

School of Doctoral Studies in Biological Sciences
University of South Bohemia in České Budějovice
Faculty of Science

Genomes of Chromerid Algae

Ph.D. Thesis

Mgr. Jan Michálek

Supervisor: Prof. Ing. Miroslav Oborník, Ph.D.
Institute of Parasitology, Biology Centre CAS v.v.i.

České Budějovice 2020

This thesis should be cited as:

Michálek J 2020. Genomes of Chromerid Algae.

Ph.D. Thesis. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic.

Annotation

This work is mainly focused on the annotation of genomes of an important group of alveolate algae, which are the closest known phototrophic relatives of sporozoan parasites. The analysis is aimed to annotate their mitochondrial and nuclear genomes and search for biosynthetic pathways essential for sporozoans by combining the *in-silico* approach with supportive experiments. The initial study describes mitochondrial genomes of *Chromera velia* and *Vitrella brassicaformis* and models their unprecedentedly divergent respiratory chains. The following study summarizes the general characteristics of nuclear genomes of the two chromerids, highlights main genomic features shared with apicomplexan parasites, and traces the evolutionary transitions from the phototrophic ancestor to the obligate apicomplexan parasites. The study was completed by a comprehensive biochemical, genetic, and evolutionary analyses of fatty acid biosynthesis in chromerids.

Declaration [in Czech]

Prohlašuji, že svoji disertační práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury. Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

České Budějovice, 20.9.2020

Jan Michálek

This thesis originated from a partnership of Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre CAS, supporting doctoral studies in the Molecular and Cell Biology and Genetics study programme.



Přírodovědecká
fakulta
Faculty
of Science

Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice



Financial support

This work was financially supported by Grant Agency of the Czech Republic projects P506-12-1522, 13-33039S and 18-13458S.

Acknowledgements

Here, I would like to thank my supervisor Míra Oborník for involving me in this remarkable work and being patient in my moments out of space. Our chats and discussions are always inspiration and fun. Thanks to his team, Laboratory of Evolutionary Protistology, a very special bunch all the time. I thank particularly to Žaba, a great pal and colleague, and Katcha for keeping the whole team and lab breathing.

I thank to my family for all kinds of support they gave me during my effort to get some respectable education in a profession I am fond of. And most importantly, thanks to Radka for keeping me aware of the fact that life is not just accomplishing the never-ending queue of tasks. Thank you!

List of papers and author's contribution

I. Flegontov P, Michálek J, Janouškovec J, Lai DH, Jirku M, Hajdušková E, Tomčala A, Otto TD, Keeling PJ, Pain A, Oborník M, Lukeš J 2015. **Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites**. *Molecular Biology and Evolution* 32: 1115-1131.

J. Michálek and P. Flegontov contributed equally to this work. J. Michálek searched for the parts of missing complex III, confirmed its absence by wet-lab experiments, performed all the phylogenetic analyses in the study, drafted the scheme of respiratory chains, and participated in the writing of the manuscript.

II. Woo YH, Ansari H, Otto TD, Klinger CM, Kolísko M, **Michálek J**, Saxena A, Shanmugam D, Tayyrov A, Veluchamy A, Ali S, Bernal A, del Campo J, Cihlář J, Flegontov P, Gornik SG, Hajdušková E, Horák A, Janouškovec J, Katris NJ, Mast FD, Miranda-Saavedra D, Mourier T, Naeem R, Nair M, Panigrahi AK, Rawlings ND, Padron-Regalado E, Ramaprasad A, Samad N, Tomčala A, Wilkes J, Neafsey DE, Doerig C, Bowler C, Keeling PJ, Roos DS, Dacks JB, Templeton TJ, Waller RF, Lukeš J, Oborník M, Pain A 2015. **Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites**. *Elife* 4: e06974.

J. Michálek identified de-novo fatty acid synthesis pathways, described the structure of putative FAS I multienzymes, and computed their phylogeny. He also identified enzymes of the non-mevalonate isoprenoid pathways in chromerids and predicted their subcellular localizations.

III. Tomčala A, Michálek J, Schneedorferová I, Füssy Z, Gruber A, Vancová M, Oborník M 2020. **Fatty Acid Biosynthesis in Chromerids**. *Biomolecules* 10: 1102.

J. Michálek and A. Tomčala contributed equally to this work. J. Michálek performed genomic searches, phylogenetic analyses, targeting predictions, interpreted results, and participated in the writing of the manuscript.

Co-author agreement:

Miroslav Oborník, the supervisor of this Ph.D. thesis and co-author of all of the publications included in this thesis, fully acknowledges the contribution of Jan Michálek in all publications as described.

Prof. Ing. Miroslav Oborník, Ph.D.

Contents

Introduction	1
Alveolata	1
<i>Chromera velia</i>	2
<i>Vitrella brassicaformis</i>	2
Complex plastids	3
Mitochondria	6
Endosymbioses and lateral gene transfer	8
Aims and Scope	9
Conclusion	10
References	15
Paper I. Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasite	26
Paper II. Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites.	60
Paper III. Fatty Acid Biosynthesis in Chromerids.	102
Curriculum vitae	128

Alveolata is a remarkable group of protists characterized by the presence of a specific cellular structure called cortical alveoli – a system of flattened vesicles located just below their cell wall. Ecologically diverse alveolates comprise of significant phototrophs, predators, and symbionts (commensals, mutualists and parasites). Alveolates (Figure 1.) are divided into the main groups of Apicomplexa (Sporozoa and Apicomonada), Dinozoa (dinoflagellates and perkinsids), Ciliophora (Cavalier-Smith 2018), and recently added Colponemida (Tikhonenkov et al. 2020). Besides predatory colpodellids, Apicomonada include also chromerids – *Chromera velia* (Moore et al. 2008) and *Vitrella brassicaformis* (Oborník et al. 2012), complex algae associated with corals. Alveolates Dinozoa and Apicomplexa, are united under the term Myzozoa, defined by the myzocytosis, phagocytosis-derived feeding mechanism. By use of various forms of an apical complex, the protist attach the prey or host, and suck out the content of the cell (Cavalier-Smith and Chao 2004). The affiliation of chromerids to Apicomplexa *sensu lato* has also been recently supported by the description of a deep branching archigregarine *Digyalum owenii*, which forms a sister lineage to all apicomplexan parasites, chromerids and colpodellids (Janouškovec et al. 2019). Attention is being paid to Sporozoa (Cavalier-Smith 2018) – obligate apicomplexan parasites, which are worldwide highly relevant to the everyday life of humans and associated fauna. The word ‘Apicomplexa’ refers to the protrusive structure of the apical complex used by most apicomplexans for a symbiont-host interaction. Some sporozoan species are among the leading scourges of humankind. Above all, malarial parasites *Plasmodium falciparum* and *Plasmodium vivax* (Talapko et al. 2019) are limited to tropical and subtropical conditions but insidious coccidian *Toxoplasma gondii* influences human behavior and overall health condition even in developed countries (e.g., Flegr and Kuba 2016; Flegr and Markoš 2014). In the course of medical research, it has been discovered that most apicomplexan parasites possess a multimembrane organelle with indisputable attributes of a non-photosynthetic relic plastid, indicating an autotrophic evolutionary ancestry of these parasites. Moreover, it has been shown that the so-called apicoplast (Maréchal and Cesbron-Delauw 2001; McFadden and Waller 1997; Oborník et al. 2009) is essential for the parasite survival thus becoming an object of high medical interest.

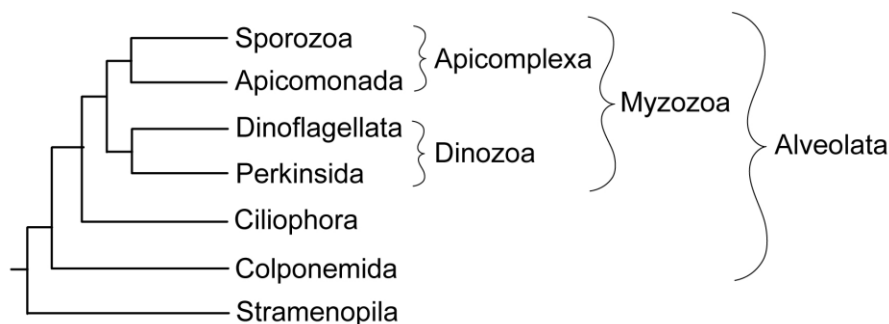


Figure 1. Phylogenetic tree showing the position of Apicomonada within alveolates. Adapted from Tikhonenkov et al. 2020.

The first discovered chromerid alga, ***Chromera velia*** (Apicomonada; Apicomplexa), was isolated from the scleractinian coral *Plesiastrea versipora* growing in shallow Australian waters in 2001. When observed closer, the rather dull-looking round brownish cells in size from 5 to 10 μm (Figure 2.) reminded a

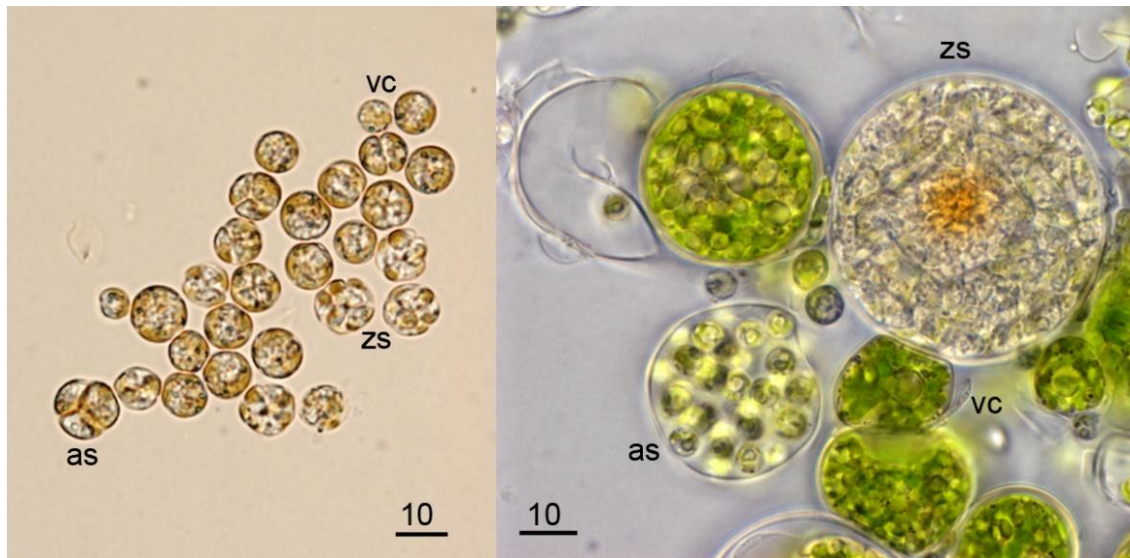


Figure 2. Light microscopy showing a typical life stages (vc – vegetative cell, zs – zoosporangium, as – autosporangium) of *C. velia* (left) and *V. brassicaformis* (right) (Füssy and Oborník 2017).

widespread dinoflagellate coral symbiont of the genus *Symbiodinium*, showing, however, many distinct and similar features. The system of cortical alveoli with a supportive microtubular corset, reduced pre-conoid, and a large plastid coated in four membranes (Figure 4.), indicate its relation to Myzozoa (Moore et al. 2008; Oborník et al. 2011). Cells show many other morphological traits that can be observed in various alveolates with the myzozoan lifestyle. An intense light exposure triggers a transformation of coccoid vegetative cells into bi-flagellated zoospores (Figure 3.), which strongly resemble related predatory colpodellids. The structure of pre-conoid in *C. velia* is homologous to the apical complex in sporozoans, and colpodellids (Oborník et al. 2011). The resemblance of colpodellids is not only coincidental, since they constitute, together with chromerids, the group Apicomonada (Cavalier-Smith 2018; Gile and Slamovits 2014; also called "chrompodellids" Janouškovec et al. 2015). The accumulating evidence clearly supported the idea that chromerids represent a state of evolution closest to the hypothetical phototrophic ancestor of apicomplexan parasites, however, bearing a plastid with preserved photosynthetic function.

The second formally described chromerid alga, ***Vitrella brassicaformis*** (Apicomonada; Apicomplexa) (Oborník et al. 2012), was isolated from another stony coral *Leptastrea purpurea* on the Great Barrier Reef. Although molecular data suggest its close relation to *C. velia* in frame of Apicomonada, numerous

differences in the life cycle, gene organization and biochemical properties provide an evidence for high mutual divergence between *V. brassicaformis* and *C. velia* (Janouškovec et al. 2010; Oborník et al. 2012). The first obvious difference is overall appearance of the cells. While *C. velia* cells are regular smooth coccoids of brownish color, the vegetative cells and sporangia of *V. brassicaformis* are green

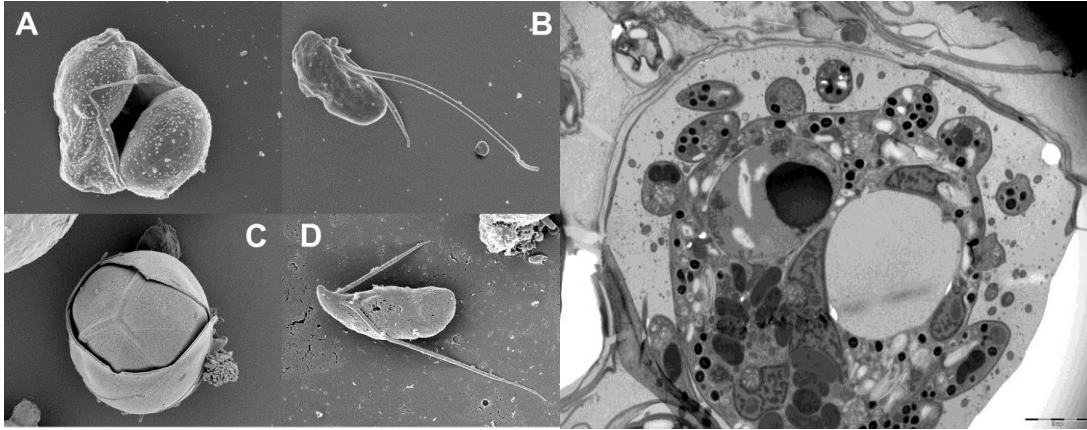


Figure 3. Scanning electron microscopy (left) showing vegetative cells with zoospores of *C. velia* (A,B) and *V. brassicaformis* (C,D). Transmission electron microscopy (right) of budding *V. brassicaformis* (Füssy and Oborník, 2017; Füssy et al. 2017).

and have a laminated cell wall resembling a cabbage head (“brassicaformis” means cabbage-shaped). While size of vegetative cells and sporangia can reach 10 μm in *C. velia*, can in the case of *V. brassicaformis* sporangia overcome 30 μm . In contrast to *C. velia* with only four immotile coccoids released from the sporangium (Oborník et al. 2011) and up to 10 zoospores in the zoosporangium (Oborník et al. 2016), *V. brassicaformis* produces dozens of autospores and zoospores in each sporangium (Figure 3.). Zoosporangia of *V. brassicaformis* are equipped with conspicuous operculum likely homologous to Stieda body in Coccidia (Füssy and Oborník 2017; Oborník and Lukeš 2013; Oborník et al. 2012). It is remarkable that *V. brassicaformis* produces two different types of zoospores, and the zoospore fusions have been observed, suggesting a possible sexual stage in the *V. brassicaformis* lifecycle. In contrast, *C. velia* seems to be asexual during entire lifecycle (Füssy et al. 2017).

Many members of the SAR group (Stramenopila, Alveolata, Rhizaria; Burki et al. 2012) contain **complex plastids**. Complex (secondary, tertiary and higher order events) endosymbiosis was proposed based on the presence of multimembrane plastid envelopes (3 membranes and more) of complex algae (Archibald 2015; Füssy and Oborník 2018; Gould et al. 2008; Keeling 2013b; Oborník 2018, 2019a). It is believed that the two inner plastidial membranes are of cyanobacterial origins, the second outermost membrane originates from the cell membrane of the engulfed rhodophyte, and the phagosomal membrane of the exosymbiont (secondary host) constitute the outermost membrane of the four-membraned complex rhodophyte-derived plastid (Cavalier-Smith 1999). Rhodophyte origin of

the SAR plastids was inferred from the characteristic pigment composition, with chlorophyll *c* as the main accessory tetrapyrrole (Cavalier-Smith 1986), and was confirmed by the accumulating phylogenetic evidence (Delwiche and Palmer 1997; Durnford, et al. 1999; Janouškovec, et al. 2010; Yoon, et al. 2002). Since algae with rhodophyte derived plastids are not monophyletic, and the algal clades contain many early branching heterotrophic species, the single acquisition of the plastid proposed by Thomas Cavalier-Smith (Cavalier-Smith 1999) has been many times questioned. Various hypotheses have been proposed to explain such discrepancy, involving either series of higher order endosymbiotic events (Burki et al. 2016; Füssy and Oborník 2018; Nowack et al. 2008; Petersen et al. 2014; Stiller et al. 2014), or separated secondary endosymbioses (Falkowski et al. 2004; Oborník 2018, 2019a). An outstanding example of tertiary plastids has been described in the dinoflagellates *Durinskia baltica* and *Kryptoperidinium foliaceum*, which host almost perfectly preserved diatom endosymbionts (Crowell et al. 2019; Imanian et al. 2010); such dinoflagellates are therefore called ‘dinotoms’.

The acquisition of a plastid endosymbiont was followed by a gradual reduction of its genome and functions. Reductive evolution affected the organelle in all lineages to various extents, reaching extremes in sporozoans (Füssy and Oborník 2017), euglenophyte *Euglena longa* (Füssy, et al. 2019), and green parasitic algae *Helicosporidium* (de Koning and Keeling 2006) and *Polytomella*. The relic plastid of the last-named species even lacks any genome (Smith and Lee 2014). So far the smallest sporozoan plastome was described in the genus *Plasmodium*, where it was found highly conserved in structure and gene repertoire – usually encoding for 30 proteins in total size of 35 kb but is still supplemented by approximately 500 nuclear-encoded genes (Arisue, et al. 2012; Ralph, et al. 2004a; Waller and McFadden 2005). The apicoplast was completely lost only in genus *Cryptosporidium* (Abrahamsen et al. 2004; Huang et al. 2004; Riordan et al. 2003; Zhu et al. 2000) and the eugregarine *Gregarina niphadroides* (Toso and Omoto 2007). In most parasitic apicomplexans, the apicoplast defended its very existence by keeping several essential functions: the biosynthesis of isoprenoids, fatty acid synthesis, synthesis of tetrapyrroles (heme) and iron-sulphur cluster assembly (Lim and McFadden 2010; Seeber and Soldati-Favre 2010; van Dooren and Hapuarachchi 2017; van Dooren et al. 2012). Therefore, the apicoplast has been repeatedly evaluated as a prospective apicomplexan-specific drug target (Low, et al. 2018; Mukherjee and Sadhukhan 2016; Ralph, et al. 2001; Uddin, et al. 2018). The plastids found in *C. velia* and *V. brassicaformis* indicated their relationship to the apicoplast, based on phylogenies, biochemical and structural features (Moore et al. 2008). Phylogenetic analyses of plastid genes showed that chromerid and sporozoan plastids share the common ancestry (Janouškovec et al. 2010; Moore et al. 2008; Janouškovec et al. 2015; Woo et al. 2015) but also pointed out a relatively high divergence between *C. velia* and *V. brassicaformis* plastids. The organization of plastid superoperon strongly suggested common origins of sporozoan, chromerid and stramenopile plastids (Janouškovec et al. 2010). The

use of non-canonical code for tryptophan (UGA) in plastid-encoded proteins of *C. velia* is an important common molecular feature of the chromerid and coccidian (e.g., *T. gondii*) plastids, which has never been found in any other plastid organelle



Figure 4. Transmission electron microscopy showing a section of nucleus (N) and large plastid (P) of *C. velia* .

(Janouškovec et al. 2010; Moore et al. 2008; Oborník and Lukeš 2013, 2015). Also, a unique composition of light-harvesting pigments in *C. velia* – chlorophyll *a*, violaxanthin and a special isoform of isofucoxanthin, in unprecedented (Moore et al. 2008). Chlorophyll *c*, which is missing from all chromerids (Moore et al. 2008; Oborník et al. 2012; Oborník and Lukeš, 2013; Oborník and Lukeš, 2015; Füssy and Oborník, 2017), is characteristic for almost all red-algal derived plastids except that of all eustigmatophytes (Sukenik et al. 1992) and the ochrophyte *Xanthonema debile* (Gardian, et al. 2011). It was even hypothesized that Apicomplexa (Sporozoa and Apicomonada) acquired plastid by replacing the original alveolate plastid by tertiary stramenopile endosymbiont (Oborník et al. 2012; Sobotka et al. 2017; Ševčíková et al. 2015). A close relationship between chromerid and sporozoan plastids is also supported by a non-canonical arrangement of the tetrapyrrole synthesis in all apicomplexans (sporozoans and apicomonads). The pathway is a hybrid between the route in primary eukaryotic heterotrophs (e.g., animals, fungi) and the synthesis known in phototrophic

eukaryotes (plants and algae): in chromerids, the route starts with the mitochondrial C4 pathway for the aminolevulinate synthesis (heterotrophic-like condensation of succinyl-CoA and glycine like in sporozoans), while the remaining steps are located to the plastid; apicomplexan parasites placed the four middle steps in the apicoplast, the pathway is finalized in the apicomplexan mitochondrion (Cihlář, et al. 2019; Kořený, et al. 2011). Also, the density of plastid genomes differs in chromerids: while the plastid genome of *C. velia* encodes only 80 proteins on 120 kb long linear molecule, conservative and much more compacted plastid genome of *V. brassicaformis* encodes similar number of 81 genes just on the 85 kb circular molecule. Genes in the *C. velia* plastid are AT-rich, photosynthetic genes *psaA* and *atpB* are split in two parts, which are individually transcribed and translated, and the protein products are then assembled to functional photosystems. Similarly to dinoflagellates (Wang and Morse 2006), some of plastid transcripts are polyuridylylated in chromerids (Janouškovec et al. 2010). The *V. brassicaformis* plastid does not use the non-canonical coding (Janouškovec et al. 2010), contain much shorter and non-split genes, and is pigmented by vaucheriaxanthin (Oborník et al. 2012) instead of isofucoxanthin in *C. velia* (Moore et al. 2008). Chromerid *C. velia* is capable of extraordinarily efficient photosynthesis (Kotabová et al. 2011; Quigg et al. 2012) despite the high reduction of proteins in photosystems. However, PSI in *C. velia* contains four yet undescribed subunits with no homology in the public databases (Sobotka et al. 2017).

Myzozoans (dinoflagellates, perkinsids, chromerids, colpodellids, and apicomplexan parasites) have evolved remarkably reduced **mitochondria**. Their genomes and functionalities were considered the last level before the mitochondria lost genomes and became mitosomes or hydrogenosomes (Tovar et al. 1999). Such an extensive reduction of a mitochondrial genome and functions usually correlates with an anaerobic lifestyle (Stairs et al. 2015). However, even the absence of mitochondrial genome does not necessarily condemn the organism to live without oxygen; endoparasitic dinoflagellates of the genus *Amoebophrya* host mitochondrion without genome but with functional oxidative phosphorylation (John et al. 2019). Among apicomplexan parasites (Sporozoa), the true mitosome was found only in the genus *Cryptosporidium* – an epicellular parasite living in the low-oxygen environment of a gut lumen of vertebrates. The general pattern observed in myzozoan mitochondrial genomes is composed of a set of three genes encoding respiratory chain enzymes – *cox1* and *cox3* (genes for subunits of cytochrome *c* oxidase), *cob* gene for cytochrome *c* reductase, and several mitochondrial ribosomal rRNA fragments (Feagin et al. 1997; Flegontov et al. 2015; Nash et al. 2007; Oborník and Lukeš 2015; Waller and Jackson 2009). The rest of the gene set coding for proteins necessary for oxidative phosphorylation is nuclear-encoded, including *cox2*, otherwise present in all other eukaryotic mitochondrial genomes (Waller and Keeling 2006). The structure of mitochondrial genomes conspicuously differs between sporozoans, apicomonads and

dinoflagellates. While sporozoan mitochondrial genomes are organized on homogeneous linear molecules spanning 5.8-11 kb, dinoflagellates and apicomonads placed their mitochondrial genes on heterogeneous linear DNA fragments encoding genes, gene fragments and fused genes (Flegontov et al. 2015; Imanian and Keeling 2007; Jackson et al. 2007; Kamikawa et al. 2007; Nash et al. 2007; Oborník and Lukeš 2015; Waller and Jackson 2009). Advanced dinoflagellates extensively use RNA editing in their mitochondria, which seems to be derived because the deep branching Dinozoa *Perkinsus* and *Oxyrrhis* lack this gene repair mechanism (Masuda et al. 2010; Slamovits et al. 2007; Zhang et al. 2011). The early branching colponemid *Acavomonas peruviana* has the mitochondrial genome organized in linear molecules terminated with telomere-like sequences at both ends, encoding much more complex set of genes than that found in myzozoan mitochondria (Janouškovec et al. 2013; Tikhonenkov et al. 2014; Tikhonenkov et al. 2020).

Although there are no doubts questioning the relationship between chromerids and sporozoans, the structure and content of chromerid mitochondrial genomes rather resemble dinoflagellates. The genomes are, like in dinoflagellates, organized on heterogeneous linear molecules, coding for gene fragments, complete genes, and fused genes, all in various combinations (Flegontov et al. 2015; Oborník and Lukeš 2015).



Figure 5. Transmission electron microscopy showing a section with mitochondria (M) and chromerosome (C) in *C. velia* (Oborník et al. 2011).

While the mitochondrial genome of *V. brassicaformis* codes for the same reduced set as most other myzozoans (*cox1*, *cox3*, and *cob*), the genome in the mitochondrion of *C. velia* (Figure 5.) is reduced even more, as it also lacks *cob*

gene. Since genes coding for remaining protein components of the respiratory complex III (cytochrome *c* reductase) are missing from the nuclear genome of *C. velia*, the respiratory chain lacks this complex. Like in other myxozoans, respiratory chains of chromerids also do not have the complex I (NADH dehydrogenase), which is substituted by numerous nuclear-encoded alternative NADH dehydrogenases. Therefore, the chain of *C. velia* is interrupted, constituting two disconnected subchains (Flegontov et al. 2015). Such unique arrangement of the respiratory chain was also recently described in the intracellular parasite *A. ceratii*, which lacks complexes I and III like *C. velia* (John et al. 2019).

Endosymbioses and lateral gene transfer

Mitochondria and primary plastids are organelles, which arose from the primary endosymbiosis of eukaryotes engulfing an alpha-proteobacteria or/and cyanobacteria. The process of complex (secondary and higher order) endosymbiosis is based on phagocytosis of a eukaryotic organism already possessing a primary or complex plastid. It is believed that the original plasmatic membrane of the engulfed eukaryotic organism was preserved, and the endosymbiont remained packed in the phagosome of the host cell. Preservation of these membranes resulted in multiple membranes enclosing the plastid (Archibald 2009; Keeling 2013a; Oborník 2019b). However, it is obvious that in the case of higher-order plastids (tertiary, quaternary, etc.), the membrane envelope has been either reduced or the plastid was obtained by myxocytosis (Cavalier-Smith 1999). The same applies to the peridinin pigmented plastid of dinoflagellates, which is surrounded by the three membranes envelope (Keeling 2013; Waller and Kořený 2017). Similar to mitochondria, plastid genomes have been gradually reduced with the consequent transfer of endosymbiont genes into the host nuclear genome. Such a phenomenon is referred to as the endosymbiotic gene transfer (Timmis et al. 2004). Therefore, we can find a high number of originally bacterial (cyanobacterial or proteobacterial) genes in nuclear genomes of all eukaryotes. Moreover, the enzymes of various origins can combine in so-called mosaic metabolic pathways, thus reflecting amazing plasticity of a cell metabolism (Cihlář et al. 2016; Curtis et al. 2012; Kořený et al. 2010; Nara et al. 2000; Oborník and Green 2005).

Protein products of nuclear-encoded genes involved in the organellar metabolism are transported from the cytosol to the place of the action thanks to the N-terminal targeting presequence. In the primary organelles, it is the transit peptide that is responsible for the translocation through the double membrane envelope of primary plastids or mitochondria. In the case of complex plastids, the proteins contain a bi-partite N-terminal presequence, composed of the signal peptide (SP) followed by the transit peptide-like domain (Bolte et al. 2009; Cavalier-Smith 1999; Patron and Waller 2007). This combination suggests a connection of the secretory pathway and complex plastids (DeRocher et al. 2012; Heiny et al. 2014; Kilian and Kroth 2003; McFadden 2011; Nassoury et al. 2003). Proteins targeted into plastids

with triple membrane have an additional hydrophobic domain (membrane anchor) following the signal and transit peptide (Durnford and Gray 2006; Patron et al. 2005). Bi-partite targeting sequence motifs display different levels of conservation in various groups of protists bearing complex plastids. Most algae hosting rhodophyte-derived complex plastids bounded by four membranes share conserved 'ASAF' motif allowing a more or less easy identification of such genes using software tools (Gruber, et al. 2015; Gruber, et al. 2007).

Transit peptides are specifically recognized by the translocon machinery responsible for the transport of proteins through organellar envelopes. The translocons on the outer and inner chloroplast membranes (TIC and TOC complexes) recognize chloroplast-specific transit peptides and deliver proteins into primary plastids (Becker, et al. 2004; Gross and Bhattacharya 2009; Vojta, et al. 2004). In the case of complex plastid with three membranes, these two complexes are preceded by the translocon complex of endoplasmic reticulum, which recognizes the signal peptide-like presequence in order to bypass the outermost membrane of the complex plastid. Signal peptide is consequently cleaved off and transit peptide exposed (Bolte et al. 2009; Kilian and Kroth 2003; Patron and Waller 2007). Plastids with four membranes have an additional envelope layer underlying the ER membrane to pass through to TIC/TOC complexes. This layer is passed through the complex 'SELMA' (Symbiont-specific ERAD-Like MAchinery) that is derived from the apparatus originally responsible for disposing of misfolded proteins from endoplasmic reticulum (Felsner et al. 2011).

Aims and scope

A discovery of a so-called "evolutionary missing link" is always an object of interest. Since the described diversity of protists is less than a tip of an iceberg, we still have almost a limitless number of questions about their evolution. The most important groups of eukaryotes are fairly described but the connections between them are still often unclear. The two species, *C. velia* and *V. brassicaformis*, which are the only known organisms of their kind yet quite diverse, give us a scientifically exciting opportunity to test all hypotheses about ancestry of a highly important group of parasites. The ability of *C. velia* to grow rapidly in artificial conditions of enriched f/2 medium (artificial seawater) allows us performing experiments to test *in silico* findings. Easy cultivation, which is not much practicable in apicomplexan parasites and colpodellids, predetermines *C. velia* for an excellent model organism. With current advanced technologies of the whole genome sequencing, mass spectrometry, and microscopy, we have aimed to describe some of the substantial aspects of nuclear and mitochondrial genomes of chromerids, and fatty acid metabolism to put them in context of the myzozoan evolution.

Conclusion

Flegontov P*, Michálek J*, Janouškovec J, Lai DH, Jirku M, Hajdušková E, Tomčala A, Otto TD, Keeling PJ, Pain A, Oborník M, Lukeš J 2015. **Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites.** *Molecular Biology and Evolution* 32: 1115-1131. (*equal contribution)

The first publication describes genomes and functions of chromerid mitochondria and puts them in context of myzozoan evolution. Using the genomic data and wet-lab experiments to support our findings we have investigated mitochondria of chromerids to see how they differ from sporozoan mitochondria and suggest how such state have evolved. Surprisingly, we have found that the mitochondrial genome of *C. velia* is even more reduced than in most apicomplexan parasites; it is composed of numerous short linear heterogeneous molecules encoding only for conserved *cox1*, highly divergent *cox3*, and fragments of rRNA genes. The mitochondrial genome of *C. velia* represents the smallest mitochondrial genome described to date. In contrast to *C. velia*, the mitochondria of *V. brassicaformis* has the same genome composition as most myzozoan mitochondria. The repertoire of nuclear encoded components of the respiratory chain in *C. velia* does not contain any proteins of complex I like apicomplexans, and complex III is missing as well. The function of the missing complex I is substituted by alternative NADH dehydrogenases, which, however, do not contribute to the proton gradient. Electrons produced by these dehydrogenases and complex II are channeled to ubiquinone and passed to the alternative oxidase, which spends them on generation of water. The next step maintained by complex III, which would normally accept the electron flow and attend on building of proton gradient, is present only in *V. brassicaformis*. Therefore, the only part creating the proton gradient (membrane potential) in *C. velia* is complex IV powered by the system of lactate dehydrogenases coupled with cytochrome c oxidoreductase. In addition, the nuclear genome of *C. velia* also encodes an outstanding number of enzymes, which are rather specific for anaerobic protists but not for most sporozoans. This finding indicates that the common ancestor of chromerids and apicomplexans was much more metabolically versatile and predisposed for survival in anaerobic conditions. While the mitochondrion of *V. brassicaformis* meets our expectations with an arrangement of the respiratory chain, *C. velia* shows very advanced reductions of mitochondrial genome and functions unusual for aerobic eukaryotes. The mitochondrial genome and respiratory chain of *V. brassicaformis*, which is mostly identical to those in most sporozoans, seems to be closer to the default pre-apicomplexan shape, since a retrograde gain of respiratory chain complexes is hardly imaginable. A model of respiratory chain without complexes I and III similar to that in *C. velia* was also recently described *A. ceratii*, an intracellular parasite with no plastid, that has a functional mitochondrion completely lacking a genome (John et al. 2019).

Woo YH, Ansari H, Otto TD, Klinger CM, Kolísko M, Michálek J, Saxena A, Shanmugam D, Tayyrov A, Veluchamy A, Ali S, Bernal A, del Campo J, Cihlář J, Flegontov P, Gornik SG, Hajdušková E, Horák A, Janouškovec J, Katris NJ, Mast FD, Miranda-Saavedra D, Mourier T, Naeem R, Nair M, Panigrahi AK, Rawlings ND, Padron-Regalado E, Ramaprasad A, Samad N, Tomčala A, Wilkes J, Neafsey DE, Doerig C, Bowler C, Keeling PJ, Roos DS, Dacks JB, Templeton TJ, Waller RF, Lukeš J, Oborník M, Pain A 2015. **Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites.** *Elife* 4: e06974.

