number S0040 with 35% purity. 24-ethylcholest-5-en-3β-ol showed a fragmentation pattern of M^+ - H – side chain (*m/z* 255), M^+- TMS-O – CH_3 (*m/z* 381), M^+ - TMS-O – H (*m/z* 396), and M^+ - CH_3 (*m/z* 471).

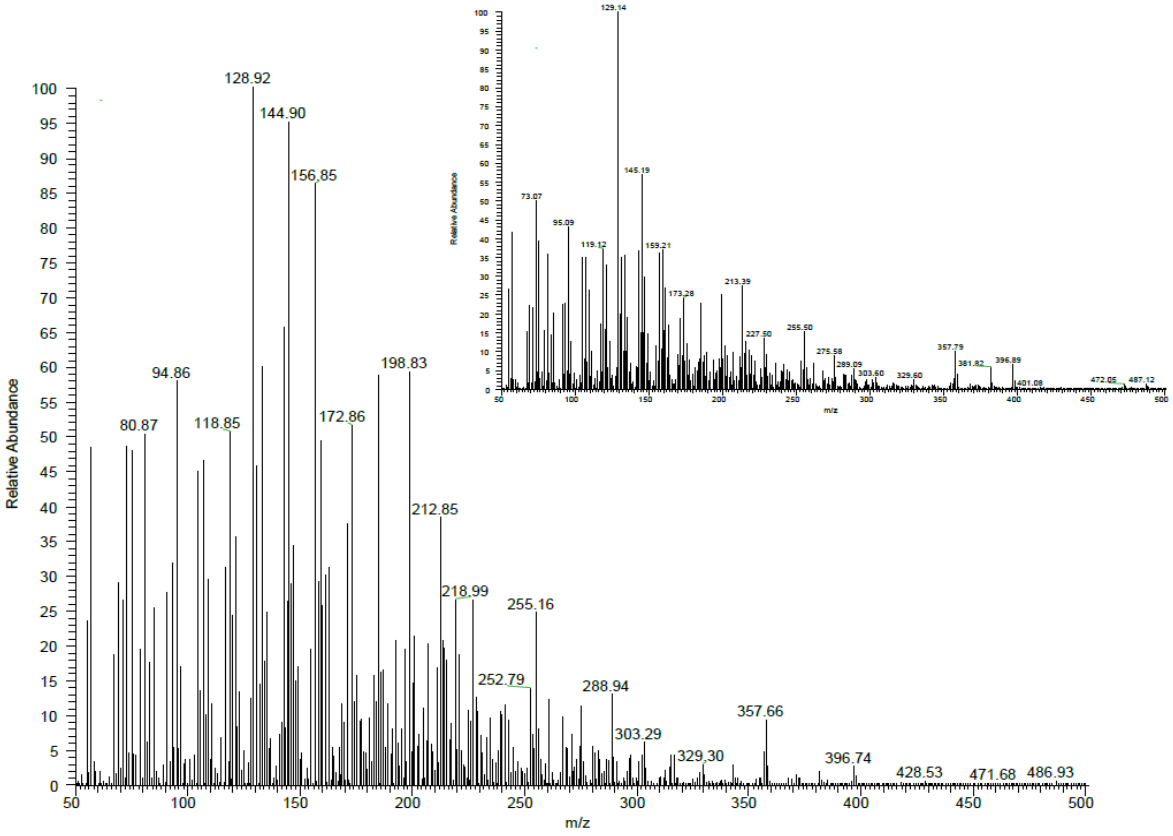


Figure 1. Mass spectra of trimethylsilyl ether (TMS) derivatives of sterol of Vitrella brassicaformis for 24-ethylcholest-5-en-3β-ol (24-ethylcholest-5-en-3β-ol standard is shown in the inset).

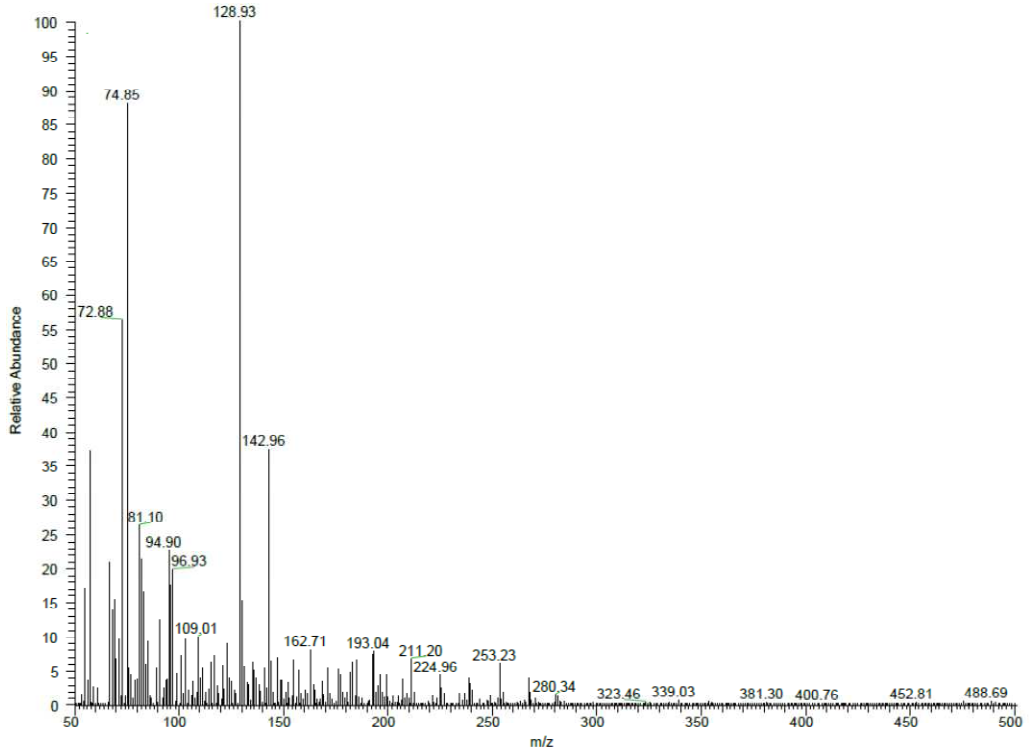


Figure 2. Mass spectra of trimethylsilyl ether (TMS) derivatives of sterol of V. brassicaformis for unknown $C_{26:2}$ sterol.

Phylogenetic analysis using protein sequence for *dxr, ispF, ispD* and *idi* genes

Phylogenetic analyses were performed using four different genes related to isoprenoid and sterol biosyntheses. The relationship was constructed using plants, red and green algae, diatoms, dinoflagellates, and apicomplexans wherever the data were available.

A maximum likelihood tree for protein sequence for the *dxr* gene (Figure 3) showed *V. brassicaformis* grouped with the peridinin-containing dinoflagellate*Pyrocystis lunula* with a bootstrap support of 100*. V. brassicaformis* and *P. lunula* further grouped with *Lepidodinium chlorophorum* (bootstrap support 100) and all those three species group with *Oxyrrhis marina* with bootstrap support value 91. The dinoflagellates and *V. brassicaformis* were closely grouped with perkinsids and then to apicomplexans, suggesting *V. brassicaformis* is more closely related to apicomplexans than *C. velia*. *C. velia* groups with red algae (rhodophytes) and heterokonts.

V. brassicaformis formed a sister clade with *C. velia* in the analyses of the *ispF* gene (Figure 4), and the *isp*D gene (Figure 5) showed *Nannochloropsis gaditana*, a heterokonts, *a*s a close relative to *V. brassicaformis* and *Guillardia theta*, a cryptophyte to *C. velia*. *V. brassicaformis* and *C. velia* formed a sister clade in maximum likelihood tree constructed using *idi* gene and closely related to *N. gaditana* (Figure 6).

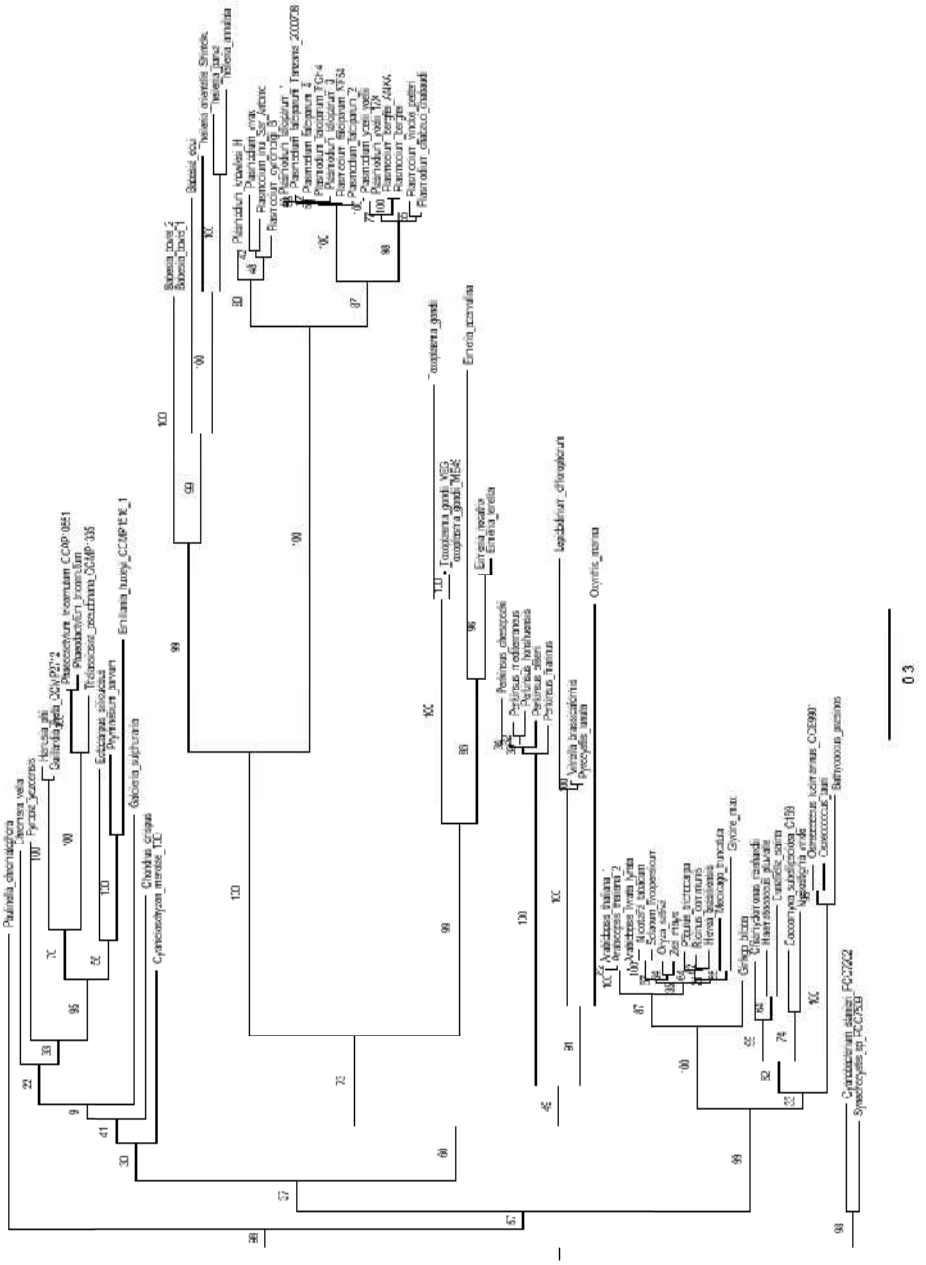


Figure 3. Maximum likelihood tree generated bby using RAxML with LG+G model for 1 deoxy-D-xylulose 5-phosphate reductoisomerase protein sequence.