The second paper represents a thorough comparative summary of nuclear genomes of *C. velia* (193.6 Mb; 26,112 predicted protein-coding genes) and *V. brassicaformis* (72.7 Mb; 22,817 predicted protein-coding genes). It appears that it took five stages from a sporozoan ancestor to highly specialized intracellular parasites. The transitions between stages are characterized mostly by almost equal count of gene gains and losses when the common ancestor of chromerids and sporozoans diverged from the alveolate ancestor, many losses at the rise of sporozoans, and further group-specific modifications with extreme reduction levels in haematozoans. Analysis of metabolic pathways, endomembrane trafficking system, and flagellar apparatus confirmed that most initial losses were accompanied by the loss of photosynthesis and sterol synthesis, while advanced parasites were characterized by further lineage-specific losses of various genes. Apart from a reduction, parts of the flagellar apparatus and some cytoskeletal components were re-purposed in the development of cell-invading structures and mechanisms such as apical complex and gliding motility. This indicates that pre-sporozoan ancestor was most probably a phototrophic protist with some more or less limited ability of myzocytosis (Oborník 2020). Origin of an extracellular parasitism can be also observed in the ancient group of gregarines and related genus *Cryptosporidium* (Bartošová-Sojková et al. 2015)

Interesting is also the high abundance of transposable elements in genome of *C. velia* (almost 30 Mb of 193.6 Mb). The role of transposable elements (TEs) in endosymbiotic gene transfer has been already described in plants. Transposable elements were found to be enhancing susceptibility of nuclear genome to the acceptance of plastid genes, the speed of their integration and the mutation rate. This was proposed as one of the main drivers shaping the structures and evolution of plant genomes (Michalovova et al. 2013; Zhang et al. 2020). The high abundance of transposable elements in *C. velia* is responsible for the increased genome size, and their activity may stay behind the disturbed respiratory chain. The initial genomic study of *Amoebophrya ceratii*, a dinoflagellate which has an analogously reduced respiratory chain and no mitochondrial genome, also found a significant activity of transposable elements (John et al. 2019). The genomes of other dinoflagellates are in general large and demonstrate very advanced levels of organelle integration often together with a high abundance of TEs, which probably served as a driver of their diversity (Song et al. 2018). It seems that the

combination of abundant TEs with a complex engulfment event were probably among the main catalyzers of myzozoan evolution. The disparity between the two chromerid genomes in TEs content is another trait highlighting their mutual divergence. The proliferation of TEs is widely found to cause genome expansions. It was shown that TEs enhance the adaptive potential by increasing genomic plasticity and this way drive a lineage specific diversification in oomycetes, fungi, plants, and animals (Hawkins et al. 2008; Raffaele et al. 2012; Seidl et al. 2017; Alioto et al. 2020; Warren et al. 2015; Woo et al. 2019).

Tomčala A*, Michálek J*, Schneedorferová I, Füssy Z, Gruber A, Vancová M, Oborník M 2020. **Fatty Acid Biosynthesis in Chromerids**. *Biomolecules* 10: 1102. (*equal contribution)

The third paper summarizes the description of fatty acid synthesis in chromerids. The synthesis of basic short-chain saturated fatty acids varies not only between different sporozoan subgroups but also between genera, species and life stages. We measured the fatty acid spectra in chromerids, and described the corresponding genes to draft a model of their biosynthesis. Both *C. velia* and *V. brassicaformis* produce short-chain saturated fatty acids by plastid-localized type II fatty acid synthesis pathway (FAS II). The products of 14-18 carbon chain lengths subsequently undergo desaturation and elongation steps in the cytosol and endoplasmic reticulum. We also described a large spectrum of genes resembling type I fatty acid synthases (FAS I), which were found responsible for fatty acid synthesis or elongation in *Cryptosporidium parvum* (Zhu et al. 2000) and *Toxoplasma gondii* (Goodman and McFadden 2007), respectively. Some of these genes might play a role in FA synthesis but more likely represent polyketide synthases responsible for the production of algal toxins similar to those found in dinoflagellates. Products of polyketide synthases have a large spectrum of biological activity – e.g., toxins (bacterial, algal, fungal), antibiotics and immunomodulators (Hussain et al. 2017; Choi et al. 2017; Mady and Haggag 2020; Park et al. 2020; Robertsen and Musiol-Kroll 2019). Their presence in chromerids could be related to an interaction with the hypothetical host by modulating its immunity. Compared to the apicomplexan parasites that have significantly reduced or almost completely lost their ability to synthesize fatty acids, chromerids resemble more of the other thoroughly described algal fatty acid producers such as the diatom *Phaeodactylum tricornutum* (Burrows et al. 2012) or the chlorophyte *Chlamydomonas reinhardtii* (Yang et al. 2015), including the boosting effect of a nitrogen deprivation. The phylogenetic analyses showed that many genes encoding enzymes involved in fatty acid biosynthesis display various origins reflecting an endosymbiotic evolutionary history of chromerids and sporozoans. As the fatty acid synthesis is one of the most energetically demanding processes in any cell, it is logical that it has been substantially reduced in sporozoans as they have evolved mechanisms to outsource fatty acids from their hosts.

The high diversity of this particular gene family can be achieved by sequential rearrangements of domain series (Jenke-Kodama et al. 2005). Similar high diversity of this gene family is characteristic for dinoflagellates just as the abundance of TEs (Beedessee et al. 2018). The activity of TEs has been shown to change activities of polyketide-based cyanobacterial toxins (Fewer et al. 2011; Christiansen et al. 2008; Jones et al. 2009). Again, we can speculate a possible role of transposable elements in the high diversity of PKS/FAS-like genes in *C. velia*. The rich genomic equipment for synthesis of fatty acids and effective mechanism for their storage show that the sporozoan ancestor was also an excellent phototroph.

The insight into the repertoires and structures of nuclear and organellar genomes of chromerids shape a reasonable conception of how the last common ancestor of chromerids and sporozoans looked like – a metabolically plastic organism with ability of myzocytosis. The large rich-equipped genome of the ancestor provided a basis for lines of specialization. As the myzocytotic apparatus of chromerids seems rather reduced, it is possible that the original state preceding the divergence of the specialized sporozoan parasites was represented by phototrophy along with a mild extracellular parasitism rather than mutualism (Obornik 2020). Similar strategy can be seen on examples among higher plants such as mistletoes (*Viscum* spp.), *Nuytsia floribunda* or *Rhinanthus* spp. It is worth mentioning that the plant hemiparasite European mistletoe (*Viscum album*) was found to lack respiratory complex I in its mitochondria and possess many alternative oxidases and dehydrogenases (Maclean et al. 2018).

The example of a dinoflagellate genus *Symbiodinium* formerly accepted as a paradigm of mutualism shows that mutualism and parasitism are continual (Lesser et al. 2013). A shift to parasitism has already been described in *Symbiodinium* inside the sponge *Cliona orientalis* after 20 days of cultivation without light (Fang et al. 2017). Another example of a gastropod host *Strombus gigas* shows that the shift can be stage specific so the beneficial symbiotic effect in the host's early stages can be reimbursed in adult organism once tissues become less translucent (Banaszak et al. 2013). An experimental horizontal transmission of *Symbiodinium* between hosts of the same species was also found to induce a shift towards parasitism (Sachs and Wilcox 2006). It appears that *Symbiodinium* is more likely a domesticated mild parasite than a pure mutualist. It reminds a pathogenic interaction which acts as a selective pressure for individuals proficient enough to establish a mutualist relationship. It is a struggle of symbiont's tendency to parasitize with the host's immunity. This initial interaction seems to stabilize the border between parasitism and mutualism in every new host generation (Mies et al. 2017). In general, the overall evolutionary age of host-parasite coexistence within a mutual niche is negatively correlated with parasite's virulence. For example, evolutionarily young Coccidia and Haemosporidia are highly virulent

for their mammalian hosts, whereas old clades of gregarines have only a mild impact on their hosts, and deep branching sporozoans *Nephromyces* and corallicolids are harmless or even mutualistic (Kwong, et al. 2019; Saffo, et al. 2010). This, again, supports an idea that the ancestor was more parasitic than previously thought.

Even though we have never witnessed the myzocytosis in chromerids in action, we can surely place them among myzozoans based on their phylogenetic traits and ultrastructure of *C. velia*. There are many levels of myzocytosis among different groups of myzozoans and yet, there is still a cryptic diversity of chromerids that have never been formally described (Janouškovec et al. 2012; Oborník 2020). Based on their close relationship with colpodellids we can guess that some of the undescribed chromerids may be more 'aggressive'.

The indisputable potential of chromerids to be perfect model organisms for studying apicomplexan biology now relies mostly on their cultivability together with sequenced genomes and transcriptomes. This fact allowed us to experimentally test our hypotheses coming out of our *in silico* work. The next step, which so far represents the hardest obstacle, is to develop a technique for their genetic transformation.

References

- Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S, Buck GA, Xu P, Bankier AT, Dear PH, Konfortov BA, Spriggs HF, Iyer L, Anantharaman V, Aravind L, Kapur V 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304: 441-445. doi: 10.1126/science.1094786
- Alioto T, Alexiou KG, Bardil A, Barteri F, Castanera R, Cruz F, Dhingra A, Duval H, Fernández i Martí Á, Frias L, Galán B. 2020. Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. *The Plant Journal* 101(2):455-72.
- Archibald JM 2015. Endosymbiosis and eukaryotic cell evolution. *Current Biology* 25: R911-R921.
- Archibald JM 2009. The puzzle of plastid evolution. *Current Biology* 19: R81-88. doi: 10.1016/j.cub.2008.11.067
- Arisue N, Hashimoto T, Mitsui H, Palacpac NM, Kaneko A, Kawai S, Hasegawa M, Tanabe K, Horii T 2012. The *Plasmodium* apicoplast genome: conserved structure and close relationship of *P. ovale* to rodent malaria parasites. *Molecular Biology and Evolution* 29: 2095-2099.
- Banaszak AT, Ramos MG, Goulet TL 2013. The symbiosis between the gastropod *Strombus gigas* and the dinoflagellate *Symbiodinium*: an ontogenic journey from mutualism to parasitism. *Journal of Experimental Marine Biology and Ecology* 449: 358-365.
- Bartošová-Sojtková P, Oppenheim RD, Soldati-Favre D, Lukeš J 2015. Epicellular apicomplexans: parasites “On the way in”. *PLoS Pathogens* 11: e1005080.
- Becker T, Jelic M, Vojta A, Radunz A, Soll J, Schleiff E 2004. Preprotein recognition by the Toc complex. *EMBO Journal* 23: 520-530.
- Beedessee G, Hisata K, Roy MC, Van Dolah FM, Satoh N, Shoguchi E 2018. Comparative genomics-first approach to understand diversification of secondary metabolite biosynthetic pathways in symbiotic dinoflagellates. bioRxiv: 376251.
- Bolte K, Bullmann L, Hempel F, Bozarth A, Zauner S, Maier UG 2009. Protein targeting into secondary plastids. *Journal of Eukaryotic Microbiology* 56: 9-15. doi: 10.1111/j.1550-7408.2008.00370.x
- Burki F, Flegontov P, Oborník M, Cihlář J, Pain A, Lukeš J, Keeling PJ 2012. Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biology and Evolution* 4: 626-635. doi: 10.1093/gbe/evs049
- Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mylnikov AP, Keeling PJ 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proceedings of the Royal Society B: Biological Sciences* 283: 20152802.

- Burrows EH, Bennette NB, Carrieri D, Dixon JL, Brinker A, Frada M, Baldassano SN, Falkowski PG, Dismukes GC 2012. Dynamics of lipid biosynthesis and redistribution in the marine diatom *Phaeodactylum tricornutum* under nitrate deprivation. *BioEnergy Research* 5: 876-885.
- Cavalier-Smith T 2018. Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. *Protoplasma* 255: 297-357. doi: 10.1007/s00709-017-1147-3
- Cavalier-Smith T 1986. The kingdom Chromista: origin and systematics. *Progress in Phycological Research* 4: 309-347.
- Cavalier-Smith T, Chao EE 2004. Protalveolate phylogeny and systematics and the origins of Sporozoa and dinoflagellates (phylum Myzozoa nom. nov.). *European Journal of Protistology* 40: 185-212. doi: 10.1016/j.ejop.2004.01.002
- Cavalier-Smith T 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree 1, 2. *Journal of Eukaryotic Microbiology* 46: 347-366.
- Cihlář J, Füssy Z, Horák A, Oborník M 2016. Evolution of the Tetrapyrrole Biosynthetic Pathway in Secondary Algae: Conservation, Redundancy and Replacement. *PLoS ONE* 11: e0166338. doi: 10.1371/journal.pone.0166338
- Cihlář J, Füssy Z, Oborník M. 2019. Evolution of tetrapyrrole pathway in eukaryotic phototrophs. In: *Advances in Botanical Research*: Elsevier. p. 273-309.
- Crowell RM, Nienow JA, Cahoon AB 2019. The complete chloroplast and mitochondrial genomes of the diatom *Nitzschia palea* (Bacillariophyceae) demonstrate high sequence similarity to the endosymbiont organelles of the dinoflagellate *Durinskia baltica*. *Journal of Phycology* 55: 352-364.
- Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, Hopkins JF, Kuo A, Rensing SA, Schmutz J, Symeonidi A, Elias M, Eveleigh RJ, Herman EK, Klute MJ, Nakayama T, Oborník M, Reyes-Prieto A, Armbrust EV, Aves SJ, Beiko RG, Coutinho P, Dacks JB, Durnford DG, Fast NM, Green BR, Grisdale CJ, Hempel F, Henrissat B, Hoppner MP, Ishida K, Kim E, Kořený L, Kroth PG, Liu Y, Malik SB, Maier UG, McRose D, Mock T, Neilson JA, Onodera NT, Poole AM, Pritham EJ, Richards TA, Rocap G, Roy SW, Sarai C, Schaack S, Shirato S, Slamovits CH, Spencer DF, Suzuki S, Worden AZ, Zauner S, Barry K, Bell C, Bharti AK, Crow JA, Grimwood J, Kramer R, Lindquist E, Lucas S, Salamov A, McFadden GI, Lane CE, Keeling PJ, Gray MW, Grigoriev IV, Archibald JM 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492: 59-65. doi: 10.1038/nature11681
- de Koning AP, Keeling PJ 2006. The complete plastid genome sequence of the parasitic green alga *Helicosporidium* sp. is highly reduced and structured. *BMC Biol* 4: 12.

- Delwiche CF, Palmer JD. 1997. The origin of plastids and their spread via secondary symbiosis. In: *Origins of algae and their plastids*: Springer. p. 53-86.
- DeRocher AE, Karnataki A, Vaney P, Parsons M 2012. Apicoplast targeting of a *Toxoplasma gondii* transmembrane protein requires a cytosolic tyrosine-based motif. *Traffic* 13: 694-704. doi: 10.1111/j.1600-0854.2012.01335.x
- Durnford D, Deane J, Tan S, McFadden G, Gantt E, Green B 1999. A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. *Journal of Molecular Evolution* 48: 59-68.
- Durnford DG, Gray MW 2006. Analysis of *Euglena gracilis* plastid-targeted proteins reveals different classes of transit sequences. *Eukaryotic Cell* 5: 2079-2091.
- Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor F 2004. The evolution of modern eukaryotic phytoplankton. *Science* 305: 354-360.
- Fang JK, Schönberg CH, Hoegh-Guldberg O, Dove S 2017. Symbiotic plasticity of *Symbiodinium* in a common excavating sponge. *Marine Biology* 164: 104.
- Feagin JE, Mericle BL, Werner E, Morris M 1997. Identification of additional rRNA fragments encoded by the *Plasmodium falciparum* 6 kb element. *Nucleic Acids Research* 25: 438-446. doi: 10.1093/nar/25.2.438
- Felsner G, Sommer MS, Gruenheit N, Hempel F, Moog D, Zauner S, Martin W, Maier UG 2011. ERAD components in organisms with complex red plastids suggest recruitment of a preexisting protein transport pathway for the periplastid membrane. *Genome Biology and Evolution* 3: 140-150.
- Fewer DP, Halinen K, Sipari H, Bernardová K, Mänttari M, Eronen E, Sivonen K 2011. Non-autonomous transposable elements associated with inactivation of microcystin gene clusters in strains of the genus *Anabaena* isolated from the Baltic Sea. *Environmental Microbiology Reports* 3: 189-194.
- Flegontov P, Michálek J, Janouškovec J, Lai DH, Jirků M, Hajdušková E, Tomčala A, Otto TD, Keeling PJ, Pain A, Oborník M, Lukeš J 2015. Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. *Molecular Biology and Evolution* 32: 1115-1131. doi: 10.1093/molbev/msv021
- Flégr J, Kuba R 2016. The relation of *Toxoplasma* infection and sexual attraction to fear, danger, pain, and submissiveness. *Evolutionary Psychology* 14: 1474704916659746.
- Flégr J, Markoš A 2014. Masterpiece of epigenetic engineering—how *Toxoplasma gondii* reprogrammes host brains to change fear to sexual attraction. *Molecular Ecology* 23: 5934-5936.
- Füssy Z, Masařová P, Kručinská J, Esson HJ, Oborník M 2017. Budding of the Alveolate Alga *Vitrella brassicaformis* Resembles Sexual and Asexual Processes in Apicomplexan Parasites. *Protist* 168: 80-91. doi: 10.1016/j.protis.2016.12.001

- Füssy Z, Oborník M 2018. Complex Endosymbioses I: From Primary to Complex Plastids, Multiple Independent Events. *Methods in Molecular Biology* 1829: 17-35. doi: 10.1007/978-1-4939-8654-5_2
- Füssy Z, Oborník M 2017. Chromerids and Their Plastids. *Advances in Botanical Research* 84: 187-218. doi: 10.1016/bs.abr.2017.07.001
- Füssy Z, Záhonová K, Tomčala A, Krajčovič J, Yurchenko V, Oborník M, Eliáš M 2019. An unprecedented combination of metabolic pathways in the cryptic plastid of a non-photosynthetic euglenophyte. *bioRxiv* 765255.
- Gardian Z, Tichý J, Vácha F 2011. Structure of PSI, PSII and antennae complexes from yellow-green alga *Xanthonema debile*. *Photosynthetic Research* 108: 25-32. doi: 10.1007/s11120-011-9647-z
- Gile GH, Slamovits CH 2014. Transcriptomic analysis reveals evidence for a cryptic plastid in the colpodellid *Voromonas pontica*, a close relative of chromerids and apicomplexan parasites. *PLoS One* 9: e96258.
- Goodman CD, McFadden GI 2007. Fatty acid biosynthesis as a drug target in apicomplexan parasites. *Current Drug Targets* 8: 15-30. doi: 10.2174/138945007779315579
- Gould SB, Waller RF, McFadden GI 2008. Plastid evolution. *Annual Review of Plant Biology* 59. Gross J, Bhattacharya D 2009. Revaluating the evolution of the Toc and Tic protein translocons. *Trends in plant science* 14: 13-20.
- Gruber A, Vugrinec S, Hempel F, Gould SB, Maier UG, Kroth PG 2007. Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant Molecular Biology* 64: 519-530. doi: 10.1007/s11103-007-9171-x
- Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T 2015. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *The Plant Journal* 81: 519-528
- Heiny SR, Pautz S, Recker M, Przyborski JM 2014. Protein traffic to the *Plasmodium falciparum* apicoplast: evidence for a sorting branch point at the Golgi. *Traffic* 15: 1290-1304.
- Huang J, Mullapudi N, Lancto CA, Scott M, Abrahamsen MS, Kissinger JC 2004. Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. *Genome Biology* 5: R88.
- Hussain H, Al-Sadi AM, Schulz B, Steinert M, Khan A, Green IR, Ahmed I 2017. A fruitful decade for fungal polyketides from 2007 to 2016: antimicrobial activity, chemotaxonomy and chemodiversity. *Future Medicinal Chemistry* 9: 1631-1648.
- Choi S-S, Nah H-J, Pyeon H-r, Kim E-S 2017. Biosynthesis, regulation, and engineering of a linear polyketide tautomycin: a novel immunosuppressant in *Streptomyces* sp. CK4412. *Journal of Industrial Microbiology & Biotechnology* 44: 555-561.

- Christiansen G, Molitor C, Philmus B, Kurmayer R 2008. Nontoxic strains of cyanobacteria are the result of major gene deletion events induced by a transposable element. *Molecular Biology and Evolution* 25: 1695-1704.
- Imanian B, Keeling PJ 2007. The dinoflagellates *Durinskia baltica* and *Kryptoperidinium foliaceum* retain functionally overlapping mitochondria from two evolutionarily distinct lineages. *BMC Evolutionary Biology* 7: 172. doi: 10.1186/1471-2148-7-172
- Imanian B, Pombert J-F, Keeling PJ 2010. The complete plastid genomes of the two 'dinotoms' *Durinskia baltica* and *Kryptoperidinium foliaceum*. *PLoS ONE* 5: e10711.
- Jackson CJ, Norman JE, Schnare MN, Gray MW, Keeling PJ, Waller RF 2007. Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. *BMC Biology* 5: 41. doi: 10.1186/1741-7007-5-41
- Janouškovec J, Horák A, Barott KL, Rohwer FL, Keeling PJ 2012. Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. *Current Biology* 22: R518-519. doi: 10.1016/j.cub.2012.04.047
- Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proceedings of the National Academy of Science of USA* 107: 10949-10954. doi: 10.1073/pnas.1003335107
- Janouškovec J, Paskerova GG, Miroliubova TS, Mikhailov KV, Birley T, Aleoshin VV, Simdyanov TG 2019. Apicomplexan-like parasites are polyphyletic and widely but selectively dependent on cryptic plastid organelles. *eLife* 8. doi: 10.7554/eLife.49662
- Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Kolísko M, Mylnikov AP, Keeling PJ 2015. Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proceedings of the National Academy of Sciences of USA* 112: 10200-10207.
- Janouškovec J, Tikhonenkov DV, Mikhailov KV, Simdyanov TG, Aleoshin VV, Mylnikov AP, Keeling PJ 2013. Colponemids represent multiple ancient alveolate lineages. *Current Biology* 23: 2546-2552.
- Jenke-Kodama H, Sandmann A, Muller R, Dittmann E 2005. Evolutionary implications of bacterial polyketide synthases. *Molecular Biology and Evolution* 22: 2027-2039. doi: 10.1093/molbev/msi193
- John U, Lu Y, Wohlrab S, Groth M, Janouškovec J, Kohli GS, Mark FC, Bickmeyer U, Farhat S, Felder M, Frickenhaus S, Guillou L, Keeling PJ, Moustafa A, Porcel BM, Valentin K, Glockner G 2019. An aerobic eukaryotic parasite with functional mitochondria that likely lacks a mitochondrial genome. *Science Advances* 5: eaav1110. doi: 10.1126/sciadv.aav1110
- Jones AC, Gu L, Sorrels CM, Sherman DH, Gerwick WH 2009. New tricks from ancient algae: natural products biosynthesis in marine cyanobacteria. *Current Opinion in Chemical Biology* 13: 216-223.

- Kamikawa R, Inagaki Y, Sako Y 2007. Fragmentation of mitochondrial large subunit rRNA in the dinoflagellate *Alexandrium catenella* and the evolution of rRNA structure in alveolate mitochondria. *Protist* 158: 239-245. doi: 10.1016/j.protis.2006.12.002
- Keeling PJ 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annual Review of Plant Biology* 64: 583-607.
- Kilian O, Kroth P 2003. Evolution of protein targeting into “complex” plastids: The “secretory transport hypothesis”. *Plant Biology* 5: 350-358.
- Kořený L, Lukeš J, Oborník M 2010. Evolution of the haem synthetic pathway in kinetoplastid flagellates: an essential pathway that is not essential after all? *International Journal for Parasitology* 40: 149-156. doi: 10.1016/j.ijpara.2009.11.007
- Kořený L, Sobotka R, Janouškovec J, Keeling PJ, Oborník M 2011. Tetrapyrrole synthesis of photosynthetic chromerids is likely homologous to the unusual pathway of apicomplexan parasites. *Plant Cell* 23: 3454-3462. doi: 10.1105/tpc.111.089102
- Kotabová E, Kana R, Jarešová J, Prášil O 2011. Non-photochemical fluorescence quenching in *Chromera velia* is enabled by fast violaxanthin de-epoxidation. *FEBS Letters* 585: 1941-1945. doi: 10.1016/j.febslet.2011.05.015
- Kwong WK, Del Campo J, Mathur V, Vermeij MJ, Keeling PJ 2019. A widespread coral-infecting apicomplexan with chlorophyll biosynthesis genes. *Nature* 568: 103-107.
- Lesser M, Stat M, Gates R 2013. The endosymbiotic dinoflagellates (*Symbiodinium sp.*) of corals are parasites and mutualists. *Coral Reefs* 32: 603-611.
- Lim L, McFadden GI 2010. The evolution, metabolism and functions of the apicoplast. *Philosophical Transactions of the Royal Society of London B Biological Sciences* 365: 749-763. doi: 10.1098/rstb.2009.0273
- Low LM, Stanisci DI, Good MF 2018. Exploiting the apicoplast: apicoplast-targeting drugs and malaria vaccine development. *Microbes and infection* 20: 477-483.
- Macleán AE, Hertle AP, Ligas J, Bock R, Balk J, Meyer EH 2018. Absence of complex I is associated with diminished respiratory chain function in European mistletoe. *Current Biology* 28: 1614-1619. e1613.
- Mady M, Haggag E 2020. Review on Fungi of Genus *Penicillium* As a Producers of Biologically Active Polyketides. *Journal of Advanced Pharmacy Research* 4: 33-45.
- Maréchal E, Cesbron-Delauw M-F 2001. The apicoplast: a new member of the plastid family. *Trends in Plant Science* 6: 200-205.
- Masuda I, Matsuzaki M, Kita K 2010. Extensive frameshift at all AGG and CCC codons in the mitochondrial cytochrome c oxidase subunit 1 gene of *Perkinsus marinus* (Alveolata; Dinoflagellata). *Nucleic Acids Research* 38: 6186-6194. doi: 10.1093/nar/gkq449

- McFadden GI 2011. The apicoplast. *Protoplasma* 248: 641-650. doi: 10.1007/s00709-010-0250-5
- McFadden GI, Waller RF 1997. Plastids in parasites of humans. *Bioessays* 19: 1033-1040.
- Mies M, Sumida PY, Rädercker N, Voolstra CR 2017. Marine invertebrate larvae associated with *Symbiodinium*: a mutualism from the start? *Frontiers in Ecology and Evolution* 5: 56.
- Michalovova M, Vyskot B, Kejnovsky E 2013. Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. *Heredity* 111: 314-320. doi: 10.1038/hdy.2013.51
- Moore RB, Oborník M, Janouškovec J, Chrudimsky T, Vancova M, Green DH, Wright SW, Davies NW, Bolch CJ, Heimann K, Slapeta J, Hoegh-Guldberg O, Logsdon JM, Carter DA 2008. A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* 451: 959-963. doi: 10.1038/nature06635
- Mukherjee A, Sadhukhan GC 2016. Anti-malarial drug design by targeting apicoplasts: new perspectives. *Journal of Pharmacopuncture* 19: 7.
- Nara T, Hshimoto T, Aoki T 2000. Evolutionary implications of the mosaic pyrimidine-biosynthetic pathway in eukaryotes. *Gene* 257: 209-222.
- Nash EA, Barbrook AC, Edwards-Stuart RK, Bernhardt K, Howe CJ, Nisbet RER 2007. Organization of the mitochondrial genome in the *dinoflagellate* *Amphidinium carterae*. *Molecular Biology and Evolution* 24: 1528-1536.
- Nassoury N, Cappadocia M, Morse D 2003. Plastid ultrastructure defines the protein import pathway in dinoflagellates. *Journal of Cell Science* 116: 2867-2874.
- Nowack EC, Melkonian M, Glöckner G 2008. Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Current Biology* 18: 410-418.
- Oborník M 2018. The Birth of Red Complex Plastids: One, Three, or Four Times? *Trends in Parasitology* 34: 923-925. doi: 10.1016/j.pt.2018.09.001
- Oborník M 2019a. Endosymbiotic Evolution of Algae, Secondary Heterotrophy and Parasitism. *Biomolecules* 9: 266. doi: 10.3390/biom9070266
- Oborník M 2019b. In the beginning was the word: How terminology drives our understanding of endosymbiotic organelles. *Microbial Cell* 6: 134-141. doi: 10.15698/mic2019.02.669
- Oborník M 2020. Photoparasitism as an Intermediate State in the Evolution of Apicomplexan Parasites. *Trends in Parasitology* doi: 10.1016/j.pt.2020.06.002
- Oborník M, Green BR 2005. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Molecular Biology and Evolution* 22: 2343-2353. doi: 10.1093/molbev/msi230

- Oborník M, Janouškovec J, Chrudimský T, Lukeš J 2009. Evolution of the apicoplast and its hosts: from heterotrophy to autotrophy and back again. *International Journal for Parasitology* 39: 1-12. doi: 10.1016/j.ijpara.2008.07.010
- Oborník M, Lukeš J 2013. Cell biology of chromerids: autotrophic relatives to apicomplexan parasites. *International Review of Cell and Molecular Biology* 306: 333-369. doi: 10.1016/B978-0-12-407694-5.00008-0
- Oborník M, Lukeš J 2015. The Organellar Genomes of *Chromera* and *Vitrella*, the Phototrophic Relatives of Apicomplexan Parasites. *Annual Review of Microbiology* 69: 129-144. doi: 10.1146/annurev-micro-091014-104449
- Oborník M, Modrý D, Lukeš M, Černotíková-Stříbrná E, Cihlák 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. doi: 10.1016/j.protis.2011.09.001
- 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. doi: 10.1016/j.protis.2010.02.004
- Park HB, Goddard TN, Oh J, Patel J, Wei Z, Perez CE, Mercado BQ, Wang R, Wyche TP, Piizzi G 2020. Bacterial Autoimmune Drug Metabolism Transforms an Immunomodulator into Structurally and Functionally Divergent Antibiotics. *Angewandte Chemie International Edition* 59: 7871-7880.
- Patron NJ, Waller RF 2007. Transit peptide diversity and divergence: A global analysis of plastid targeting signals. *Bioessays* 29: 1048-1058. doi: 10.1002/bies.20638
- Patron NJ, Waller RF, Archibald JM, Keeling PJ 2005. Complex protein targeting to dinoflagellate plastids. *Journal of Molecular Biology* 348: 1015-1024. doi: 10.1016/j.jmb.2005.03.030
- Petersen J, Ludewig A-K, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H 2014. *Chromera velia*, endosymbioses and the rhodoplex hypothesis—plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biology and Evolution* 6: 666-684.
- Quigg A, Kotabová E, Jarešová J, Kana R, Šetlík J, Šediva B, Komárek O, Prášil O 2012. Photosynthesis in *Chromera velia* represents a simple system with high efficiency. *PLoS ONE* 7: e47036. doi: 10.1371/journal.pone.0047036
- Ralph SA, D'Ombrain MC, McFadden GI 2001. The apicoplast as an antimalarial drug target. *Drug Resistance Update* 4: 145-151. doi: 10.1054/drup.2001.0205
- Ralph SA, Van Dooren GG, Waller RF, Crawford MJ, Fraunholz MJ, Foth BJ, Tonkin CJ, Roos DS, McFadden GI 2004a. Metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nature Reviews Microbiology* 2: 203-216

- Riordan CE, Ault JG, Langreth SG, Keithly JS 2003. *Cryptosporidium parvum* Cpn60 targets a relict organelle. *Current Genetics* 44: 138-147.
- Robertsen HL, Musiol-Kroll EM 2019. Actinomycete-Derived Polyketides as a Source of Antibiotics and Lead Structures for the Development of New Antimicrobial Drugs. *Antibiotics* 8: 157.
- Sachs JL, Wilcox TP 2006. A shift to parasitism in the jellyfish symbiont *Symbiodinium microadriaticum*. *Proc Biol Sci* 273: 425-429. doi: 10.1098/rspb.2005.3346
- Saffo MB, McCoy AM, Rieken C, Slamovits CH 2010. *Nephromyces*, a beneficial apicomplexan symbiont in marine animals. *Proceedings of the National Academy of Sciences* 107: 16190-16195.
- Seeber F, Soldati-Favre D 2010. Metabolic Pathways in the Apicoplast of Apicomplexa. *International Review of Cell and Molecular Biology* 281: 161-228. doi:10.1016/s1937-6448(10)81005-6
- Slamovits CH, Saldarriaga JF, Larocque A, Keeling PJ 2007. The highly reduced and fragmented mitochondrial genome of the early-branching dinoflagellate *Oxyrrhis marina* shares characteristics with both apicomplexan and dinoflagellate mitochondrial genomes. *Journal of Molecular Biology* 372: 356-368.
- Smith DR, Lee RW 2014. A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. *Plant Physiol* 164: 1812-1819.
- Sobotka R, Esson HJ, Koník P, Trsková E, Moravcová L, Horák A, Dufková P, Oborník M 2017. Extensive gain and loss of photosystem I subunits in chromerid algae, photosynthetic relatives of apicomplexans. *Scientific Reports* 7: 1-13.
- Song B, Chen S, Chen W 2018. Dinoflagellates, a unique lineage for retrogene research. *Frontiers in Microbiology* 9: 1556.
- Stairs CW, Leger MM, Roger AJ 2015. Diversity and origins of anaerobic metabolism in mitochondria and related organelles. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370: 20140326. doi: 10.1098/rstb.2014.0326
- Stiller JW 2014. Toward an empirical framework for interpreting plastid evolution. *Journal of Phycology* 50: 462-471.
- Sukenik A, Livne A, Neori A, Yacobi YZ, Katcoff D 1992. Purification and characterization of a light-harvesting chlorophyll-protein complex from the marine Eustigmatophyte *Nannochloropsis* sp. *Plant and Cell Physiology* 33: 1041-1048.
- Ševčíková T, Horák A, Klimeš V, Zbránková V, Demir-Hilton E, Sudek S, Jenkins J, Schmutz J, Přibyl P, Fousek J 2015. Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Scientific Reports* 5: 10134.

- Talapko J, Škrlec I, Alebić T, Jukić M, Včev A 2019. Malaria: the past and the present. *Microorganisms* 7: 179.
- Tikhonenkov DV, Janouškovec J, Mylnikov AP, Mikhailov KV, Simdyanov TG, Aleoshin VV, Keeling PJ 2014. Description of *Colponema vietnamica* sp.n. and *Acavomonas peruviana* n. gen. n. sp., two new alveolate phyla (Colponemidia nom. nov. and Acavomonidia nom. nov.) and their contributions to reconstructing the ancestral state of alveolates and eukaryotes. *PLoS ONE* 9: e95467. doi: 10.1371/journal.pone.0095467
- Tikhonenkov DV, Strassert JFH, Janouškovec J, Mylnikov AP, Aleoshin VV, Burki F, Keeling PJ 2020. Predatory colponemids are the sister group to all other alveolates. *Molecular Phylogenetics and Evolution* 149: 106839. doi: 10.1016/j.ympev.2020.106839
- Timmis JN, Ayliffe MA, Huang CY, Martin W 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* 5: 123-135. doi: 10.1038/nrg1271
- Tomčala A, Michálek J, Schneedorferová I, Füssy Z, Gruber A, Vancová M, Oborník M 2020. Fatty Acid Biosynthesis in Chromerids. *Biomolecules* 10: 1102.
- Toso MA, Omoto CK 2007. *Gregarina niphandrodes* may lack both a plastid genome and organelle. *Journal of Eukaryotic Microbiology* 54: 66-72. doi: 10.1111/j.1550-7408.2006.00229.x
- Tovar J, Fischer A, Clark CG 1999. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Molecular Microbiology* 32: 1013-1021. doi: 10.1046/j.1365-2958.1999.01414.x
- Uddin T, McFadden GI, Goodman CD 2018. Validation of putative apicoplast-targeting drugs using a chemical supplementation assay in cultured human malaria parasites. *Antimicrobial Agents in Chemotherapy* 62.
- van Dooren GG, Hapuarachchi SV. 2017. The dark side of the chloroplast: biogenesis, metabolism and membrane biology of the apicoplast. *Advances in Botanical Research* 84: 145-185.
- van Dooren GG, Kennedy AT, McFadden GI 2012. The use and abuse of heme in apicomplexan parasites. *Antioxidand Redox Signal* 17: 634-656. doi: 10.1089/ars.2012.4539
- Waller RF, Jackson CJ 2009. Dinoflagellate mitochondrial genomes: stretching the rules of molecular biology. *Bioessays* 31: 237-245. doi: 10.1002/bies.200800164
- Waller RF, Keeling PJ 2006. Alveolate and chlorophycean mitochondrial cox2 genes split twice independently. *Gene* 383: 33-37.
- Waller RF, Kořený L. 2017. Plastid complexity in dinoflagellates: A picture of gains, losses, replacements and revisions. *Advances in Botanical Research* 84: 105-143.
- Waller RF, McFadden GI 2005. The apicoplast: a review of the derived plastid of apicomplexan parasites. *Current Issues in Molecular Biology* 7: 57-80.