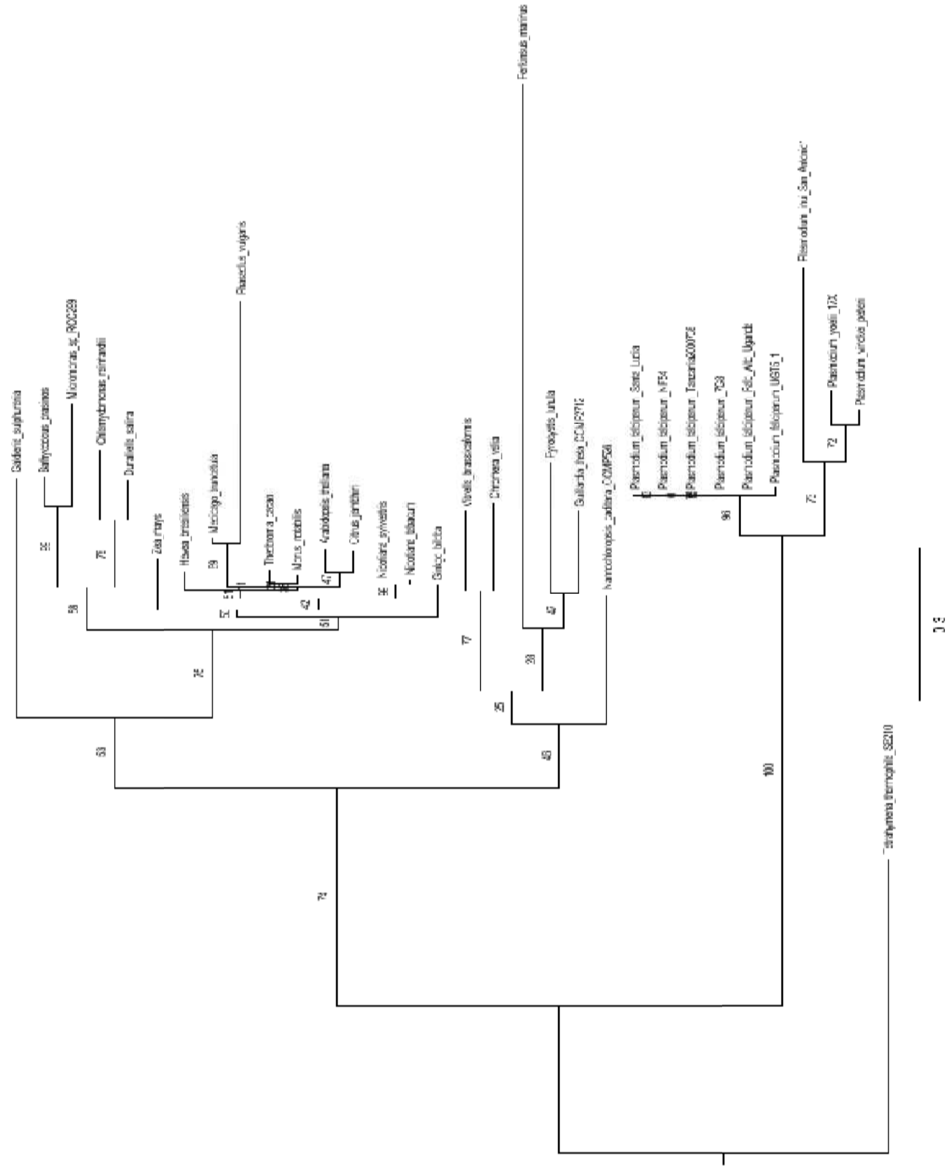


Figure 4. Maximum likelihood tree generated by using RAxML with LG+G model for 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase protein sequence.

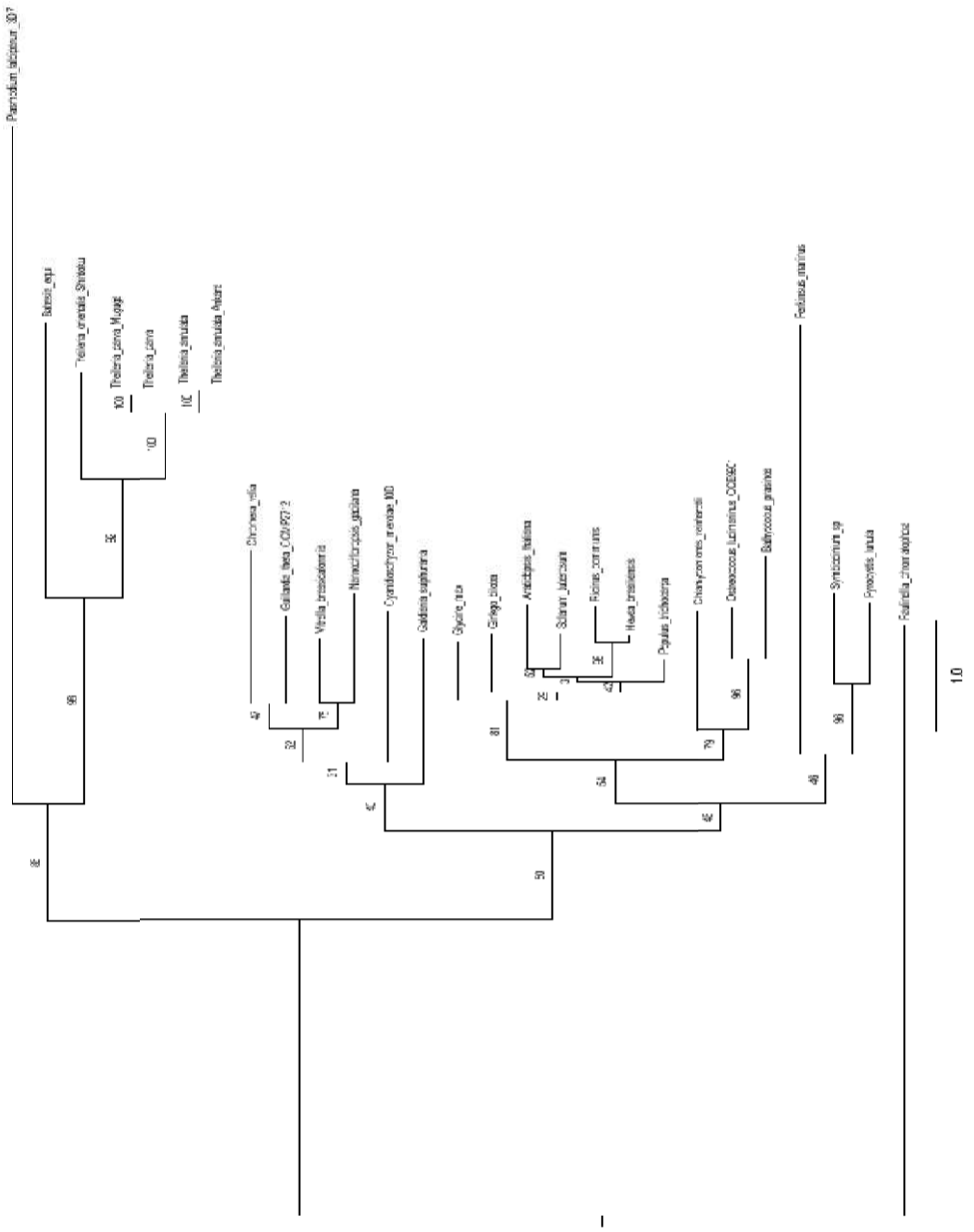


Figure 5. Maximum likelihood tree generated by using RAxML with LG+G model for 2 C-methyl-D-erythritol 4-phosphate cytidyltransferase protein sequence.

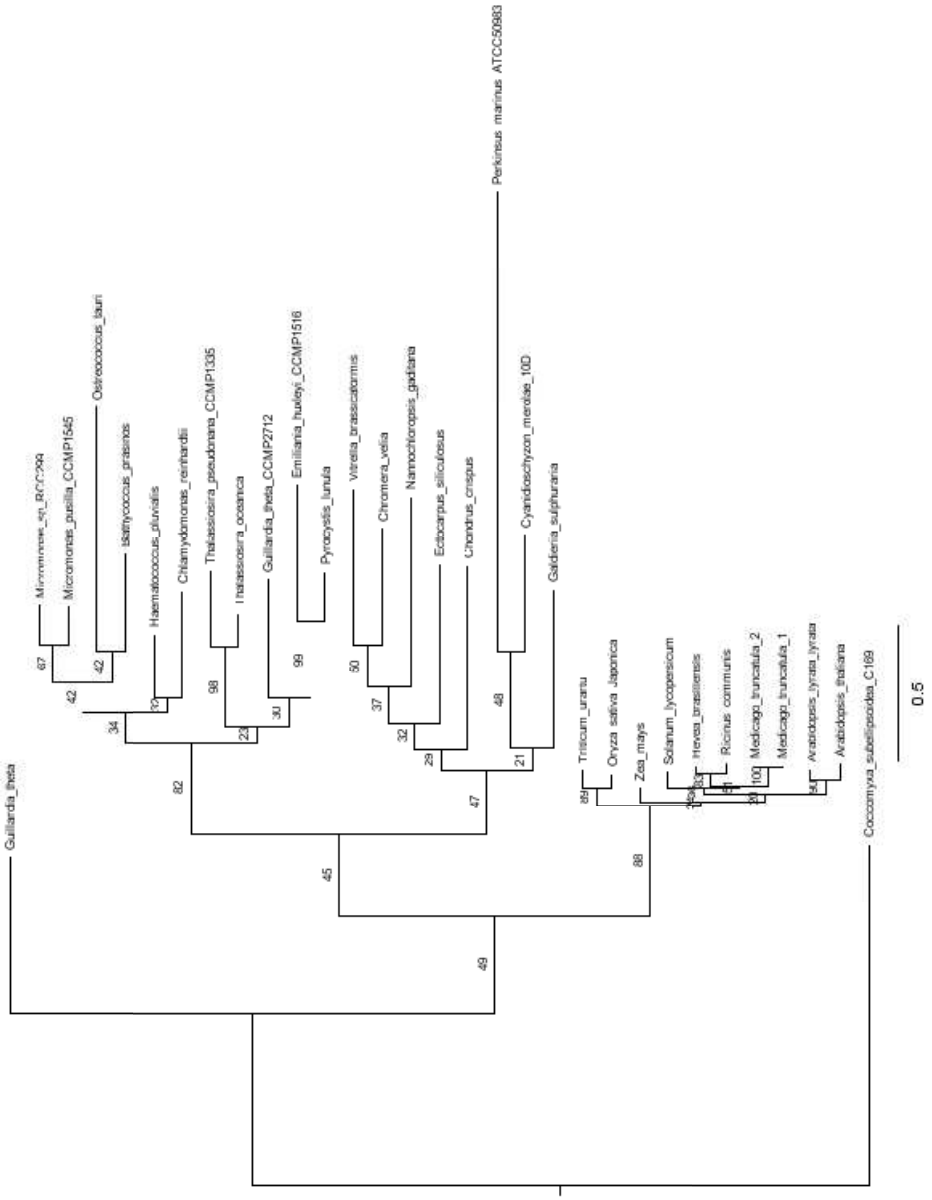


Figure 6. Maximum likelihood tree generated by using RAxML with LG+G model for isopentenyl diphosphate isomerase protein sequence

In summary, the four maximum likelihood trees using four different isoprenoid and sterol biosynthetic genes demonstrated rhodophytes as a possible ancestor of sterol biosynthetic genes in *V. brassicaformis* and *C. velia*. *V. brassicaformis* and *C. velia* form a sister clade in the phylogenetic tree constructed using *ispF* and *idi* genes but phylogenetic analyses using *dxr* and *ispD* genes showed grouping of *V. brassicaformis* and *C. velia* to *P. lunula* and *G. theta.*

DISCUSSION

The phylogenetic (Janouškovec et al. 2010) and ultrastructure (Oborník et al. 2012) studies previously carried out on *V. brassicaformis* were focused on evolutionary relationships between it and *C. velia*, and between these two chromerids and other groups of algae and non-photosynthetic apicomplexans *V. brassicaformis* and *C. velia* have been classified under the phylum Chromerida as based on ultrastructural features and phylogenetic analyses of nuclear and plastid genes, and both species have been shown to be closely related to apicomplexans (Moore et al. 2008; Janouškovec et al. 2010; Oborník et al. 2012). Apicomplexans are classified under the group Alveolata and are characterized by parasitic mode of life (Adl et al. 2012). This group consists of species that cause devastating diseases like malaria and toxoplasmosis in humans and coccidiosis in poultry.

Apicomplexans lack the ability to synthesize sterols *de novo*, although they possess an enzyme to convert cholesterol to cholesteryl esters (Bottè et al. 2013; Nishikawa et al. 2005). Thus, a comparison on sterol composition could not be made between apicomplexans and the chromerids, *V. brassicaformis* and *C. velia*. Dinoflagellates, a closely related group as based on ultrastructural and phylogenetic features (Janouškovec et al. 2010; Moore et al. 2008; Oborník et al. 2012), however, synthesize a wide variety of sterols *de novo* (Patterson 1971; Patterson 1991) and can be a pertinent comparison in terms of sterol composition.

Of the two different sterols produced *de novo* during autotrophic growth of *V. brassicaformis*, one of them was identified as 24-ethylcholest-5-en-3β-ol and the other was an unidentified $C_{26:2}$ sterol. 24-ethylcholest-5-en-3β-ol is the predominant sterol in both *V. brassicaformis* and *C. velia*; however, *C. velia* produces a an additional array of sterols that are commonly found in plant and other groups of algae but not in dinoflagellates (Leblond et al. 2012). For example, *P. lunula*, a peridinin-containing dinoflagellate produces cholesterol and 4,24-dimethyl-5α-cholestan-3β-ol as its major sterols (Dahmen et al. 2011), the non-peridinin-containing, bloom- forming dinoflagellate, *Karenia brevis*, produces two predominant sterols, (24S)-4α-methyl-5α-ergosta-8(14),22-dien-3β-ol (gymnodinosterol) and 27-nor-(24S)-4α-methyl-5α-ergosta-8(14),22-dien-3β-ol (brevesterol) that are not produced by other dinoflagellates (Leblond et al. 2002).,

Rhodophytes typically produce cholesterol and several other sterols such 5α-cholestan-3β-ol (cholestanol), cholest-5,24-diene-3β-ol (desmosterol), 22-cholest-5,7-dien-3β-ol (dehydrocholesterol) (Patterson 1991). However, several classes of algae, such as seemingly unrelated to chromerids and with green algal plastid ancestry, such as glaucocystophytes (Leblond et al. 2011), and chrysophytes (Volkman 1993), and chlorarachniophytes (Leblond et al. 2005) have been reported to produce sitosterol.

Sterols are synthesized via two pathways- MVA and MEP. And both synthesize a common sterol biosynthetic intermediate, isopentenyl diphosphate, although they utilize different precursors to arrive at this point. The MVA (Bloch 1965) is used by many eukaryotes, and is found in the cytosol of higher plants (Kuzuyama 2002). The mevalonate pathway leads to formation of isoprene unit, isopentenyl diphosphate after completing six enzymatic steps. Acetyl CoA is first converted into acetotatoacetyl CoA by an enzyme acetoacetyl CoA thiolase which is then converted into 3-hydroxy-3-methylglutaryl CoA (HMG CoA) by 3-hydroxy-3-methylglutaryl CoA synthase. HMG CoA is converted to mevalonate and then to phosphomevalonate by the enzymes HMG CoA reductase and phosphomevalonate kinase respectively. Phosphomevalonate kinase also converts phosphomevalonate into diphosphomevalonate which is ultimately converted to isopentenyl diphosphate by an enzyme mevalonate diphophate decarboxylase (Nes 2011). Conversely, the MEP (Rohmer et al. 1996) is

commonly found in eubacteria, green algae, and the chloroplasts of higher plants (Kuzuyama 2002), utilizes glyceraldehyde 3-phosphate and pyruvate as precursors to synthesize isopentenyl diphophate. Glyceraldehyde 3-phosphate and pyruvate is condensed to 1-deoxy-D-xylulose-5-phopsphate (DOXP) by an enzyme DOXP synthase. DOXP is further converted to 2-C-methyl-D-erythritol-4-phosphate (MEP) and then to cytidine 5'- diphosphate-2-C-methyl-D-erythritol (CDP-ME) by two enzymes- DOXP-reductoisomerase and CDP-ME-synthase respectively. CDP-ME is further converted to CDP-methyl-D-eryhritol-2-phosphate (CDP-ME2P), 2-C-methyl-D-erythritol-2,4-cyclo-diphosphate (MEcPP), 4-hydroxy-3-methyl-2-(E)-butenyl-diphosphate (HMBPP), and then to isopentenyl diphosphate by four enzymes CDP-ME-kinase, MEcPP-synthase, HMBPP-synthase, and HMBPP-reductase respectively (Lichtenthaler 2010). The two molecules of isopentenyl diphosphate are condensed to form an isomer, dimethylallyl diphosphate, which is further cyclized to form squalene 2,3-oxide.

Beyond this step, Benveniste (1986) divides the overall pathway of sterol biosynthesis into four basic steps: the lanosterol-cycloartenol bifurcation, the alkylation reaction that adds methyl and ethyl groups at C-24 position of sterol, opening of the cyclopropane ring of cycloeucalenol and obtusifoliol, and the later stage characterized by removal of methyl group at C-4 and C-14 position of sterol, conversion of Δ^8 into Δ^5 sterols. The lanosterol pathway is common in non-photosynthetic eukaryotes, such as animals and fungi, whereas the

cycloartenol pathway is commonly found in photosynthetic eukaryotes, including algae, bryophytes, and tracheophytes (Benveniste 1986). Lohr et al. (2012) present a tabular summary for the presence of MVA and MEP pathways in various groups of algae and protists, among which the apicomplexa and dinoflagellates possess the MEP pathway, whereas other groups of algae such as the rhodophytes, glaucocystophyte, streptophytes, euglenophytes, chlorarachniophtyes, heterokontophytes, haptophytes, and cryptophytes possess both the pathways.

The MEP pathway comprises seven genes: 1-deoxy-D-xylulose 5-phosphate synthase (*dxs*), 1-deoxy-D-xylulose 5-phosphate reductoisomerase (*ispC*/*dxr*), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (*ispD*/*ygbP*), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (*ispE*/*ychB*), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (*ispF*/*ygbB*), 4-hydroxy-3-methylbut-2-enyl diphosphate synthase (*hds*), and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (*hdr)* (as reviewed by Lichtenthaler 2010; Rohdich et al. 2001) that lead to the biosynthesis of isopentenyl diphosphate. The transcriptome analysis of *V. brassicaformis* identified all the seven genes related to MEP pathway however, no genes related to MVA pathway were identified (Appendix A) suggesting that the sterol biosynthesis in *V. brassicaformis* occurs through MEP pathway.

Protein sequences related to isoprenoid biosynthesis (*ispF*, *ispD*, *ispC* and *idi*) found in the transcriptomes of both *C. velia* and *V. brassicaformis* were utilized for generating maximum likelihood trees with bootstrap support to see the evolutionary relationship as it relates to sterol biosynthesis between *C. velia* and *V. brassicaformis* and these chromerids with other groups of algae. According to chromoalveolate hypothesis (Cavalier-Smith 1999) four major eukaryotic lineages- cryptophytes, haptophytes, heterokonts, and alveolates- that contain chlorophyll *c* originated from a single endosymbiotic event between the ancestors of those groups and a red alga(e). Several phylogenetic studies have however challenged this hypothesis suggesting the multiple endosymbiotic events in chromalveolates (Baurin et al. 2010); these are evidenced by the occurrence of ancient recruited green algal-derived genes in diatoms and other chromalveolates (Moustafa et al. 2009; Frommolt et al. 2008). Phylogenetic analyses of various genes such assignal recognition docking protein, folate biopterin transporter, vitamin k epoxide reductase, and fructose-bisphosphate aldolase in *C. velia,* have shown sharing of both red and green algal-derived genes (Burki et al. 2012; Woehle, et al. 2011). Thus, in performing gene-by-gene analyses of isoprenoid biosynthetic genes in *V. brassicaformis*, it can be determined if there are any green algal derived genes in isoprenoid biosynthetic pathway as seen in *C. velia* and other chromalveolates.

The maximum likelihood tree generated using the protein sequence of 1-deoxy-D-xylulose-5-phosphate reductoisomerase grouped *V. brassicaformis* with dinoflagellates, with *P. lunula* forming a sister clade that was supported with a high bootstrap value of 100 (Figure 3). *Vitrella brassicaformis* formed a sister clade with perkinsids, and perkinsids along with *V. brassicaformis* and dinoflagellates formed a sister clade with apicomplexans. *C. velia* grouped with red alga rhodophytes and heterokonts. Thus the maximum likelihood tree using 1-deoxy-D-xylulose-5-phosphate reductoisomerase protein sequence (Figure 3) suggests that *V. brassicaformis* is more closely related to apicomplexans than *C. velia* and both of these species have a red algal-derived *ispC* gene.

V. brassicaformis and *C. velia* formed a sister clade with bootstrap support of 77 in the maximum likelihood tree constructed using the 2-C-methyl-D-erythritol 2, 4-cyclodiphosphate synthase protein sequence (Figure 4). *V. brassicaformis* forms sister clade with an eustigmatophytes, *N. gaditana* (bootstrap support 75) and *C. velia* form sister clade with *Guillardia theta* (bootstrap support 47) as shown by the likelihood tree generated using 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase protein sequence (Figure 5).

The likelihood tree generated using isopentenyl diphosphate isomerase sequence groups *V. brassicaformis* and *C. velia* (bootstrap support 50) together forming a sister clade which further groups with *N. gaditana* and rhodophytes (Figure 6). The phylogenetic grouping of *V. brassicaformis* with *N.gaditana* is

consistent with finding of Oborník et al. 2012 that suggests morphological similarity and identical pigmentation between *V. brassicaformis* and eustigmatophytes. All the maximum likelihood trees support the red algal origin of the analysed genes and none of the trees showed relatedness of *V. brassicaformis* with green algae.

In conclusion, *C. velia* produces a larger diversity of sterols than *V. brassicaformis*. Sitosterol produced by *V. brassicaformis* is also found in other group of algae and higher plants and thus cannot be of biomarker utility. The phylogenetic analyses of isoprenoid biosynthetic genes in *V. brassicaformis* and *C. velia* showed their red algal origin.

REFERENCES

Adl SM. 2012. The revised classification of eukaryotes. J. Eukaryot. Microbiol. 59: 429-493.

Benveniste P. 1986. Sterol biosynthesis. Ann. Rev. Plant Physiol. 37: 275-308.

Benveniste P. 1999. Biosynthesis and accumulation of sterols. Ann. Rev. Plant Biol. 55: 429-457.

Bloch K. 1965. The biological synthesis of cholesterol. Science. 150: 1928.

Boon JJ, Rijpstra WIC, Lange FD, Leeuw JWD, Yoshioka M, Shimizu Y. 1979. Black sea sterol- a molecular fossil for dinoflagellate blooms. Nature. 277: 125-127.

Botté CY et al. 2013. Atypical lipid composition in the purified relict plastid (apicoplast) of malaria parasites. Proc. Natl. Acad. Sci., 110: 75-6-7511.

Burge CB, Karlin S. 1998. Finding the genes in genomic DNA. Curr. Opin. Struct. Biol. 8: 346-354.

Burki F, Flegontov P, Oborník M, Cihlář J, Pain A, Lukeš J, Keeling PJ. 2012. Re-evaluating the green versus red signal in eukaryoes with secondary plastid of red algal origin. Genome Biol. Evol. 4: 626-635.

Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: Euglenoid, dinoflagellate, and sporozoan plastid origin and the eukaryote family tree. J. Eukaryot. Microbiol. 46: 347-366.

Coppens I. 2006. Contribution of host lipids to *Toxoplasma* pathogenesis. Cell. Microbiol. 8: 1-9.

Dahmen JL, Leblond JD. 2011. Free sterol composition of species in the dinoflagellate genus *Pyrocystis lunula*: a spectrum of sterol diversity. J. Eukaryot. Microbiol. 58: 475-479.

Dufourc EJ. 2008. Sterols and membrane dynamics. J. Chem. Biol. 1: 53-77.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792-1797.

Eisenreich W, Bacher A, Brigoni D, Rohdich F. 2004. Biosynthesis of isoprenoids via the non mevalonate pathway. Cell. Mol. Life Sci. 61: 1401-1426.

Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zauner S, Maier UG, Grossman AR, Bhattacharya D. 2008. Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. Mol. Biol. Evol. 25: 2653-2667.

Grunwald C. 1971. Effects of free sterols, steryl ester, and steryl glycoside on membrane permeability. Plant Physiol. 48: 653-655.

Hartmann MA. 1998. Plant sterols and the membrane environment. Trends in Plant Sci. 3: 170-175.

Khadka M, Dahmen JL, Mohamed S, Leblond JD. 2014. Comparative study of galactolipid composition and biosynthetic genes for galactolipid synthases

in *Vitrella brassicaformis* and *Chromera velia*, two recently identified chromerids with red algal derived plastids. Algol. Stud. 144: 73-93.

Kuzuyama T. 2002. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. Biosci. Biotechnol. Biochem. 66: 1619-1627.

Leblond JD, Chapman PJ. 2002. A survey of the sterol composition of the marine dinoflagellates *Karenia brevis*, *Karenia mikimotoi*, and *Karlodinium micrum*: distribution of sterols within other members of the class dinophyceae. J. Phycol. 38: 670-682.

Leblond JD, Lasiter AD. 2012. Sterols of the green-pigmented, aberrant plastid dinoflagellate, *Lepidodinium chlorophorum* (Dinophyceae). Protist. 163: 38-46.

Leblond JD, Dahmen JL, Seipelt RL, Elrod-Erickson MJ, Kinaid R, Howard JC, Evens TJ, Chapman PJ. 2005. Lipid composition of chlorarachniophytes (Chlorarachniophyceae) from the genera *Bigelowiella*, *Gymnochlora*, and *Lotharella*. J. Phycol. 41: 311-321.

Leblond JD, Dodson J, Khadka M, Holder S, Seipelt RL. 2012. Sterol composition and biosynthetic genes of the recently discovered photosynthetic alveolate, *Chromera velia* (Chromerida), a close relative of apicomplexans. J. Eukaryot. Microbiol. 0: 1-7.

Leblond JD, Timofte HI, Roche SA, Porter NM. 2011. Sterols of glaucocystophytes. Phycolog. Res. 59: 129-134.

Leblond JD, Lasiter AD, Li C, Logares R, Rengefors K, Evens TJ. 2010. A data mining approach to dinoflagellate clustering according to sterol composition: Correlations with evolutionary history. Int. J. Data Min. Bioinformatics. 4: 431-451.

Lichtenthaler HK. 1999. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 50: 47-65.

Lichtenthaler HK. 2010. The non-mevalonate DOXP/MEP (Deoxyxylulose 5-phosphate/methylerythritol 4-phosphate) pathway of chloroplast isoprenoid and pigment biosynthesis. Reveiz CA, Benning C, Bohnert H, Daniell H, Hoober JK, Lichtenthaler HK, Portis AR, Tripathy BC editors. The chloroplast: Basics and applications. Advances in photosynthesis and respiration. Springer Science. 31: 95-118.

Lohr M, Schwender J, Polle JEW. 2012. Isoprenoid biosynthesis in eukaryotic phototrophs: A spotlight on algae. Plant Sci. 185-186: 9-22.

Maddison WP, Maddison DR. 2011. Mesquite: a modular system for evolutionary analysis. Version 2.75. Available from: http://mesquiteproject.org.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In proceedings of the gateway

computing environments workshop (GCE). 14 Nov. 2010. New Orleans, LA. Pp 1-8.

Moldowan JM, Fago FJ, Lee CY, Jacobson SR, Watt DS, Slougui NE, Jeganathan A, Young DC. 1990. Sedimentary 24-n-propylcholestanes, molecular fossils diagnostic of marine algae. Science. 247: 309-312.

Mooney BD, Nichols PD, de Salas MF, Hallegraeff GM. 2007. Lipid, fatty acid, and sterol composition of eight species of Kareniaceae (Dinophyta): chemotaxonomy and putative lipid phycotoxins. J. Phycol. 43: 101-111.

Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kanehisa M. 2007. KAAS: An automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35: W182-W185.

Moustafa, A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009. Genomi footprints of a cryptic plastid endosymbiosis in diatoms. Science. 324: 1724-1726.