- Wang Y, Morse D 2006. Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate *Lingulodinium*. *Nucleic Acids Research* 34: 613-619. doi: 10.1093/nar/gkj438
- Vojta A, Alavi M, Becker T, Hörmann F, Kuchler M, Soll J, Thomson R, Schleiff E 2004. The protein translocon of the plastid envelopes. *Journal of Biological Chemistry* 279: 21401-21405.
- Woo YH, Ansari H, Otto TD, Klinger CM, Kolísko M, Michálek J, Saxena A, Shanmugam D, Tayyrov A, Veluchamy A, Ali S, Bernal A, del Campo J, Cihlár J, Flegontov P, Gornik SG, Hajdušková E, Horák A, Janouškovec J, Katris NJ, Mast FD, Miranda-Saavedra D, Mourier T, Naeem R, Nair M, Panigrahi AK, Rawlings ND, Padron-Regalado E, Ramaprasad A, Samad N, Tomčala A, Wilkes J, Neafsey DE, Doerig C, Bowler C, Keeling PJ, Roos DS, Dacks JB, Templeton TJ, Waller RF, Lukeš J, Oborník M, Pain A 2015. Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *eLife* 4: e06974. doi: 10.7554/eLife.06974
- Yang D, Song D, Kind T, Ma Y, Hoefkens J, Fiehn O 2015. Lipidomic analysis of *Chlamydomonas reinhardtii* under nitrogen and sulfur deprivation. *PLoS ONE* 10: e0137948.
- Yoon HS, Hackett JD, Pinto G, Bhattacharya D 2002. The single, ancient origin of chromist plastids. *Journal of Phycology* 38: 40-40.
- Zhang GJ, Dong R, Lan LN, Li SF, Gao WJ, Niu HX 2020. Nuclear Integrants of Organellar DNA Contribute to Genome Structure and Evolution in Plants. *International Journal of Molecular Sciences* 21: 707. doi: 10.3390/ijms21030707
- Zhang H, Campbell DA, Sturm NR, Dungan CF, Lin S 2011. Spliced leader RNAs, mitochondrial gene frameshifts and multi-protein phylogeny expand support for the genus *Perkinsus* as a unique group of alveolates. *PLoS ONE* 6: e19933. doi: 10.1371/journal.pone.0019933
- Zhu G, Marchewka MJ, Woods KM, Upton SJ, Keithly JS 2000. Molecular analysis of a Type I fatty acid synthase in *Cryptosporidium parvum*. *Molecular and Biochemical Parasitology* 105: 253-260. doi: 10.1016/s0166-6851(99)00183-8

All supplementary data related to author's papers are available on-line at:

Fatty Acid Biosynthesis in Chromerids

<https://www.mdpi.com/2218-273X/10/8/1102>

Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites

<https://academic.oup.com/mbe/article/32/5/1115/1130018>

Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites

<https://elifesciences.org/articles/06974>

Paper I

Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasite

This is a pre-copyedited, author-produced version of an article accepted for publication in Molecular Biology and Evolution following peer review. The version of record [Flegontov P, Michálek J, Janouškovec J, et al, Divergent Mitochondrial Respiratory Chains in Phototrophic Relatives of Apicomplexan Parasites, Molecular Biology and Evolution 2015; 32 (5): 1115–1131, doi:10.1093/molbev/msv021] is available online at: <https://academic.oup.com/mbe/article/32/5/1115/1130018>

Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites

Pavel Flegontov^{1,2,◇}, Jan Michálek^{1,3,◇}, Jan Janouškovec^{4,5,†}, De-Hua Lai^{1,‡}, Milan Jirků^{1,3}, Eva Hajdušková¹, Aleš Tomčala¹, Thomas D. Otto⁶, Patrick J. Keeling^{4,5}, Arnab Pain⁷, Miroslav Oborník^{1,3,8,*} & Julius Lukeš^{1,3,5,*}

¹ Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic

² Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

³ Faculty of Sciences, University of South Bohemia, České Budějovice, Czech Republic

⁴ Department of Botany, University of British Columbia, Vancouver, Canada

⁵ Canadian Institute for Advanced Research, Toronto, Canada

⁶ Wellcome Trust Sanger Institute, Hinxton, United Kingdom

⁷ Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia.

⁸ Institute of Microbiology, Czech Academy of Sciences, Třeboň, Czech Republic.

◇ These authors contributed equally to this work.

*Correspondence to: obornik@paru.cas.cz; jula@paru.cas.cz.

† Present address: San Diego State University, San Diego, USA

‡ Present address: Center for Parasitic Organisms, School of Life Sciences, Sun Yat-Sen University, Guangzhou, People's Republic of China

Abstract

Four respiratory complexes and ATP-synthase represent central functional units in mitochondria. In some mitochondria and derived anaerobic organelles, a few or all of these respiratory complexes have been lost during evolution. We show that the respiratory chain of *Chromera velia*, a phototrophic relative of parasitic apicomplexans, lacks complexes I and III, making it a uniquely reduced aerobic mitochondrion. In *Chromera*, putative lactate:cytochrome *c* oxidoreductases are predicted to transfer

electrons from lactate to cytochrome *c*, rendering complex III unnecessary. The mitochondrial genome of *Chromera* has the smallest known protein-coding capacity of all mitochondria, encoding just *cox1* and *cox3* on heterogeneous linear molecules. In contrast, another photosynthetic relative of apicomplexans, *Vitrella brassicaformis*, retains the same set of genes as apicomplexans and dinoflagellates (*cox1*, *cox3* and *cob*).

Introduction

Although all extant mitochondria are believed to originate from a single endosymbiotic event, mitochondrial genomes have evolved remarkable diversity (Burger et al. 2003). Most mitochondria retain a genome, a remnant of an ancestral α -proteobacterial chromosome. It is usually organized on a multicopy, circular DNA molecule or a circularly permuted linear array that encodes genes for rRNAs, tRNAs, and proteins involved in the electron transport chain, ATP synthesis, and a few housekeeping functions (Burger et al. 2003). The most gene-rich mitochondrial genomes identified to date have been found in jakobids (Burger et al. 2013), while all other eukaryotes have gene complement reduced to varying extent (Gray et al. 2004). In several cases, for instance in the kinetoplastid flagellates, the mitochondrial genome has expanded into highly complex forms (Lukeš et al. 2005; Gray et al. 2010). Moreover, in several eukaryotic lineages including diplomonads, parabasalids and microsporidians, the genome was lost altogether from their highly reduced organelles (Müller et al. 2012). The most reduced mitochondrial genomes are found in the alveolates, or more specifically the myzozoans, which include apicomplexans, dinoflagellates, and their relatives. Myzozoan mitochondrial genomes encode only subunits 1 and 3 of cytochrome *c* oxidase (*cox1* and *cox3*), one subunit of cytochrome *c* reductase (*cob*), and truncated but apparently functional fragments of small (SSU) and large subunit (LSU) mitoribosomal rRNA genes (Feagin et al. 1997; Nash et al. 2008; Waller and Jackson 2009). The gene coding for *cox2*, universally present in other mitochondrial genomes, has been split and transferred to the nucleus (Waller and Keeling 2006).

In apicomplexan parasites, mitochondrial genes are encoded on a linear molecule ranging in length from 5.8 kb in *Parahaemoproteus* to 11.0 kb in *Babesia*, which can be either monomeric or is composed of circularly permuted linear arrays (Feagin et al. 1997). In dinoflagellates, the mitochondrial genome is broken into small linear fragments containing genes with extensively edited transcripts, as well as many pseudogenes and apparently non-functional gene fragments (Imanian and Keeling 2007; Jackson et al. 2007; Kamikawa et al. 2007; Nash et al. 2007; Nash et al. 2008; Kamikawa et al. 2009; Waller and Jackson 2009).

The deep-branching relatives of dinoflagellates, *Perkinsus* and *Oxyrrhis*, lack RNA editing, and in the former species translational frame shifting is required for expression of its mitochondrial genes (Slamovits et al. 2007; Masuda et al. 2010; Zhang et al. 2011). Third major group of alveolates are the ciliates, which usually harbor a gene-rich (~50 genes) mitochondrial genome (Gray et al. 2004), with their anaerobic members featuring a reduced gene set (Akhmanova et al. 1998). The nature of mitochondrial genome of *Acanthamoeba peruviana*, a heterotrophic alveolate branching on the root of myzozoans, demonstrates that the ancestral state for myzozoan mitochondria was a linear chromosome with telomeric sequences at both ends (Janouškovec et al. 2013a; Tikhonenkov et al. 2014).

Since the diversity of mitochondrial genomes of dinoflagellates and myzozoans allows reconstruction of a putative ancestral state, we wondered how informative the deep-branching relatives of apicomplexans will be in this respect. Since the discovery in 2008 of the coral-associated *Chromera velia* and the more recent description of related *Vitrella brassicaformis*, these photosynthetic algae have received much attention (Moore et al. 2008; Janouškovec et al. 2010, 2013b; Woehle et al. 2011; Burki et al. 2012; Oborník et al. 2012; Oborník and Lukeš 2013; Petersen et al. 2014). While *Chromera* and *Vitrella* have been used to solve a few riddles about apicomplexan plastid genome evolution (Janouškovec et al. 2010; Janouškovec et al. 2013b), it remained unclear how their mitochondrial genomes compare to the equally unusual apicomplexan mitochondrion, which lacks some signature elements of the canonical mitochondrial metabolism. *Chromera* has been proposed to harbor just the *cox1* gene in its mitochondrion (Petersen et al. 2014). Here we report the mitochondrial genomes and transcriptomes of *Chromera* and *Vitrella*, supplemented with the analysis of mitochondrial metabolic pathways based on the data from high-quality assemblies of their nuclear genomes (Pain A, Otto, TD, Keeling PJ, Lukeš L, unpublished data), which reveal an unexpectedly altered respiratory chain in the former alga. The reduction of the respiratory chain found in the mitochondrion of *Chromera* is unprecedented in an aerobic eukaryote.

Results

Mitochondrial genome of Chromera

We searched a high-quality whole genome assembly and transcriptome, as well as an assembly derived from a mitochondrial DNA-enriched fraction (see Materials and Methods for details) for the five mitochondrial genes expected in the myzozoan lineage, namely *cox1*, *cox3*, *cob*, genes coding for SSU and LSU rRNAs, and other typical mitochondrial genes. We identified a single conserved gene coding for *cox1*, a highly divergent *cox3* gene, and

numerous rRNA fragments. While *cox1* can be easily found by BLAST search, *cox3* is too divergent to be detected by homology-based searching algorithms. However, when the mitochondrial sequences were investigated for the presence of ORFs, one ORF with about 250 amino acids of *cox3*-like sequence was identified as a putative *cox3* (see Materials and Methods for details). This gene contains 42% hydrophobic residues and 7 predicted transmembrane segments (supplementary S1 and S2; Supplementary Material online), and has a very weak sequence similarity with the *cox3* genes when investigated with HMM searches. However, the percentage of hydrophobic residues, number of predicted transmembrane domains (6 in *cox3* of *Plasmodium*), presence of short motifs conserved in myzozoan *cox3* proteins, and similar divergence of *cox3* in dinoflagellates (see alignment in supplementary fig. S1; Supplementary Material online), all point to the fact that the gene identified in *Chromera* is *cox3*.

The *cob* gene, which was so far present in all other aerobic mitochondrial genomes, was surprisingly absent from all assemblies and raw reads from *Chromera*. Phylogenetic analysis of *cox1* shows the *Chromera* gene on a long branch with low support (bootstrap 38%, posterior probability 0.92) as a sister clade to the comparably divergent sequence from *Perkinsus*. The highly divergent *cox1* gene from *Vitrella* appears to be weakly related (bootstrap 45%, posterior probability 0.99) to its *Oxyrrhis* homolog, both branching inside the highly supported myzozoan clade (fig. S3). Analysis of total and mitochondrial DNA-enriched sequences also showed that the two protein-coding genes are found in multiple genomic contexts in the mitochondrion of *Chromera*. The full-length *cox1* genes are flanked by various sequences, and the same applies for a number of gene fragments (supplementary tables S1 and S2; Supplementary Material online), yet only a single full-length transcript is formed (fig 1A). The putative *cox3* is fused with an upstream *cox1* fragment (amino acids 1-192; supplementary table S2) and, as shown by qPCR, a fusion transcript is produced, which apparently undergoes a subsequent cleavage (fig. 1A; supplementary fig. S4; Supplementary Material online).

The mitochondrial genome of *Vitrella* contains a fused *cob-cox1* gene and a divergent *cox3* gene (supplementary fig. S5, Supplementary Material online). In *Chromera*, we have also identified expressed oligoadenylated rRNA gene fragments *rnl1*, *rnl8*, *rnl12*, *rns1*, *rns5*, and *rns7* (fig. B-D) previously found in myzozoans (Imanian and Keeling 2007; Slamovits et al. 2007; Nash et al. 2008; Waller and Jackson 2009). Multiple lines of evidence including electron microscopy of the mitochondrial DNA-enriched fraction (fig. 2) show that the *Chromera* mitochondrial DNA is composed of numerous, short linear molecules

(supplementary fig. S6; Supplementary Material online). No indications of RNA editing have been found in the mitochondrial-encoded transcripts of *Chromera* and *Vitrella*.

Respiratory chain of Chromera

Searching for the nuclear-encoded subunits of the respiratory chain did not identify any traces of proteins associated with complexes I (NADH:ubiquinone oxidoreductase) and III (ubiquinol: cytochrome *c* oxidoreductase). In contrast, subunits of complexes II (membrane-bound succinate dehydrogenase) and IV (cytochrome *c* oxidase), ATP synthase, and most enzymes associated with the respiratory chain in the related apicomplexans and dinoflagellates (electron-transfer flavoprotein, alternative NADH dehydrogenases, alternative oxidase, soluble fumarate reductase etc.) (Danne et al. 2013) were found in the nuclear genome of *Chromera* (fig. 3A, table 1).

Detailed reconstruction of the *Chromera* metabolism revealed that the entire respiratory chain and associated pathways are uniquely divided into two independently operating subchains (fig. 3A, table 1). The initial subchain is represented by alternative NADH dehydrogenases, which substitute for the electron transport function of the lost complex I. Electrons from alternative NADH dehydrogenases, complex II, electron-transfer flavoprotein: ubiquinone oxidoreductase and other donors (fig. 3A, table 1) are channeled to ubiquinone, which passes them to alternative oxidase, an electron sink, without any consequent proton pumping and link to the respiration whatsoever. Our model proposes that complex IV and ATP synthase operate independently (fig. 3A), with complex IV being the only complex that is able to pump protons into the mitochondrial intermembrane space and form a membrane potential. Orthologs of both L- and D-lactate: cytochrome *c* oxidoreductases were identified in the nuclear genome (see Materials and Methods for additional details). These proteins are supposed to mediate the transfer of electrons unidirectionally from lactate to cytochrome *c*, thus uniquely bypassing the missing complex III (fig. 3A). Analysis of L-lactate: cytochrome *c* oxidoreductases shows that in eukaryotes these genes have an extremely patchy phylogenetic distribution, as they are present in fungi, some dinoflagellates, *Perkinsus*, *Chromera* (2 genes), and *Vitrella* (2 genes; see alignment in supplementary fig. S7). Phylogenetic analysis of D-lactate: cytochrome *c* oxidoreductases reveals a complex evolutionary history, as the gene from *Chromera* clusters with high support along with fungal homologues in the eukaryotic clade (fungi, plants, *Ectocarpus*, amoebae and rhodophytes; see supplementary fig. S8). In addition to this clade, other eukaryotes appear in the company of various bacteria. However, none of these bacterial-derived clades contain the *Chromera*

sequence.

Regeneration of lactate from pyruvate is accomplished by a bidirectional NADH-dependent D-lactate dehydrogenase of likely ancient eukaryotic origin (fig. 4; table 2). The canonical conversion of lactate to pyruvate via a NADH-dependent L-lactate dehydrogenase is missing from *Chromera*, although the canonical enzyme is present in *Vitrella* and the apicomplexans (table 1). This suggests that the common ancestor of apicomplexans, *Chromera* and *Vitrella* possessed both enzymes. Phylogenetic analysis shows that the unique bidirectional NADH-dependent D-lactate dehydrogenase clusters with its homologs from various diatoms, the coccolithophore *Emiliania huxleyi* and the chlorarachniophyte *Bigeloviella natans*, forming a separate clade from bacterial sequences (fig. 4). This suggests that the bidirectional NADH-dependent D-lactate dehydrogenase may have been present in the ancestor of the SAR supergroup and was differentially lost from some of its descendants. It also appears that organisms using this enzyme lack the canonical NADH-dependent L-lactate dehydrogenase.

We have also performed phylogenetic analyses of the following alternative ubiquinone reductases in the respiratory chains of *Chromera* and *Vitrella* (fig. 3): alternative NADH dehydrogenase (NDH2; fig. 5), electron-transfer flavoprotein: ubiquinone oxidoreductase (ETFQO; supplementary fig. S9), glycerol-3-phosphate: ubiquinone oxidoreductase (G3PDH; fig. 6, supplementary fig. S10), dihydroorotate: ubiquinone oxidoreductase (DHODH; supplementary fig. S11), sulfide: ubiquinone oxidoreductase (SQO; supplementary fig. S12), and of another alternative cytochrome *c* reductase, galacto-1,4-lactone: cytochrome *c* oxidoreductase (G14LDH; supplementary fig. S13). Most of the investigated enzymes, particularly NDH2 (fig. 5), G3PDH (fig. 6), ETFQO (supplementary fig. S9), and G14LDH (supplementary fig. S13) show possible ancient eukaryotic origins, since in these trees eukaryotes form a distinct clade, separate from α -proteobacteria or cyanobacteria. We cannot, however, exclude some horizontal gene transfer from unspecified bacteria to an ancient eukaryote. NDH2 from *Chromera* and *Vitrella* appears within a clade composed of eukaryotes with plastid. Since hosts of primary and secondary plastids constitute this clade, we suppose these genes originated in the primary host nucleus (nucleus of the algal endosymbiont in secondary endosymbiosis); genes in this large clade likely passed through several duplication events (fig. 5). However, one additional gene from *Vitrella*, which is absent from *Chromera*, appears among apicomplexan parasites, branching specifically with coccidians in a clade supposed to originate from secondary host (exosymbiont) nucleus (fig. 5). The phylogeny of G3PDH (fig. 6, supplementary fig. S10) shows the branching of *Chromera* and *Vitrella*, as

well as *Perkinsus* in a sister position to apicomplexan parasites in frame of eukaryotes. The ETFQO tree displays similar grouping with *Chromera* and *Vitrella* genes sitting among homologs from alveolates (supplementary fig. S9). Since G14LDH is absent from apicomplexans, its sister position to kinetoplastid flagellates has, due to the limited taxon sampling, no phylogenetic relevance (supplementary fig. S13). We can only conclude that *Chromera* and *Vitrella* carry G14LDH of eukaryotic origin.

The second group of trees, in particular DHODH and SQO, show possible horizontal gene transfer from a bacterial donor. The SQO from *Chromera*, *Vitrella* and the centric diatom *Thalassiosira pseudonana* branch between Aquificales and α -proteobacteria (supplementary fig. S12). Another SQO homolog from *Vitrella*, which is absent from *Chromera*, is branching among various fungi; however, the clade also contains homologs from *Tetrahymena* and *Nannochloropsis* (supplementary fig. S12). Both *Chromera* and *Vitrella* contain two genes coding for DHODH: one is branching together with Apicomplexa and other eukaryotes with high support (supplementary fig. S11). This gene is probably of mitochondrial origin because α -proteobacteria (together with γ -proteobacteria) cluster on the root of this eukaryotic clade. The second gene appears in a sister position to distant homologs from *Prunus* and a haptophyte, all sitting nonspecifically among bacterial sequences (supplementary fig. S11).

In contrast to *Chromera*, the putative respiratory chain of *Vitrella* is similar to that found in apicomplexans. However, it also contains L- and D-lactate: cytochrome *c* oxidoreductases and NADH-dependent L-lactate dehydrogenase, whereas NADH-dependent D-lactate dehydrogenase is missing (fig. 3B). To confirm the absence of functional complex III, *Chromera* and *Vitrella* were incubated with a specific inhibitor of this respiratory complex, the fungicide azoxystrobin (Balba 2007). It should be noted that *Chromera* grows much faster than *Vitrella* and the latter alga forms large cell aggregations, which makes individual cells uncountable. For this reason, when evaluating the comparative effect of azoxystrobin on both algae, cumulative weight was used instead of growth curves to quantify the cultures. As shown in fig. 7C, growth of *Vitrella* was inhibited, whereas *Chromera* remained unaffected even by very high concentrations of the drug. Growth curves also demonstrate that cyanide (KCN), a potent inhibitor of complex IV, killed *Chromera* (fig. 7A). Similarly, exposure to salicylhydroxamic acid (SHAM), an inhibitor of alternative oxidase, resulted in growth inhibition under a prolonged incubation (fig. 7A). However, when an inhibitory effect on respiration by SHAM was investigated in a short time frame of 4 minutes, no decline of respiration was recorded (fig. 7B).

Discussion

Four respiratory complexes and the proton-driven ATP-synthase perform core mitochondrial functions that are dispensable only in anaerobic or sugar-rich environments, often as a consequence of parasitism (Müller et al. 2012). Indeed, any major modifications to the respiratory chain are generally infrequent and almost exclusively associated with anaerobes. Only a few cases of its reduction have been encountered in aerobic eukaryotes. The respiratory complex I is missing from the mitochondrion of *Plasmodium* spp. and other apicomplexans (van Dooren et al. 2006; Vaidya and Mather 2009; Sheiner et al. 2013), as well as from some fungi (e.g. *Saccharomyces*) where alternative NADH dehydrogenases substitute its electron transfer function (Marcet-Houben et al. 2009). A pattern common to many eukaryotic anaerobes is the concurrent loss of complexes III and IV, while complexes I, II, and V have been retained, yet act in a different way: complex I donates electrons to rhodoquinone, which passes them to a membrane-bound fumarate reductase, whereas complex II-like enzymes operate in reverse, producing succinate (Müller et al. 2012). This type of respiratory chain occurs in anaerobic stages of some parasitic metazoans (a fluke *Fasciola* and a roundworm *Ascaris*), in marine benthic species experiencing periods of severe anoxia (a mussel *Mytilus* and an annelid *Arenicola*) and in an excavate alga *Euglena gracilis* grown in anaerobic conditions (Müller et al. 2012). Complexes III and IV were lost secondarily from the plant-pathogenic kinetoplastid *Phytomonas*, likely as a result of a deletion in the mitochondrial genome and subsequent adaptation to the sugar-rich environment of plant sap (Porcel et al. 2014). A similar pattern involving the loss of complexes III, IV, and V is found in the ciliate *Nyctotherus* and the stramenopile *Blastocystis* (Müller et al. 2012). In its bloodstream stage, the kinetoplastid *Trypanosoma brucei* shuts down its respiratory chain, activates glycerol-3-phosphate: ubiquinone dehydrogenase and an alternative oxidase, and to generate proton gradient switches its complex V into the reverse (Tielens and van Hellemond 2009). A 'petite' mutant of *T. brucei* termed *T. b. evansi* lost all complexes containing the mitochondrial-encoded subunits except for complex V, where mutations in the nuclear-encoded subunits and an ATP/ADP transporter compensate for the loss of its mitochondrial-encoded subunit *atp6* (Dean et al. 2013). Last but not least, numerous anaerobic (e.g. *Entamoeba*, *Giardia*, *Mikrocytos*, and *Trichomonas*) and facultative anaerobic parasitic protists (e.g. *Cryptosporidium* and microsporidia) retained mitochondrion-derived hydrogenosomes and mitosomes, from which the respiratory chain has been lost altogether (Müller et al. 2012; Burki et al. 2013).

It should be noted that, except for *Chromera*, complex III is always lost along with complex IV. Moreover, reductions of the respiratory chain occur almost exclusively in anaerobic organisms, of which the most studied are parasites, yet *Chromera* is a fully aerobic, free-living phototroph, which can live in a simple inorganic medium just in the presence of light (Moore et al. 2008; Oborník et al. 2009). Using a selective inhibitor of complex III, we provide experimental evidence that this complex is indeed absent from *Chromera* but active in *Vitrella*. Moreover, since the SHAM treatment had no effect on the oxygen consumption of *Chromera*, the alternative oxidase is, in agreement with our hypothesis, disconnected from the respiratory function. However, the inhibitory effect of a long exposure to SHAM may be a consequence of increased production of radical oxygen species, since a blocked alternative oxidase cannot function as a sink for electrons generated by complex II, which may consequently cause damage to the mitochondrial membrane.

Although *Chromera* and *Vitrella* grow under aerobic conditions, they also possess an extensive set of enzymes characteristic for anaerobic protists (table 1 and fig. 3): a [FeFe]-hydrogenase with a complete set of three hydrogenase maturases, pyruvate: NADP oxidoreductase, pyruvate-formate lyase, NADH oxidase, acetaldehyde/alcohol dehydrogenase, acetate: succinate CoA-transferase, and acetyl-CoA synthetase (ADP-forming), to name the most prominent ones. These enzymes are absent in nearly all major lineages of the parasitic apicomplexans, with the exception of the mitosome-bearing human parasite *Cryptosporidium*. This suggests that the life history of *Chromera* may include previously unknown periods under anaerobic conditions or, more importantly, that the common ancestor of apicomplexans was substantially more metabolically diverse than are nearly all of its extant descendants. Such versatility of the mitochondrion has recently emerged in the aerobes *Naegleria* and *Chlamydomonas*, which contain a large set of 'anaerobic' enzymes (Müller et al. 2012). Dinoflagellates and perkinsids (Danne et al. 2013), and now even more prominently *Chromera* and *Vitrella*, are joining this club. In fact, *Chromera* qualifies as the most trophically versatile protist, capable of phototrophy, aerobic and strictly anaerobic osmotrophy, and maybe even predation (Oborník et al. 2012; Oborník and Lukeš 2013).

When the respiratory chains of *Chromera* and *Vitrella* are seen in an evolutionary context (fig. 8), it is obvious that the evolution of mitochondrial genomes of alveolates is reductive. We suppose that the mitochondrial genome of an ancestral alveolate underwent linearization, and at a later stage was reduced in all myzozoan lineages through massive loss of protein-coding genes and rRNA fragmentation. In *Chromera*, *Vitrella* and dinoflagellates,

the genome has entered a fragmented and scrambled state due to extensive recombination, but in most apicomplexans it retained its composition from monomeric linear molecules.

Chromera displays a uniquely reduced mitochondrial genome and possesses a respiratory chain interrupted in an unprecedented fashion. Phylogeny of L and D lactate cytochrome *c* oxidoreductases likely donating electrons to complex IV shows that these enzymes are very probably eukaryotic in origin. We can see the same pattern in the evolution of most of alternative ubiquinone reductases (NDH2, ETFQO, G3PDH, and G14LDH). One of the two DHODHs and bacterial SQO in *Chromera* and *Vitrella* are the only genes likely of mitochondrial origins among the detected alternative ubiquinone reductases. We have found only a single candidate gene for horizontal gene transfer from bacteria to the *Chromera* and *Vitrella* nuclear genome. This implies that the entire unusual enzymatic equipment found in *Chromera* was already present in an ancient eukaryote.

Materials and Methods

The *Chromera* genome was sequenced by combining the following reads (supplementary table S3, Supplementary Material online): single 454 reads (average length ~300 nt), paired-end Illumina MiSeq reads (filtered and trimmed reads <250 nt in length), mate pair Illumina HiSeq reads (filtered and trimmed reads <100 nt in length, insert size 1.7-5.5 kb), and additional paired-end Illumina reads for a fraction enriched for mitochondrial DNA by centrifugation in a CsCl-Hoechst 33258 density gradient (supplementary fig. S14, Supplementary Material online). All the reads were assembled simultaneously using GS De Novo Assembler (Newbler) v.2.9 (see assembly statistics in supplementary table S4, Supplementary Material online), and mitochondrial contigs were identified using BLAST with the apicomplexan *cox1*, *cox3* and *cob* proteins and rRNA fragments. The Newbler assembler defines contigs as unambiguous parts of a read overlap graph, hence some contigs can be shorter than the reads that make them, with reads flowing from one contig into another (reads can also flow between the ends of long contigs). Therefore, extension of initial gene- or gene fragment-containing contigs was possible by following alternative paths in a graph of direct contig links produced by Newbler (see a detailed example of this approach in supplementary fig. S15, Supplementary Material online). Mitochondrial origin and integrity of the resulting ‘supercontigs’ was verified by mapping on them reads of the mitochondrial DNA-enriched fraction (see an example in supplementary fig. S16, Supplementary Material online) and mate pair reads with long insert size (data not shown), respectively.

The scrambled structure of the *Chromera* mitochondrial genome revealed by 454 and Illumina sequencing was further investigated by PCR and Sanger sequencing (supplementary fig. S17, Supplementary Material online). *Vitrella* mitochondrial contigs were assembled from Illumina HiSeq paired-end and mate pair reads (filtered and trimmed reads <100 nt in length) with Velvet and CLC Genomics Workbench v. 6.5 (see contig sequences in supplementary fig. S5, Supplementary Material online). Transcription of *Chromera* and *Vitrella* mitochondrial contigs was assessed by mapping strand-specific transcriptome reads of poly-A fraction on them (supplementary table S3, Supplementary Material online). The list of transcripts identified in the extended mitochondrial contigs of *Chromera* is shown in supplementary table S2 (Supplementary Material online).

Components of the respiratory chain and related enzymes in *Chromera*, *Vitrella*, and in the reference species set (table 1) were identified using at least one hidden Markov model (HMM) per protein, either based on MUSCLE alignments of annotated proteins in alveolates, or taken from the Pfam database or from NCBI CDD. HMM searches were performed with the HMMER3 software using protein-specific E-value cut-offs, depending on protein length and the need to distinguish closely related protein families. Sources of HMM models and E-value cut-offs for the enzymes most discussed in the paper are shown in table 3. NADH-dependent L-lactate and malate dehydrogenases were distinguished using characteristic residues.

Multiple alignments of *cox1*, L- and D -lactate: cytochrome *c* oxidoreductases, galacto-1,4-lactone: cytochrome *c* oxidoreductase, NADH-dependent D-lactate dehydrogenase, alternative NADH dehydrogenases, electron-transfer flavoprotein: ubiquinone oxidoreductase, glycerol-3-phosphate: ubiquinone oxidoreductase, sulfide: ubiquinone oxidoreductase, and dihydroorotate: ubiquinone oxidoreductase were performed using MUSCLE (Edgar 2004). Alignments were edited, gaps and ambiguously aligned regions were excluded from further analyses. Trees were computed using Maximum Likelihood, ML (RAxML; Stamatakis 2014), and Bayesian inference (MrBayes 3.2; Ronquist et al. 2012). ML trees were constructed under the LG model (all proteins except *cox1*), or cpREV model (*cox1*), according to ProtTest (Darriba et al. 2011). MrBayes was used to assess Bayesian topologies and posterior probabilities (PP) using two independent Monte-Carlo Markov chains run under the default settings for 2 million generations. We used 500,000 generations as a burn-in and omitted them from topology reconstruction and PP calculation. Bayesian trees were computed using the WAG model. Putative localizations of enzymes of interest

(table 2) were predicted with TargetP (Emanuelsson et al. 2000), SignalP (Petersen et al. 2011) and MitoProt (Claros and Vincens 1996) programs.

The mitochondrial DNA-enriched sample was prepared for electron microscopy by the cytochrome c method (Ferguson and Davis, 1978). Plasmid pGEM served as a standard. The samples were examined with a Philips CM12 electron microscope at 80 kV, with precise magnification determined using replica grating.

Azoxystrobin (Sigma) was dissolved in DMSO (stock solution 10 mg/ml), and a series of five concentrations in the *f/2* medium ranging from 10 to 2,000 ng/ml was prepared. Each concentration and negative controls were assayed in three replicates. Cultivation flasks containing 20 ml of medium were inoculated with 0.5 ml of starting culture. Three replicates of *Chromera* culture were observed every other day for 2 weeks by measuring absorbance changes (OD₆₀₀, fig. 3A) three times. Finally, after 4 weeks the resulting biomass from equal volume of all five replicates was harvested, dried and weighted (fig. 3C). Due to extremely tight clustering in clumps, *Vitrella* is suitable for neither absorbance measurement, nor cell counting. Also, in contrast to *Chromera*, *Vitrella* culture grows much slower. Therefore, only total biomass after 9 weeks was determined (fig. 3C).

For measurement of respiration in intact cells, exponentially growing *C. velia* culture was harvested by centrifugation, cells were washed and resuspended in the *f/2* medium at a concentration of approximately 3×10^8 cells/ml. Oxygen consumption at 26°C was determined using a Clarktype electrode (Theta 90, Czech Republic) in the dark. After 4 min stabilization, salicylhydroxamic acid (SHAM) was added to final concentrations of 3 and 18 mM (fig. 7B). Cells were incubated for 2 - 4 min with each concentration of the drug and O₂ consumption was then converted to micromol/min. Respiration with no inhibitors was taken as 100%.

Sequences reported in this paper have been deposited in the NCBI BioProject (Accession No. PRJEB6670, PRJEB667) database, in Short Read Archive (Accession No. ERR558149, ERR558150, ERR558151, ERR558152, ERR558194, ERR558195, ERR558196, ERR558197, ERR558198, ERR558199, ERR558200, ERR558201, ERR558202, ERR571482, ERR571483, ERR571484, ERR571485), and in NCBI Genbank (see supplementary table S1, Supplementary Material online). The annotated nuclear genomes of *C. velia* and *V. brassicaformis* have been deposited in EuPathDB (eupathdb.org).

Acknowledgements

We thank the KAUST Bioscience Core Laboratory personnel for sequencing Illumina

libraries used in this project, Evgeny S. Gerasimov (Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow) and Martin Kolísko (University of British Columbia, Vancouver) for help with sequence analysis, Oldřich Benada (Institute of Microbiology, Prague) for help with electron microscopy, Anton Horváth (Comenius University, Bratislava), Dave Speijer (University of Amsterdam, Amsterdam) and Ivan Hrdý (Charles University, Prague) for useful comments. This work was supported by Czech Science Foundation grants P506/12/1522, 13-33039S (support for JM) and P501/12/G055 (support for AT) to M.O., project Algatech (CZ.1.05/2.1.00/03.0110) to M.O., FP7 agreement 316304 to M.O. and J.L., the KAUST award FIC/2010/09 to A.P., M.O. and J. L., and by a grant from the Canadian Institutes for Health Research to P. J. K. and J. L. is also supported by the Praemium Academiae, and P. F. by a grant of the Moravian-Silesian region (project MSK2013-DT1) and by Russian Foundation for Basic Research (project 14-04-31936). P. J. K. and J. L. are Senior Fellows and J. J. is a Global Scholar at the Canadian Institute for Advanced Research, and P. J. K. was supported by a Fellowship from the John Simon Guggenheim Foundation.

References

- Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampl V, et al. 2012. The revised classification of eukaryotes. *J Eukaryot Microbiol.* 59:429-493.
- Akhmanova A, Voncken F, van Alen T, van Hock A, Boxma B, Vogels G, Veenhuist M, Hackstein JHP. 1998. A hydrogenosome with a genome. *Nature* 396:527-528.
- Balba H. 2007. Review of strobilurin fungicide chemicals. *J Environ Sci Health B* 42:441-451.
- Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: anything goes. *Trends Genet.* 12:709-716.
- Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. *Genome Biol Evol.* 5:418-438.
- Burki F, Flegontov P, Oborník M, Cihlář J, Pain A, Lukeš J, Keeling PJ. 2012. Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biol Evol.* 4:626-635.
- Burki F, Corradi N, Sierra R, Pawlowski J, Meyer GR, Abbott CL, Keeling PJ. 2013.

- Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in rhizaria. *Curr Biol.* 23:1541-1547.
- Claros MG, Vincens P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem.* 241:779-786.
- Danne JC, Gornik SG, Macrae JI, McConville MJ, Waller RF. 2013. Alveolate mitochondrial metabolic evolution: dinoflagellates force reassessment of the role of parasitism as a driver of change in apicomplexans. *Mol Biol Evol.* 30:123-139.
- Dean S, Gould MK, Dewar CE, Schnauffer AC. 2013. Single point mutations in ATP synthase compensate for mitochondrial genome loss in trypanosomes. *Proc Natl Acad Sci U S A.* 110:14741-14746.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792-1797.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* 300:1005-1016.
- Feagin JE, Mericle BL, Werner E, Morris M. 1997. Identification of additional components of the fragmented rRNAs of the two *Plasmodium falciparum* 6 kb element. *Nucleic Acids Res.* 25:438-446.
- Ferguson J, Davis RW. 1978. A new electron microscopic technique for establishing the positions of genes: an analysis of the yeast ribosomal RNA coding region. *J Mol Biol.* 123:417-430.
- Ginger ML, Fritz-Laylin LK, Fulton C, Cande WZ, Dawson SC. 2010. Intermediary metabolism in protists: a sequence-based view of facultative anaerobic metabolism in evolutionarily diverse eukaryotes. *Protist* 161:642-671.
- Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. *Annu Rev Genet.* 38:477-524.
- Gray MW, Lukeš J, Archibald JM, Keeling PJ, Doolittle WF. 2010. Cell biology. Irremediable complexity? *Science* 330:920-921.
- Imanian B, Keeling PJ. 2007. The dinoflagellates *Durinskia baltica* and *Kryptoperidinium foliaceum* retain functionally overlapping mitochondria from two evolutionarily distinct lineages. *BMC Evol Biol.* 7:172.
- Jackson CJ, Norman JE, Schnare MN, Gray MW, Keeling PJ, Waller RF. 2007. Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. *BMC Biol.* 5:41.
- 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.

- Janouškovec J, Horák A, Barott KL, Rohwer FL, Keeling PJ. 2012. Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. *Curr Biol.* 22:R518-R519.
- Janouškovec J, Tikhonenkov DV, Mikhailov KV, Simdyanov TG, Aleoshin VV, Mylnikov AP, Keeling PJ. 2013a. Colponemids represent multiple ancient alveolate lineages. *Curr Biol.* 23:2546-2552.
- Janouškovec J, Sobotka R, Lai DH, Flegontov P, Koník P, Komenda J, Ali S, Prášil O, Pain A, Oborník M, et al. 2013b. Split photosystem protein, linear-mapping topology, and growth of structural complexity in the plastid genome of *Chromera velia*. *Mol Biol Evol.* 30:2447-2462.
- Kamikawa R, Inagaki Y, Sako Y. 2007. Fragmentation of mitochondrial large subunit rRNA in the dinoflagellate *Alexandrium catenella* and the evolution of rRNA structure in alveolate mitochondria. *Protist* 158:239-245.
- Kamikawa R, Nishimura H, Sako Y. 2009. Analysis of the mitochondrial genome, transcripts, and electron transport activity in the dinoflagellate *Alexandrium catenella* (Gonyaulacales, Dinophyceae). *Phycol Res.* 57:1-11.
- Lukeš J, Hashimi H, Zíková A. Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. 2005. *Curr Genet.* 48:277-299.
- Masuda I, Matsuzaki M, Kita K. 2010. Extensive frameshift at all AGG and CCC codons in the mitochondrial cytochrome c oxidase subunit 1 gene of *Perkinsus marinus* (Alveolata; Dinoflagellata). *Nucleic Acids Res.* 38:6186-6194.
- Marcet-Houben M, Marceddu G, Gabaldon T. 2009. Phylogenomics of the oxidative phosphorylation in fungi reveals extensive gene duplication followed by functional divergence. *BMC Evol Biol.* 9:295.
- Moore RB, Oborník M, Janouškovec J, Chrudimský T, Vancová M, Green DH, Wright SW, Davies NW, Bolch CJ, Heimann K, et al. 2008. A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* 451:959-963.
- Müller M, Mentel M, van Hellemond JJ, Henze K, Woehle C, Gould SB, Yu RY, van der Giezen M, Tielens AG, Martin WF. 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol Mol Biol Rev.* 76:444-495.
- Nash EA, Barbrook AC, Edwards-Stuart RK, Bernhardt K, Howe CJ, Nisbet ER. 2007. Organization of the mitochondrial genome in the dinoflagellate *Amphidinium carterae*. *Mol Biol Evol.* 24:1528-1536.
- Nash EA, Nisbet RER, Barbrook AC, Howe CJ. 2008. Dinoflagellates: a mitochondrial

- genome all at sea. *Trends Genet.* 24:328-335.
- Oborník M, Modrý D, Lukeš M, Černotíková-Stříbrná E, Cihlár 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, Lukeš J. 2013. Cell biology of chromerids: autotrophic relatives to apicomplexan parasites. *Int Rev Cell Mol Biol.* 306:333-369.
- Oborník M, Janouškovec J, Chrudimský T, Lukeš J. 2009. Evolution of the apicoplast and its hosts: from heterotrophy to autotrophy and back again. *Int J Parasitol.* 39:1-12.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785-786.
- Petersen J, Ludewig AK, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H. 2014. *Chromera velia*, endosymbioses and the rhodoplex hypothesis-plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biol Evol.* 6:666-684.
- Porcel BM, Denoëud F, Opperdoes F, Noel B, Madoui MA, Hammarton TC, Field MC, Da Silva C, Couloux A, Poulain J, et al. 2014. The streamlined genome of *Phytomonas* spp. relative to human pathogenic kinetoplastids reveals a parasite tailored for plants. *PLoS Genet.* 10:e100400.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61:539-542.
- Sheiner L, Vaidya AB, McFadden GI. 2013. The metabolic roles of the endosymbiotic organelles of *Toxoplasma* and *Plasmodium* spp. *Curr Opin Microbiol.* 16:452-458.
- Slamovits CH, Saldarriaga JF, Larocque A, Keeling PJ. 2007. The highly reduced and fragmented mitochondrial genome of the early-branching dinoflagellate *Oxyrrhis marina* shares characteristics with both apicomplexan and dinoflagellate mitochondrial genomes. *J Mol Biol.* 372:356-368.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313.
- Tielens AG, van Hellemond JJ. 2009. Surprising variety in energy metabolism within Trypanosomatidae. *Trends Parasitol.* 25:482-490.
- Tikhonenkov DV, Janouškovec J, Mylnikov AP, Mikhailov KV, Simdyanov G, Aleoshin VV, Keeling PJ. 2014. Description of *Colponema vietnamica* sp.n. and *Acavomonas peruviana*

- n. gen. n. sp., two new alveolate phyla (Colponemidia nom. nov. and Acavomonidia nom. nov.) and their contributions to reconstructing the ancestral state of alveolates and eukaryotes. *PLoS One* 9:e95467.
- Vaidya AB, Mather MW. 2009. Mitochondrial evolution and functions in malaria parasites. *Annu Rev Microbiol.* 63:249-267.
- Van Dooren GG, Stimmler LH, McFadden GI. 2006. Metabolic maps and functions of the Plasmodium mitochondrion. *FEMS Microbiol Rev.* 30:596-630.
- Waller RF, Jackson CJ. 2009. Dinoflagellate mitochondrial genomes: stretching the rules of molecular biology. *Bioessays* 31:237-245.
- Waller RF, Keeling PJ. 2006. Alveolate and chlorophycean mitochondrial *cox2* genes split twice independently. *Gene* 383:33-37.
- 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 Evol.* 3:1220-1230.
- Yurchenko V, Hobza R, Benada O, Lukeš J. 1999. *Trypanosoma avium*: large minicircles in the kinetoplast DNA. *Exp Parasitol.* 92:215-218.
- Zhang H, Campbell DA, Sturm NR, Dungan CF, Lin S. 2011. Spliced leader RNAs, mitochondrial gene frameshifts and multi-protein phylogeny expand support for the genus *Perkinsus* as a unique group of alveolates. *PLoS One* 6:e19933.

Figure legends

Fig. 1. Transcription pattern of *Chromera velia* nuclear and mitochondrial genes or gene fragments revealed by Northern blot analysis of total RNA. RNA was detected by γ -ATP radiolabeled antisense oligonucleotides. In panels A-C, Ethidium bromide-stained gels with a marker (M) and total RNA (T) are shown to the left. (A) Transcripts of a nuclear GAPDH (glyceraldehyde-3-phosphate dehydrogenase) gene, the *cox1* and *cox3* genes (mature transcripts), primary transcript of *cox3* containing a part of *cox1* (*cox1-3*), and regions flanking the *cox1* gene (clones C and L). (B) Transcripts of the mitochondrial rRNA gene fragments (*rnl1*, *rnl8*, *rnl12*, and *rns1*) resolved on a high resolution 5% polyacrylamide gel. (C) Transcription pattern of rRNA gene fragments *rnl8* (detected with oligonucleotides *rnl8A*, A2 and A3) and *rns1* (detected with oligonucleotides *rns1A* and A2), as revealed in a medium resolution 1% agarose gel. (D) Positions of labeled oligonucleotides are marked with bars; polyadenylation sites, indicated as -AAAA, were determined by 3' polyA RACE.

Fig. 2. Electron microscopy of the *Chromera velia* mitochondrial DNA prepared from fractions H1 and H2 obtained from ultracentrifugation of total DNA in the CsCl-Hoechst 33258 gradient (see supplementary fig. S14) and examined in a transmission electron microscope as described previously (Yurchenko et al. 1999).

Fig. 3. The reconstructed respiratory chains in (A) *Chromera velia* and (B) *Vitrella brassicaformis*. Protein localization is shown based on predicted targeting and on available data. Proteins with alternative or uncertain localization are indicated in gray with a question mark. In *Chromera* the respiratory chain is split into two disconnected parts. Complex II, alternative NADH dehydrogenase (NDH2), glycerol 3-phosphate dehydrogenase (G3PDH), electron transfer flavoprotein:ubiquinone oxidoreductase (ETFQO) and other enzymes are donating electrons to ubiquinone, which passes them to alternative oxidase (AOX). D-lactate:cytochrome c oxidoreductase (D-LDH), L-lactate:cytochrome c oxidoreductase (L-LDH, also known as cytochrome b₂), and L-galactono-1,4-lactone dehydrogenase (G14LDH) are donating electrons to soluble cytochrome c, which is oxidized by cytochrome c oxidase (complex IV). The NADH-dependent D-lactate dehydrogenase (D-LDH2) recycles pyruvate, generated by the cytochrome-dependent D-LDH and L-LDH, into lactate, and consumes NADH. The respiratory chain of *Vitrella* is continuous, with alternative ubiquinone reductases, D-LDH, L-LDH, and G14LDH also present. However, D-LDH2 is absent and AOX lacks a robust mitochondrial import signal. The complexes are not depicted to scale.

Fig. 4. A maximum likelihood phylogenetic tree including *Chromera velia* NADH-dependent D-lactate dehydrogenase (D-LDH2), with its eukaryotic and bacterial orthologs. Bayesian PP/ML bootstraps are in the table. Accession number of sequences in the tree: Fracy1_207885; Psemu1_183545; Phatr2_46664; Thaps3_268453; EMIHUDRAFT_450888; Bigna1_67620 (sequences downloaded from JGI <http://jgi.doe.gov/>); WP_012178639.1; WP_008193115.1; WP_033467713.1; WP_028227993.1; WP_003774459.1; CED58655.1; WP_011218908.1; WP_022589137.1; KGK78572.1; WP_012195855.1; WP_004902775.1; WP_029486096.1; WP_014082344.1; WP_013349576.1; WP_028245171.1; WP_019617597.1 (sequences downloaded from NCBI <http://www.ncbi.nlm.nih.gov/>)

Fig. 5. Bayesian phylogenetic tree as inferred from alternative NADH dehydrogenase amino

acid sequences. Proteins from *Chromera* and *Vitrella*, which are supposed to be involved in the respiratory chain are marked by *. Numbers above branches indicate Bayesian PP/ML bootstrap support (1000 replicates).

Fig. 6. A part of the Bayesian phylogenetic tree as inferred from glycerol-3-phosphate dehydrogenase amino acid sequences. (see full tree in supplementary fig. S10). Numbers above branches indicate Bayesian PP/ML bootstrap support (1000 replicates).

Fig. 7. Growth of *Chromera velia* and *Vitrella brassicaformis* in the presence of specific inhibitors of selected mitochondrial proteins. (A) Growth curves of *Chromera* cultivated in the presence of salicylhydroxamic acid (SHAM), potassium cyanide (KCN), and azoxystrobin. KCN, an inhibitor of cytochrome c oxidase (complex IV), slows down the growth. Lack of inhibition by azoxystrobin supports the absence of cytochrome c reductase (complex III). (B) SHAM, an inhibitor of alternative oxidase (AOX), is efficiently inhibiting the growth of *Chromera* at concentrations that did not affect respiration in a short-term treatment. (C) Graphs showing resulting biomass after growth curve experiments. *Chromera* was harvested after reaching the stationary growth phase at 4 weeks. Due to much slower growth, *Vitrella* was harvested after reaching the stationary growth phase at 9 weeks. Biomass for each concentration was calculated as average of five repetitions (standard deviations are shown). High concentration of DMSO in the media slightly decreases the final biomass.

Fig. 8. Overview of the mitochondrial evolution in alveolates.

Table 1. Respiratory chain components and other enzymes typical for anaerobic species in *Chromera* and *Vitrella*, in apicomplexan, dinoflagellate, and perkinsid datasets. Proteins typical for anaerobic species (Ginger et al. 2010; Müller et al. 2012) are marked with a star. A number of paralogs detected in each genome is shown, and presence/absence only is shown for transcriptomic or proteomic data (presence is denoted by a filled circle). If proteins show strong evidence of lateral gene transfer from bacteria (data not shown), the respective cells are boxed. Commonly used abbreviations and EC numbers are shown, when applicable. Phylogenetic patterns of protein occurrence are color-coded in the following way: orange, proteins lost in *Chromera* only; light-green, proteins occurring exclusively in *Chromera* (proteins found in *Chromera* and in dinoflagellates/perkinsids are included into this group as well, as transcriptomic data cannot robustly demonstrate absence of a certain protein, the same applies to all groups below); blue-green, proteins occurring in both *Chromera* and *Vitrella*; gray, in *Chromera*, *Vitrella*, and in *Cryptosporidium*, a mitosome-bearing apicomplexan; violet, in *Chromera*, *Vitrella*, and apicomplexans (but lost in at least 2 of 5 major apicomplexan lineages); pink, in dinoflagellates/perkinsids only.

name	common abbreviation	EC number	respiratory chain complex or metabolic process	chromerids						Apicomplexa				Perkinsella marina	Oxyrrhis marina	Dinophyceae: Gonyaulacales
				Chromera proteome	Chromera transcriptome	Chromera genome	Vitrella transcriptome	Vitrella genome	Eimeria	Cryptosporidium	Neospora /Toxoplasma	Babesia /Theileria	Plasmodium			
NADH dehydrogenase																
mitochondrial alternative NADH dehydrogenase																
electron transfer flavoprotein																
electron transfer flavoprotein:ubiquinone oxidoreductase																
glycerol 3-phosphate:ubiquinone oxidoreductase																
sulfide:ubiquinone oxidoreductase *																
dihydroorotate:ubiquinone oxidoreductase																
ubiquinone biosynthesis enzymes																
succinate dehydrogenase																
soluble fumarate reductase *																
alternative oxidase *																
cytochrome <i>b</i>																
cytochrome <i>c</i>																
mitochondrial Rieske protein																
14 kDa protein																
hinge subunit																
cytochrome <i>c</i> 1 heme lyase																
L-lactate dehydrogenase (cytochrome) (cytochrome <i>b</i> 2)																
D-lactate dehydrogenase (cytochrome)																
L-galactonolactone dehydrogenase																
glucose-methanol-choline oxidoreductase																
cytochrome <i>c</i>																
cytochrome <i>c</i> oxidase																
ATP synthase																
glycolytic enzymes																
pyruvate:orthophosphate dikinase *																
2-oxoglutarate dehydrogenase																
aconitase																
citrate synthase *																
fumarase *																
isocitrate dehydrogenase (NAD)																
isocitrate dehydrogenase (NADP)																
malate dehydrogenase *																
succinyl-CoA synthetase (succinate-CoA ligase, GDP or ADP-forming) *																
phosphoenolpyruvate carboxykinase (ATP) *																
phosphoenolpyruvate carboxykinase (GTP) *																
pyruvate carboxylase																
malic enzyme *																
D-lactate dehydrogenase *																
L-lactate dehydrogenase *																
pyruvate decarboxylase *																
acetaldehyde/alcohol dehydrogenase *																
alcohol dehydrogenase, NAD-dependent *																
pyruvate dehydrogenase complex *																
pyruvate:NADP oxidoreductase *																
pyruvate:formate lyase *																
pyruvate:formate lyase activating enzyme *																
acetate:succinate CoA-transferase *																
acetyl-CoA synthetase, ADP forming (acetate-CoA ligase) *																
acetyl-CoA synthetase, AMP forming (acetate-CoA ligase) *																
phosphate acetyl transferase (phosphotransacetylase) *																
acetate kinase *																
propionate:succinate CoA-transferase *																
propionyl-CoA carboxylase *																
[FeFe] hydrogenase maturase																
[FeFe] hydrogenase maturase																
[FeFe] hydrogenase maturase																
[FeFe] hydrogenase *																
nuclear prelamin A recognition factor-like protein																
coenzyme F420 hydrogenase																
alpha-ketoacyl reductase *																
alpha-ketoacyl synthase *																
NAD(P)H:ferredoxin oxidoreductase *																
flavoiron protein *																
NADH oxidase																
nitrite reductase *																
sulphite reductase																
alanine racemase *																
alanine aminotransferase *																
aspartate aminotransferase *																
glutamate dehydrogenase *																
opine dehydrogenases *																

Table 2. Putative localizations of enzymes of interest predicted with the TargetP, SignalP and MitroProt programs. S- ER signal peptide; C- chloroplast transit peptide; M – mitochondrial transit peptide; Anch – signal anchor (transmembrane domain).

gene ID	enzyme name	signal peptide	length, aa	transit peptide (uncut signal peptide)	length, aa	transit peptide (cut signal peptide)	length, aa	MitroProt II
Cvel_24584.t1	NADH-dependent D-lactate dehydrogenase	S 0.986	18	S 0.149	18	M 0.525	19-48	0
Vbra_22711.t1	NADH-dependent L-lactate dehydrogenase	0	0	0	0	0	0	0
Vbra_4978.t1	NADH-dependent L-lactate dehydrogenase	0	0	C 0.292 / M 0.225	5	0	0	0
Vbra_16839.t1	NADH-dependent L-lactate dehydrogenase	0	0	0	0	0	0	0
Cvel_6082.t1	D-lactate: cytochrome <i>c</i> oxidoreductase	S 0.941	17	C 0.561	13	M 0.815	18-24	0.8995
Cvel_20317.t1	D-lactate: cytochrome <i>c</i> oxidoreductase	Anch 0.587	46	C 0.759 / M 0.706	52	C 0.759 / M 0.706	1-52	0.9959
Vbra_3212.t1	D-lactate: cytochrome <i>c</i> oxidoreductase	S 0.315	21	M 0.796	93	M 0.796	1-93	0.9946
Vbra_9909.t1	D-lactate: cytochrome <i>c</i> oxidoreductase	0	0	M 0.806	18	0	0	0.8978
Cvel_4425.t1	L-lactate: cytochrome <i>c</i> oxidoreductase	0	0	0	0	0	0	0
Cvel_5763.t1	L-lactate: cytochrome <i>c</i> oxidoreductase	0	0	0	0	0	0	0
Vbra_7868.t1	L-lactate: cytochrome <i>c</i> oxidoreductase	0	0	0	0	0	0	0
Vbra_19079.t1	L-lactate: cytochrome <i>c</i> oxidoreductase	0	0	0	0	0	0	0
Vbra_2234.t1	galactose-1-4-lactone: cytochrome <i>c</i> oxidoreductase	S 0.993	21	M 0.595	55	M 0.900	22-56	0.9863
Cvel_9992.t1	galactose-1-4-lactone: cytochrome <i>c</i> oxidoreductase	S 0.999	27	S 0.794	27	M 0.738	24-54	0
Cvel_6529.t1	alternative oxidase	0	0	M 0.913	35	0	0	0.9747
Cvel_21608.t1	alternative oxidase	S 0.575	34	M 0.849	29	0	0	0.9918
Vbra_7072.t1	alternative oxidase	S 0.635	30	M 0.782	38	0	0	0.966
Vbra_14859.t1	alternative oxidase	S 0.843	23	M 0.578	57	C 0.588 / M 0.602	24-57	0.9786
Vbra_15987.t1	alternative oxidase	S 0.848	35	M 0.544	45	0	0	0.9836
Cvel_3201.t1	alternative NADH dehydrogenase	0	0	0	0	0	0	0
Cvel_12889.t1	alternative NADH dehydrogenase	0	0	0	0	0	0	0
Cvel_17721.t1	alternative NADH dehydrogenase	S 1.000	18	M 0.411	41	M 0.824	19-41	0.951
Vbra_528.t1	alternative NADH dehydrogenase	0	0	0	0	0	0	0
Vbra_4873.t1	alternative NADH dehydrogenase	S 0.998	24	C 0.189 / M 0.134 / S 0.153	20	C 0.795	25-75	0.9844
Vbra_8108.t1	alternative NADH dehydrogenase	0	0	0	0	0	0	0
Vbra_13793.t1	alternative NADH dehydrogenase	S 0.997	25	S 0.608	19	C 0.850	20-62	0.709
Vbra_21650.t1	alternative NADH dehydrogenase	S 1.000	24	S 0.869	23	C 0.965	24-54	0.6591
Vbra_21162.t1	glycerol 3-phosphate: ubiquinone oxidoreductase	S 0.812	25	M 0.721	111	0	0	0.4891
Cvel_27966.t1	glycerol 3-phosphate: ubiquinone oxidoreductase	S 0.492	25	M 0.872	82	0	0	0.8563
Cvel_25193.t1	dihydroorotate: ubiquinone oxidoreductase	S 0.986	24	S 0.698	23	S 0.510	17	0
Vbra_7220.t1	dihydroorotate: ubiquinone oxidoreductase	Anch 0.717	44	0	0	0	0	0
Cvel_9173.t1	electron transfer flavoprotein: ubiquinone oxidoreductase	0	0	M 0.732	52	-	-	0.9261
Vbra_4697.t1	electron transfer flavoprotein: ubiquinone oxidoreductase	0	0	0	0	0	0	0
Cvel_17813.t1	sulfide: ubiquinone oxidoreductase	S 0.906	28	M 0.439	42	0	0	0.1324
Vbra_5597.t1	sulfide: ubiquinone oxidoreductase	0	0	M 0.614	109	0	0	0.7835
Vbra_15197.t1	sulfide: ubiquinone oxidoreductase	S 0.999	17	C 0.247	64	C 0.795	47	0.3424

Table 3. Sources of HMMs and E-value cut-offs used for identification of alternative ubiquinone and cytochrome *c* reductases, and lactate dehydrogenases with the HMMER3 software.

enzyme	source of HMM: taxa or database (accession)	HMM length, aa	E-value cut-off
L-lactate: cytochrome <i>c</i> oxidoreductase (cytochrome <i>b</i> ₂)	fungi, <i>Perkinsus</i> , <i>Karolodinium</i>	502	10 ⁻⁶⁰
D-lactate: cytochrome <i>c</i> oxidoreductase	NCBI CDD (PLN02805)	563	10 ⁻¹⁶⁰
galacto-1,4-lactone: cytochrome <i>c</i> oxidoreductase	Pfam (ALO)	260	10 ⁻⁵⁰
NADH-dependent D-lactate dehydrogenase	NCBI CDD (PRK11183)	567	10 ⁻⁵⁰
NADH-dependent L-lactate dehydrogenase	Apicomplexa*	322	10 ⁻¹⁰⁰
alternative NADH dehydrogenase	Sarcocystidae	553	10 ⁻⁸⁰
electron-transfer flavoprotein: ubiquinone oxidoreductase	Pfam (ETF_QO)	110	10 ⁻³⁰
glycerol-3-phosphate: ubiquinone oxidoreductase	Aconoidasida	625	10 ⁻¹⁰⁰
sulfide: ubiquinone oxidoreductase	prokaryotes and eukaryotes	418	10 ⁻³⁰
dihydroorotate: ubiquinone oxidoreductase	Apicomplexa	442	1e ⁻⁸⁰

* The model for NADH-dependent L-lactate dehydrogenase detects malate dehydrogenase as well, and distinguishing these enzymes requires analysis of characteristic residues in the alignment.

Figures

Fig. 1.

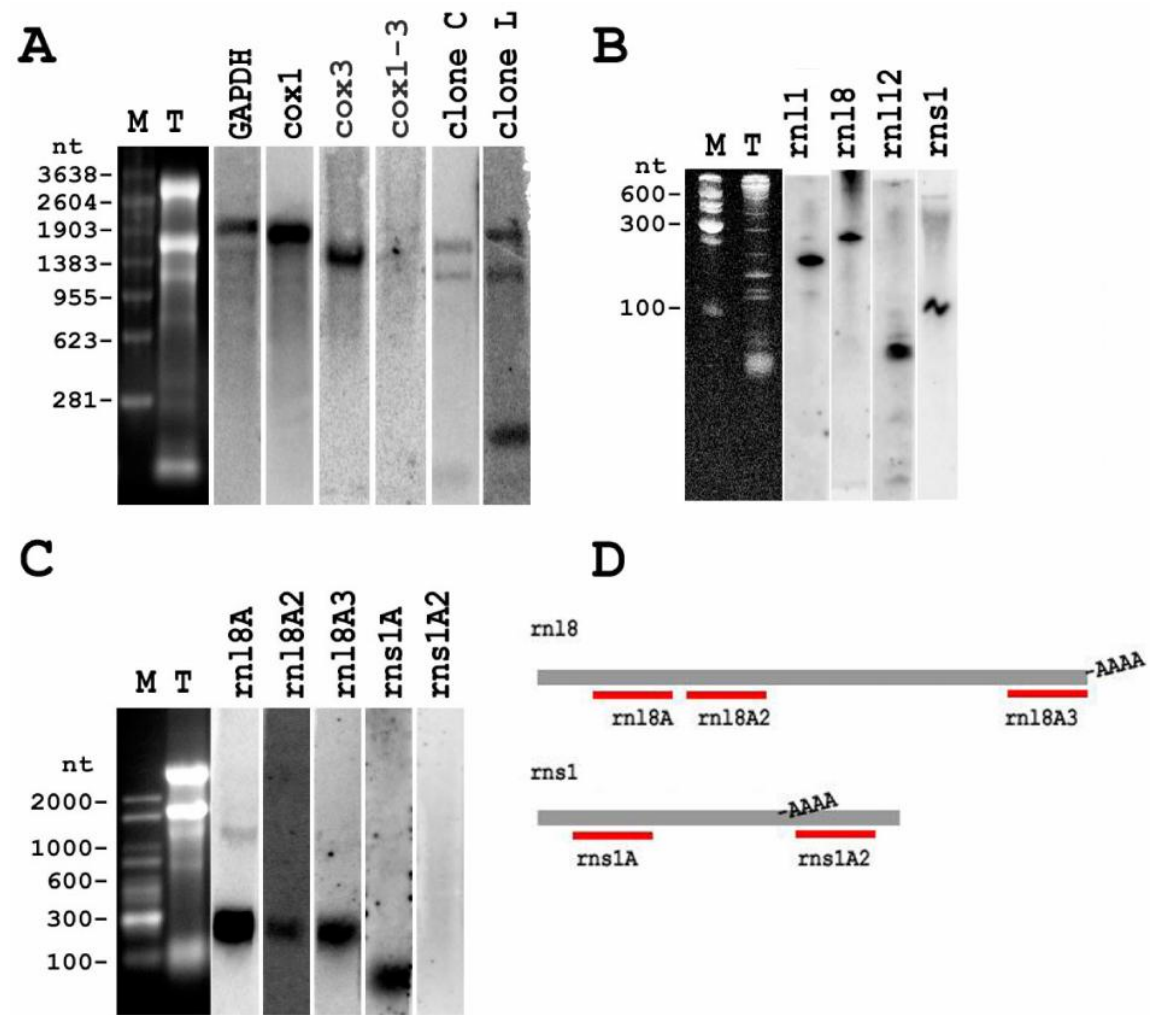


Fig. 2.

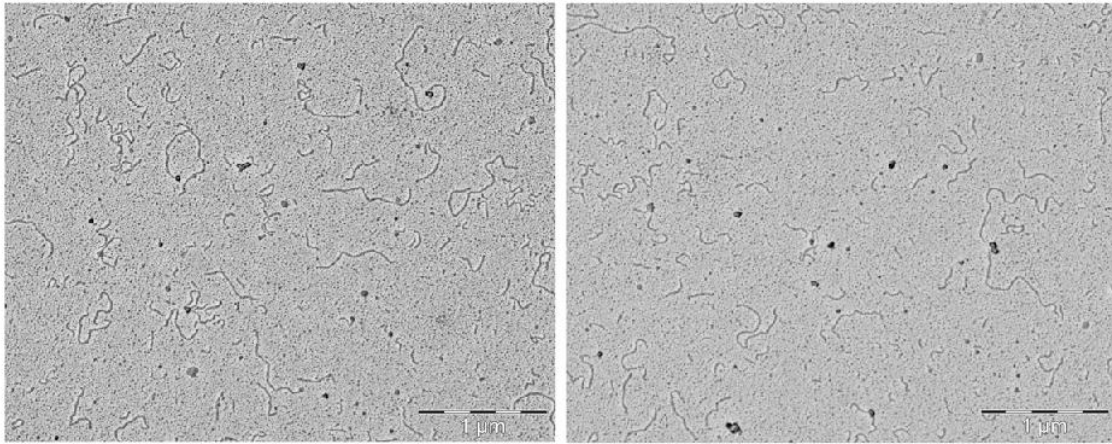


Fig. 3.

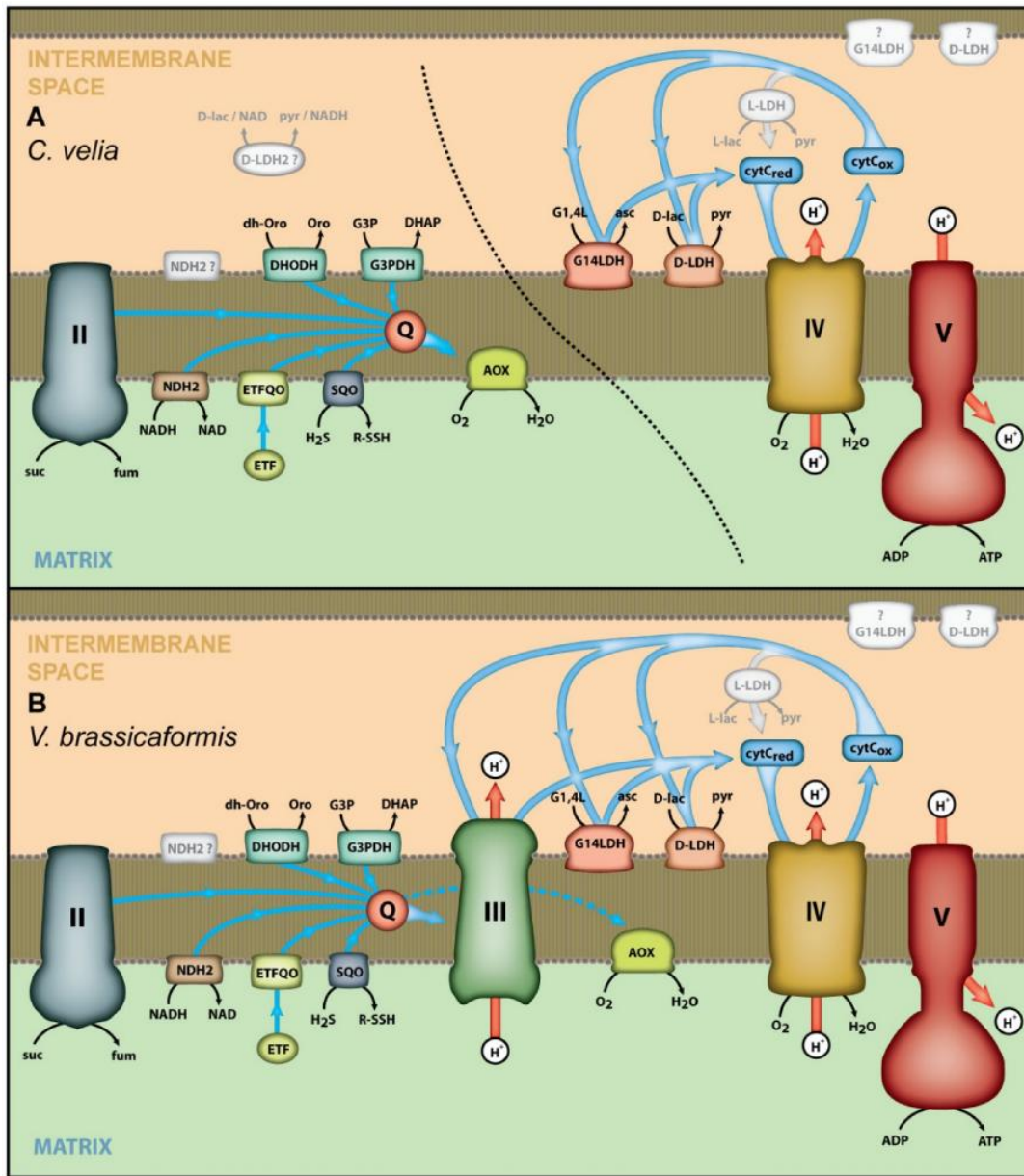


Fig. 4.