Nes WD. 2011. Biosynthesis of cholesterol and other sterols. Chem. Rev. 11: 6423-6451.

Nishikawa Y, Quittnat F, Stedman TT, Voelker DR, Choi JY, Zahn M, Marc Pypaert, Joiner KA, Coppens I. 2005. Host cell lipids control cholesteryl ester synthesis and storage in intracellular *Toxoplasma*. Cell. Microbiol. 7: 849-867.

Oborník M, Modrý D, Lukeš M, Černotíkova-Stříbrná E, Cihlář J, Tesařová M, Kotabová E, Vancová M, Prášil O, Lukeš J. 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.

Patterson GW. 1971. The distribution of sterols in algae. Lipids. 6: 120-127..

Patterson GW. 1991. Sterols of algae. Patterson GW, Nes WD editors. Physiology and biochemistry of sterols. American Oil Chemists' Society, Illinois. 5: 118-157.

Rampen SW, Abbas BA, Schouten S, Damsté JSS. 2010. A comprehensive study of sterols in marine diatoms (Bacillariophyta): Implications for their use as tracers for diatom productivity. Limnol. Oceanogr. 55: 91-105.

Rohdich F, Kis K, Bacher A, Eisenreich W. 2001. The non-mevalonate pathway of isoprenoids: Genes, enzymes and intermediates. Curr. Opin. Chem. Biol. 5: 535-540.

Rohmer M, Seemann M, Horbach S, Bringer-Meyer S, Sahm H. 1996. Glyceraldehyde 3-phosphate and pyruvate as precursors of isoprenic units in an alternative non-mevalonate pathway for terpenoid biosynthesis. J. Am. Chem. Soc. 118: 2564-2566.

Sayers EW et al. 2009. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 37: D5-D15.

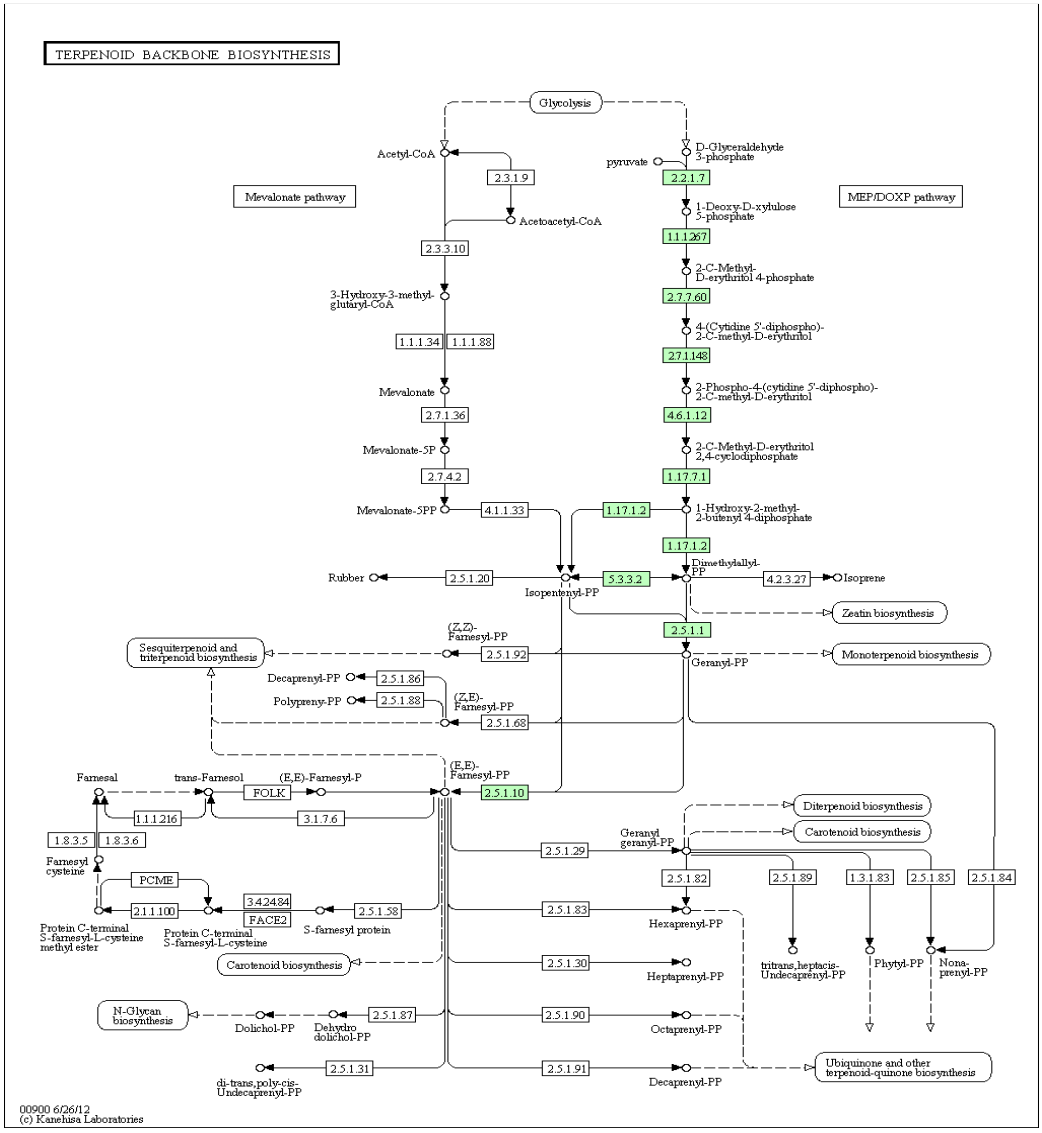
Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics., doi: 10.1093/bioinformatics/btu033.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725-2729.

Volkman JK. 1993. Sterols in microorganisms. Appl. Microbiol. Biotechnol. 60: 495-506.

Volkman JK, Barrett SM, Dunstan DA, Jeffrey SW. 1993. Geochemical significance of the occurrence of dinosterol and other 4-methyl sterols in a marine diatom. Org. Geochem. 20: 7-15.

APPENDICES



Appendix A. Kyoto encyclopedia of gene and genome (KEGG) map showing various genes for *V. brassicaformis* related to non mevalonate/methylerythritol pathway

CHAPTER IV

MONO- AND DIGALACTOSYLDIACYLGLYCEROL COMPOSITION OF DINOFLAGELLATES. BIOCHEMICAL AND GENOMIC COMPARISON OF GALACTOLIPID BIOSYNTHESIS BETWEEN *CHROMERA VELIA* (CHROMERIDA), A PHOTOSYNTHETIC ALVEOLATE WITH RED ALGAL PLASTID ANCESTRY, AND THE DINOFLAGELLATE, *LINGULODINIUM POLYEDRUM*

INTRODUCTION

Chromera velia is a recently discovered, photosynthetic alveolate that is the closest known relative (based on LSU rDNA, SSU rDNA and concatenated nuclear gene datasets) to non-photosynthetic, obligately parasitic apicomplexans, such as *Babesia*, *Plasmodium*, and *Toxoplasma* (Moore et al. 2008; Janouškovec et al. 2010). *Chromera velia* is also closely related to perkinsids, which are non-photosynthetic parasites of shellfish, and to photosynthetic and non-photosynthetic dinoflagellates (Moore et al. 2008). Both Moore *et al*. (2008) and Janouškovec *et al*. (2010), who published the complete chloroplast genome sequence of *C. velia*, found that the plastid of *C. velia* shares ancestry and traits with the plastid of peridinin-containing dinoflagellates. Based on the *psb*A gene, the plastid of *C. velia* is more closely related to peridinin-containing dinoflagellates than to other heterokonts (Moore et al. 2008). The

chloroplast of *C. velia* has also been found to be related to the apicoplast, a non-photosynthetic relic plastid of apicomplexan parasites which also shares a red algal ancestry (Janouškovec et al. 2010).

Mono- and digalactosyldiacylglycerol (MGDG and DGDG, respectively) are galactolipids that comprise the majority of the lipid components of chloroplast membranes in photosynthetic life forms and play an important structural role in photosynthesis (see multiple chapters in Wada & Murata, 2009). Recently, peridinin-containing dinoflagellates, the largest group of photosynthetic dinoflagellates in terms of the number of known genera, were found via positive-ion electrospray/mass spectrometry (ESI/MS) analysis to be divided into two clusters according to the forms of MGDG and DGDG present (Gray et al. 2009). Cluster 1 dinoflagellates possessed major forms of MGDG and DGDG that contained C_{18} fatty acids in the *sn*-1 and *sn*-2 positions (i.e. C_{18}/C_{18}), whereas Cluster 2 dinoflagellates possessed C_{20}/C_{18} (*sn*-1/*sn*-2) major forms of MGDG and DGDG. More specifically, Cluster 1 dinoflagellates generally possessed 18:5(n-3, octadecapentaenoic acid)/18:4(n-3, octadecatetraenoic acid) MGDG, 18:5/18:5 MGDG (double bond notation of n-3 is omitted in *sn*-1/*sn*-2 designations here and the following text for simplicity), 18:4/18:4 DGDG, and 18:5/18:4 DGDG. Cluster 2 dinoflagellates possessed 20:5(n-3, eicosapentaenoic acid)/18:4 MGDG, 20:5/18:5 MGDG, 20:5/18:4 DGDG, and 20:5/18:5 DGDG as the major galactolipids. In addition to these two types of

peridinin-containing dinoflagellates, a recent study on the forms of MGDG and DGDG present in *Karenia brevis*, *Kryptoperidinium foliaceum*, and *Lepidodinium chlorophorum* – three non-peridinin-containing dinoflagellates of aberrant plastid ancestry –showed that these algae reflect to different degrees the fatty acid composition of their plastid ancestors, and in some cases show distinct differences from peridinin-containing species (Leblond & Lasiter, 2009). A previous study (Leblond et al. 2010a) has also elucidated how species of *Pyrocystis*, which are Cluster 2 dinoflagellates, modulate their galactolipid-associated fatty acids in response to temperature. Thus, through the use of ESI/MS (and ESI/MS/MS) analyses, we now have a more complete picture of MGDG and DGDG composition in dinoflagellates than at any other time in the long history of lipid research.

To date, despite the evolutionary importance of *C. velia*'s plastid, there is a noticeable gap in knowledge of its lipid biochemistry, with only a single published study on the composition of *C. velia*'s forms of MGDG and DGDG by Botté *et al*. (2011) and a single published study on its sterol composition by Leblond *et al*. (2012). In their recent study, Botté *et al*. (2011) found major forms of MGDG and DGDG containing 20:5 at both the *sn*-1 and *sn*-2 positions, and minor forms containing 20:5 and 16:0 (regiochemistry undetermined) . In addition, they found putative MGDG and DGDG synthases (enzymes responsible for incorporating galactose into the lipids) that were similar to the synthases in other algae with red

algal plastid origin, and they commented that an abundance of 20:5 is consistent with a red algal plastid ancestry for *C. velia*.

Our objectives in this work were (1) to identify the molecular forms of MGDG and DGDG in *C. velia* grown under two temperatures, 20 and 30℃, in orde r to examine the effect of temperature on the fatty acid composition of these lipids; and (2) to utilize gene annotations and conceptual models to make comparisons with galactolipid biosynthesis in peridinin-containing dinoflagellates, specifically the C_{20}/C_{18} member, *Lingulodinium polyedrum*.

MATERIALS AND METHODS

Cultures, growth conditions, and galactolipid processing

Chromera velia CCMP 2878 was acquired from the National Center for Marine Algae and Microbiota (formerly the Provasoli–Guillard Center for the Culture of Marine Phytoplankton), East Boothbay, ME, USA, and was grown phototrophically in triplicate in 1 litre of L1 medium (Guillard & Hargraves, 1993) at 20℃ and 30℃ under a 14 : 10 h light–dark cycle at an irradiance of approximately 50 µmol photons·m^{-2}·s^{-1}. Cells were harvested during the exponential phase of growth by filtration onto a precombusted 934-AH Whatman glass fibre filter. Lipids were extracted from the filtered biomass and MGDG and DGDG were separated from other lipid classes according to the techniques described by Leblond & Chapman (2000). Briefly, the total lipid extracts were

separated into five component lipid fractions on columns of activated Unisil silica (1.0 g, 100-200 mesh, activated at 120℃, Clarkson Chromatography, South Williamsport, PA, USA). The following solvent regime was used to separate lipids according to polarity, with fraction 5 eluting the most polar lipids (Leblond & Chapman, 2000): (1) 12 ml methylene chloride (sterol esters), (2) 15 ml 5% acetone in methylene chloride with 0.05% acetic acid (free sterols, tri- and diacylglycerols, and free fatty acids), (3) 10 ml 20% acetone in methylene chloride (monoacylglycerols), (4) 45 ml acetone (MGDG, DGDG, and sulfoquinovosyldiacylglycerol [SQDG]), and (5) 15 ml methanol with 0.1% acetic acid (polar lipids, including non-phosphorus-containing lipids).

Mass spectrometry of galactolipids

Fractionated MGDG and DGDG were dissolved in a solution of methanol, chloroform, and 50 mM sodium acetate according to Welti *et al*. (2002) to produce positively charged sodium adducts $[M+Na]^+$. These adducts were subjected to a positive-ion electrospray ionization/mass spectrometry (ESI/MS) scan from *m/z* 100–2000 via direct injection of a 5 µl sample volume into a methylene chloride carrier solvent at 0.5 ml min^{-1} into a Finnigan DecaXP ion trap mass spectrometer (Waltham, Massachusetts, USA). The relative abundance of each lipid was determined from the peak heights of individual sodium adducts of MGDG and DGDG.