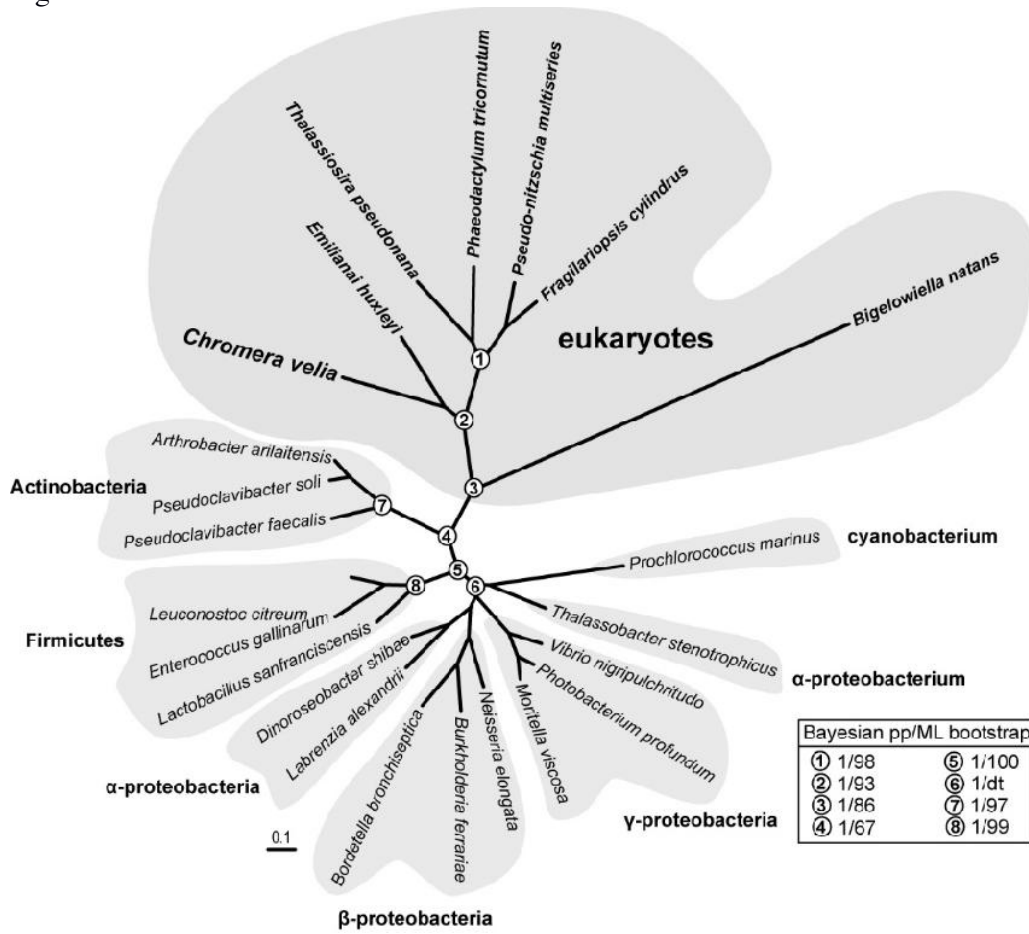


Fig. 5.

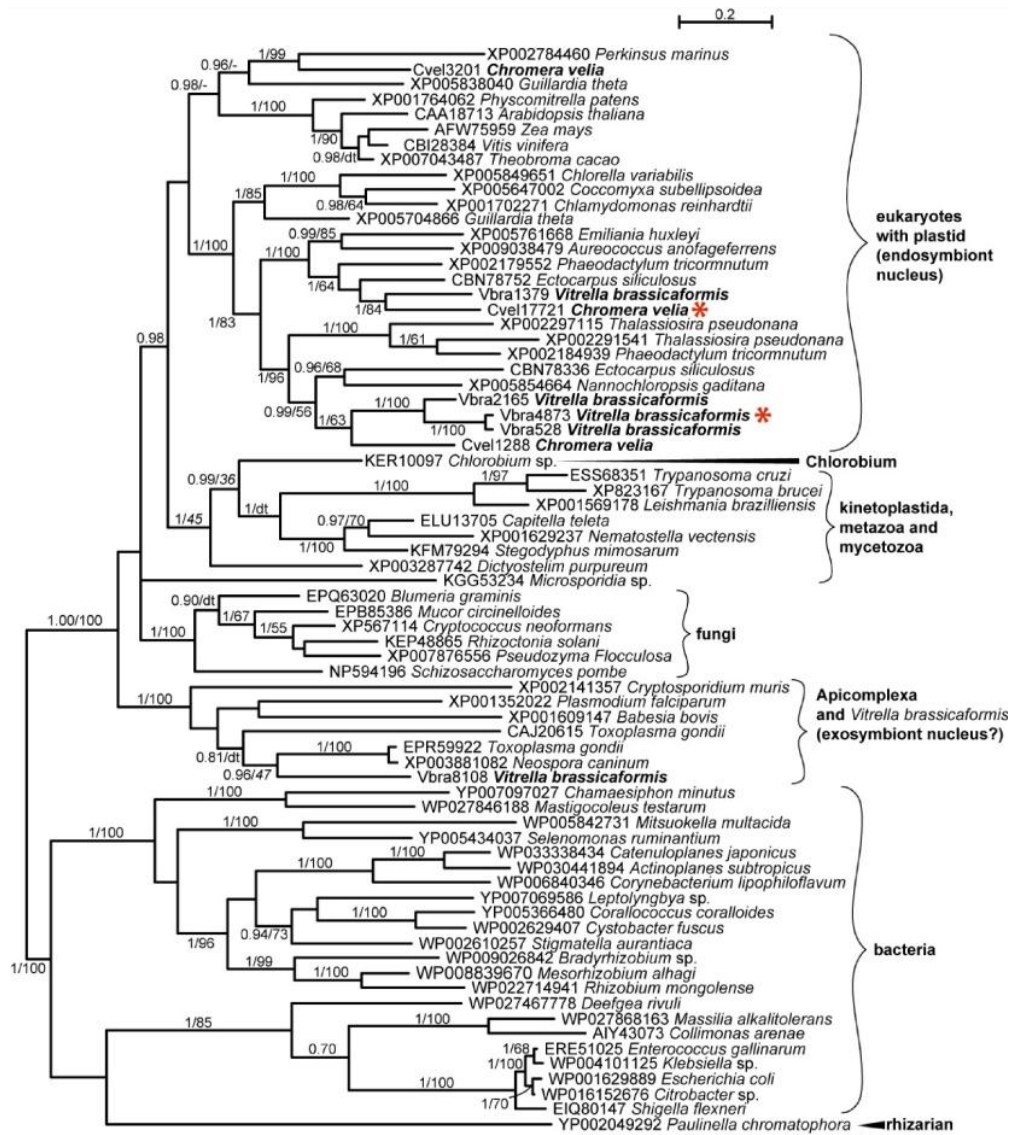


Fig. 6.

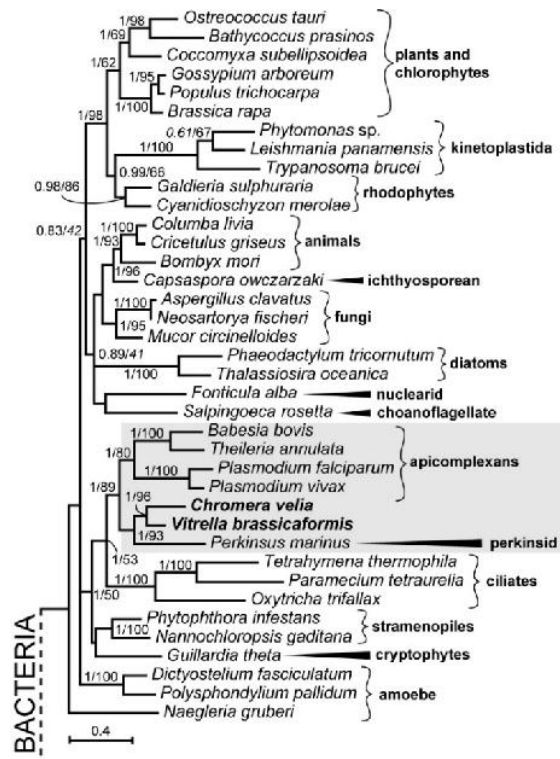


Fig. 7.

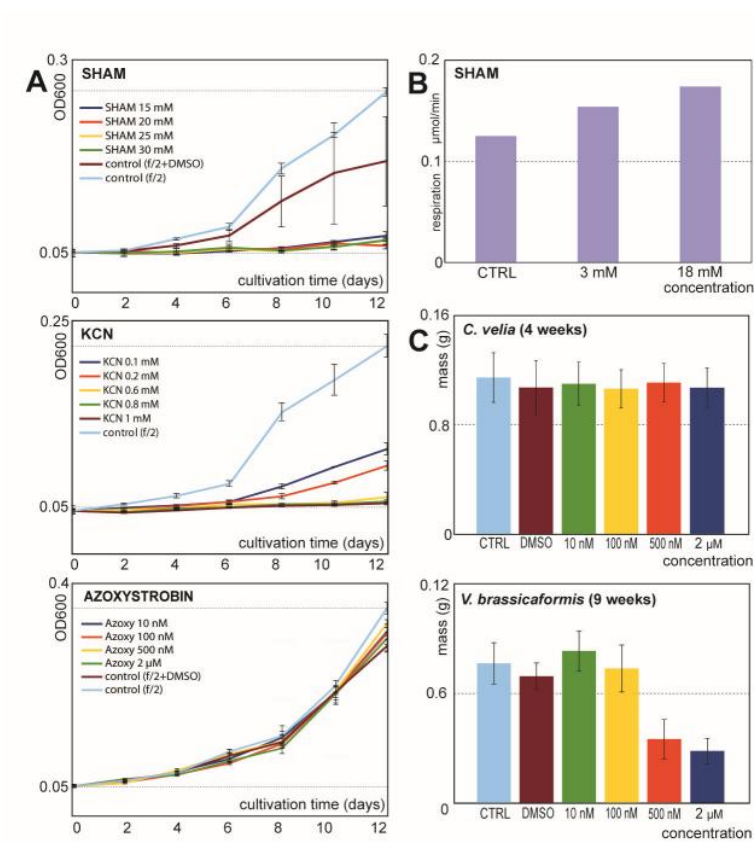
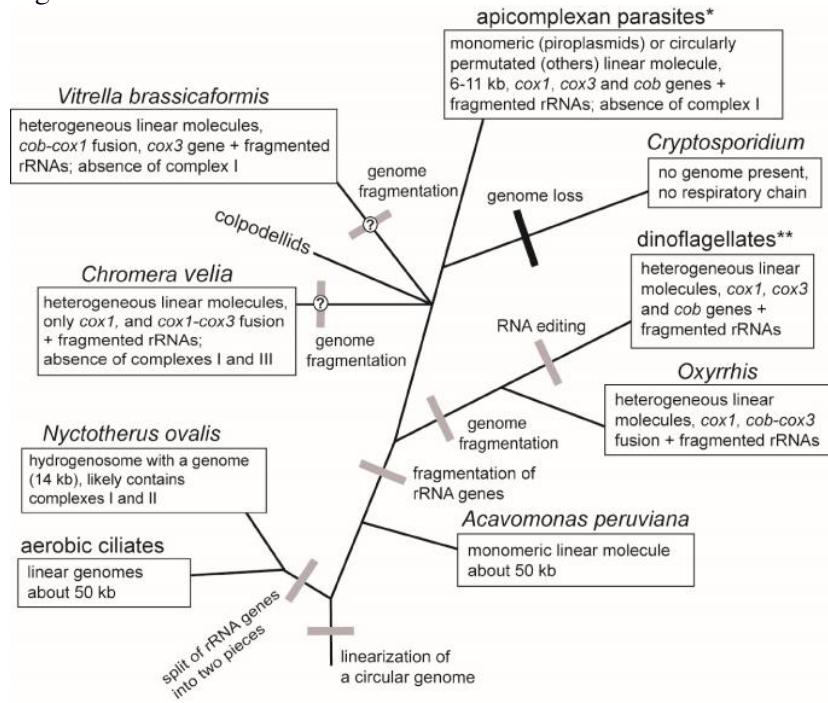


Fig. 8.



Paper II

Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites

Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites

Yong H Woo^{1*}, Hifzur Ansari¹, Thomas D Otto², Christen M Klinger^{3†}, Martin Kolisko^{4†}, Jan Michálek^{5,6†}, Alka Saxena^{1†‡}, Dhanasekaran Shanmugam^{7†}, Annageldi Tayyrov^{1†}, Alaguraj Veluchamy^{8†§}, Shahjahan Ali^{9¶}, Axel Bernal¹⁰, Javier del Campo⁴, Jaromír Cihlář^{5,6}, Pavel Flegontov^{5,11}, Sebastian G Gornik¹², Eva Hajdušková⁵, Aleš Horák^{5,6}, Jan Janoušek⁴, Nicholas J Katris¹², Fred D Mast¹³, Diego Miranda-Saavedra^{14,15}, Tobias Mourier¹⁶, Raece Naeem¹, Mridul Nair¹, Aswini K Panigrahi⁹, Neil D Rawlings¹⁷, Eriko Padron-Regalado¹, Abhinay Ramaprasad¹, Nadira Samad¹², Aleš Tomčala^{5,6}, Jon Wilkes¹⁸, Daniel E Neafsey¹⁹, Christian Doerig²⁰, Chris Bowler⁸, Patrick J Keeling⁴, David S Roos¹⁰, Joel B Dacks³, Thomas J Templeton^{21,22}, Ross F Waller^{12,23}, Julius Lukeš^{5,6,24}, Miroslav Oborník^{5,6,25}, Arnab Pain^{1*}

*For correspondence: yong.woo@kaust.edu.sa (YHW); arnab.pain@kaust.edu.sa (AP)

†These authors contributed equally to this work

Present address: [†]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Institute, Seattle, United States; [§]Biological and Environmental Sciences and Engineering Division, Center for Desert Agriculture, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; [¶]The Samuel Roberts Noble Foundation, Ardmore, United States

Competing interests: The authors declare that no competing interests exist.

Funding: See page 17

Received: 16 February 2015

Accepted: 16 June 2015

Published: 15 July 2015

Reviewing editor: Magnus Nordborg, Vienna BioCenter, Austria

© Copyright Woo et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

¹Pathogen Genomics Laboratory, Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; ²Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, United Kingdom; ³Department of Cell Biology, University of Alberta, Edmonton, Canada; ⁴Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, Canada; ⁵Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic; ⁶Faculty of Sciences, University of South Bohemia, České Budějovice, Czech Republic; ⁷Biochemical Sciences Division, CSIR National Chemical Laboratory, Pune, India; ⁸Ecology and Evolutionary Biology Section, Institut de Biologie de l'École Normale Supérieure, CNRS UMR8197 INSERM U1024, Paris, France; ⁹Bioscience Core Laboratory, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; ¹⁰Department of Biology, University of Pennsylvania, Philadelphia, United States; ¹¹Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava, Czech Republic; ¹²School of Botany, University of Melbourne, Parkville, Australia; ¹³Seattle Biomedical Research Institute, Seattle, United States; ¹⁴Centro de Biología Molecular Severo Ochoa, CSIC/Universidad Autónoma de Madrid, Madrid, Spain; ¹⁵IE Business School, IE University, Madrid, Spain; ¹⁶Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark; ¹⁷European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom; ¹⁸Wellcome Trust Centre For Molecular Parasitology, Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom; ¹⁹Broad Genome Sequencing and Analysis Program, Broad Institute of MIT and Harvard, Cambridge, United States; ²⁰Department of Microbiology, Monash University, Clayton, Australia; ²¹Department of Microbiology and Immunology, Weill Cornell Medical College, New York, United States; ²²Department of Protozoology, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan; ²³Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom; ²⁴Canadian Institute for Advanced Research, Toronto, Canada; ²⁵Institute of Microbiology, Czech Academy of Sciences, České Budějovice, Czech Republic

Abstract The eukaryotic phylum *Apicomplexa* encompasses thousands of obligate intracellular parasites of humans and animals with immense socio-economic and health impacts. We sequenced nuclear genomes of *Chromera velia* and *Vitrella brassicaformis*, free-living non-parasitic photosynthetic algae closely related to apicomplexans. Proteins from key metabolic pathways and from the endomembrane trafficking systems associated with a free-living lifestyle have been progressively and non-randomly lost during adaptation to parasitism. The free-living ancestor contained a broad repertoire of genes many of which were repurposed for parasitic processes, such as extracellular proteins, components of a motility apparatus, and DNA- and RNA-binding protein families. Based on transcriptome analyses across 36 environmental conditions, *Chromera* orthologs of apicomplexan invasion-related motility genes were co-regulated with genes encoding the flagellar apparatus, supporting the functional contribution of flagella to the evolution of invasion machinery. This study provides insights into how obligate parasites with diverse life strategies arose from a once free-living phototrophic marine alga.

DOI: [10.7554/eLife.06974.001](https://doi.org/10.7554/eLife.06974.001)

eLife digest Single-celled parasites cause many severe diseases in humans and animals. The apicomplexans form probably the most successful group of these parasites and include the parasites that cause malaria. Apicomplexans infect a broad range of hosts, including humans, reptiles, birds, and insects, and often have complicated life cycles. For example, the malaria-causing parasites spread by moving from humans to female mosquitoes and then back to humans.

Despite significant differences amongst apicomplexans, these single-celled parasites also share a number of features that are not seen in other living species. How and when these features arose remains unclear. It is known from previous work that apicomplexans are closely related to single-celled algae. But unlike apicomplexans, which depend on a host animal to survive, these algae live freely in their environment, often in close association with corals.

Woo et al. have now sequenced the genomes of two photosynthetic algae that are thought to be close living relatives of the apicomplexans. These genomes were then compared to each other and to the genomes of other algae and apicomplexans. These comparisons reconfirmed that the two algae that were studied were close relatives of the apicomplexans.

Further analyses suggested that thousands of genes were lost as an ancient free-living algae evolved into the apicomplexan ancestor, and further losses occurred as these early parasites evolved into modern species. The lost genes were typically those that are important for free-living organisms, but are either a hindrance to, or not needed in, a parasitic lifestyle. Some of the ancestor's genes, especially those that coded for the building blocks of flagella (structures which free-living algae use to move around), were repurposed in ways that helped the apicomplexans to invade their hosts. Understanding this repurposing process in greater detail will help to identify key molecules in these deadly parasites that could be targeted by drug treatments. It will also offer answers to one of the most fascinating questions in evolutionary biology: how parasites have evolved from free-living organisms.

DOI: [10.7554/eLife.06974.002](https://doi.org/10.7554/eLife.06974.002)

Introduction

The phylum *Apicomplexa* is comprised of eukaryotic, unicellular, obligate intracellular parasites, infecting a diverse range of hosts from marine invertebrates, amphibians, reptiles, birds to mammals including humans. More than 5000 species have been described to date, and over 1 million apicomplexan species are estimated to exist (Adl et al., 2007; Pawlowski et al., 2012). Clinically and economically important apicomplexan pathogens, for example, *Babesia*, *Cryptosporidium*, *Eimeria*, *Neospora*, *Theileria*, *Toxoplasma* (Tenter et al., 2000), and the malaria-causing parasite *Plasmodium* wreak profound negative impacts on animal and human welfare.

Despite their diverse host tropism (Roos, 2005) and life cycle strategies, apicomplexans possess several unifying molecular and cellular features, including the abundance of specific classes of nucleic acid-binding

proteins with regulatory functions in parasitic processes (Campbell et al., 2010; Flueck et al., 2010; Radke et al., 2013; Kafsack et al., 2014; Sinha et al., 2014), extracellular proteins for interactions with the host (Templeton et al., 2004a; Anantharaman et al., 2007), an apical complex comprising a system of cytoskeletal elements and secretory organelles (Hu et al., 2006), an inner membrane complex (IMC) derived from the alveoli (Eisen et al., 2006; Kono et al., 2012; Shoguchi et al., 2013), and a non-photosynthetic secondary plastid, termed the apicoplast (McFadden et al., 1996). How and when these features arose is unclear, owing to the lack of suitable outgroup species for comparative analyses.

Chromerids comprise single-celled photosynthetic colpodellids closely associated (and likely symbiotic) with corals (Cumbo et al., 2013; Janouškovec et al., 2013). Phylogenetic analysis demonstrates that these algae are closely related to Apicomplexa (Janouškovec et al., 2013), confirming the long-standing hypothesis that apicomplexan parasites originated from a free-living, photosynthetic alga (McFadden et al., 1996; Moore et al., 2008). Two known chromerid species, *Chromera velia* and *Vitrella brassicaformis* (Moore et al., 2008; Oborník et al., 2011, 2012), can be cultivated in the laboratory, and their plastid (Janouškovec et al., 2010) and mitochondrial genomes (Flegontov et al., 2015) have been described. We explored whole nuclear genomes of *Chromera* and *Vitrella* to understand how obligate intracellular parasitism has evolved in Apicomplexa.

Results and discussion

Genome assembly and annotation

A shotgun approach was used to sequence and assemble the *Chromera* and *Vitrella* nuclear genome into 5953 and 1064 scaffolds totaling 193.6 and 72.7 million base-pairs (Mb). The disparity in genome size is attributable largely to the presence of transposable elements (TEs) totaling ~30 Mb in *Chromera* vs only 1.5 Mb in *Vitrella*, as the predicted number of protein-coding genes is almost the same at 26,112 and 22,817, respectively. Detailed characterizations of the two genomes and their gene structures are described in Appendix 1 and **Supplementary files 1, 2**.

Ancestral gene content of free-living and parasitic species

We constructed a phylogenetic tree of 26 species, comprising *Chromera*, *Vitrella*, 15 apicomplexans, 2 dinoflagellates, 2 ciliates, 4 stramenopiles, and a green alga. On the phylogenetic tree (Figure 1A), *Chromera* and *Vitrella* formed a group closest to the apicomplexan clade, consistent with previous phylogenies (Moore et al., 2008; Janouškovec et al., 2010, 2013, 2015; Oborník et al., 2012). The long branches from their common node are consistent with drastic differences in morphology, life cycle (Oborník et al., 2012), plastid (Janouškovec et al., 2010) and mitochondrial genomes (Flegontov et al., 2015) between the two chromerids (Figure 1A). Likewise, despite common origins, apicomplexans show extensively diverse lifestyles, including host tropism and invasion phenotypes (Figure 1B).

We reconstructed the parsimonious gene repertoires for the ancestors of the 26 species, at the nodes of the phylogenetic tree (Figure 2A; Figure 2—figure supplement 1). We note five key nodes on the evolutionary paths to present-day apicomplexans: the alveolate ancestor; the common ancestor of Apicomplexa and chromerids, termed the proto-apicomplexan ancestor; the apicomplexan ancestor; the ancestor of apicomplexan lineages, for example, coccidia and hematozoa; and extant apicomplexans (Figure 2A). Protein-coding genes from the 26 species were clustered by OrthoMCL (Li et al., 2003) into groups of homologous genes, hereafter defined as orthogroups. We note that an orthogroup could have homologous genes from different species (putative orthologs) or from the same species (putative paralogs arising from gene duplications). Gains or losses of orthogroups are displayed as green or red sections of a pie on the phylogenetic tree in Figure 2A. Divergence of the proto-apicomplexan ancestor from the alveolate ancestor (Stage I) was accompanied by losses of 1668 and gains of 2197 orthogroups (sum of the two 'pies' in Stage I). Transition of the free-living proto-apicomplexan ancestor to the apicomplexan ancestor (Stage II) is accompanied by many gene losses (3862 orthogroups) but few gains (81 orthogroups) (Figure 2A). Divergence of coccidians, for example, *Toxoplasma gondii*, from the apicomplexan ancestor (Stage III) is characterized by modest changes (537 losses; 414 gains), whereas divergence of hematozoans, for example, *Plasmodium* spp., is marked by drastic losses (1384 losses; 77 gains) (Figure 2A). Further divergence of apicomplexan taxa beyond Stage III is characterized by modest, lineage-specific gains (Figure 2A). Functional composition of gained genes at various stages will be discussed in later sections. Paucity of gained genes (81 orthogroups) during Stage II indicates that the genome of the

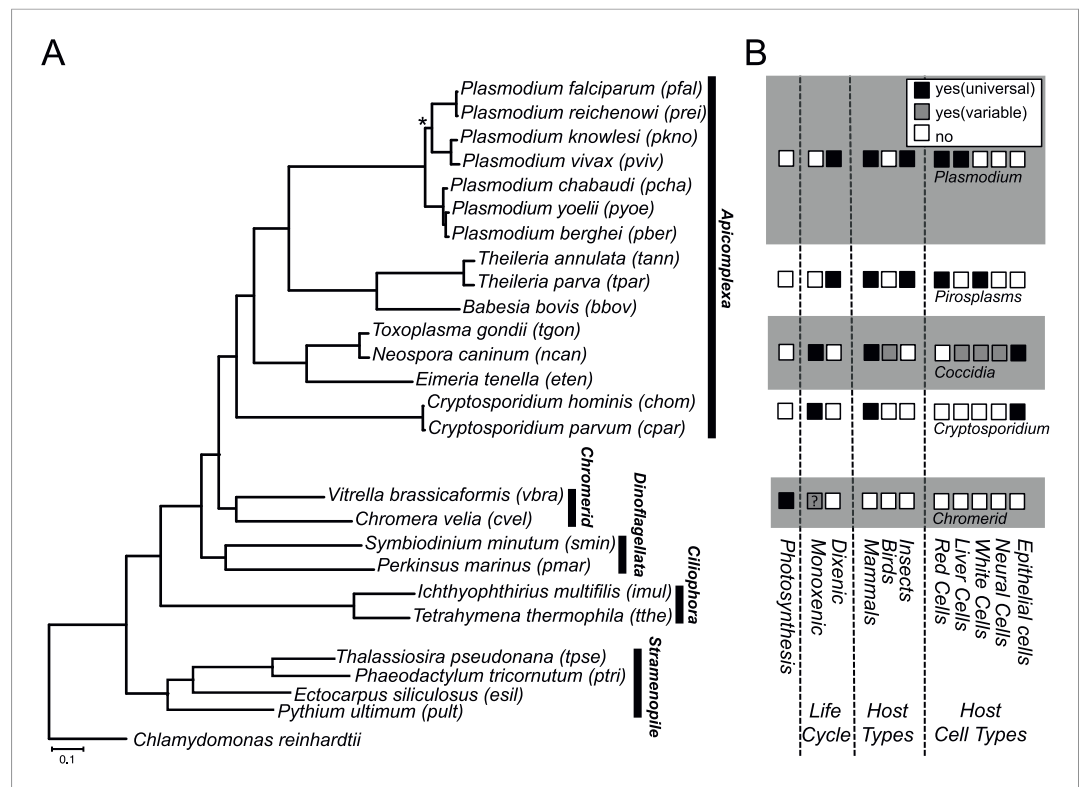


Figure 1. Phylogenetic, parasitological, and genomic context of chromerids. **(A)** Phylogenetic tree of 26 alveolate and outgroup species (see **Figure 1—source data 1** for the list of species). Multiple sequence alignments of 101 genes, which have 1:1 orthologs across all species (**Figure 1—source data 2**) were concatenated to a single matrix of 33,997 aligned amino acids. A maximum likelihood tree was inferred using RAXML with 1000 bootstraps, with *Chlamydomonas reinhardtii* as an outgroup. All clades are supported with bootstrap values of 100% except one node (*) with 99%, and also with 1.00 posterior probability from a bayesian phylogenetic tree based on PhyloBayes (**Lartillot and Philippe, 2004**) (CAT-GTR). **(B)** Lifestyles of the apicomplexan and chromerid species under investigation. '?': uncertainty due to lack of relevant data.

DOI: [10.7554/eLife.06974.003](https://doi.org/10.7554/eLife.06974.003)

The following source data are available for figure 1:

Source data 1. List of 24 species excluding *Chomera* and *Vitrella* used in this study and their data sources.

DOI: [10.7554/eLife.06974.004](https://doi.org/10.7554/eLife.06974.004)

Source data 2. A list of 101 shared orthogroups with a single gene in all of the 26 species, used for the species phylogenetic tree.

DOI: [10.7554/eLife.06974.005](https://doi.org/10.7554/eLife.06974.005)

free-living ancestor possessed most of the genes that were present in the common ancestor of apicomplexans and survived in their present-day descendants.

Progressive, lineage-specific losses during apicomplexan evolution

Parasite evolution has been associated with genome reduction across several branches of the tree of life (**Keeling, 2004; Sakharkar et al., 2004; Morrison et al., 2007**). Examples also exist, however, where parasite genomes are not reduced (**Pombert et al., 2014**) but expanded (**Raffaele and Kamoun, 2012**), underscoring the fact that the genome reduction process during parasite evolution is not completely understood. We sought to characterize in detail the dynamics of gene loss across apicomplexan evolution, particularly for components of molecular processes that are hallmarks of free-living lifestyle. We performed a systematic analysis of the cellular components involved in: (1) cellular metabolic pathways; (2) the endomembrane trafficking systems, regulating the movement of molecules across intracellular compartments in eukaryotes (**Leung et al., 2008**); and (3) the flagellum, a highly conserved apparatus for motility in aqueous environment (**Silflow and Lefebvre, 2001**).

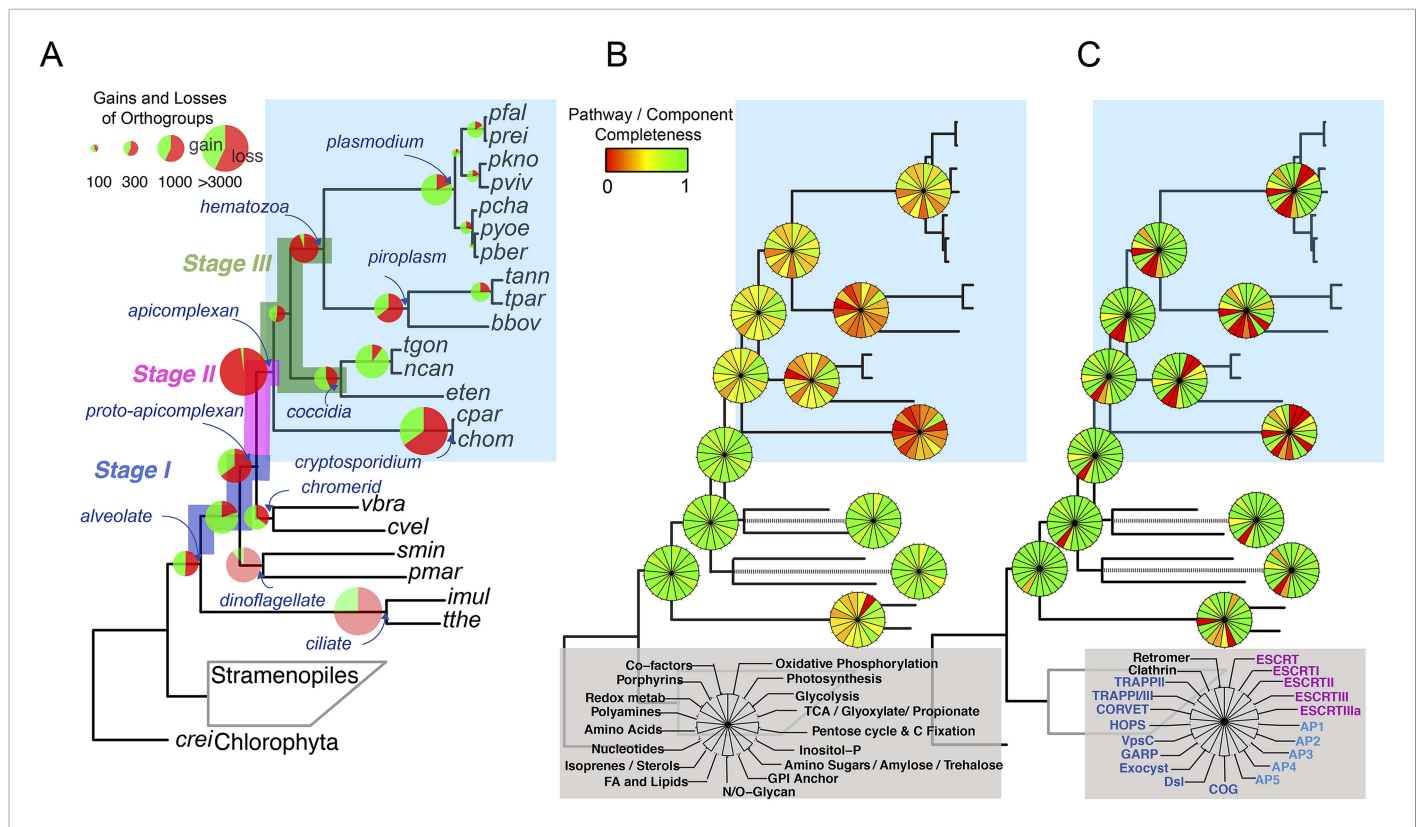


Figure 2. Gene content changes during apicomplexan evolution. **(A)** Gains and losses of orthogroups inferred based on Dollo parsimony (Csuros, 2010). Analysis based on a gene birth-and-death model provided similar results (Figure 2—figure supplement 1A). Stages I, II, and III (shown in blue, pink and green, respectively) represent groups of branches from the alveolate ancestor to apicomplexan lineage ancestors. Stage III could not be determined for *Cryptosporidium* lineage because of sparse taxon sampling. The area of a green or red section in a pie is proportional to the number of gained or lost orthogroups, respectively. **(B, C)** Overview of metabolic capabilities **(B)** and endomembrane components **(C)** in apicomplexan and chromerid ancestors. Gains and losses of enzymes and components were inferred, based on Dollo parsimony (Csuros, 2010). The pie charts are color-coded based on the fraction of enzymes or components present. Additional results from analysis of individual components and enzymes can be found in Figure 2—figure supplements 2,3,4,5, Supplementary file 3. Individual components and enzymes are listed in Figure 2—source data 1, 2. Similar analyses were performed for components encoding flagellar apparatus (Figure 2—figure supplement 5B).

DOI: 10.7554/eLife.06974.006

The following source data and figure supplements are available for figure 2:

Source data 1. Distribution of enzymes based on KEGG.

DOI: 10.7554/eLife.06974.007

Source data 2. Genes encoding subunits of the endomembrane trafficking system.

DOI: 10.7554/eLife.06974.008

Figure supplement 1. Gene gains and losses across the hypothetical ancestors of the 26 species under study.

DOI: 10.7554/eLife.06974.009

Figure supplement 2. Overview of chromerid Carbamoyl Phosphate Synthetase (CPS) and Fatty Acid Synthase I (FAS I).

DOI: 10.7554/eLife.06974.010

Figure supplement 3. Summary of metabolic pathways based on KEGG Assignments.

DOI: 10.7554/eLife.06974.011

Figure supplement 4. An overview of endomembrane trafficking components.

DOI: 10.7554/eLife.06974.012

Figure supplement 5. Evolutionary history of genes encoding cytoskeleton across 26 species.

DOI: 10.7554/eLife.06974.013

The following source data is available for figure 2s5:

Figure supplement 5—source data 1. Genes encoding components of the flagellar apparatus in the 26 species.

DOI: 10.7554/eLife.06974.014

The inferred proto-apicomplexan ancestor, like present-day chromerids, possessed complete metabolic pathways for sugar metabolism, assimilation of nitrate and sulfite, and photosynthesis-related functions (**Figure 2B**, **Figure 2—figure supplement 3**, Appendix 2, and **Supplementary file 3**). Unlike in other photosynthetic algae, both *Chromera* and *Vitrella* initiate heme synthesis in the mitochondrion using aminolevulinic synthase (C4 pathway), which thus far has been found only in a few eukaryotic heterotrophs, such as *Euglena gracilis*, dinoflagellates, and apicomplexans (Kořený et al., 2011; van Dooren et al., 2012; Danne et al., 2013) (Appendix 2 and **Supplementary file 4**). Both chromerids and apicomplexans encode modular multi-domain fatty acid synthase I (FASI)/polyketide synthase enzymes and single-domain FASII components (**Figure 2—figure supplement 2A,B**). Treatment of *Chromera* with a FASII inhibitor triclosan showed decreased production of long chain fatty acids (**Figure 2—figure supplement 2C** and Appendix 2), suggesting that *Chromera* synthesizes short-chain saturated fatty acids using the FASI pathway, which are then elongated using the FASII pathway. This was previously demonstrated in *Toxoplasma*, an apicomplexan that possesses both FASI and FASII (Mazumdar and Striepen, 2007). Likely, the proto-apicomplexan ancestor was a phototrophic alga harboring characteristic metabolic features previously found only in apicomplexan parasites, especially with regard to plastid-associated metabolic functions (see above and other examples in Appendix 2) (Kořený et al., 2011; van Dooren et al., 2012; Danne et al., 2013).

Transition to an apicomplexan ancestor (Stage II) was accompanied by the loss of metabolic processes including photosynthesis and sterol biosynthesis (**Figure 2B** and **Figure 2—figure supplement 3**). The apicomplexan ancestor appeared to possess a significant complement of enzymes in various pathways (**Figure 2B**) (Lim and McFadden, 2010). The differentiation of apicomplexan lineages (Stage III) was accompanied by further lineage-specific losses: for example, loss of FASI in *Plasmodium* spp, loss of FASII in *Cryptosporidium* spp., which has also lost the apicoplast, and loss of enzymes mediating polyamine biosynthesis in all lineages except *Plasmodium* (**Figure 2B** and **Figure 2—figure supplement 3**). These support the notion that enzymes involved in cellular metabolism critical for free-living organisms were not completely lost during the transition to the apicomplexan ancestor, but were further lost during subsequent differentiation and host-adaptation of apicomplexan lineages.

The proto-apicomplexan had a nearly complete repertoire of the endomembrane trafficking complexes, and much of this repertoire persisted through to the apicomplexan ancestor (Stage II) (Hager et al., 1999; Klinger et al., 2013a) (**Figure 2C**, **Figure 2—figure supplement 4** and Appendix 3). Differentiation of apicomplexan lineages (Stage III) was accompanied by lineage-specific losses, for example, loss of the Endosomal Sorting Complex Required for Transport II (ESCRTII) in all lineages except in piroplasms, whereas some components were retained across all lineages, such as the retromer complex components and clathrin, both systems implicated in invasion processes (Pieperhoff et al., 2013; Tomavo et al., 2013) (**Figure 2C**, **Figure 2—figure supplement 4** and Appendix 3). These lineage-specific losses have led to diverse, reduced sets of endomembrane trafficking components in present-day apicomplexans (Hager et al., 1999; Klinger et al., 2013a). Some of these components that were present in chromerids were absent in specific apicomplexan lineages as well as in dinoflagellates and ciliates, further clarifying that these losses are independent, lineage-specific events rather than ancient, shared events.

All known components of flagella were present in the proto-apicomplexan ancestor (**Figure 2—figure supplement 5A,B**). Most of the components were retained in the apicomplexan ancestor (Stage II), but losses occurred as apicomplexan lineages differentiated (Stage III). Components of intraflagellar transport, which are typically essential for assembling flagella, were lost in the other lineages except in coccidians (**Figure 2—figure supplement 5A,B**). The basal body proteins, which support an organizing center for microtubules, were lost from piroplasms. Some striated fiber assembly (SFA) proteins, typically associated with basal body rootlets, were maintained in all apicomplexan lineages including piroplasms (**Figure 2—figure supplement 5A,B,D**); their presence has been hailed as evidence that some flagellar-proteins are repurposed for new functions in apicomplexans (see below) (Francia et al., 2012).

In summary, one of the major events during apicomplexan evolution is progressive, continued loss of components important for free-living organisms. While Stage II was accompanied by a massive loss of such components including those implicated in photosynthesis, the apicomplexan ancestor still possessed many proteins, which were lost later during differentiation of lineages with diverse life strategies.

Emergent features of apicomplexans

Evolution of present-day apicomplexan parasites was accompanied not only by gene losses as noted above (**Figure 2**) but also by gene gains. We sought to determine if genes gained at a particular stage

of apicomplexan evolution, as depicted by the gray violin in **Figure 3**, would be over-represented with those involved in parasitic processes such as intracellular invasion into and egress from host cells. For *Plasmodium falciparum* and *T. gondii*, we compiled three classes of protein-coding genes directly or indirectly involved in parasitic processes of apicomplexans based on in silico prediction or information from previous functional studies ('Materials and methods'). Extracellular proteins are secreted by the apicomplexans for various parasitic processes, for example, some of them are targeted to the host cytoplasm, nucleus, and plasma membrane to modulate parasite–host interactions

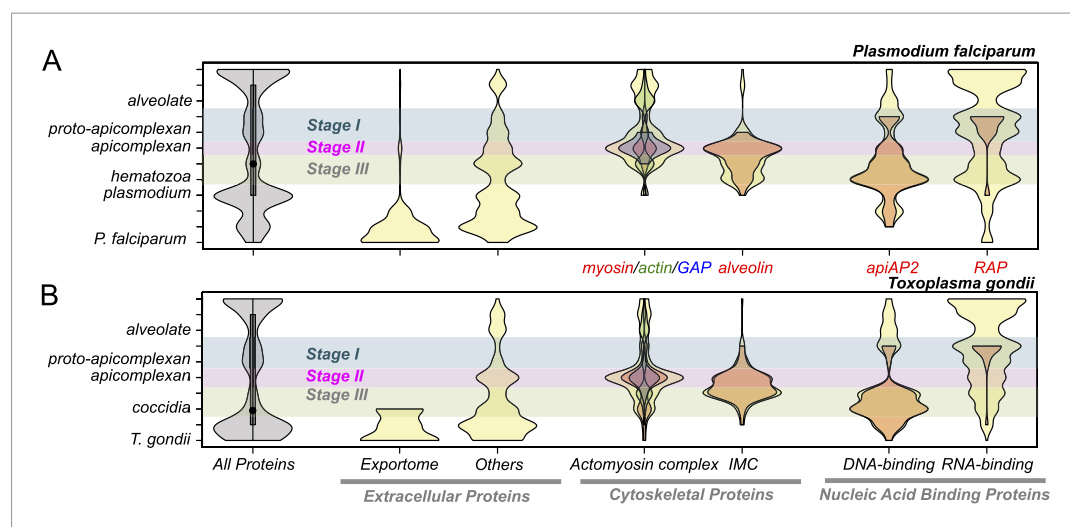


Figure 3. Evolutionary history of *Plasmodium falciparum* and *Toxoplasma gondii* genes. Violin plots showing distribution of evolutionary ages of genes (Y-axis: from species-specific (bottom) to deeply conserved (top)) in *P. falciparum* (A) and *T. gondii* (B). Evolutionary age of a gene is defined as the earliest node on the evolutionary path of the phylogenetic tree where homolog can be detected ('Materials and methods'). The horizontal thickness of a violin is proportional to the number of genes (gray) or the fraction of genes (yellow) in a functional category (X-axis) out of all with the same evolutionary age. Selected functional sub-categories are overlaid with red, green, or blue violin plots. The maximum width of each violin is scaled to be uniform across categories. Inner boxes in the gray violins indicate inter-quartile ranges and circles indicate medians. Colored shades along the X-axis indicate Stages I–III (**Figure 2**). Extracellular proteins include proteins targeted to host cytoplasm, nucleus, and plasma membrane ('exportome') and all other proteins, which are secreted or localized on the parasite surface ('others'). Cytoskeletal proteins include proteins associated with 'actomyosin motor complex' and 'IMC'. All extracellular and cytoskeletal proteins are listed in **Figure 3—source data 1, 2**. Nucleic acid-binding proteins are predicted in silico based on presence of DNA-binding domains (DBDs) and RNA-binding domains (RBDs). See 'Materials and methods' for details on how these genes are defined and compiled. Domain architectures of representative extracellular proteins in apicomplexans and chromerids are displayed as schematics in **Figure 3—figure supplement 4**. Sequence homology networks (**Figure 2—figure supplement 5E** and **Figure 3—figure supplements 1B, 2B, 3B**) and gene gains and losses on the phylogenetic tree (**Figure 3—figure supplements 1A, 2A, 3A**) provide complementary views on the evolutionary history of these genes.

DOI: [10.7554/eLife.06974.015](https://doi.org/10.7554/eLife.06974.015)

The following source data and figure supplements are available for figure 3:

Source data 1. Genes encoding extracellular proteins in *P. falciparum* and *T. gondii*.

DOI: [10.7554/eLife.06974.016](https://doi.org/10.7554/eLife.06974.016)

Source data 2. Genes encoding cytoskeletal components in the 26 species.

DOI: [10.7554/eLife.06974.017](https://doi.org/10.7554/eLife.06974.017)

Figure supplement 1. Evolutionary history of apiAP2 genes.

DOI: [10.7554/eLife.06974.018](https://doi.org/10.7554/eLife.06974.018)

Figure supplement 2. Evolutionary history of alveolins.

DOI: [10.7554/eLife.06974.019](https://doi.org/10.7554/eLife.06974.019)

Figure supplement 3. Evolutionary history of RAP genes.

DOI: [10.7554/eLife.06974.020](https://doi.org/10.7554/eLife.06974.020)

Figure supplement 4. Domain architectures of extracellular proteins in chromerids and apicomplexans.

DOI: [10.7554/eLife.06974.021](https://doi.org/10.7554/eLife.06974.021)

(Mundwiler-Pachlatko and Beck, 2013; Bougdour et al., 2014). Cytoskeletal proteins provide structural support to the cell and also the molecular machinery for motility and intracellular invasion (Baum et al., 2006; Soldati-Favre, 2008). Proteins with DNA-binding domains (DBDs) or RNA-binding domains (RBDs) can regulate various molecular processes of apicomplexan parasites. Indeed, proteins with AP2 (apiAP2) DBD have been shown to act as genetic control switches for diverse apicomplexan processes (Balaji et al., 2005; Campbell et al., 2010; Flueck et al., 2010; Radke et al., 2013; Sinha et al., 2014; Kaneko et al., 2015).

Genes encoding extracellular proteins exported into the host environments were over-represented among those gained after Stage III (Figure 3), suggesting that adaptation to specific hosts was accompanied by expansion of extracellular proteins mediating host–parasite interactions (Templeton et al., 2004a; Anantharaman et al., 2007). Stage III was accompanied by gains of those encoding DBD proteins, mostly apiAP2 proteins (Figure 3 and Figure 3—figure supplement 1A,B), suggesting extensive regulatory changes mediated by apiAP2 proteins during lineage differentiation. We note that losses of other canonical DBD proteins, for example, proteins with HSF_DNA-bind (Pfam: PF00447) domain during transition to apicomplexan ancestor (Stage II) and proteins with Tub (Pfam: PF01167) domain along the piroplasm lineage, contribute to further dominance of apiAP2 among the DBD proteins (Figure 3—figure supplement 1C). Stage II was accompanied by over-represented gains of various cytoskeletal components, including alveolins, those of the actomyosin complex (e.g., myosins) and glideosome-associated proteins with multiple membrane spans 1 and 3 (GAPM1 and GAPM3), suggesting that the molecular machinery powering gliding motility, which is essential for host cell invasion arose during evolution to apicomplexans (Frenal et al., 2010) (Figure 3, Figure 3—figure supplement 2, and Appendix 4). Gene gains during Stage I were over-represented by proteins with ‘RBD abundant in Apicomplexans’ (RAP, Pfam: PF08373) (Lee and Hong, 2004), many of which were conserved as one-to-one orthologs across descending lineages, suggesting development of evolutionarily conserved functions before apicomplexans and chromerids diverged (Figure 3, and Figure 3—figure supplement 3). Chromerid genomes encode many orthologs of apicomplexan cytoskeletal proteins (Appendix 4), including GAPM2, a member of an important protein family for apicomplexan cytoskeletal structure and gliding motility (Bullen et al., 2009), and the IMC sub-compartment protein family (ISP), implicated in establishing apical polarity and coordinating the unique cell cycle of apicomplexans (Poulin et al., 2013) (Figure 2—figure supplement 5E). These data suggest that some components existed in the free-living proto-apicomplexan ancestor and were subsequently repurposed for parasitic processes of apicomplexans.

The *Chromera* and *Vitrella* genomes encode many proteins that are specific to chromerids yet contain functional domains implicated in molecular processes of apicomplexan parasites. For example, there are chromerid-specific proteins with domain architectures similar to those in apicomplexan extracellular proteins, including those previously implicated in host interactions and described in apicomplexans only (Figure 3—figure supplement 4 and Appendix 5, and Supplementary file 5). Presence of such chromerid proteins implies some commonality in extracellular recognition and cross-species interactions and this correlates well with the presumed associations with the coral holobiont (Janoušková et al., 2012, 2013; Cumbo et al., 2013). Importantly, chromerid genomes encode numerous apiAP2 proteins, more abundant than dinoflagellates, suggesting that they have expanded in the proto-apicomplexan ancestor after it split from dinoflagellates (Figure 3—figure supplement 1D). Many of the chromerid apiAP2 proteins belong to putative paralogous clusters, suggesting that their expansion was driven by gene duplication (Figure 3—figure supplement 1D; Appendix 6). Only a small subset of the apiAP2 proteins are shared across apicomplexans, suggesting that the large apiAP2 complement in the proto-apicomplexan ancestor has diversified independently in descending lineages (Figure 3—figure supplement 1A).

In summary, genes encoding critical components of the parasitic lifestyle of apicomplexans were gained at different stages of apicomplexan evolution, some implying subsequent specialization to particular host niches, but others suggesting early adaptations before committing to parasitic lifestyle. This is evident by chromerid orthologs of many such proteins, for example, RAP proteins and specialized cytoskeletal components. Further, chromerid genomes encode chromerid-specific proteins that are not detected as orthologs of apicomplexan proteins but still have functional domains implicated in parasitic processes in apicomplexans. Together, these data imply that a molecular transition had occurred in free-living ancestors of apicomplexans, providing a foundation for host–parasite interactions and further adaptation.

Conserved gene expression programs in the proto-apicomplexan ancestor

Chromera and *Vitrella* genomes allowed us to reconstruct the gene content of the free-living ancestor of apicomplexans. To infer their putative functions using genome-wide gene expression information (Hu et al., 2010), we cultured *Chromera* under 36 different combinations of temperatures, iron and salt concentrations, and generated their gene expression profiles by RNA-seq (Box et al., 2005). In addition, we have obtained a publicly available growth perturbation data set for *P. falciparum* (Hu et al., 2010). There were 1918 orthogroups shared between the two species. We identified pairs of orthogroups that are co-expressed, that is, showing similar expression patterns across the various conditions, in both species ('Materials and methods') (Figure 4—figure supplement 1A). Such an orthogroup pair, that is, those with conserved co-expression between the two species, would include candidate genes that have been co-regulated together during apicomplexan evolution, from the free-living ancestor to present-day parasites due to conserved functions. This approach, successfully utilized by several studies in the past (Stuart et al., 2003; Mutwil et al., 2011; Gerstein et al., 2014), led to the following two observations in this study.

Many RAP genes appeared during Stage I and have been conserved across the descending phyla (Figure 3 and Figure 3—figure supplement 3), but their precise cellular roles are unknown. For 11 out of 12 orthogroups with RAP domains, co-expressed orthogroups overlapped significantly (Fisher's exact test, $p < 0.05$) between *P. falciparum* and *Chromera*, suggesting involvement of RAP proteins in cellular processes evolutionarily conserved across apicomplexans and chromerids (Figure 4A). RAP and their co-expressed orthogroups encode proteins with putative mitochondrial import signals more often than expected by chance in *Chromera* and *P. falciparum* (Fisher's exact test, $p < 0.05$) (Figure 4B), and also in other apicomplexans and chromerids (Figure 4—figure supplement 1B). We have randomly chosen three *Toxoplasma* RAP genes with predicted mitochondrial localization signals (Supplementary file 6) and confirmed experimentally by 3' endogenous gene-tagging with reporter epitopes that all three are localized to the organelle (Figure 4C). Some of the orthogroups co-expressed with orthogroups containing RAP domains encode protein products predicted to be metabolic enzymes, implying possible involvement of RAPs in mitochondrial metabolism (Figure 4—figure supplement 1C). Consistent with this, the *Cryptosporidium* lineage that has a highly reduced mitochondrion lacking both the genome and most canonical metabolic pathways (Abrahamsen et al., 2004; Xu et al., 2004) is the only apicomplexan group to have also lost its RAP repertoire (Figure 4—figure supplement 1D). Loss of RAPs along with a set of mitochondrial functions in this lineage is consistent with a mitochondrial role for RAPs. We speculate that the free-living proto-apicomplexan ancestor possessed within its mitochondrion a regulatory process mediated by RNA-binding activities of the RAP proteins, which has been retained by the extant apicomplexans and chromerids.

As discussed earlier, the proto-apicomplexan ancestor appears to have possessed genes implicated in invasion processes of present-day apicomplexans (Figure 3). Among the 1918 orthogroups, we identified 80 orthogroups comprising genes functionally annotated as implicated in invasion processes. The frequency of co-expression amongst them in the free-living *Chromera* was significantly higher than expected by chance ($p < 0.0005$), suggesting pre-existing functional relationships before transitioning to parasites (Figure 4D). We identified several modules or groups of co-expressed orthogroups (Figure 4E). In one of the co-expression modules (numbered 1 in Figures 4E), 9 out of 10 orthogroups are co-expressed with a gene encoding SFA (Cvel_872), a key protein for organizing the basal bodies of the flagellar apparatus in algae and the apical complexes in apicomplexans (Kawase et al., 2007; Francia et al., 2012) (Figure 4F). We note that SFAs are the only flagellar components found in all apicomplexans tested (Figure 2—figure supplement 5A). Also in this module, for 9 out of 10 orthogroups, their co-expressed orthogroups in *Chromera* overlapped significantly with those in *P. falciparum* (Fisher's exact test, $p < 0.05$), indicating that their regulatory programs have been evolutionarily conserved (Figure 4G). This module include various types of genes implicated in host cell invasion processes of apicomplexans such as genes encoding rhoptyr protein ROP9, apical sushi protein ASP, and gliding motility components GAP40 and GAPM2. The apical complex has been postulated to have emerged from the flagellar apparatus and associated cellular transport systems in free-living algae, based on ultrastructural evidence (Okamoto and Keeling, 2014; Portman et al., 2014). These results suggest that, in the free-living ancestor, some of the genes

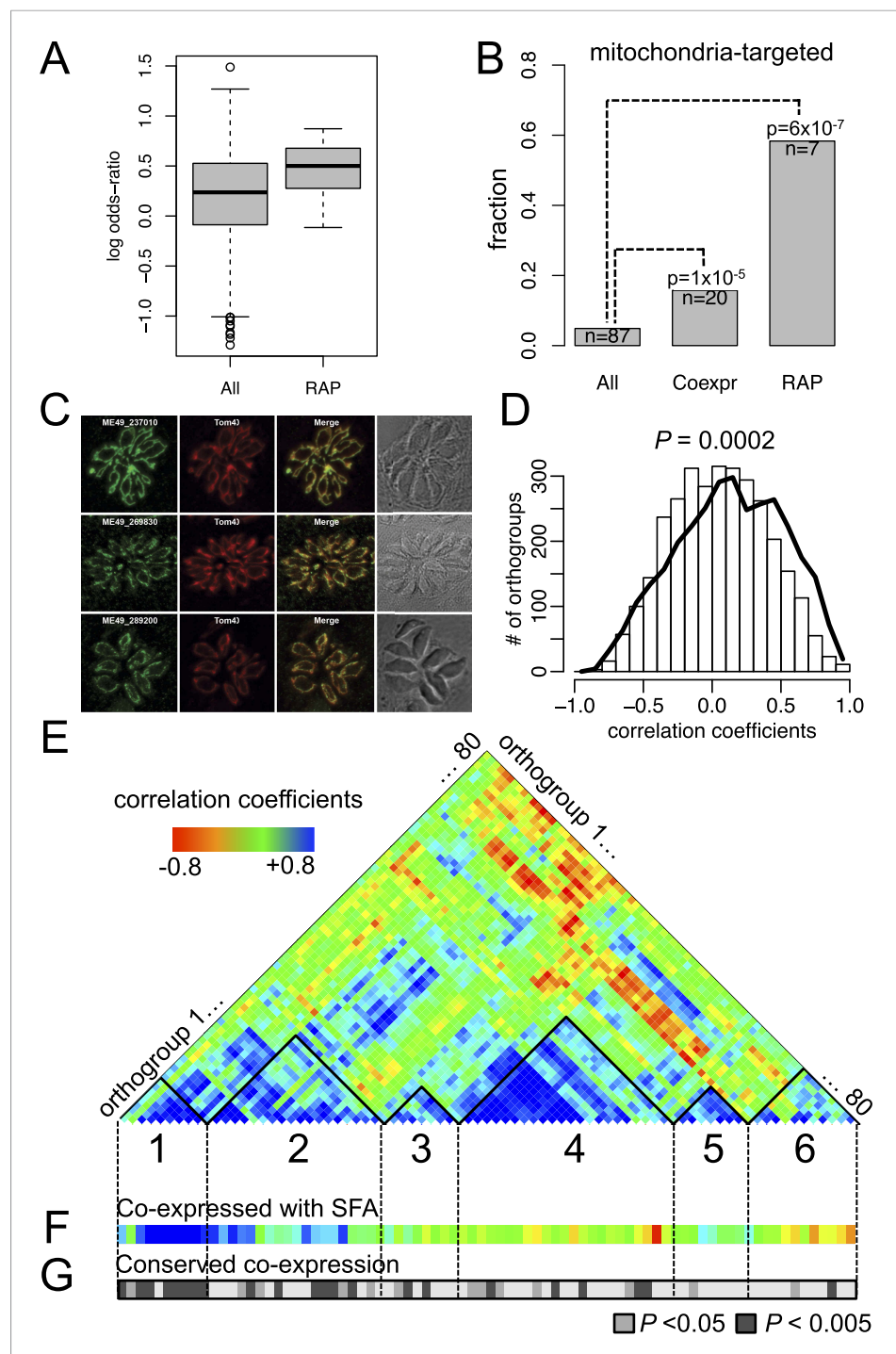


Figure 4. Conserved transcriptional programs in apicomplexans and chromerids. **(A)** Boxplot showing the extent of evolutionary conservation of transcriptional programs for all orthogroups or those with RAP domains. X-axis: 'All' (all orthogroups excluding RAP); 'RAP' (orthogroups with RAP domains). Y-axis: log-transformed odds-ratio, representing, for each orthogroup, the degree of overlap between its co-expressed orthogroups in *Chromera* and those in *P. falciparum*. **(B)** Bar chart showing the fraction of orthogroups (Y-axis) predicted to be targeted to mitochondria in both species ('Materials and methods'). The number of genes are displayed below each bar. X-axis: 'All' (all orthogroups excluding the other two categories); 'Coexpr' (orthogroups co-expressed with RAP in both species); 'RAP' (orthogroups with RAP domains). The fractions in 'Coexpr' and 'RAP' groups were compared against the fraction in 'All', and p-values based Fisher's exact test are displayed above the bar. Files deposited in European Nucleotide Archive are listed in [Figure 4—source data 1](#) with corresponding conditions. **(C)** Sub-cellular

Figure 4. continued on next page

Figure 4. Continued

localization of RAP proteins encoded by TGME49_237010, TGME49_269830, and TGME49_289200 was tested in *T. gondii* by 3' tagging of the endogenous genes with the coding sequence for the hemagglutinin epitope, together with a mitochondrial marker Tom40. See **Supplementary file 6** for details of the localization predictions.

(D) Distributions of Spearman's rank correlation coefficients of gene expression between all possible pairs from the 80 orthogroups implicated in invasion processes in apicomplexans (black outline) were compared against those from 80 randomly selected ones (histogram). The p value indicates statistical significance of the difference based on 10,000 random samplings. The 80 orthogroups and corresponding genes in *Chromera* and *P. falciparum* are listed in **Figure 4—source data 2**. (E) Heatmap showing a matrix of correlation coefficients amongst the 80 orthogroups. Based on a hierarchical clustering, we classified them into six co-expression modules, labeled as numeral 1–6.

(F) Heatmap showing correlation coefficients with striated fiber assemblin (SFA) (Cvel_872). The color scheme is the same as in (E). (G) Heatmap indicating statistical significance of conserved transcriptional program, that is, the odds-ratio as defined in (A) (Fisher's exact test, $p < 0.05$ (gray); $p < 0.005$ (black)).

DOI: [10.7554/eLife.06974.022](https://doi.org/10.7554/eLife.06974.022)

The following source data and figure supplement are available for figure 4:

Source data 1. RNA-seq libraries of *Chromera velia* under various growth conditions.

DOI: [10.7554/eLife.06974.023](https://doi.org/10.7554/eLife.06974.023)

Source data 2. List of genes implicated in invasion processes in apicomplexans.

DOI: [10.7554/eLife.06974.024](https://doi.org/10.7554/eLife.06974.024)

Source data 3. Evolutionary conservation of 12 orthogroups with RAP domains (for 'RAP' category in **Figure 4A**).

DOI: [10.7554/eLife.06974.034](https://doi.org/10.7554/eLife.06974.034)

Figure supplement 1. Mitochondrial targeting of RAP and its putative role in mitochondrial metabolism.

DOI: [10.7554/eLife.06974.025](https://doi.org/10.7554/eLife.06974.025)

implicated in the invasion process of present-day apicomplexans were functionally associated with those implicated in flagellar motility, providing the much-needed genetic evidence for the postulate. We speculate that a group of functionally related proteins associated with the flagellar apparatus was repurposed as a module of the apical complex and became a foundation for the invasion machinery.

Conclusion

Analysis of *Chromera* and *Vitrella* genomes has enabled insights into how apicomplexan parasites have evolved from free-living ancestors. The transition to parasitism was accompanied by massive genomic loss that continued as its descendants became specialized intracellular parasites infecting diverse hosts. The genome of free-living photosynthetic ancestors encodes many component proteins previously assumed to be restricted to the parasitic apicomplexan lineages. Such pre-existing components, including those of what would later become part of the invasion machinery, were co-opted during evolution to facilitate a successful parasitic lifestyle in multiple hosts. The genome of the proto-apicomplexan ancestor served as a molecular blueprint for evolution of the most successful group of eukaryotic parasites known to date.

Data access

Sequencing data have been deposited in the European Bioinformatics Institute under the European Nucleotide Archive (ENA) sample accession number ERP006228 for *C. velia* and ERP006229 for *V. brassicaformis* for all DNA- and RNA-seq experiments. The assembly and the annotations were submitted under accession numbers CDMZ01000001-CDMZ01005953 for *C. velia* and CDMY01000001-CDMY01001064 for *V. brassicaformis*. Some of the *Vitrella* DNA-seq experiments were done at Broad Institute and are deposited at Short Read Archive under accession numbers SRX152523 and SRX152525. The annotations and assemblies can be viewed and queried in EupathDB (<http://cryptodb.org/cryptodb/>).

Materials and methods

DNA preparation and sequencing

Genomic DNA of *C. velia* CCMP2878 (subsequently referred to as *Chromera*) and *V. brassicaformis* CCMP3155 (subsequently referred to as *Vitrella*) was extracted and then sheared into short fragment

size libraries (300–500 base pair (bp)) and large fragment size libraries (3–8 kbp fragments) by focused-ultrasonication (Covaris Inc., Woburn, USA). The last 3–8 kb libraries were prepared following Nextera mate pair protocol, following manufacturer's instructions. We used three different methods to generate the library: the Illumina (Illumina, San Diego, CA) TruSeq DNA protocol LT Sample Prep Kit (catalog no. #FC-121-2001), an amplification-free method (*Kozarewa et al., 2009*) (TruSeq DNA PCR-Free LT Sample Preparation Kit catalog no. #FC-121-3001) and the Illumina Nextera Mate Pair Sample Preparation Kit (catalog no. #FC-132-1001). The libraries were sequenced on an Illumina HiSeq2000 platform following the manufacturers standard cluster generation and sequencing protocols (*Bentley et al., 2008*; *Quail et al., 2012*). Image analysis, base-calling, and quality filtering were processed by Illumina software.

RNA preparation and sequencing

For isolation of RNAs, *Chromera* and *Vitrella* were grown under standard culture conditions (*Oborník et al., 2012*). Total RNA was extracted from the cells using TRIzol. The polyA⁺ RNA fraction was selected using oligo(dT) beads, and RNA-seq libraries were prepared using TruSeq RNA Sample Prep kit (catalog no. FC-122-1001). Strand-specific RNA-seq libraries were prepared using TruSeq Stranded mRNA LT Sample Prep Kit (catalog no. RS-122-2101) and sequenced as paired-end (2 x 100 bp) reads on a HiSeq2000 platform.

We performed additional RNA sequencing of *Chromera* subject to various environmental perturbations, to construct a global gene expression network based on transcriptomes under various perturbation conditions during in vitro growth. *Chromera* cultures were exposed to a combination of stresses (**Figure 4—figure supplement 1C**). First, six different media were prepared from the combinations of salt concentration (16.7 g/l, 33.3 g/l, 66.6 g/l) and iron deficiency by chelation (*Sutak et al., 2010*). After seeding, the cultures were maintained in the normal temperature and light condition for eleven days (*Oborník et al., 2011*). After randomization, the cultures were incubated at 26°C, 37°C, or 14°C for 0 (control), 0.5, or 2 hr. There were two biological replicates of each, in total 66 flasks of the cultures. Then, the cultures were processed with centrifugation at 3500 RPM for 15 min at 4°C to precipitate the cells. Total RNA was extracted from the 66 cultures after the treatments using Norgen RNA Extraction kit based on manufacturer's protocol (Norgen Biotek Corporation, Canada). RNA quality was assessed using Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). RNA concentration was determined with a Qubit (Invitrogen, Carlsbad, CA). Strand-specific RNA-seq libraries were prepared from extracted high-quality RNAs (RIN ≥ 8.0 as measured on an Agilent Bioanalyzer 2100) using the Illumina TrueSeq LT stranded RNA sample kit according to manufacturer's instructions. Prior to cluster generation, concentration and size of libraries were assayed using the Agilent DNA1000 kit. Libraries from all samples were sequenced as single-end (1 x 50 bp) reads on the Illumina HiSeq 2000. The RNA-seq reads were aligned to the reference genome using tophat (version 2.0.8, default parameters) and cufflinks (version 2-1.0.2, default parameters) (*Trapnell et al., 2012*). The FPKM values were \log_2 normalized with an offset of 1 and were further corrected for different distributions across the samples using the quantile normalization method (*Bolstad et al., 2003*).

Genome assembly

For *Vitrella*, the reads were corrected and assembled followed by several base correction, scaffolding and gap filling steps as briefly described below. As first step, the short insert libraries were corrected with SGA (*Simpson and Durbin, 2012*) (version 0.9.19). The corrected reads were assembled with velvet (*Zerbino and Birney, 2008*) (version 1.2.08). Iterating through different parameter settings, we choose a k-mer of 75 bp as the best parameter set. The resulting scaffolds (larger than 1 kb) were further scaffolded with SSPACE (*Boetzer et al., 2011*) using first the Illumina library (insert = 550 bp) and larger insert (1 kb) Illumina library reads. Sequencing gaps were closed with Gapfiller (*Boetzer and Pirovano, 2012*) (version 1.1.1) with two iterations, using the bowtie mapping option and PCR-Free libraries. Base pair call errors were corrected in three iterations of ICORN (*Otto et al., 2010*), using the amplification-free library. Furthermore, sequencing gaps were closed, using IMAGE (*Tsai et al., 2010*) with the amplification-free library. The assembly was quality-controlled using REAPR (*Hunt et al., 2013*), breaking the contigs at possible miss-assemblies, using the mate pair libraries. This was followed by another scaffolding step. We systematically removed 620 scaffolds containing 25.65 Mb representing the bacterial contamination. The *Vitrella* CCMP3155 assembly contains 72.7

Mb (including 931,689 N's) in 1064 scaffolds (ENA accession numbers CDMY01000001-CDMY01001064). The scaffolds were constructed from 4177 contigs.

For *Chromera*, the assembly pipeline and the algorithms used were the same as *Vitrella*, but due to the larger size, higher amount of low-complexity regions, and difficulties in generating high-quality large insert size libraries, additional steps were included to the assembly process. First, the reads of the PCR-Free library were corrected with SGA ([Simpson and Durbin, 2012](#)) and then assembled with velvet and using a k-mer of 71 (version August 2011). Next, the contigs were scaffolded, gapfilled, and corrected with ICORN, as described earlier. We mapped the reads of all large insert size libraries using SMALT (<ftp://ftp.sanger.ac.uk/pub/resources/software/smalt/>). We excluded scaffolds smaller than 1 kb. Different iterations with SSPACE were undertaken and the assembly was quality-checked with REAPR. After scaffolding, gapfiller and IMAGE were run as above, followed by ICORN. The 1725 scaffolds (spanning 16.02 Mb) representing bacterial contamination were removed. The final assembly of *Chromera* CCMP2878 contains 193.66 Mb (including 582,995 N's) in 5953 scaffolds (ENA accession numbers CDMZ01000001-CDMZ01005953). The scaffolds are constructed from 13,987 contigs.

Gene prediction

We used Augustus ([Stanke et al., 2006](#)) (version 2.5.5) for gene prediction. We manually curated 716 and 245 gene models for *Chromera* and *Vitrella*, respectively, using BLAST similarity-based approaches, and we also generated automated gene models using Cufflinks ([Trapnell et al., 2012](#)) from RNA-seq data sets, in order to use them as a 'training gene model set' for Augustus prediction. The strand-specific RNA-Seq, mapped with TopHat2 ([Kim et al., 2013](#)), was used as evidence in Augustus for intron evidence.