Subsequent ESI/MS/MS was performed using a collision energy between 37.5 and 48%, and major cleaved fatty acids were identified by the differences between the masses of the original ions and their fragments. The positions of the acyl chains (*sn*-1 or *sn*-2) were determined by their relative percent compositions according to a variation of the procedure (i.e. similar conditions on a different instrument) established by Guella *et al*. (2003).

In silico and phylogenetic analysis of galactolipid gene sequences of *C. velia* and *L. polyedrum*

Genes were annotated from both *C. velia* and *L. polyedrum* by using genes encoding galactolipid synthesis enzymes from the red alga *Cyanidioschyzon merolae*, the plant *Arabidopsis thaliana*, the green alga *Chlamydomonas reinhardtii*, and the apicomplexan *Toxoplasma gondii*. Briefly, genes related to galactolipid synthesis from these organisms were used in a tBLASTn search of the expressed sequence tag (EST) and transcriptome assembly databases of both *C. velia* and *L. polyedrum* (Woehle et al. 2011; Roy & Morse, 2012). The putative sequences found from *C. velia* and *L. polyedrum* were then used in a BLASTp search against the NCBI database to identify the closest homologues and confirm that the *in silico* annotations were correct (Altschul et al. 1997). The amino acid sequences of various algal and plant MGDG and DGDG synthases were aligned in MAFFT using local pair analysis (L-INS-I) (Katoh et al. 2005).

Phylogenetic trees were built in PhyML 3.0 using the LG model and SPR parameters with 100 bootstrap replications (Guindon et al. 2010).

RESULTS

Galactolipid composition of *C. velia* grown at different temperatures

Examination of the glycolipid fractions of triplicate cultures of 20℃-grown *C. velia* via positive-ion ESI/MS revealed the presence of two forms of MGDG, namely 20:5/18:1 MGDG (*m/z* 825 as sodium adduct) and 20:5/20:5 MGDG (*m/z* 845 as sodium adduct), and one form of DGDG, 20:5/20:5 DGDG (*m/z* 1007 as sodium adduct), in each culture (Table 1). Of these three forms, 20:5/20:5 MGDG and 20:5/20:5 DGDG were much more abundant than 20:5/18:1 MGDG, with relative percentage abundances of approximately 35 and 60%, respectively (Table 1). At a growth temperature of 30℃, 20:5/20:5 MGDG and DGDG were still the most abundant galactolipids; however, in addition 20:5/20:4 forms of these lipids, which were not observed at 20℃, were detected, at relative abundances of approximately 5 and 8%, respectively. The presence of these less unsaturated forms of C_{20}/C_{20} MGDG and DGDG was accompanied by a commensurate decrease in the abundances of 20:5/20:5 MGDG and DGDG, along with a slight decrease in 20:5/18:1 MGDG (Table 1). Also present in *C. velia* grown at 30℃, but not at 20℃, were 20:50/16:0 and 20:4/16:0 DGDG in

low abundance (Table 1). Our results are consistent with those of Botté *et al.* (2011), who found 20:5/20:5 MGDG and DGDG as the dominant galactolipid forms.

Table 1. Relative abundance (in % of total fragment height using listed masses) of *Chromera velia* galactolipids at 20 and 30℃, as determined via positive-ion ESI/MS. Data are mean ± S.D. (*n* = 3).

Lipid	***m/z*[1]**	**20℃ .**	**30℃**
20:5/18:1 MGDG	825	4.2 ± 1.6	1.6 ± 1.4
20:5/20:5 MGDG	845	35.3 ± 5.2	30.2 ± 7.8
20:5/20:4 MGDG	847	–	4.6 ± 1.2
20:5/16:0 DGDG	961	–	3.1 ± 0.6
20:4/16:0 DGDG	963	–	1.8 ± 1.8
20:5/20:5 DGDG	1007	60.4 ± 4.3	51.0 ± 7.8
20:5/20:4 DGDG	1009	–	7.7 ± 1.2
20:5/20:4 DGDG	1009	–	7.7 ± 1.2

[1]Mass rounded down to the nearest integer.

Annotation of genes involved in galactolipid synthesis in *C. velia*

We attempted to reconstruct the galactolipid synthesis pathway in *C. velia in silico* from transcriptomic sequence data (Woehle et al. 2011) and several genes potentially involved in galactolipid synthesis were identified (Table 2). In building this pathway, genes previously characterized and/or annotated from, *Arabidopsis*

thaliana, *Chlamydomonas reinhardtii*, *Cyanidioschyzon merolae* and *Toxoplasma gondii* were used to identify potential galactolipid gene candidates in *C. velia*. Genes involved in the *de novo* synthesis of medium chain fatty acids (C_{14}-C_{18}) from malonyl-CoA were detected in the genome database (Table 2). These included malonyl-CoA-acyl carrier protein (ACP) transacylase (FabD) and the components of the fatty acid synthase II (FASII) complex, including 3-ketoacyl-ACP synthase (FabB), 3-ketoacyl-ACP synthase III (FabH), 3-ketoacyl-ACP reductase (FabG), 3–hydroxyacyl-ACP dehydratase (FabZ), and enoyl-ACP reductase (FabI) (Fig. 1). Each of the putative FASII genes showed homology to diatoms and apicomplexans when a search of the NCBI database was conducted with these putative *C. velia* galactolipid synthesis genes (Table 3). Using the prediction program ChloroP, we determined that each of the components of the FASII complex except 3-hydroxyacyl-ACP-dehydratase contained a chloroplast transit peptide (CTP) sequence, which targets proteins to the plastid (Emanuelsson et al. 1999). However, malonyl-CoA-acyl carrier protein transacylase and 3-hydroxyacyl-ACP-dehydratase did not contain a CTP sequence. One reason for the absence of this sequence may be due to the sequence information being incomplete. Consistent with this idea, malonyl-CoA-acyl carrier protein is targeted to the chloroplast, based on the PSORT program, and its closest homologue contains a CTP. 3-Hydroxylacyl-ACP-dehydratase is also targeted to either the mitochondria or chloroplast, based on the PSORT

program; however, the sequence does not contain a plastidial or mitochondrial signal, indicating that more sequence information is needed to predict the subcellular location of this protein.

Table 2. Annotated genes involved in galactolipid synthesis from *Chromera velia* and indication of a chloroplast transit peptide (CTP). Reaction steps refer to Figs 1 and 2.

Reaction step	**Enzyme**	**Gene**	**GenBank accession**	**CTP**
2	Malonyl-CoA: ACP transacylase	*FabD*	JO786709.1	No
3	3-ketoacyl-ACP synthase III	*FabH*	JO786666.1	Yes
	FASII complex			
4A	3-ketoacyl-ACP reductase	*FabG*	JO786674.1	Yes
4B	3-Hydroxyacyl-ACP dehydratase	*FabZ*	JO800691.1	No
4C	Enoyl-ACP-reductase	*FabI*	JO786847.1	Yes
4D	3-ketoacyl-ACP synthase	*FabB*	JO813297.1	Yes
	Extraplastidial enzymes			
5a	Elongase	*Elo*	JO804831.1	
5b	Elongase	*Elo*	JO801545.1	
6a	Δ9 Desaturase	*Des9*	JO792025.1	

The gene encoding stearoyl-ACP desaturase, which converts hexadecanoic acid (16:0) to hexadecenoic acid (16:1) and octadecanoic acid (18:0) to

octadecenoic acid (18:1) in the plastid was not detected in *C. velia*. Galactolipids in algae and plants, which are comprised of C_{18} and C_{20} carbon fatty acids in the *sn*-2 position are known to have modifications which are made in the endoplasmic reticulum (ER) (Browse et al. 1986). These modifications are made by extraplastidial desaturases and elongases. Two elongases were detected in the *C. velia* transcriptome and showed homology to other algae, including diatoms (Table 3). These elongases may be involved in fatty acid elongation resulting in the C_{20} fatty acids found in galactolipids. However, there was only one desaturase detected in *C. velia*, which had an unexpected homology to bacterial Δ9 phospholipid-CoA desaturases (Table 3). Other desaturases possibly contributing to the synthesis of 20:5 were not identified, which may be due to the transcriptome database of *C. velia* lacking full coverage of all expressed genes.

Table 3. The four most similar homologues of proteins encoded by putative galactolipid synthesis genes in *Chromera velia.*

Enzyme	**E value**	**Organism type**	**Species**	**Accession Number**
Malonyl-ACP:CoA transacylase	2e-132	Stramenopiles	*Thalassiosira pseudonana*	XP002290601.1
FabD	1e-128	Stramenopiles	*Phaeodactylum tricornutum*	XP002181767.1
	7e-128	Stramenopiles	*Ectocarpus siliculosus*	CBN74809.1
	3e-114	Green plants	*Micromonas* sp.	XP002507066.1
3-Ketoacyl-ACP synthase	2e-44	Apicomplexa	*Neospora caninum*	CBZ49555.1
FabB	5e-44	Apicomplexa	*Toxoplasma gondii*	XP002370161.1
	2e-40	Stramenopiles	*Ectocarpus siliculosus*	CBN77813.1
	7e-39	Bacteria	*Thermincola potens*	YP003640817.1

Table 3. Continued: The four most similar homologues of proteins encoded by putative galactolipid synthesis genes in *Chromera velia.*

Enzyme	**E value**	**Organism type**	**Species**	**Accession Number**
3-Ketoacyl-ACP synthase III	3e-107	Stramenopiles	*Ectocarpus siliculosus*	CBN77814.1
FabH	2e-97	Stramenopiles	*Thalassiosira pseudonana*	XP002295320.1
	2e-89	Cyanobacteria	*Microcystis aeruginosa*	CCI05386.1
	5e-88	Cyanobacteria	*Nostoc punctiforme*	YP001863843.1
3- Ketoacyl-reductase	1e-87	Bacteria	*Bacillus methanolicus*	EIJ79549.1
FabG	4e-86	Apicomplexa	*Plasmodium falciparum*	XP001352100.1
	7e-85	Bacteria	*Caldalkalibacillus thermarum*	ZP08533272.1
	1e-84	Bacteria	*Blautia hansenii*	ZP05854013.1

Table 3. Continued: The four most similar homologues of proteins encoded by putative galactolipid synthesis genes in *Chromera velia.*

Enzyme	**E value**	**Organism type**	**Species**	**Accession Number**
3-Hydroxyacyl-ACP dehydratase	8e-76	Apicomplexa	*Toxoplasma gondii*	EEE19304.1
FabZ	3e-62	Stramenopiles	*Phaeodactylum tricornutum*	XP002176151.1
	3e-58	Apicomplexa	*Plasmodium cynomolgi*	GAB67832.1
	1e-52	Green plants	*Helianthus annuus*	ADL60215.1
Enoyl-ACP-reductase	9e-165	Stramenopiles	*Phaeodactylum tricornutum*	XP002177931.1
FabI	5e-162	Stramenopiles	*Thalassiosira pseudonana*	XP002288236.1
	9e-161	Stramenopiles	*Ectocarpus siliculosus*	CBN77155.1
	6e-160	Apicomplexa	*Toxoplasma gondii*	XP002367432.1

Table 3. Continued: The four most similar homologues of proteins encoded by putative galactolipid synthesis genes in *Chromera velia.*

Enzyme	**E value**	**Organism type**	**Species**	**Accession Number**
Δ9 Desaturase	7e-22	Bacteria	*Gemmata obsuriglobus*	ZP02732702.1
	2e-20	Bacteria	*Hydrogenophaga sp.*	ZP10155081.1
	2e-20	Bacteria	*Bradyrhizobium japonicum*	YP005609939.1
	2e-19	Bacteria	*Lentisphaera araneosa*	ZP01874416.1
Elongase	1e-94	Stramenopiles	*Ectocarpus siliculosus*	CBJ26568.1
	4e-93	Stramenopiles	*Phaeodactylum tricornutum*	AAW70157.1
	1e-88	Stramenopiles	*Thalassiosira pseudonana*	XP0228848.1
	3e-87	Stramenopiles	*Albugo laibaechii*	CCA25115.1

Table 3. Continued: The four most similar homologues of proteins encoded by putative galactolipid synthesis genes in *Chromera velia.*

Enzyme	**E value**	**Organism type**	**Species**	**Accession Number**
Elongase	7e-15	Stramenopiles	*Isochrysis galbana*	ADD51571.1
	9e-14	Haptophytes	*Rebecca salina*	XP002367432.1
	6e-9	Stramenopiles	*Thalassiosira pseudonana*	XP002293395.1
	1e-9	Excavates	*Leishmania braziliensis*	XP001563363.1

Table 4. Annotated genes involved in galactolipid synthesis in *Lingulodinium polyedrum.* Reaction steps refer to Figs 1 and 2.

Reaction Step/Subcategory	**Enzyme**	**Gene**	**Accession Number**
1	ACCase	*ACCase*	JO135861.1
2	Malonyl-CoA:ACP transacylase	*FabD*	JO739746.1
3	3-ketoacyl-ACP synthase III	*FabH*	JO722316.1

Table 4. Continued: Annotated genes involved in galactolipid synthesis in *Lingulodinium polyedrum.* Reaction steps refer to Figs 1 and 2.