In summary, from the *Chromera* and *Vitrella* genome, we ab initio predicted 30,478 and 23,503 protein-coding genes, respectively, of which 18,829 and 18,240 were detected as being expressed from RNA-seq evidence as poly A+ transcripts ([Supplementary file 1](#)). Excluding putative TEs, 26,112 and 22,817 genes were predicted as protein-coding genes in *Chromera* and *Vitrella*. We annotated partial genes, when a gene probably spans more than one scaffold, located at the borders of a scaffold. We demarcated and annotated as pseudo genes if they contain in frame stop codons. We flagged gene models as transposon elements, if they overlap with the predicted TE regions and had no more than three and two intron for *Chromera* and *Vitrella*, respectively. To annotate untranslated regions (UTRs) of the predicted protein-coding genes, we used CRAIG ([Bernal et al., 2007](#)) with default parameters with mapping of the RNA-Seq data as computed by GSNAP ([Wu and Nacu, 2010](#)) (version 2013-08-19, default parameters). The annotation of both genomes has the ENA accession numbers CDMZ01000001-CDMZ01005953 and CDMY01000001-CDMY01001064 and is also available in EuPathDB ([Aurrecochea et al., 2013](#)).

Functional annotations

The predicted genes were assigned putative functions based on BLASTP (E value $<10^{-6}$) matches against UNIPROT (version March 2012). The predicted protein products were assigned protein domains using *hmmsearch* (HMMER 3.1b1, May 2013) for Pfam A v26.0. Statistical threshold defined by the Pfam ([Finn et al., 2014](#)) database was used. We aligned AP2 sequences in apicomplexan species based on PfamA AP2 (PF00847), and built apicomplexan-specific AP2 (apiAP2) hidden Markov model (HMM), and scanned the predicted protein-coding genes for apiAP2 domains; we annotated api-AP2 DNA-binding transcription factor genes with both domain and sequence E values to be less than 10^{-3} . The following Pfam RBDs were used to define RNA-binding proteins: 'CAT_RBD', 'dsRNA_bind', 'S1', 'DEAD', 'KH_1', 'KH_2', 'KH_3', 'KH_4', 'KH_5', 'RRM_1', 'RRM_2', 'RRM_3', 'RRM_4', 'RRM_5', 'RRM_6', 'SET', 'PUF', and 'RAP'. The list of DBDs was downloaded from a database of DBDs ([Wilson et al., 2008](#)). Transmembrane domains and signal peptides were assigned with the tools TMHMM 2.0 ([Krogh et al., 2001](#)) and signalP 4.0 ([Petersen et al., 2011](#)), respectively, with default parameters.

We collected several categories of genes implicated in parasitic processes in apicomplexans for two archetypal apicomplexan parasites, *Toxoplasma* and *Plasmodium*. We primarily obtained annotations from PlasmoDB ([Bahl et al., 2003](#)) and ToxoDB ([Gajria et al., 2008](#)). Information for sub-cellular localization of genes is obtained from GeneDB ([Logan-Klumpler et al., 2012](#)) and ApiLoc, a database of published protein sub-cellular localization for apicomplexan species

(<http://apiloc.biochem.unimelb.edu.au/apiloc/apiloc>). Some putative parasite genes were inferred based on orthology by OrthoMCL clustering (Li *et al.*, 2003) with closely related species with results from functional studies. We performed exhaustive literature searches to manually curate individual genes, to define rules for in silico searches across the proteomes of this study, and to categorize the identified genes based on their localization and function. The categories of parasite genes are defined as follows.

Cytoskeleton

The cytoskeleton of an organism provides the necessary structural framework for the maintenance of cell shape and integrity. We compiled two groups of cytoskeletal proteins, IMC associated proteins and actomyosin complex. First, IMC associated proteins, comprises alveolin proteins, a membrane occupation and recognition nexus protein (MORN), which associate with IMC and spindle poles and are indispensable for asexual and sexual development (Ferguson *et al.*, 2008). IMC sub-compartment proteins (ISPs) are critical for establishing apical polarity in the parasite (Poulin *et al.*, 2013). Second, components of actomyosin motor complex, which powers the characteristic gliding motility (Soldati-Favre, 2008), comprises actin, myosin, tubulin, gliding associated proteins (GAPs), aldolase, and various actin-regulatory proteins, which will assist actin in the process of quick polymerization–depolymerization cycles between F-actin and G-actin during this process. Examples of actin-regulatory proteins are Arp2/3 complex and formins (FH2) for nucleation; F-actin capping for filament regulation; coronin for cross-linking/bundling and profilin, CAP, cofilin/ADF and gelsolin for monomer treadmill (Baum *et al.*, 2006).

Extracellular proteins

Extracellular proteins are defined as parasite proteins, which are localized either on the surface or secreted off the parasite. They are released in a concerted manner to ensure successful adhesion to the surface, entry into the host cell, multiplication, and escape. Extracellular proteins can be categorized as (1) 'exportome' are proteins translocated to the host cytoplasm, membranes, and nucleus crossing the boundary membrane parasitophorous vacuole (PV); and (2) 'others', which stay on the parasite surface or released from the parasite, but not into the host intracellular space. The exportome genes are released mostly from the parasite's secretory organelles such as rhoptries and dense granules (Ravindran and Boothroyd, 2008; Trecek *et al.*, 2011; Mundwiler-Pachlatko and Beck, 2013; Bougdour *et al.*, 2014). Some of these genes possess host targeting or also known as the Plasmodium export element (PEXEL). Many PEXEL-negative proteins have been identified too (Hsiao *et al.*, 2013; Mundwiler-Pachlatko and Beck, 2013). These genes are sorted and targeted through a specialized structure known as Maurer's cleft formed in the host cytoplasm (Mundwiler-Pachlatko and Beck, 2013). These genes are mostly kinases, proteases, and surface molecules, which modulate the host and hijack the host machinery in favor of parasitic growth and host immune evasion (Trecek *et al.*, 2011; Li *et al.*, 2012; Bougdour *et al.*, 2014). The 'other' extracellular proteins consist of surface antigens (e.g., MSPs), SERAs, TRAPs, AMA-1, microneme proteins, ROPs and RONs etc.

TEs

Repeat annotation was done by using the REPET pipeline (Flutre *et al.*, 2011) and LTR finder (Xu and Wang, 2007). The overall pipeline comprises of two steps: de novo detection and classification. In the first step, the scaffolds are split into smaller batches (~1000 batches of 200 kb each). These genomic fragments were aligned against each other to detect the HSPs (High-scoring pairs) using BLASTER (Quesneville *et al.*, 2003). HSPs are then clustered using a combination of three methods such as GROUPE (Quesneville *et al.*, 2003), RECON (Bao and Eddy, 2002), and PILER (Edgar and Myers, 2005). Structure-based LTR retrotransposons (RTs) detection tools such as LTRharvest (Ellinghaus *et al.*, 2008) and LTR finder, which are based on 100–1000 bp long terminal repeats with a 1 kb–15 kb separation and target site duplication site at vicinity of 60 bp to the two terminal repeats. These LTRs detected are clustered using BlastClust. Multiple sequence alignment of each cluster was performed using MUSCLE (Edgar, 2004). Each cluster aligned was searched against Repbase (Jurka *et al.*, 2005) using BLASTER (Quesneville *et al.*, 2003) and HMMER (Johnson *et al.*, 2010). A consensus feature was detected for each aligned cluster. Further PASTEC (Flutre *et al.*, 2011), which is based on the Wicker classification, was used for consensus classification.

The repeats were annotated as follows. The genomic chunks were randomized and HSPs were detected using BLASTER (Quesneville et al., 2003), CENSOR (Jurka et al., 1996), and RepeatMasker (Tempel, 2012). These HSPs were filtered and combined. Again, full-length genomic scaffolds were compared to Repbase using MATCHER. Satellite and simple repeats were detected using the mreps (Kolpakov et al., 2003), TRF (Benson, 1999), RMSR (RepeatMasker). Finally, a long-join procedure was followed to combine the nested repeats. The whole annotation was exported to a genome-browser readable GFF3 file.

Clustering homologous genes

OrthoMCL 2.0 (Li et al., 2003) was used with a default inflation parameter ($I = 1.5$) (Chen et al., 2006) to generate groups of homologous genes (defined as orthogroups), which could have homologs from different species (putative orthologs) or from the same species (putative paralogs from gene duplications). For some genes of high interest, we manually inspected the alignments of the protein sequences within the orthogroup, which were done with MAFFT (Kato and Standley, 2013). We assigned Pfam domains to an orthogroup if more than half of the genes in an orthogroup were assigned the Pfam domains.

Sub-cellular localization prediction

There are several tools available for a general eukaryotic sub-cellular localization prediction (Du et al., 2011), but they are not applicable to alveolates due to its unique chloroplast membrane arising from secondary endosymbiosis. Therefore, HECTAR (Gschloessl et al., 2008), which was developed for the bipartite sub-cellular prediction, was used. There is no stand-alone version of HECTAR, and the online version allows only one sequence at a time. We implemented a modified HECTAR algorithm as a PERL script for batch prediction of the whole proteomes. Each protein sequence was predicted for signal sequence using SignalP 3.0 (Bendtsen et al., 2004), the signal sequence is cleaved, and the remaining amino acid sequence was used as input for the transit peptide prediction by TargetP (Emanuelsson et al., 2000). Sequences with both signal peptide and the transit peptide (either chloroplast or mitochondria) are predicted to be in the chloroplast. Sequences without the signal peptide but with the transit peptide (either chloroplast or mitochondria) are predicted to be in mitochondria. Sequences with signal peptide, without transit peptide, and predicted by TargetP to be secretory are classified as secretory proteins.

For the RAP proteins, we tested the validity of our sub-cellular localization prediction in two ways. First, we compared our in-house algorithm with other published tools: TargetP (Emanuelsson et al., 2000), MitoProt2 (Claros and Vincens, 1996), iPSORT (Bannai et al., 2002), and PredSL (Petsalaki et al., 2006) (Supplementary file 6, only mitochondrial prediction is shown). We found that our mitochondrial prediction for RAP genes is in concordance with other methods. Second, we experimentally verified mitochondrial localization in *T. gondii* by 3' tagging of the endogenous genes with the coding sequence for the hemagglutinin epitope for three RAP proteins that were predicted to target to mitochondria with high probability.

Statistical analysis

A statistical environment software R was used for most of the analyses and generating parts of figures. An R package *vioplot* was used to generate the violin plot (Hintze and Nelson, 1998). A ward algorithm on the distance matrix based on (1- correlation coefficients) in an R function *hclust* was used for all hierarchical clustering of gene expression patterns unless noted otherwise.

Evolutionary analysis

We compiled the reference proteomes of 26 alveolate and stramenopile species (Figure 1—source data 2) from public databases such as EupathDB (Aurrecochea et al., 2013) and NCBI Genome database (<http://www.ncbi.nlm.nih.gov/genome/>).

We generated a phylogenetic species tree using a data set composed of 101 one-to-one orthologs across the 26 species (see Figure 1—source data 1 for gene IDs). Amino acid sequences were aligned using MAFFT (Kato and Standley, 2013), highly variable sites were edited by trimAL (Capella-Gutierrez et al., 2009) and after manual inspection. The resulting alignment of 33,997 amino acid positions was used to construct trees by a maximum likelihood

(ML) method and Bayesian inference. The ML tree was computed using RAXML 8.1.16 by gamma corrected LG4X model (Stamatakis, 2014; Le et al., 2012). Robustness of the tree was estimated by bootstrap analysis in 1000 replicates. Bayesian tree was constructed by PhyloBayes (Lartillot and Philippe, 2004) using two-infinite mixture model CAT-GTR as implemented in PhyloBayes 3.3f. Two independent chains were run until they converged (i.e., maximum observed discrepancy was lower than 0.2), and the effective number of model parameters was at least 100 after the first 1/5 generation was omitted from topology and posterior probability inference. All clades in the tree were supported with posterior probability 1.00 and 100% bootstraps, except for one node, which representing the common ancestor of human *Plasmodium* spp. was supported by 99% bootstrap.

We performed the gene gain and loss analysis based on Dollo parsimony using Count software (Csuros, 2010). This approach allows reconstructing gene contents at observed species and at hypothetical ancestors, and gene gains and losses at branching points. The Dollo parsimony strictly prohibits multiple gains of genes. To test for validity of this assumption, we repeated analyses based on parsimony settings allowing multiple gene gains or on a phylogenetic birth-and-death model (Csuros, 2010) and reached the same conclusion (Figure 2—figure supplement 1). We have also repeated the analysis using Wagner's parsimony, allowing multiple gains per tree with gain penalty of 2 or greater, and obtained similar results (data not shown). For the analysis of metabolic enzymes, endomembrane trafficking system components, and flagellar apparatus components, the ancestral presence was inferred based on Dollo parsimony from the presence of components in the observed species. For the endomembrane trafficking component analysis, we assumed that the last common ancestor had a complete repertoire of the components.

We have inferred the evolutionary age of *P. falciparum* and *T. gondii* genes as the early node on the phylogenetic tree where the most distant species have genes with significant sequence homology (reciprocal BLASTP E value $<10^{-10}$ and clustering with OrthoMCL).

Comparison of gene expression network between *Chromera velia* and *Plasmodium falciparum*

We studied if orthologs of *Chromera* and *P. falciparum* show similar gene expression changes to physiologically equivalent growth conditions. Identifying equivalent conditions is difficult as the two species have completely different lifestyles and live in different environments. Instead, we tested if a given gene and its ortholog would show correlated expression patterns with the same set of genes (and orthologs), allowing a way to compare gene expression behavior measured under different conditions. To uncover gene-to-gene co-expression relationships, the organisms from whom transcriptomes are sampled must be exposed to various growth conditions. This approach has been successfully used in other eukaryotes (Stuart et al., 2003; Hu et al., 2010; Mutwil et al., 2011). For *Chromera*, we generated RNA-seq-based transcriptome under combinations of varying salt concentrations, iron concentrations, and temperature changes, resulting in 36 unique combinations (see 'Materials and methods' and Figure 4—figure supplement 1C). For *P. falciparum*, we obtained previously published microarray-based gene expression data sets of 144 unique conditions from 23 time series, representing stresses from various growth-inhibiting compounds (Hu et al., 2010). It has been shown that gene expression data generated using different molecular platforms are reproducible and accurate enough for cross-platform comparisons (Woo et al., 2004). Based on each data set, we calculated Spearman correlation coefficients ρ between all possible pairs from the 1918 orthogroups shared between *Chromera* and *P. falciparum* (1918×1918 matrix). We also calculated a 1918×1918 weighted adjacency matrix using CLR algorithm (Faith et al., 2007) as implemented in an R package *minet* (with parameters of method = 'clr', estimator = 'mi.shrink', and disc = 'equalfreq') (Meyer et al., 2008). Expression level of multiple genes in a given orthogroup was averaged. To rule out any potential systematic biases associated with averaging expression levels of homologous, yet distinct genes, we repeated some of the analyses with 1560 orthogroups that have one-to-one orthologs between the two species and reached the same conclusions (data not shown). A pair of genes (or orthogroup) were determined as co-expressed if the Spearman's correlation coefficient ρ is greater than 0.3 and if the value from the weighted adjacency matrix of the network is greater than 0.01. We calculated an odds-ratio to measure the

extent of conservation of co-expressed genes: (# of genes co-expressed in both species) × (# of genes co-expressed in none of the species)/[(# of genes co-expressed in *P. falciparum* only) × (# of genes co-expressed in *C. velia* only)], and Fisher's exact test was used to assess the statistical significance. For calculation of the odds-ratios, co-expression was determined based on correlation coefficient to minimize count granularity in the two-by-two table.

Acknowledgements

We thank the KAUST Bioscience Core Laboratory personnel for sequencing specific Illumina libraries used in this project, KAUST Computational Bioscience Research Center for providing computing resources, Gordon Langsley (Institut Cochin, Inserm U1016, Paris) and Anthony Holder (The Francis Crick Institute, London) for comments on the manuscript draft. AV and CB thank Florian Maumus for advice regarding TE annotation. The primary funding for this work was provided by KAUST award FIC/2010/09 to JL, MO, and AP.

Additional information

Funding

Funder	Grant reference	Author
King Abdullah University of Science and Technology (KAUST)	FIC/2010/09	Aswini K Panigrahi, Julius Lukeš, Miroslav Oborník, Arnab Pain
Council of Scientific and Industrial Research	BSC0124	Dhanasekaran Shanmugam
National Institute of Allergy and Infectious Diseases (NIAID)	HHSN272200900018C	Daniel E Neafsey
Australian Research Council (ARC)	DP120100599	Ross F Waller
Monash University		Christian Doerig
National Health and Medical Research Council (NHMRC)		Christian Doerig
Czech Science Foundation (Grantová agentura České republiky)	P506/12/1522, 13-33039S, P501/12/G055	Jan Michálek, Jaromír Cihlár, Aleš Tomčala, Julius Lukeš, Miroslav Oborník


The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

YHW, Performed gene annotations; environmental perturbation and transcriptome profiling; invasion pathway and apical complex analysis; DNA- and RNA-binding protein analysis; cross-species transcriptome analysis, Coordinated the genome and transcriptome analyses, Wrote the initial manuscript, Wrote the final manuscript; HA, Performed gene annotations; invasion pathway and apical complex analysis; subcellular targeting prediction; curation of extracellular and cytoskeletal genes in apicomplexans; TDO, Performed genome assembly and gene prediction; gene annotations; gene family analysis; CMK, Performed endomembrane trafficking system analysis; MK, JC, Performed genome analysis; JM, Performed fatty acid biosynthesis; AS, Performed generation and maintenance of specimen, DNA and RNA extractions, library preparation and sequencing; validation of predicted genes; manual curation for gene predictions; DS, Performed global metabolic analysis, Commented and edited on versions of the draft manuscript; AT, Performed generation and maintenance of specimen, DNA and RNA extractions, library preparation and sequencing; manual curation for gene predictions; environmental perturbation and transcriptome profiling; AV, Performed transposable element analysis; SA, EH, JJ, MN, Performed generation and maintenance of specimen, DNA and RNA extractions, library preparation and sequencing; AB, Performed gene annotation validations; JC, Performed generation and maintenance

of specimen, DNA and RNA extractions, library preparation and sequencing; heme pathway and phylogenetic analysis; PF, Performed analysis of chromerid metabolism; commented and edited on versions of the draft manuscript; SGG, performed generation and maintenance of specimen, DNA and RNA extractions, library preparation and sequencing; commented and edited on versions of the draft manuscript; AH, Performed urea pathway and phylogenetic analyses; NJK, NS, Performed validation of RAP's localization in mitochondria; FDM, DM-S, NDR, JW, Performed comparative genome analysis; TM, Performed gene structure analysis; RN, Performed genome validation, annotation, and submission; AKP, Conceived the project; validation of predicted genes; EP-R, AR, Performed manual curation for gene predictions; AT, Performed MS and gas chromatography on fatty acid synthesis; DEN, Performed generation and maintenance of specimen, DNA and RNA extractions, library preparation and sequencing; contributed some raw sequencing reads data; CD, Performed comparative genome analysis, Commented and edited on versions of the draft manuscript; CB, Performed transposable element analysis, Commented and edited on versions of the draft manuscript; PJK, Performed genome analysis, Commented and edited on versions of the draft manuscript; DSR, Global metabolic analysis, Commented and edited on versions of the draft manuscript; JBD, Performed endomembrane trafficking system analysis, Wrote the initial manuscript, Commented and edited on versions of the draft manuscript; TJT, Performed extracellular protein analysis, Wrote the initial manuscript, Commented and edited on versions of the draft manuscript; RFW, Performed validation of RAP's localization in mitochondria, Wrote the initial manuscript, Commented and edited on versions of the draft manuscript; JL, Conceived the project, Commented and edited on versions of the draft manuscript; MO, Conceived the project, Analysis of chromerid metabolism, Commented and edited on versions of the draft manuscript; AP, Conceived the project, Wrote the initial manuscript, Commented and edited on versions of the draft manuscript, Co-ordinated the project

Author ORCIDs

Yong H Woo,  <http://orcid.org/0000-0002-0338-6493>

Javier del Campo,  <http://orcid.org/0000-0002-5292-1421>

Arnab Pain,  <http://orcid.org/0000-0002-1755-2819>

Additional files

Supplementary files

- Supplementary file 1. Summary of the genome assembly and the annotated genes of *Chromera velia*, *Vitrella brassicaformis*. Details of transposable elements on the genome are shown in **Supplementary file 2**.
DOI: [10.7554/eLife.06974.026](https://doi.org/10.7554/eLife.06974.026)
- Supplementary file 2. Summary of transposable elements on the *Chromera velia* and *Vitrella brassicaformis* genomes.
DOI: [10.7554/eLife.06974.027](https://doi.org/10.7554/eLife.06974.027)
- Supplementary file 3. Genes encoding proteins involved in forming photosystems in *Chromera velia* and *Vitrella brassicaformis*.
DOI: [10.7554/eLife.06974.028](https://doi.org/10.7554/eLife.06974.028)
- Supplementary file 4. Genes encoding enzymes involved in heme biosynthesis in chromerids.
DOI: [10.7554/eLife.06974.029](https://doi.org/10.7554/eLife.06974.029)
- Supplementary file 5. Domains of extracellular proteins and example genes in chromerids. (a) Species abbreviations: *Perkinsus marinus*, *P. mar*; *Chromera velia*, *C. vel*; *Vitrella brassicaformis*, *V. bra*; and *Cryptosporidium parvum*, *C. par*. (b) Domain accession identifiers. Domain information can be retrieved at the NCBI Conserved Domain website: (<http://www.ncbi.nlm.nih.gov/cdd>). (c) At the time of publication this accession identifier was valid but the relevant entry could not be retrieved via the NCBI Conserved Domain website: (<http://www.ncbi.nlm.nih.gov/cdd>). (d) A domain having two cysteines and thus far found only as tandem arrays in proteins of *Chromera velia* (for example, Cvel_967). (e) Cysteine-rich domain found in *Cryptosporidium* oocyst wall proteins (COWP) and in coccidians. (f) Archaeal protease-type repeats first described in the *Cryptosporidium* predicted EC protein, cgd7_4560. The domain was previously described as 'A small domain with characteristically spaced cysteine residues that is fused to a papain-like protease domain in the secreted protein

AF1946 from *Archaeoglobus fulgidus* (Templeton et al., 2004a). (g) The domain was previously described as 'Domain typically with 6 cysteines, seen thus far mainly in animals with a few occurrences in plants. It is found in the sea anemone toxin metridin and fused to animal metal proteases, plant prolyl hydroxylases and is vastly expanded in the genome of *Caenorhabditis elegans* (Templeton et al., 2004a)'. (h) The domain was previously described as 'β-strand rich domain, predicted to form a β-sandwich structure that is found in bacterial secreted levanases and glucosidases (Templeton et al., 2004a)'.

DOI: [10.7554/eLife.06974.030](https://doi.org/10.7554/eLife.06974.030)

• Supplementary file 6. Mitochondrial localization predictions of selected RAP genes. Various algorithmic methods were used to identify candidates for experimental validations in *Toxoplasma*. Classifications are given in the column 'Loc' as M-mitochondria; S- secreted; O-others.

DOI: [10.7554/eLife.06974.031](https://doi.org/10.7554/eLife.06974.031)

Major datasets

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset ID and/or URL	Database, license, and accessibility information
Yong H Woo et al,	2015	Chromera velia DNA and RNA sequencing reads	http://www.ebi.ac.uk/ena/data/view/ERP006228	Publicly available at the EBI European Nucleotide Archive (Accession no: ERP006228).
Yong H Woo et al,	2015	Vitrella brassicaformis DNA and RNA sequencing reads	http://www.ebi.ac.uk/ena/data/view/ERP006229	Publicly available at the EBI European Nucleotide Archive (Accession no: ERP006229).
Yong H Woo et al,	2015	Chromera Velia Genome Assembly	http://www.ebi.ac.uk/ena/data/view/CDMZ00000000	Publicly available at the EBI European Nucleotide Archive (Accession no: CDMZ00000000).
Yong H Woo et al,	2015	Vitrella brassicaformis Genome Assembly	http://www.ebi.ac.uk/ena/data/view/CDMY01000000	Publicly available at the EBI European Nucleotide Archive (Accession no: CDMY01000000).

The following previously published dataset was used:

Author(s)	Year	Dataset title	Dataset ID and/or URL	Database, license, and accessibility information
Hu G, Cabrera A, Kono M, Mok S, Chaal BK, Haase S, Engelberg K, Cheemadan S, Spielmann T, Preiser PR, Gilberger TW, Bozdech Z	2010	Perturbation Transcriptome of Plasmodium falciparum	http://www.nature.com/nbt/journal/v28/n1/extended_data_tables/nbt.1597-S2.xls	Publicly available as a part of published dataset.

References

- Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S, Buck GA, Xu P, Bankier AT, Dear PH, Konfortov BA, Spriggs HF, Iyer L, Anantharaman V, Aravind L, Kapur V. 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**:441–445. doi: [10.1126/science.1094786](https://doi.org/10.1126/science.1094786).
- Adl SM, Leander BS, Simpson AG, Archibald JM, Anderson OR, Bass D, Bowser SS, Brugerolle G, Farmer MA, Karpov S, Kolisko M, Lane CE, Lodge DJ, Mann DG, Meisterfeld R, Mendoza L, Moestrup O, Mozley-Standridge SE, Smirnov AV, Spiegel F. 2007. Diversity, nomenclature, and taxonomy of protists. *Systematic Biology* **56**: 684–689. doi: [10.1080/10635150701494127](https://doi.org/10.1080/10635150701494127).
- Allen AE, Dupont CL, Obornik M, Horak A, Nunes-Nesi A, Mccrow JP, Zheng H, Johnson DA, Hu H, Fernie AR, Bowler C. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **473**: 203–207. doi: [10.1038/nature10074](https://doi.org/10.1038/nature10074).
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**:403–410. doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Anantharaman V, Iyer LM, Balaji S, Aravind L. 2007. Adhesion molecules and other secreted host-interaction determinants in *Apicomplexa*: insights from comparative genomics. *International Review of Cytology* **262**:1–74. doi: [10.1016/S0074-7696\(07\)62001-4](https://doi.org/10.1016/S0074-7696(07)62001-4).

- Aurrecoechea C**, Barreto A, Brestelli J, Brunk BP, Cade S, Doherty R, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Hu S, Iodice J, Kissinger JC, Kraemer ET, Li W, Pinney DF, Pitts B, Roos DS, Srinivasamoorthy G, Stoeckert CJ Jr, Wang H, Warrenfeltz S. 2013. EuPathDB: the eukaryotic pathogen database. *Nucleic Acids Research* **41**:D684–D691. doi: [10.1093/nar/gks1113](https://doi.org/10.1093/nar/gks1113).
- Bahl A**, Brunk B, Crabtree J, Fraunholz MJ, Gajria B, Grant GR, Ginsburg H, Gupta D, Kissinger JC, Labo P, Li L, Mailman MD, Milgram AJ, Pearson DS, Roos DS, Schug J, Stoeckert CJ Jr, Whetzel P. 2003. PlasmoDB: the *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic Acids Research* **31**:212–215. doi: [10.1093/nar/gkg081](https://doi.org/10.1093/nar/gkg081).
- Balaji S**, Babu MM, Iyer LM, Aravind L. 2005. Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Research* **33**:3994–4006. doi: [10.1093/nar/gki709](https://doi.org/10.1093/nar/gki709).
- Bannai H**, Tamada Y, Maruyama O, Nakai K, Miyano S. 2002. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* **18**:298–305. doi: [10.1093/bioinformatics/18.2.298](https://doi.org/10.1093/bioinformatics/18.2.298).
- Bao Z**, Eddy SR. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research* **12**:1269–1276. doi: [10.1101/gr.88502](https://doi.org/10.1101/gr.88502).
- Baum J**, Gilberger TW, Frischknecht F, Meissner M. 2008. Host-cell invasion by malaria parasites: insights from *Plasmodium* and *Toxoplasma*. *Trends in Parasitology* **24**:557–563. doi: [10.1016/j.pt.2008.08.006](https://doi.org/10.1016/j.pt.2008.08.006).
- Baum J**, Papenfuss AT, Baum B, Speed TP, Cowman AF. 2006. Regulation of apicomplexan actin-based motility. *Nature Reviews. Microbiology* **4**:621–628. doi: [10.1038/nrmicro1465](https://doi.org/10.1038/nrmicro1465).
- Bendtsen JD**, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**:783–795. doi: [10.1016/j.jmb.2004.05.028](https://doi.org/10.1016/j.jmb.2004.05.028).
- Benson G**. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**: 573–580. doi: [10.1093/nar/27.2.573](https://doi.org/10.1093/nar/27.2.573).
- Bentley DR**, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Catenazzi E, Chiara M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtkova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczky C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurler ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**:53–59. doi: [10.1038/nature07517](https://doi.org/10.1038/nature07517).
- Bernal A**, Crammer K, Hatzigeorgiou A, Pereira F. 2007. Global discriminative learning for higher-accuracy computational gene prediction. *PLoS Computational Biology* **3**:e54. doi: [10.1371/journal.pcbi.0030054](https://doi.org/10.1371/journal.pcbi.0030054).
- Boetzer M**, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**:578–579. doi: [10.1093/bioinformatics/btq683](https://doi.org/10.1093/bioinformatics/btq683).
- Boetzer M**, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biology* **13**:R56. doi: [10.1186/gb-2012-13-6-r56](https://doi.org/10.1186/gb-2012-13-6-r56).
- Bolstad BM**, Irizarry RA, Astrand M, Speed TP. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**:185–193. doi: [10.1093/bioinformatics/19.2.185](https://doi.org/10.1093/bioinformatics/19.2.185).
- Bougdour A**, Tardieux I, Hakimi MA. 2014. *Toxoplasma* exports dense granule proteins beyond the vacuole to the host cell nucleus and rewires the host genome expression. *Cellular Microbiology* **16**:334–343. doi: [10.1111/cmi.12255](https://doi.org/10.1111/cmi.12255).
- Box GE**, Stuart Hunter J, Hunter WG. 2005. *Statistics for experimenters: design, innovation, and discovery*. 2nd ed., Wiley series in probability and statistics. Hoboken: Wiley-Interscience.
- Bullen HE**, Tonkin CJ, O'Donnell RA, Tham WH, Papenfuss AT, Gould S, Cowman AF, Crabb BS, Gilson PR. 2009. A novel family of Apicomplexan glideosome-associated proteins with an inner membrane-anchoring role. *The Journal of Biological Chemistry* **284**:25353–25363. doi: [10.1074/jbc.M109.036772](https://doi.org/10.1074/jbc.M109.036772).
- Campbell TL**, de Silva EK, Olszewski KL, Elemento O, Llinas M. 2010. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathogens* **6**: e1001165. doi: [10.1371/journal.ppat.1001165](https://doi.org/10.1371/journal.ppat.1001165).
- Capella-Gutierrez S**, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973. doi: [10.1093/bioinformatics/btp348](https://doi.org/10.1093/bioinformatics/btp348).
- Caron F**, Meyer E. 1989. Molecular basis of surface antigen variation in paramecia. *Annual Review of Microbiology* **43**:23–42. doi: [10.1146/annurev.mi.43.100189.000323](https://doi.org/10.1146/annurev.mi.43.100189.000323).

- Chaudhry F**, Little K, Talarico L, Quintero-Monzon O, Goode BL. 2010. A central role for the WH2 domain of Srv2/CAP in recharging actin monomers to drive actin turnover in vitro and in vivo. *Cytoskeleton* **67**:120–133. doi: [10.1002/cm.20429](https://doi.org/10.1002/cm.20429).
- Chen F**, Mackey AJ, Stoeckert CJ Jr, Roos DS. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Research* **34**:D363–D368. doi: [10.1093/nar/gkj123](https://doi.org/10.1093/nar/gkj123).
- Claros MG**, Vincens P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. *European Journal of Biochemistry* **241**:779–786. doi: [10.1111/j.1432-1033.1996.00779.x](https://doi.org/10.1111/j.1432-1033.1996.00779.x).
- Coppens I**, Sinai AP, Joiner KA. 2000. *Toxoplasma gondii* exploits host low-density lipoprotein receptor-mediated endocytosis for cholesterol acquisition. *The Journal of Cell Biology* **149**:167–180. doi: [10.1083/jcb.149.1.167](https://doi.org/10.1083/jcb.149.1.167).
- Csuros M**. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**:1910–1912. doi: [10.1093/bioinformatics/btq315](https://doi.org/10.1093/bioinformatics/btq315).
- Cumbo VR**, Baird AH, Moore RB, Negri AP, Neilan BA, Salih A, van Oppen MJ, Wang Y, Marquis CP. 2013. *Chromera velia* is endosymbiotic in larvae of the reef corals *Acropora digitifera* and *A. tenuis*. *Protist* **164**:237–244. doi: [10.1016/j.protis.2012.08.003](https://doi.org/10.1016/j.protis.2012.08.003).
- Danne JC**, Gornik SG, Macrae JI, McConville MJ, Waller RF. 2013. Alveolate mitochondrial metabolic evolution: dinoflagellates force reassessment of the role of parasitism as a driver of change in apicomplexans. *Molecular Biology and Evolution* **30**:123–139. doi: [10.1093/molbev/mss205](https://doi.org/10.1093/molbev/mss205).
- Du P**, Li T, Wang X. 2011. Recent progress in predicting protein sub-cellular locations. *Expert Review of Proteomics* **8**:391–404. doi: [10.1586/epr.11.20](https://doi.org/10.1586/epr.11.20).
- Edgar RC**. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**:1792–1797. doi: [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340).
- Edgar RC**, Myers EW. 2005. PILER: identification and classification of genomic repeats. *Bioinformatics* **21**(Suppl 1): i152–i158. doi: [10.1093/bioinformatics/bti1003](https://doi.org/10.1093/bioinformatics/bti1003).
- Eisen JA**, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM, Tallon LJ, Delcher AL, Salzberg SL, Silva JC, Haas BJ, Majoros WH, Farzad M, Carlton JM, Smith RK Jr, Garg J, Pearlman RE, Karrer KM, Sun L, Manning G, Elde NC, Turkewitz AP, Asai DJ, Wilkes DE, Wang Y, Cai H, Collins K, Stewart BA, Lee SR, Wilamowska K, Weinberg Z, Ruzzo WL, Wloga D, Gaertig J, Frankel J, Tsao CC, Gorovsky MA, Keeling PJ, Waller RF, Patron NJ, Cherry JM, Stover NA, Krieger CJ, del Toro C, Ryder HF, Williamson SC, Barbeau RA, Hamilton EP, Orias E. 2006. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biology* **4**:e286. doi: [10.1371/journal.pbio.0040286](https://doi.org/10.1371/journal.pbio.0040286).
- Ellinghaus D**, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**:18. doi: [10.1186/1471-2105-9-18](https://doi.org/10.1186/1471-2105-9-18).
- Emanuelsson O**, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**:953–971. doi: [10.1038/nprot.2007.131](https://doi.org/10.1038/nprot.2007.131).
- Emanuelsson O**, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**:1005–1016. doi: [10.1006/jmbi.2000.3903](https://doi.org/10.1006/jmbi.2000.3903).
- Faith JJ**, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. 2007. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biology* **5**:e8. doi: [10.1371/journal.pbio.0050008](https://doi.org/10.1371/journal.pbio.0050008).
- Fedoroff NV**. 2012. Presidential address. Transposable elements, epigenetics, and genome evolution. *Science* **338**:758–767. doi: [10.1126/science.338.6108.758](https://doi.org/10.1126/science.338.6108.758).
- Fehrenbacher K**, Huckaba T, Yang HC, Boldogh I, Pon L. 2003. Actin comet tails, endosomes and endosymbionts. *The Journal of Experimental Biology* **206**:1977–1984. doi: [10.1242/jeb.00240](https://doi.org/10.1242/jeb.00240).
- Ferguson DJ**, Sahoo N, Pinches RA, Bumstead JM, Tomley FM, Gubbels MJ. 2008. MORN1 has a conserved role in asexual and sexual development across the Apicomplexa. *Eukaryot Cell* **7**:698–711. doi: [10.1128/EC.00021-08](https://doi.org/10.1128/EC.00021-08).
- Field HI**, Coulson RMR, Field MC. 2013. An automated graphics tool for comparative genomics: the Coulson plot generator. *BMC Bioinformatics* **14**:141. doi: [10.1186/1471-2105-14-141](https://doi.org/10.1186/1471-2105-14-141).
- Finn RD**, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Research* **42**:D222–D230. doi: [10.1093/nar/gkt1223](https://doi.org/10.1093/nar/gkt1223).
- Finn RD**, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* **39**:W29–W37. doi: [10.1093/nar/gkr367](https://doi.org/10.1093/nar/gkr367).
- Flegontov P**, Michalek J, Janouskovec J, Lai DH, Jirku M, Hajduskova E, Tomcala A, Otto TD, Keeling PJ, Pain A, Obornik M, Lukeš J. 2015. Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. *Molecular Biology and Evolution* **32**:1115–1131. doi: [10.1093/molbev/msv021](https://doi.org/10.1093/molbev/msv021).
- Flueck C**, Bartfai R, Niederwieser I, Witmer K, Alako BT, Moes S, Bozdech Z, Jenoe P, Stunnenberg HG, Voss TS. 2010. A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS Pathogens* **6**:e1000784. doi: [10.1371/journal.ppat.1000784](https://doi.org/10.1371/journal.ppat.1000784).
- Flutre T**, Duprat E, Feuillet C, Quesneville H. 2011. Considering transposable element diversification in de novo annotation approaches. *PLoS ONE* **6**:e16526. doi: [10.1371/journal.pone.0016526](https://doi.org/10.1371/journal.pone.0016526).
- Folch J**, Lees M, Sloane Stanley GH. 1957. A simple method for the isolation and purification of total lipides from animal tissues. *The Journal of Biological Chemistry* **226**:497–509.
- Foth BJ**, Goedecke MC, Soldati D. 2006. New insights into myosin evolution and classification. *Proceedings of the National Academy of Sciences of USA* **103**:3681–3686. doi: [10.1073/pnas.0506307103](https://doi.org/10.1073/pnas.0506307103).
- Francia ME**, Jordan CN, Patel JD, Sheiner L, Demerly JL, Fellows JD, de Leon JC, Morrisette NS, Dubremetz JF, Stripen B. 2012. Cell division in Apicomplexan parasites is organized by a homolog of the striated rootlet fiber of algal flagella. *PLoS Biology* **10**:e1001444. doi: [10.1371/journal.pbio.1001444](https://doi.org/10.1371/journal.pbio.1001444).