Reaction Step/Subcategory	**Enzyme**	**Gene**	**Accession Number**
	FASII Complex		
4A	3-ketoacyl ACP reductase	*FabG*	JO750886.1
4B	3-Hydroxyacyl ACP dehydratase	*FabZ*	JO719639.1
4C	Enoyl ACP reductase	*FabI*	JO762004.1
4D	3-ketoacyl ACP synthase	*FabB*	JO722316.1
	Extraplastidial Enzymes		
5a	Elongase	*Elo*	JO723605.1
5b	Elongase	*Elo*	JO698611.1
6a	Δ9 Desaturase	*Des9*	JO741757.1
6b	Δ 5 Desaturase	*Des5*	JO732212.1
6c	Δ 6 Desaturase	*Des6*	JO706276.1
6d	Δ 12 Desaturase	*Des12*	JO735750.1

Comparison with dinoflagellate genes

In order to compare galactolipid synthesis in *C. velia* with that in peridinin-containing dinoflagellates, putative genes involved in galactolipid biosynthesis were identified in the C_{20}/C_{18} dinoflagellate *Lingulodinium polyedrum*. Table 4 lists the genes detected, which may be involved in *de novo* biosynthesis of fatty acids in the plastid from acetyl-CoA. Genes encoding the acetyl-CoA carboxylase (ACCase) complex and each of the components of the FASII complex were found in the transcriptome database of *L. polyedrum*. As in *C. velia*, the stearoyl-ACP desaturase was not identified. Several extraplastidial enzymes involved in modifying fatty acids were found, including two fatty acid elongases and four fatty acid desaturases. The desaturases were found to have homology with Δ5, Δ 6, Δ 9, and Δ12 desaturases, which are known to contribute to the biosynthesis of 20:5. These desaturases and elongases correspond with the conversion of 16:0 or 18:0 to 20:5 in the ER of *L. polyedrum* which could be later incorporated into galactolipids.

Table 5. Genes encoding MGDG and DGDG synthases in *Lingulodinium polyedrum.* Reaction steps refer to Figs 1 and 2.

Reaction Step	Enzyme	Accession Number
7	MGDG Synthase	JO714796.1
8a	DGDG Synthase 1	JO706156.1
8b	DGDG Synthase 2	JO709979.1

Phylogenetic analysis of putative MGDG and DGDG synthases

Potential MGDG and DGDG synthases were identified in the peridinin dinoflagellate, *L. polyedrum* by searching the transcriptome database with the MGDG and DGDG synthases previously found in *C. velia* by Botté *et al*. (2011) (Table 5). In an attempt to obtain further information on the origins of these newly found MGDG and DGDG synthases from *L. polyedrum*, a phylogenetic tree was constructed with several plant and algal MGDG synthases and showed similar groupings of plastids originating from plants, green algae and red algae, as observed by Botté *et al*. (2011) (Fig. 2 panel A). *Lingulodinium polyedrum* showed homology to the brown alga, *E. siliculosus*, which has a plastid of red algal origin (LeCorquille et al. 2009).

A phylogenetic tree was also constructed containing various DGDG synthases, which included those from a number of plants, green algae, red algae, and algae containing red alga–derived plastids (Fig. 2, panel B). As observed previously, the DGDG synthase from *C. velia* showed homology to a group of red algal ancestors including diatoms and brown algae. Two DGDG synthases were identified from *L. polyedrum*, both of which had homology to the haptophyte, *Emiliania huxleyi*. These results are in agreement with the previous hypothesis that the alveolate plastid found in dinoflagellates is of red algal origin.

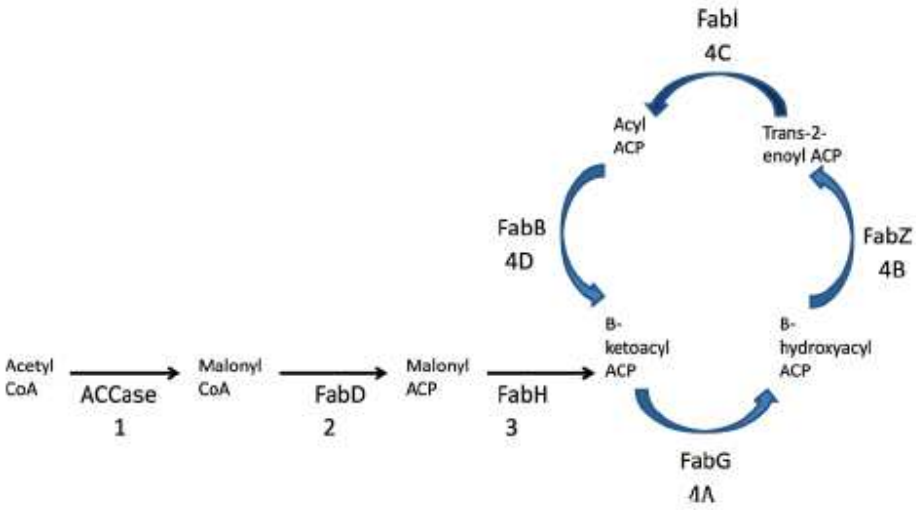


Figure 1. Pathway of *de novo* fatty acid biosynthesis in plastid utilizing FASII complex. Numbers next to arrows correlated to the reaction steps on Tables 2, 4 and 5.

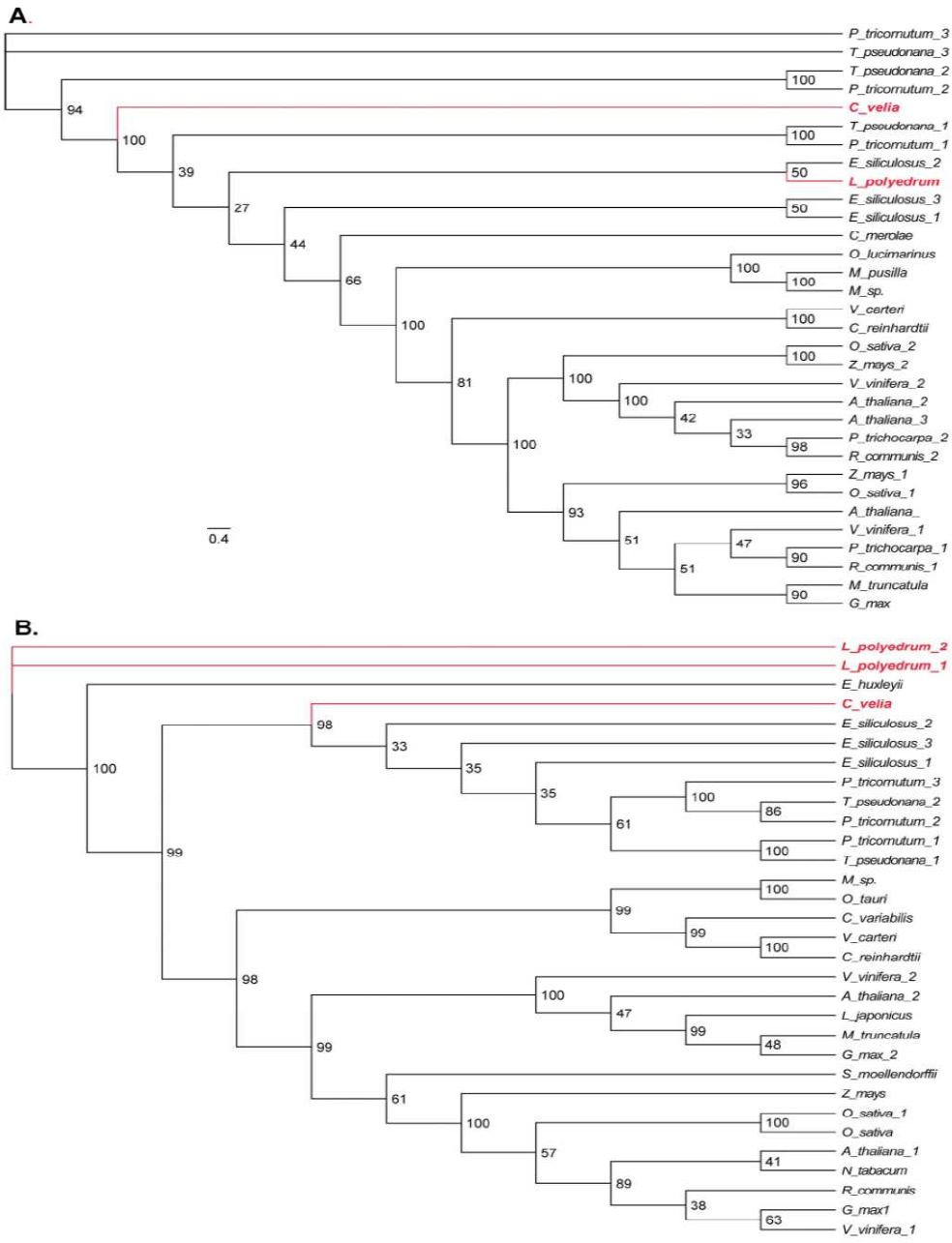


Figure 2. Phylogenetic analysis of MGDG synthases (A) and DGDG synthases (B): maximum likelihood tree with bootstrap confidence levels.

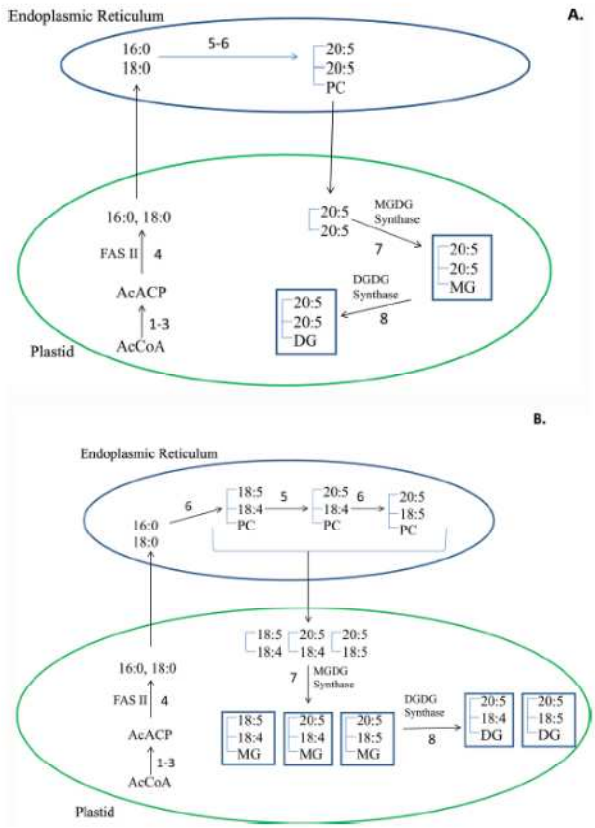


Figure 3. Comparison of pathways in galactolipid synthesis in *Chromera velia* (A) and *Lingulodinium polyedrum* (B). Numbers next to arrows correlate to the reactions steps in Tables 2, 4 and 5.

DISCUSSION

Galactolipid composition of *C. velia*, comparison with red algae, and the effect of growth temperature

Previous research by Janouškovec *et al*. (2010) revealed that *C. velia* and peridinin-containing dinoflagellates share a common red algal plastid ancestor,

and there are both distinct differences and similarities between *C. velia* and dinoflagellates with regard to cellular ultrastructure, morphology, and pigment composition (Moore et al. 2008; Oborník et al. 2011). In a comparison between *C. velia* and Cluster 2 dinoflagellates, including *Lingulodinium polyedrum*, Gray *et al*. (2009) showed that these dinoflagellates possess 20:5 at the *sn*-1 position in the most abundant forms of MGDG and DGDG. Conversely, the presence of 20:5 at the *sn*-2 position to form 20:5/20:5 MGDG and 20:5/20:5 DGDG was found to be very rare in peridinin-containing dinoflagellates, with only a few organisms possessing either lipid and only at relative abundances of less than 10% of the total galactolipids (Gray et al. 2009). Rather, in both of the clusters of peridinin-containing dinoflagellates identified by Gray *et al*. (2009), it was observed that either 18:5 or 18:4 was present at the *sn*-2 position in the most abundant forms of MGDG and DGDG. In contrast, *C. velia* possessed 20:5, but neither 18:5 nor 18:4, at the *sn*-2 position.

There are very few studies which offer insight into the fatty acid regiochemistry of the forms of MGDG and DGDG possessed by red algae, especially with regard to the effect of temperature. For example, Ackman & McLahlan (1977) have shown members of the Rhodophyceae to be enriched in 20:5, but no information is available as to the regiochemical distribution of this and other fatty acids within MGDG and DGDG. However, a comparison can be made between *C. velia*, peridinin-containing dinoflagellates, and the model red alga,

Porphyridium cruentum, based on the results of Adlerstein *et al*. (1997), who examined the molecular composition of *P. cruentum*'s MGDG and DGDG at three growth temperatures. Adlerstein *et al*. found that at the lowest growth temperature tested, 20℃, the most abundant form of MGDG was 20:5/20:5 MGDG, with 20:5/16:0 being the next most abundant form. As the growth temperature was increased to 30℃, 20:5/16:0 MGDG became more abundant than 20:5/20:5 MGDG and 20:4/16:0 went from being a minor form to one of the most abundant. At both 20 and 30℃, 20:5/20:5 DGDG was a minor form, whereas 20:5/16:0 and 20:4/16:0 were quite abundant. It should be noted that polyunsaturated C_{20} fatty acids are not typically abundant in the red algae, the Cyanidiales – which contains *P. cruentum* – being among the exceptions (Khozin et al. 1997; Sato et al. 2007).

Leblond *et al*. (2010a) observed, in three different species of the C_{20}/C_{18} peridinin-containing dinoflagellate *Pyrocystis*, that fatty acid modulation according to growth temperature occurred consistently within the *sn*-2 fatty acid of DGDG. Within the C_{20}/C_{20} forms of MGDG and DGDG produced by *C. velia*, the same modulation of *sn*-2 fatty acid unsaturation level occurred, with a $C_{20:4}$ fatty acid being produced at 30℃ but not at 20℃ (Table 1). Thus, even though a C_{18} fatty acid was not a major component of MGDG and DGDG in *C. velia*, this organism behaved similarly to *Pyrocystis* in that the *sn*-2 fatty acid became less unsaturated at the higher growth temperature. This type of *sn*-2 fatty acid

modulation was not observed in the study on *Porphyridium cruentum* by Adlerstein *et al*. (1997), but it should be noted that they did not use positive-ion ESI/MS/MS analysis to provide regiochemical association of particular fatty acids with particular forms of MGDG and DGDG. Re-examination of the effect of temperature on the galactolipids of *P. cruentum* and other red algae is warranted, using a modern mass spectrometry technique such as positive-ion ESI/MS/MS.

Although we cannot discount the participation of lipid biosynthesis genes derived from another group(s) of algae and acquired by *C. velia* at some point in its evolutionary history [see example study by Woehle *et al*. (2011)], our hypothesis is that *C. velia* has inherited its ability to synthesize 20:5/20:5 MGDG and DGDG from its red algal plastid ancestor. We make a similar hypothesis for the presence of 20:5 at the *sn*-1 position of the major forms of MGDG and DGDG in the Cluster 2 dinoflagellates identified by Gray *et al*. (2009): that this trait has been obtained from its red algal plastid ancestor. However, with the exception of radiolabelling studies by Khozin *et al*. (1997) and Sato *et al*. (2007), little is known about the genes that encode the enzymes involved in the synthesis of galactolipids in the potential red algal ancestors of the plastids of *C. velia* and peridinin-containing dinoflagellates. In this study we therefore attempted to elucidate galactolipid synthesis in both *C. velia* and *L. polyedrum* through an *in silico* analysis of their transcriptomes (Woehle et al. 2011; Roy & Morse, 2012).

Annotation of genes involved in galactolipid synthesis in *C. velia* and biosynthetic models

The genes annotated in *C. velia* illustrate that fatty acids are synthesized through the plastidial FASII complex to convert malonyl-CoA to medium-chain fatty acids; this is similar to what is found in many other algae and apicomplexans (Riekhof et al. 2005; Mazumdar & Striepen, 2007; Sato et al. 2007; Fig. 1). The absence of the plastidial stearoyl-ACP desaturase in *C. velia* is consistent with what has been found in the apicoplasts of apicomplexans and red algal plastids (Sato et al. 2007; Ramakrishnan et al. 2011). These medium-chain fatty acids are then transported to the endoplasmic reticulum (ER), where they are modified by elongases and desaturases, resulting in the 20:5 fatty acid. Radiolabelling studies have focused on the synthesis of 20:5 for incorporation into galactolipids (Khozin et al. 1997; Khozin-Goldberg et al. 2002). From genes annotated in *C. velia* and what is known about algal galactolipid biosynthesis, it appears that 16:0 and/or 18:0 are produced in the plastid through *de novo* synthesis by the FASII complex and subsequently transported to the ER, incorporated into phospholipids, and modified to 20:5 with extraplastidial elongases and desaturases (Fig. 3). Diacylglycerol moieties of 20:5/20:5 would then be transported to the plastid where they would be incorporated into MGDG and DGDG by MGDG and DGDG synthases, respectively. Two elongases were identified in *C. velia*, which may be involved in the elongation of a C_{16} or C_{18} to a

C_{20} fatty acid. However, only one desaturase was identified in our search, this having homology to a Δ9 CoA desaturase. Since it is unlikely that the desaturase identified can desaturate a fatty acid at five positions; the lack of desaturases found may be due to the lack of complete coverage of the *C. velia* transcriptome.

The genes annotated for galactolipid synthesis in our study, together with the MGDG and DGDG synthases identified by Botté *et al*. (2011), provide further evidence that *C. velia*'s plastid was inherited through a red algal ancestry. Future studies focusing on radiolabelled pulse-chase studies, coupled with functional characterization of the identified ER- desaturase and elongases along with further identification of additional desaturases, will be needed to ascertain whether this model is correct and decipher why *C. velia* incorporates 20:5 into both *sn*-1 and *sn*-2 positions of MGDG, instead of producing C_{20}/C_{18} forms like Cluster 2 dinoflagellates. Future work should also examine, in a manner similar to a recent study on n-3 polyunsaturated fatty acid biosynthesis in *E. huxleyi* by Sayanova *et al*. (2011), the genes that direct biosynthesis and incorporation of 20:5 into MGDG and DGDG in both *C. velia* and Cluster 2 dinoflagellates. Such genetic proof will be needed to determine whether red algae-derived genes are indeed responsible for synthesizing and incorporating 20:5(n-3) into MGDG and DGDG, and should elucidate whether any of the responsible enzymes are targeted to the plastid. It should be noted that 20:5/20:5 MGDG and DGDG have been observed in chlorarachniophytes and glaucocystophytes (Leblond & Roche,

2009; Leblond et al. 2010b). Although these algae are seemingly unrelated to *C. velia* in terms of plastid ancestry (because chlorarachniophytes have a green algal secondary endosymbiont and glaucocystophytes have a cyanobacterial primary endosymbiont: references are listed by Leblond & Roche, 2009, and Leblond et al. 2010b), they may serve as useful systems for future comparison. The 20:5/16:0 fatty acid combination observed in *C. velia*, however, could also fit our model illustrated in Fig._3. In agreement with previous studies, 16:0 would be produced through the prokaryotic pathway *de novo* in the plastid. The 20:5 fatty acid would be synthesized through the eukaryotic pathway, which involves *de novo* synthesis in the plastid and further modifications (i.e. desaturations and elongations) in the ER. Previous studies have shown that 20:5 produced through the eukaryotic pathway and 16:0 synthesized specifically in the plastid through the prokaryotic pathway can be added to the same MGDG molecule (Khozin-Goldberg et al. 2002). Thus, in the case of *C. velia*, the presence of 16:0 at the *sn*-2 position and 20:5 in the *sn*-1 position of galactolipids would indicate a mixture of the prokaryotic pathway and eukaryotic pathways, respectively. It should also be noted that Adlerstein *et al*. (1997) observed 18:2/16:0 MGDG and DGDG in *P. cruentum*; therefore, we cannot eliminate a prokaryotic pathway root to MGDG and DGDG in *C. velia*, especially when minor amounts of 20:5/16:0 MGDG and DGDG were observed in this study in *C. velia* grown at 30℃ and by Botte *et al.* (2011).

For dinoflagellates containing C_{20} and C_{18} fatty acids in their galactolipids such as *L*. *polyedrum*, we have developed a potential model for galactolipid synthesis based on *L. polyedrum*'s galactolipid composition (Gray et al. 2009) and gene annotations for galactolipid synthesis related genes (Roy & Morse, 2012). The identification of genes involved in the FASII synthase complex indicates that fatty acids are produced in the plastid through *de novo* synthesis, similar to what has been found in *C. velia* and other red algal plastids (Sato et al. 2007) (Fig. 1). Furthermore, as in the apicomplexans *T. gondii* and *P. falciparum*, the red alga *C. merolae*, and *C. velia*, no stearoyl-ACP Δ9 desaturase has been identified in the plastid (Sato et al. 2007; Ramakrishnan et al. 2012). Thus we hypothesize that 16:0 is transported out of the plastid and incorporated into phospholipid(s), such as phosphatidylcholine, and subsequently elongated and desaturated within the ER. A number of elongases and desaturases identified in *L. polyedrum* and those annotated in this study would likely be involved in medium-chain fatty acids to produce 20:5 (Hashimoto et al. 2008). A homologue of a Δ15 desaturase used in the conversion to 16:0 or 18:0 to 20:5 in *L polyedrum* was not found. . However, it should be noted that the protist *Acanthamoeba castellanii* has a desaturase having dual functions as both a Δ12 and Δ15 desaturase, and this could be the case too in *L. polyedrum* (Sayanova et al. 2006). It is reasonable to suggest that 16:0 is elongated and desaturated to 18:3 and/or 18:4, which are abundant in the galactolipids in Cluster 2 dinoflagellates such as *L. polyedrum*.

We hypothesize that 18:3 and/or 18:4 could be further elongated and desaturated to form 20:5 (Khozin et al. 1997). Diacylgycerol moieties of 20:5/18:5, 20:5/18:4, and 18:5/18:4 would then be transferred back to the plastid and there incorporated into galactolipids by the MGDG and DGDG synthases Evidence for this model also derives from previous observations by Leblond & Chapman (2000), who observed 18:4 and 18:5 in the phospholipid fraction of a number of peridinin-containing dinoflagellates, albeit in low amounts. It should be noted that Leblond & Chapman (2000) did not find 20:4 in their survey of phospholipid-associated fatty acids in peridinin-containing dinoflagellates; therefore, it is likely that 20:5 results from an elongation of 18:5, rather than elongation of 18:4 to 20:4 followed by desaturation to 20:5 (Hashimoto et al. 2008). Further radiolabelling studies using [1-^{14}C] linoleic acid will better elucidate how fatty acids are modified from medium chain fatty acids to long chain polyunsaturated fatty acids in the ER in a dinoflagellate such as *L. polyedrum* (Khozin-Goldberg et al. 2002).

Phylogenetic analysis of putative MGDG and DGDG synthases

The phylogenetic analysis of MGDG and DGDG synthases identified from the transcriptome database of *L. polyedrum* gives insights into their evolutionary history. The MGDG synthase showed strong homology to that of the brown alga, *Ectocarpus siliculosus* (LeCorquille et al. 2009), whose plastid is hypothesized to

have red algal ancestry. The DGDG synthase likewise showed strong homology to the red algal lineage, including *Emiliania huxleyi* (Sánchez Puerta et al. 2005). Hence galactolipid synthesis in *L. polyedrum*, and possibly other peridinin-containing dinoflagellates, is likely derived from a red algal lineage. However, in order to further improve our knowledge of the evolutionary history of galactolipid synthesis in dinoflagellates, more complete transcriptomes of additional dinoflagellates and red algae need to be obtained and analysed.

REFERENCES

Ackman RG, McLachlan J. 1977. Fatty acids in some Nova Scotian marine seaweeds: A survey for octadecapentaenoic and other biochemically novel fatty acids. Proc. Nova Scotian Inst. Sci. 28: 47-64.

Adlerstein D, Bigogno C, Khozin I, Cohen Z. 1997. The effect of growth temperature and culture density on the molecular species composition of the galactolipids in the red microalga *Porphyridium cruentum* (Rhodophyta). J. Phycol. 33: 975-979.

Altschul SF, Madden TL, Schafer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Botté CY, Yamaryo-Botté Y, Janouškovec J, Rupasinghe T, Keeling PJ, Crellin P, Coppel RL, Maréchal E, McConville MJ, McFadden GI. 2011. Identification of plant-like galactolipids in *Chromera velia*, a photosynthetic relative of malaria parasites. J. Biol. Chem. 286: 29893–29903.

Gray CG, Lasiter AD, Li C, Leblond JD. 2009. Mono- and digalactosyldiacylglycerol composition of dinoflagellates. I. Peridinin-containing taxa. Eu. J. Phycol. 44: 191–197.

Guella G, Frassanito R, Mancini I. 2003. A new solution for an old problem: The regiochemical distribution of the acyl chains in galactolipids can be

established by electrospray ionization tandem mass spectrometry. Rapid Commun. Mass Sp.17: 1982–1984.

Guillard RR, Hargraves PE. 1993. *Stichochrysis immobilis* is a diatom, not a chrysophyte. Phycologia. 32: 234–236.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. 59: 307–321.

Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ. 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc. Natl. Acad. Sci. USA. 107: 10949–10954.

Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment describes ancestral versions of the G-INS-i, L-INS-i and E-INS-i strategies. Nucleic Acids Res. 33: 511–518.

Khozin I, Adlerstein D, Bigongo C, Heimer YM, Cohen Z. 1997. Elucidation of the biosynthesis of eicosapentaenoic acid in the microalga *Porphyridium cruentum*. II. Studies with radiolabeled precursors. Plant Physiol. 114: 223–230.

Khozin-Goldberg I, Didi-Cohen S, Shayakhmetova I, Cohen Z. 2002. Biosynthesis of eicosapentaenoic acid (EPA) in the freshwater

eustigmatophyte *Monodus subterraneus* (Eustigmatophyceae). J. Phycol. 38: 745–756.

Leblond JD, Chapman PJ. 2000. Lipid class distribution of highly unsaturated long-chain fatty acids in marine dinoflagellates. J. Phycol. 36: 1103–1108.

Leblond JD, Lasiter AD. 2009a. Mono- and digalactosylacylglycerol composition of dinoflagellates. II. *Lepidodinium chlorophorum*, *Karenia brevis*, and *Kryptoperidinium foliaceum*, three dinoflagellates with aberrant plastids. Eu. J. Phycol. 44: 199–205.

Leblond JD, Roche SA. 2009b. Mono- and digalactosyldiacylglycerol composition of chlorarachniophytes (Chlorarachniophyceae): Production of a novel lauric acid (12:0)-containing form of monogalactosyldiacylglycerol (MGDG). Phycologia. 48: 101–104.

Leblond JD, Dahmen JL, Evens TJ. 2010a. Mono- and digalactosyldiacylglycerol composition of dinoflagellates. IV. Temperature-induced modulation of fatty acid regiochemistry as observed by electrospray ionization/mass spectrometry. European J. Phycol. 45: 13–18.

Leblond JD, Timofte HI, Roche SA, Porter NM. 2010b. Mono- and digalactosyldiacylglycerol composition of glaucocystophytes (Glaucophyta): A modern interpretation using positive-ion electrospray/mass spectrometry/mass spectrometry. Phycol. Res. 58: 222–229.

Leblond JD, Dodson J, Khadka M, Holder S, Seipelt RS. 2012. Sterol composition and biosynthetic genes of the recently discovered photosynthetic alveolate, *Chromera velia* (Chromerida), a close relative of apicomplexans. J. Eukaryot. Microbiol. 59: 191–197.

Le Corguillé G, Pearson G, Valente M, Viegas C, Gschloessl B, Corre E, Bailly X, Peters AF, Jubin C, Vacherie B, Cock JM, Leblanc C. 2009. Plastid genomes of two brown algae, *Ectocarpus siliculosus* and *Fucus vesiculosus*: further insights on the evolution of red-algal derived plastids. BMC Evol. Biol. 9: 253.

Mazumdar J, Striepen B. 2007. Make it or take it: Fatty acid metabolism of apicomplexan parasites. Eukaryot. Cell. 6: 1727–1735.

Moore RB et al. 2008. A photosynthetic alveolate closely related to apicomplexan parasites. Nature. 451: 959–963.

Oborník M, Vancová M, Lai DH, Janouškovec J, Keeling PJ, Lukeš J. 2011. Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of apicomplexa, *Chromera velia*. Protist. 162: 115–130.

Ramakrishnan S, Docampo MD, MacRae JI, Pujol FM, Brooks CF, Van Dooren GG, Hiltunen JK, Kastaniotis AJ, McConville MJ, Striepen B. 2012. Apicoplast and endoplasmic reticulum cooperate in fatty acid biosynthesis

in apicomplexan parasite *Toxoplasma gondii*. J. Biol. Chem. 287: 4957–4971.

Riekhof WR, Sears BB, Benning C. 2005. Annotation of genes involved in glycerolipid biosynthesis in *Chlamydomonas reinhardtii*: Discovery of the betaine lipid synthase $BTA1_{CR.}$ Eukaryot. Cell. 4: 242–252.

Roy S, Morse D. 2012. A full suite of histone and histone modifying genes are transcribed in the dinoflagellate *Lingulodinium*. Plos One. 7: E34340.

Sánchez Puerta MV, Bachvaroff TR, Delwiche CF. 2005. The complete plastid genome sequence of the haptophyte *Emiliania huxleyi*: A comparison to other plastid genomes. DNA Res. 12: 151–156.

Sato N, Moriyama T. 2000. Genomic and biochemical analysis of lipid biosynthesis in the unicellular rhodophyte *Cyanidioschyzon merolae*: Lack of a plastidic desaturation pathway results in the couple pathway of galactolipid synthesis. Eukaryot. Cell. 6: 1006–1017.

Sato N, Sugimoto K, Meguro A, Tsuzuki M. 2003. Identification of a gene for UDP sulfoquiovose synthase of a green alga *Chlamydomonas reinhardtii* and its phylogeny. DNA Res. 10: 229–237.

Sayanova O, Haslam R, Guschina I, Lloyd D, Christie WW, Harwood JL, Napier, JA. 2006. A bifunctional Delta12,Delta15-desaturase from *Acanthamoeba castellanii* directs the synthesis of highly unusual n-1 series unsaturated fatty acids. J. Biol. Chem. 281: 36533–36541.

Sayanova O, Haslam RP, Venegas CM, Ruiz López N, Worthy C, Rooks P, Allen MJ, Napier JA. 2011. Identification and functional characterisation of genes encoding the omega-3 polyunsaturated fatty acid biosynthetic pathway from the coccolithophore *Emiliania huxleyi*. Phytochem. 72: 594–600.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731–2739.

Wada H, Murata N. 2009. Lipids in photosynthesis: Essential and regulatory functions. Dordrecht, Netherlands: Springer.

Welti R, Li W, Li M, Sang Y, Biesiada H, Zhou H, Rajashekar C, Williams T, Wang X. 2002. Profiling membrane lipids in plant stress responses: Role of phospholipase D in freezing-induced lipid changes in *Arabidopsis*. J. Biol. Chem. 277: 31994–32002.

Woehle C, Dagan T, Martin WF, Gould SB. 2011. Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related *Chromera velia*. Genome Biol. and Evol. 3: 1220–1230.

PROJECT CONCLUSION

The phylum Chromerida contains two photosynthetic, marine microalgal species,*Chromera velia* and *Vitrella brassicaformis*, that are close relatives of non-photosynthetic apicomplexans. Apicomplexans are a group of parasitic protists that cause devastating diseases like malaria in humans. They possess a relic, red algae-derived plastid called an apicoplast that has lost photosynthetic ability. The apicoplast is essential for the parasites' survival, and performs essential functions like biosynthesis of isoprene units required for the synthesis of haem, dolichols, menaquinone, geranylgeranylated proteins, and farnesylated proteins. The apicomplexan parasites have lost the ability to synthesize sterols *de novo* and galactolipids.

This research project was accomplished utilizing biochemical and bioinformatics approaches to compare the sterols and galactolipid composition, and thebiosynthetic genes, of chromerids to apicomplexans, dinoflagellates, and red algae. As these species are closely related to apicomplexans, the relevancy of this study was to propose possible sterol and galactolipid compositions possessed by apicomplexans before they lost their biosynthetic abilities.

By performing mass spectrometric and transcriptomic analyses, I found that *C. velia* produces greater diversity of sterols than *V. brassicaformis*. The phylogenetic analyses utilizing sterol biosynthetic genes showed the red algal origin of Chromerida. *V. brassicaformis* is more closely related to the

dinoflagellate, *Pyrocystis lunula*, and perkinsids and then to apicomplexans based on *dxr* gene phylogeny. Chromerid sterols were found in unrelated classes of photoautotrophic microalgae, such as chlorarachniophytes, glaucocystophytes, and plants, thus making them unsuitable to use as biomarker molecule. I also conclude that apicomplexans could have possibly synthesized similar sterols in the past utilizing the non-mevalonate pathway. However, further experiments should utilize feeding of chromerids with [6^{13}-C] mevalonate and [1^{13}-C] glucose to elucidate the sterol biosynthetic pathways' enzymatic steps.

Studying the galactolipid composition of chromerids and their biosynthetic genes, I found that *V. brassicaformis* produces a greater diversity of galactolipids than *C. velia*. Both MGDG and DGDG in *V. brassicaformis* have a mixture of C_{14}, C_{16}, C_{18} fatty acids in the *sn*-2 position, indicating the use of both prokaryotic and eukaryotic pathways. *C. velia* produces galactolipids with C_{20} and C_{18} fatty acids at the *sn*-2 position at 20℃ and galactolipid with C_{16} fatty acid at *sn*-2 position at 30℃ suggesting the use of prokaryotic and eukaryotic pathways. Phylogenetic analyses using galactolipid biosynthetic genes underscored the red algal origin of plastid lipid biosynthetis in the Chromerida. As the apicomplexans do not synthesize galactolipids, here I propose an idea that these parasites might have synthesized the galactolipids in the past via prokaryotic and eukaryotic pathways with C_{20}, C_{18}, and C_{16} fatty acids attached to them.

REFERENCES

Adl SM et al. 2012. The revised classification of eukaryotes. J. Eukaryot. Microbiol. 59(5): 429-514.

Benveniste P. 1986. Sterol biosynthesis. Ann. Rev. Plant Physiol. 37: 275-308.

Botté CY et al. 2013. Atypical lipid composition in the purified relict plastid (apicoplast) of malaria parasites. Proc. Natl. Acad. Sci. 110: 75-6-7511.

Coppens I, Sinai AP, Joiner KA. 2000. Toxoplasma gondii exploits host low-density lipoprotein receptor-mediated endocytosis for cholesterol acquisition. J. Cell Biol. 149(1): 167-180.

Eisenreich W, Bacher A, Brigoni D, Rohdich F. 2004. Biosynthesis of isoprenoids via the non mevalonate pathway. Cell. Mol. Life Sci. 61(12): 1401-1426.

Goodwin TW. 1979. Biosynthesis of terpenoids. Annu. Rev. Plant Physiol. 30: 369-404.

Grunwald C. 1975. Plant sterol. Annu. Rev. Plant Physiol. 26: 209-236.

Guo J, Zhang Z, Bi Y, Yang W, Xu Y, Zhang L. 2005. Decreased stability of photosystem I in dgd1 mutant of Arabidopsis thaliana. FEBS Lett. 579(17): 4959-4968.

Hölzl G, Dörmann P. 2007. Structure and function of glycoglycerolipid in plants and bacteria. Prog. Lipid Res. 46: 225-243.

Hunter WN. 2007. The non-mevalonate pathway of isoprenoid precursor biosynthesis. J. Biol. Chem. 282: 21573-21577.

Ivanov A, Hendrickson L, Krol M, Selstam E, Oquist G, Hurry B, Huner N. 2006. Digalactosyl-diacylglycerol deficiency impairs the capacity for photosynthetic intersystem electron transport and state transitions in Arabidopsis thaliana due to photosystem I acceptor side limitations. Plant Cell Environ. 47(8): 1146-1157.

Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ. 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc. Natl. Acad. Sci. USA, 107: 10949-10954.

Kobayashi K, Kondo M, Gukuda H, Nishimura M, Ohta H. 2007. Galactolipid synthesis in chloroplast inner envelope is essential for proper thylakoid biogenesis, photosynthesis, and embryogenesis. Proc. Natl. Acad. Sci. USA. 104(43): 17216-17221.

Kokke WCMC, Fenical W, Bohlin L, Djerassi C. 1981. Sterol synthesis by cultured zooxanthallae: Implications concerning sterol metabolism in the host-symbiont association in Caribbean gorgonians. Comp. Biochem. Physiol. 68B: 281-287.

Labaied M, Jayabalasingham B, Bano N, Cha SJ, Sandoval J. Guan, G, Coppens I. 2011. Plasmodium salvages cholesterol internalized by LDL and synthesized de novo in the liver. Cell. Microbiol. 13(4): 569-586.

Lichtenthaler HK. 1999. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 50: 47-65.

Lichtenthaler HK. 2010. The non-mevalonate DOXP/MEP (Deoxyxylulose 5-phosphate/methylerythritol 4-phosphate) pathway of chloroplast isoprenoid and pigment biosynthesis. Reveiz CA, Benning C, Bohnert H, Daniell H, Hoober JK, Lichtenthaler HK, Portis AR, Tripathy BC, editors. The chloroplast: Basics and applications. Advances in photosynthesis and respiration. Springer Science. 31: 95-118.

Maréchal E, Azzouz N, de Macedo CS, Block MA, Feagin JE. Schwarz RT, Joyard J. 2002. Synthesis of chloroplast galactolipids in apicomplexan parasites. Eukaryot. Cell. 1(4): 653-656.

McFadden GI, Reith ME, Munholland J, Lang-Unnasch N. 1996. Plastid in human parasites. Nature. 381(6582): 482.

Moore RB. 2008. A photosynthetic alveolate closely related to apicomplexan parasites. Nature. 451: 959-963.

Nes WD. 2011. Biosynthesis of cholesterol and other sterols. Chem. Rev. 11: 6423-6451.

Nishikawa Y, Quittnat F, Stedman TT, Voelker DR, Choi JY, Zahn M, Yang M, Pypaert M, Joiner KA, Coppens I. 2005. Host cell lipids control cholesteryl

ester synthesis and storage in intracellular Toxoplasma. Cell. Microbiol. 7(6): 849-867.

Oborník M, Modrý D, Lukeš M, Černotíková-Stříbrná E, Cihlář J, Tesařová M, Kotabová E, Vancová M, Prášil O, Lukeš, J. 2012. Morphology, ultrastructure and life cycle of *Vitrella brassicaformis* n. sp., n. gen., a novel chromerid from the great barrier reef. Protist. 163(2): 306-323.

Oborník M, Vancová M, Lai DH, Janouškovec J, Keeling PJ, Lukeš J. 2011. Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of apicomplexa, *Chromera velia*. Protist. 162(1): 115-130.

Pearson A, Budin M, Brocks JJ. 2003. Phylogenetic and biochemical evidence for sterol synthesis in the bacterium *Gemmata obscuriglobus*. Proc. Natl. Acad. Sci. USA. 100(26): 15352-15357.

Rodriguez-Concepcion M, Campos N, Lois LM, Maldonado C, Hoeffler JF, Grosdemange-Billiard C, Rohmer M, Boronat A. 2000. Genetic evidence of branching in the isoprenoid pathway for the production of isopentenyl diphosphate and dimethylallyl diphosphate in *Escherichia coli*. FEBS Lett. 473(3): 328-332.

Rohmer M. 1999. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. Nat. Prod. Rep. 16(5): 565-574.

Wanke M, Skorupinska-Tudek K, Swiezewska E. 2001. Isoprenoid biosynthesis via 1-deoxy-D-xylulose 5-phosphate/2-C-methyl-D-erythritol 4-phosphate (DOXP/MEP) pathway. Acta Biochim. Pol. 48(3): 663-672.