- Frenal K, Polonais V, Marq JB, Stratmann R, Limenitakis J, Soldati-Favre D. 2010. Functional dissection of the apicomplexan glideosome molecular architecture. *Cell Host & Microbe* **8**:343–357. doi: [10.1016/j.chom.2010.09.002](https://doi.org/10.1016/j.chom.2010.09.002).
- Frenal K, Soldati-Favre D. 2009. Role of the parasite and host cytoskeleton in Apicomplexa parasitism. *Cell Host & Microbe* **5**:602–611. doi: [10.1016/j.chom.2009.05.013](https://doi.org/10.1016/j.chom.2009.05.013).
- Gajria B, Bahl A, Brestelli J, Dommer J, Fischer S, Gao X, Heiges M, Iodice J, Kissinger JC, Mackey AJ, Pinney DF, Roos DS, Stoeckert CJ Jr, Wang H, Brunk BP. 2008. ToxoDB: an integrated *Toxoplasma gondii* database resource. *Nucleic Acids Research* **36**:D553–D556. doi: [10.1093/nar/gkm981](https://doi.org/10.1093/nar/gkm981).
- Gandhi M, Jangi M, Goode BL. 2010. Functional surfaces on the actin-binding protein coronin revealed by systematic mutagenesis. *The Journal of Biological Chemistry* **285**:34899–34908. doi: [10.1074/jbc.M110.171496](https://doi.org/10.1074/jbc.M110.171496).
- Gerstein MB, Rozowsky J, Yan KK, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ, Pei B, Harmanci AO, Duff MO, Djebali S, Alexander RP, Alver BH, Auerbach R, Bell K, Bickel PJ, Boeck ME, Boley NP, Booth BW, Cherbas L, Cherbas P, Di C, Dobin A, Drenkow J, Ewing B, Fang G, Fastuca M, Feingold EA, Frankish A, Gao G, Good PJ, Guigo R, Hammonds A, Harrow J, Hoskins RA, Howald C, Hu L, Huang H, Hubbard TJ, Huynh C, Jha S, Kasper D, Kato M, Kaufman TC, Kitchen RR, Ladewig E, Lagarde J, Lai E, Leng J, Lu Z, MacCoss M, May G, McWhirter R, Merrihew G, Miller DM, Mortazavi A, Murad R, Oliver B, Olson S, Park PJ, Pazin MJ, Perrimon N, Pervouchine D, Reinke V, Reymond A, Robinson G, Samsonova A, Saunders GI, Schlesinger F, Sethi A, Slack FJ, Spencer WC, Stoiber MH, Strasbourger P, Tanzer A, Thompson OA, Wan KH, Wang G, Wang H, Watkins KL, Wen J, Wen K, Xue C, Yang L, Yip K, Zaleski C, Zhang Y, Zheng H, Brenner SE, Graveley BR, Celniker SE, Gingeras TR, Waterston R. 2014. Comparative analysis of the transcriptome across distant species. *Nature* **512**:445–448. doi: [10.1038/nature13424](https://doi.org/10.1038/nature13424).
- Gordon JL, Sibley LD. 2005. Comparative genome analysis reveals a conserved family of actin-like proteins in apicomplexan parasites. *BMC Genomics* **6**:179. doi: [10.1186/1471-2164-6-179](https://doi.org/10.1186/1471-2164-6-179).
- Gournier H, Goley ED, Niederstrasser H, Trinh T, Welch MD. 2001. Reconstitution of human Arp2/3 complex reveals critical roles of individual subunits in complex structure and activity. *Molecular Cell* **8**:1041–1052. doi: [10.1016/S1097-2765\(01\)00393-8](https://doi.org/10.1016/S1097-2765(01)00393-8).
- Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27**:221–224. doi: [10.1093/molbev/msp259](https://doi.org/10.1093/molbev/msp259).
- Gschloessl B, Guermeur Y, Cock JM. 2008. HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinformatics* **9**:393. doi: [10.1186/1471-2105-9-393](https://doi.org/10.1186/1471-2105-9-393).
- Hafner MS, Sudman PD, Villablanca FX, Spradling TA, Demastes JW, Nadler SA. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* **265**:1087–1090. doi: [10.1126/science.8066445](https://doi.org/10.1126/science.8066445).
- Hager KM, Striepen B, Tilney LG, Roos DS. 1999. The nuclear envelope serves as an intermediary between the ER and Golgi complex in the intracellular parasite *Toxoplasma gondii*. *Journal of Cell Science* **112**:2631–2638. doi: [10.2307/2685478](https://doi.org/10.2307/2685478).
- Hintze JL, Nelson RD. 1998. Violin plots: a box plot-density trace synergism. *American Statistician* **52**:181–184. doi: [10.2307/2685478](https://doi.org/10.2307/2685478).
- Hirst J, Barlow LD, Francisco GC, Sahlender DA, Seaman MNJ, Dacks JB, Robinson MS. 2011. The fifth adaptor protein complex. *PLOS Biology* **9**:e1001170. doi: [10.1371/journal.pbio.1001170](https://doi.org/10.1371/journal.pbio.1001170).
- Hsiao CH, Luisa Hiller N, Haldar K, Knoll LJ. 2013. A HT/PEXEL motif in *Toxoplasma dense granule* proteins is a signal for protein cleavage but not export into the host cell. *Traffic* **14**:519–531. doi: [10.1111/tra.12049](https://doi.org/10.1111/tra.12049).
- Hu G, Cabrera A, Kono M, Mok S, Chaal BK, Haase S, Engelberg K, Cheemadan S, Spielmann T, Preiser PR, Gilberger TW, Bozdech Z. 2010. Transcriptional profiling of growth perturbations of the human malaria parasite *Plasmodium falciparum*. *Nature Biotechnology* **28**:91–98. doi: [10.1038/nbt.1597](https://doi.org/10.1038/nbt.1597).
- Hu K, Johnson J, Florens L, Fraunholz M, Suravajjala S, DiLullo C, Yates J, Roos DS, Murray JM. 2006. Cytoskeletal components of an invasion machine—the apical complex of *Toxoplasma gondii*. *PLOS Pathogens* **2**:e13. doi: [10.1371/journal.ppat.0020013](https://doi.org/10.1371/journal.ppat.0020013).
- Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD. 2013. REAPR: a universal tool for genome assembly evaluation. *Genome Biology* **14**:R47. doi: [10.1186/gb-2013-14-5-r47](https://doi.org/10.1186/gb-2013-14-5-r47).
- Janouškovec J, Horák A, Barott KL, Rohwer FL, Keeling PJ. 2012. Global analysis of plastid diversity reveals apicomplexan-related lineages in coral reefs. *Current Biology* **22**:R518–R519. doi: [10.1016/j.cub.2012.04.047](https://doi.org/10.1016/j.cub.2012.04.047).
- Janouškovec J, Horák A, Barott KL, Rohwer FL, Keeling PJ. 2013. Environmental distribution of coral-associated relatives of apicomplexan parasites. *The ISME Journal* **7**:444–447. doi: [10.1038/ismej.2012.129](https://doi.org/10.1038/ismej.2012.129).
- Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ. 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proceedings of the National Academy of Sciences of USA* **107**:10949–10954. doi: [10.1073/pnas.1003335107](https://doi.org/10.1073/pnas.1003335107).
- Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Kolisko M, Mylnikov AP, Keeling PJ. 2015. Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proceedings of the National Academy of Sciences of USA*. doi: [10.1073/pnas.1423790112](https://doi.org/10.1073/pnas.1423790112).
- Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**:431. doi: [10.1186/1471-2105-11-431](https://doi.org/10.1186/1471-2105-11-431).
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110**:462–467. doi: [10.1159/000084979](https://doi.org/10.1159/000084979).
- Jurka J, Klonowski P, Dagman V, Pelton P. 1996. CENSOR—a program for identification and elimination of repetitive elements from DNA sequences. *Computers and Chemistry* **20**:119–121. doi: [10.1016/S0097-8485\(96\)80013-1](https://doi.org/10.1016/S0097-8485(96)80013-1).
- Kafsack BF, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, Williams AE, Drought LG, Kwiatkowski DP, Baker DA, Cortes A, Llinas M. 2014. A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* **507**:248–252. doi: [10.1038/nature12920](https://doi.org/10.1038/nature12920).

- Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* **42**:D199–D205. doi: [10.1093/nar/gkt1076](https://doi.org/10.1093/nar/gkt1076).
- Kaneko I, Iwanaga S, Kato T, Kobayashi I, Yuda M. 2015. Genome-wide identification of the target genes of AP2-O, a plasmodium AP2-family transcription factor. *PLOS Pathogens* **11**:e1004905. doi: [10.1371/journal.ppat.1004905](https://doi.org/10.1371/journal.ppat.1004905).
- Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretailade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M, Weissenbach J, Vivares CP. 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**:450–453. doi: [10.1038/35106579](https://doi.org/10.1038/35106579).
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**:772–780. doi: [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- Kawase O, Nishikawa Y, Bannai H, Zhang H, Zhang G, Jin S, Lee EG, Xuan X. 2007. Proteomic analysis of calcium-dependent secretion in *Toxoplasma gondii*. *Proteomics* **7**:3718–3725. doi: [10.1002/pmic.200700362](https://doi.org/10.1002/pmic.200700362).
- Keeling PJ. 2004. Reduction and compaction in the genome of the apicomplexan parasite *Cryptosporidium parvum*. *Developmental Cell* **6**:614–616.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**:R36. doi: [10.1186/gb-2013-14-4-r36](https://doi.org/10.1186/gb-2013-14-4-r36).
- Klinger CM, Klute MJ, Dacks JB. 2013a. Comparative genomic analysis of multi-subunit tethering complexes demonstrates an ancient pan-eukaryotic complement and sculpting in Apicomplexa. *PLOS ONE* **8**:e76278. doi: [10.1371/journal.pone.0076278](https://doi.org/10.1371/journal.pone.0076278).
- Klinger CM, Nisbet RE, Ouologuem DT, Roos DS, Dacks JB. 2013b. Cryptic organelle homology in apicomplexan parasites: insights from evolutionary cell biology. *Current Opinion in Microbiology* **16**:424–431. doi: [10.1016/j.mib.2013.07.015](https://doi.org/10.1016/j.mib.2013.07.015).
- Kolpakov R, Bana G, Kucherov G. 2003. mreps: efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Research* **31**:3672–3678. doi: [10.1093/nar/gkg617](https://doi.org/10.1093/nar/gkg617).
- Kono M, Herrmann S, Loughran NB, Cabrera A, Engelberg K, Lehmann C, Sinha D, Prinz B, Ruch U, Heussler V, Spielmann T, Parkinson J, Gilberger TW. 2012. Evolution and architecture of the inner membrane complex in asexual and sexual stages of the malaria parasite. *Molecular Biology and Evolution* **29**:2113–2132. doi: [10.1093/molbev/mss081](https://doi.org/10.1093/molbev/mss081).
- Kořený L, Oborník M. 2011. Sequence evidence for the presence of two tetrapyrrole pathways in *Euglena gracilis*. *Genome Biology and Evolution* **3**:359–364. doi: [10.1093/gbe/evr029](https://doi.org/10.1093/gbe/evr029).
- Kořený L, Sobotka R, Janouskovec J, Keeling PJ, Oborník M. 2011. Tetrapyrrole synthesis of photosynthetic chromerids is likely homologous to the unusual pathway of apicomplexan parasites. *The Plant Cell* **23**:3454–3462. doi: [10.1105/tpc.111.089102](https://doi.org/10.1105/tpc.111.089102).
- Koumandou VL, Dacks JB, Coulson RM, Field MC. 2007. Control systems for membrane fusion in the ancestral eukaryote; evolution of tethering complexes and SM proteins. *BMC Evolutionary Biology* **7**:29. doi: [10.1186/1471-2148-7-29](https://doi.org/10.1186/1471-2148-7-29).
- Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nature Methods* **6**:291–295. doi: [10.1038/nmeth.1311](https://doi.org/10.1038/nmeth.1311).
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**:567–580. doi: [10.1006/jmbi.2000.4315](https://doi.org/10.1006/jmbi.2000.4315).
- Kucera K, Koblansky AA, Saunders LP, Frederick KB, De La Cruz EM, Ghosh S, Modis Y. 2010. Structure-based analysis of *Toxoplasma gondii* profilin: a parasite-specific motif is required for recognition by Toll-like receptor 11. *Journal of Molecular Biology* **403**:616–629. doi: [10.1016/j.jmb.2010.09.022](https://doi.org/10.1016/j.jmb.2010.09.022).
- Kursula I, Kursula P, Ganter M, Panjekar S, Matuschewski K, Schuler H. 2008. Structural basis for parasite-specific functions of the divergent profilin of *Plasmodium falciparum*. *Structure* **16**:1638–1648. doi: [10.1016/j.str.2008.09.008](https://doi.org/10.1016/j.str.2008.09.008).
- Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* **21**:1095–1109. doi: [10.1093/molbev/msh112](https://doi.org/10.1093/molbev/msh112).
- Le SQ, Dang CC, Gascuel O. 2012. Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Molecular Biology and Evolution* **29**:2921–2936. doi: [10.1093/molbev/mss112](https://doi.org/10.1093/molbev/mss112).
- Lee I, Hong W. 2004. RAP—a putative RNA-binding domain. *Trends in Biochemical Sciences* **29**:567–570. doi: [10.1016/j.tibs.2004.09.005](https://doi.org/10.1016/j.tibs.2004.09.005).
- Leonardi R, Zhang YM, Rock CO, Jackowski S. 2005. Coenzyme A: back in action. *Progress in Lipid Research* **44**:125–153. doi: [10.1016/j.plipres.2005.04.001](https://doi.org/10.1016/j.plipres.2005.04.001).
- Leung KF, Dacks JB, Field MC. 2008. Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage. *Traffic* **9**:1698–1716. doi: [10.1111/j.1600-0854.2008.00797.x](https://doi.org/10.1111/j.1600-0854.2008.00797.x).
- Li H, Child MA, Bogoy M. 2012. Proteases as regulators of pathogenesis: examples from the Apicomplexa. *Biochimica et Biophysica Acta* **1824**:177–185. doi: [10.1016/j.bbapap.2011.06.002](https://doi.org/10.1016/j.bbapap.2011.06.002).
- Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**:2178–2189. doi: [10.1101/gr.1224503](https://doi.org/10.1101/gr.1224503).
- Lim L, McFadden GI. 2010. The evolution, metabolism and functions of the apicoplast. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **365**:749–763. doi: [10.1098/rstb.2009.0273](https://doi.org/10.1098/rstb.2009.0273).
- Liu J, Guo W. 2012. The exocyst complex in exocytosis and cell migration. *Protoplasma* **249**:587–597. doi: [10.1007/s00709-011-0330-1](https://doi.org/10.1007/s00709-011-0330-1).

- Logan-Klumper FJ**, de Silva N, Boehme U, Rogers MB, Velarde G, McQuillan JA, Carver T, Aslett M, Olsen C, Subramanian S, Phan I, Farris C, Mitra S, Ramasamy G, Wang H, Tivey A, Jackson A, Houston R, Parkhill J, Holden M, Harb OS, Brunk BP, Myler PJ, Roos D, Carrington M, Smith DF, Hertz-Fowler C, Berriman M. 2012. GeneDB—an annotation database for pathogens. *Nucleic Acids Research* **40**:D98–D108. doi: [10.1093/nar/gkr1032](https://doi.org/10.1093/nar/gkr1032).
- Machesky LM**, Atkinson SJ, Ampe C, Vandekerckhove J, Pollard TD. 1994. Purification of a cortical complex containing two unconventional actins from *Acanthamoeba* by affinity chromatography on profilin-agarose. *The Journal of Cell Biology* **127**:107–115. doi: [10.1083/jcb.127.1.107](https://doi.org/10.1083/jcb.127.1.107).
- Magnani E**, Sjolander K, Hake S. 2004. From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. *The Plant Cell* **16**:2265–2277. doi: [10.1105/tpc.104.023135](https://doi.org/10.1105/tpc.104.023135).
- Mazumdar J**, Striepen B. 2007. Make it or take it: fatty acid metabolism of apicomplexan parasites. *Eukaryot Cell* **6**:1727–1735. doi: [10.1128/EC.00255-07](https://doi.org/10.1128/EC.00255-07).
- McFadden GI**, Reith ME, Munholland J, Lang-Unnasch N. 1996. Plastid in human parasites. *Nature* **381**:482. doi: [10.1038/381482a0](https://doi.org/10.1038/381482a0).
- Meyer PE**, Lafitte F, Bontempi G. 2008. minet: a R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics* **9**:461. doi: [10.1186/1471-2105-9-461](https://doi.org/10.1186/1471-2105-9-461).
- Miranda K**, Pace DA, Cintron R, Rodrigues JCF, Fang J, Smith A, Rohloff P, Coelho E, de Haas F, de Souza W, Coppens I, Sibley LD, Moreno SNJ. 2010. Characterization of a novel organelle in *Toxoplasma gondii* with similar composition and function to the plant vacuole. *Molecular Microbiology* **76**:1358–1375. doi: [10.1111/j.1365-2958.2010.07165.x](https://doi.org/10.1111/j.1365-2958.2010.07165.x).
- Moore RB**, Obornik M, Janouškovec J, Chrudimský T, Vancová M, Green DH, Wright SW, Davies NW, Bolch CJ, Heimann K, Slapeta J, Hoegh-Guldberg O, Logsdon JM, Carter DA. 2008. A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**:959–963. doi: [10.1038/nature06635](https://doi.org/10.1038/nature06635).
- Moriya Y**, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35**:W182–W185. doi: [10.1093/nar/gkm321](https://doi.org/10.1093/nar/gkm321).
- Morrison HG**, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, Davids BJ, Dawson SC, Elmendorf HG, Hehl AB, Holder ME, Huse SM, Kim UU, Lasek-Nesselquist E, Manning G, Nigam A, Nixon JE, Palm D, Passamaneck NE, Prabhu A, Reich CI, Reiner DS, Samuelson J, Svard SG, Sogin ML. 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* **317**:1921–1926. doi: [10.1126/science.1143837](https://doi.org/10.1126/science.1143837).
- Morrisette NS**, Sibley LD. 2002. Cytoskeleton of apicomplexan parasites. *Microbiology and Molecular Biology Reviews* **66**:21–38. doi: [10.1128/MMBR.66.1.21-38.2002](https://doi.org/10.1128/MMBR.66.1.21-38.2002).
- Mullins RD**, Stafford WF, Pollard TD. 1997. Structure, subunit topology, and actin-binding activity of the Arp2/3 complex from *Acanthamoeba*. *The Journal of Cell Biology* **136**:331–343. doi: [10.1083/jcb.136.2.331](https://doi.org/10.1083/jcb.136.2.331).
- Mundwiler-Pachlatko E**, Beck HP. 2013. Maurer's clefts, the enigma of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of USA* **110**:19987–19994. doi: [10.1073/pnas.1309247110](https://doi.org/10.1073/pnas.1309247110).
- Mutwil M**, Klie S, Tohge T, Giorgi FM, Wilkins O, Campbell MM, Fernie AR, Usadel B, Nikoloski Z, Persson S. 2011. PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. *The Plant Cell* **23**:895–910. doi: [10.1105/tpc.111.083667](https://doi.org/10.1105/tpc.111.083667).
- Nevin WD**, Dacks JB. 2009. Repeated secondary loss of adaptin complex genes in the *Apicomplexa*. *Parasitology International* **58**:86–94. doi: [10.1016/j.parint.2008.12.002](https://doi.org/10.1016/j.parint.2008.12.002).
- Obornik M**, Modrý D, Lukeš M, Cernotiková-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. doi: [10.1016/j.protis.2011.09.001](https://doi.org/10.1016/j.protis.2011.09.001).
- Obornik 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. doi: [10.1016/j.protis.2010.02.004](https://doi.org/10.1016/j.protis.2010.02.004).
- Okamoto N**, Keeling PJ. 2014. The 3D structure of the apical complex and association with the flagellar apparatus revealed by serial TEM tomography in *Psammoma pacifica*, a distant relative of the *Apicomplexa*. *PLOS ONE* **9**:e84653. doi: [10.1371/journal.pone.0084653](https://doi.org/10.1371/journal.pone.0084653).
- Otto TD**, Sanders M, Berriman M, Newbold C. 2010. Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics* **26**:1704–1707. doi: [10.1093/bioinformatics/btq269](https://doi.org/10.1093/bioinformatics/btq269).
- Pawlowski J**, Audic S, Adl S, Bass D, Belbahri L, Berney C, Bowser SS, Cepicka I, Decelle J, Dunthorn M, Fiore-Donno AM, Gile GH, Holzmann M, Jahn R, Jirku M, Keeling PJ, Kostka M, Kudryavtsev A, Lara E, Lukeš J, Mann DG, Mitchell EA, Nitsche F, Romeralo M, Saunders GW, Simpson AG, Smirnov AV, Spouge JL, Stern RF, Stoeck T, Zimmermann J, Schindel D, de Vargas C. 2012. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLOS Biology* **10**:e1001419. doi: [10.1371/journal.pbio.1001419](https://doi.org/10.1371/journal.pbio.1001419).
- Pelletier L**, Stern C, Pypaert M, Sheff D. 2002. Golgi biogenesis in *Toxoplasma gondii*. *Nature* **418**:1–5. doi: [10.1038/nature00946](https://doi.org/10.1038/nature00946).
- Petersen TN**, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**:785–786. doi: [10.1038/nmeth.1701](https://doi.org/10.1038/nmeth.1701).
- Petsalaki EI**, Bagos PG, Litou ZI, Hamodrakas SJ. 2006. PredSL: a tool for the N-terminal sequence-based prediction of protein subcellular localization. *Genomics, Proteomics & Bioinformatics* **4**:48–55. doi: [10.1016/S1672-0229\(06\)60016-8](https://doi.org/10.1016/S1672-0229(06)60016-8).
- Pieperhoff MS**, Schmitt M, Ferguson DJ, Meissner M. 2013. The role of clathrin in post-Golgi trafficking in *Toxoplasma gondii*. *PLOS ONE* **8**:e77620. doi: [10.1371/journal.pone.0077620](https://doi.org/10.1371/journal.pone.0077620).
- Pombert JF**, Blouin NA, Lane C, Boucias D, Keeling PJ. 2014. A lack of parasitic reduction in the obligate parasitic green alga *Helicosporidium*. *PLOS Genetics* **10**:e1004355. doi: [10.1371/journal.pgen.1004355](https://doi.org/10.1371/journal.pgen.1004355).

- Pollard TD**, Borisy GG. 2003. Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **112**: 453–465. doi: [10.1016/S0092-8674\(03\)00120-X](https://doi.org/10.1016/S0092-8674(03)00120-X).
- Portman N**, Foster C, Walker G, Slapeta J. 2014. Evidence of intraflagellar transport and apical complex formation in a free-living relative of the Apicomplexa. *Eukaryot Cell* **13**:10–20. doi: [10.1128/EC.00155-13](https://doi.org/10.1128/EC.00155-13).
- Poulin B**, Patzewitz EM, Brady D, Silvie O, Wright MH, Ferguson DJ, Wall RJ, Whipple S, Guttery DS, Tate EW, Wickstead B, Holder AA, Tewari R. 2013. Unique apicomplexan IMC sub-compartment proteins are early markers for apical polarity in the malaria parasite. *Biology Open* **2**:1160–1170. doi: [10.1242/bio.20136163](https://doi.org/10.1242/bio.20136163).
- Quail MA**, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**:341. doi: [10.1186/1471-2164-13-341](https://doi.org/10.1186/1471-2164-13-341).
- Quesneville H**, Nouaud D, Anxolabehere D. 2003. Detection of new transposable element families in *Drosophila melanogaster* and *Anopheles gambiae* genomes. *Journal of Molecular Evolution* **57**(Suppl 1):S50–S59. doi: [10.1007/s00239-003-0007-2](https://doi.org/10.1007/s00239-003-0007-2).
- Quigg A**, Kotabova E, Jaresova J, Kana R, Setlik J, Sediva B, Komarek O, Prasil O. 2012. Photosynthesis in *Chromera velia* represents a simple system with high efficiency. *PLOS ONE* **7**:e47036. doi: [10.1371/journal.pone.0047036](https://doi.org/10.1371/journal.pone.0047036).
- Radke JB**, Lucas O, de Silva EK, Ma Y, Sullivan WJ Jr, Weiss LM, Llinas M, White MW. 2013. ApiAP2 transcription factor restricts development of the *Toxoplasma* tissue cyst. *Proceedings of the National Academy of Sciences of USA* **110**:6871–6876. doi: [10.1073/pnas.1300059110](https://doi.org/10.1073/pnas.1300059110).
- Raffaele S**, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews. Microbiology* **10**:417–430. doi: [10.1038/nrmicro2790](https://doi.org/10.1038/nrmicro2790).
- Ravindran S**, Boothroyd JC. 2008. Secretion of proteins into host cells by Apicomplexan parasites. *Traffic* **9**: 647–656. doi: [10.1111/j.1600-0854.2008.00723.x](https://doi.org/10.1111/j.1600-0854.2008.00723.x).
- Reid AJ**, Blake DP, Ansari HR, Billington K, Browne HP, Bryant J, Dunn M, Hung SS, Kawahara F, Miranda-Saavedra D, Malas TB, Mourier T, Naghra H, Nair M, Otto TD, Rawlings ND, Rivaille P, Sanchez-Flores A, Sanders M, Subramaniam C, Tay YL, Woo Y, Wu X, Barrell B, Dear PH, Doerig C, Gruber A, Ivens AC, Parkinson J, Rajandream MA, Shirley MW, Wan KL, Berriman M, Tomley FM, Pain A. 2014. Genomic analysis of the causative agents of coccidiosis in domestic chickens. *Genome Research* **24**:1676–1685. doi: [10.1101/gr.168955.113](https://doi.org/10.1101/gr.168955.113).
- Roiko MS**, Carruthers VB. 2009. New roles for perforins and proteases in apicomplexan egress. *Cellular Microbiology* **11**:1444–1452. doi: [10.1111/j.1462-5822.2009.01357.x](https://doi.org/10.1111/j.1462-5822.2009.01357.x).
- Roos DS**. 2005. Genetics. Themes and variations in apicomplexan parasite biology. *Science* **309**:72–73. doi: [10.1126/science.1115252](https://doi.org/10.1126/science.1115252).
- Russell K**, Hasenkamp S, Emes R, Horrocks P. 2013. Analysis of the spatial and temporal arrangement of transcripts over intergenic regions in the human malarial parasite *Plasmodium falciparum*. *BMC Genomics* **14**:267. doi: [10.1186/1471-2164-14-267](https://doi.org/10.1186/1471-2164-14-267).
- Rybakin V**, Clemen CS. 2005. Coronin proteins as multifunctional regulators of the cytoskeleton and membrane trafficking. *Bioessays* **27**:625–632. doi: [10.1002/bies.20235](https://doi.org/10.1002/bies.20235).
- Sakharkar KR**, Dhar PK, Chow VT. 2004. Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. *International Journal of Systematic and Evolutionary Microbiology* **54**: 1937–1941. doi: [10.1099/ijs.0.63090-0](https://doi.org/10.1099/ijs.0.63090-0).
- Shoguchi E**, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, Koyanagi R, Takeuchi T, Hisata K, Tanaka M, Fujiwara M, Hamada M, Seidi A, Fujie M, Usami T, Goto H, Yamasaki S, Arakaki N, Suzuki Y, Sugano S, Toyoda A, Kuroki Y, Fujiyama A, Medina M, Coffroth MA, Bhattacharya D, Satoh N. 2013. Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current Biology* **23**:1399–1408. doi: [10.1016/j.cub.2013.05.062](https://doi.org/10.1016/j.cub.2013.05.062).
- Stifflow CD**, Lefebvre PA. 2001. Assembly and motility of eukaryotic cilia and flagella. Lessons from *Chlamydomonas reinhardtii*. *Plant Physiology* **127**:1500–1507. doi: [10.1104/pp.010807](https://doi.org/10.1104/pp.010807).
- Simpson JT**, Durbin R. 2012. Efficient de novo assembly of large genomes using compressed data structures. *Genome Research* **22**:549–556. doi: [10.1101/gr.126953.111](https://doi.org/10.1101/gr.126953.111).
- Singh BK**, Sattler JM, Chatterjee M, Huttu J, Schuler H, Kursula I. 2011. Crystal structures explain functional differences in the two actin depolymerization factors of the malaria parasite. *The Journal of Biological Chemistry* **286**:28256–28264. doi: [10.1074/jbc.M111.211730](https://doi.org/10.1074/jbc.M111.211730).
- Sinha A**, Hughes KR, Modrzynska KK, Otto TD, Pfander C, Dickens NJ, Religa AA, Bushell E, Graham AL, Cameron R, Kafsack BF, Williams AE, Llinas M, Berriman M, Billker O, Waters AP. 2014. A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature* **507**:253–257. doi: [10.1038/nature12970](https://doi.org/10.1038/nature12970).
- Skillman KM**, Diraviyam K, Khan A, Tang K, Sept D, Sibley LD. 2011. Evolutionarily divergent, unstable filamentous actin is essential for gliding motility in apicomplexan parasites. *PLOS Pathogens* **7**:e1002280. doi: [10.1371/journal.ppat.1002280](https://doi.org/10.1371/journal.ppat.1002280).
- Soldati-Favre D**. 2008. Molecular dissection of host cell invasion by the apicomplexans: the glideosome. *Parasite* **15**:197–205.
- Stamatakis A**. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313. doi: [10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033).
- Stanke M**, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* **34**:W435–W439. doi: [10.1093/nar/gkl200](https://doi.org/10.1093/nar/gkl200).
- Stevens JM**, Galyov EE, Stevens MP. 2006. Actin-dependent movement of bacterial pathogens. *Nature Reviews. Microbiology* **4**:91–101. doi: [10.1038/nrmicro1320](https://doi.org/10.1038/nrmicro1320).
- Struck NS**, Herrmann S, Schmuck-Barkmann I, de Souza Dias S, Haase S, Cabrera AL, Treeck M, Bruns C, Langer C, Cowman AF, Marti M, Spielmann T, Gilberger TW. 2008. Spatial dissection of the cis- and trans-Golgi

- compartments in the malaria parasite *Plasmodium falciparum*. *Molecular Microbiology* **67**:1320–1330. doi: [10.1111/j.1365-2958.2008.06125.x](https://doi.org/10.1111/j.1365-2958.2008.06125.x).
- Stuart JM**, Segal E, Koller D, Kim SK. 2003. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**:249–255. doi: [10.1126/science.1087447](https://doi.org/10.1126/science.1087447).
- Sutak R**, Slapeta J, San Roman M, Camadro JM, Lesuisse E. 2010. Nonreductive iron uptake mechanism in the marine alveolate *Chromera velia*. *Plant Physiology* **154**:991–1000. doi: [10.1104/pp.110.159947](https://doi.org/10.1104/pp.110.159947).
- Tempel S**. 2012. Using and understanding RepeatMasker. *Methods in Molecular Biology* **859**:29–51. doi: [10.1007/978-1-61779-603-6_2](https://doi.org/10.1007/978-1-61779-603-6_2).
- Templeton TJ**, Iyer LM, Anantharaman V, Enomoto S, Abrahante JE, Subramanian GM, Hoffman SL, Abrahamsen MS, Aravind L. 2004a. Comparative analysis of Apicomplexa and genomic diversity in eukaryotes. *Genome Research* **14**:1686–1695. doi: [10.1101/gr.2615304](https://doi.org/10.1101/gr.2615304).
- Templeton TJ**, Lancto CA, Vigdorovich V, Liu C, London NR, Hadsall KZ, Abrahamsen MS. 2004b. The *Cryptosporidium* oocyst wall protein is a member of a multigene family and has a homolog in *Toxoplasma*. *Infection and Immunity* **72**:980–987. doi: [10.1128/IAI.72.2.980-987.2004](https://doi.org/10.1128/IAI.72.2.980-987.2004).
- Tenter AM**, Heckerth AR, Weiss LM. 2000. *Toxoplasma gondii*: from animals to humans. *International Journal for Parasitology* **30**:1217–1258. doi: [10.1016/S0020-7519\(00\)00124-7](https://doi.org/10.1016/S0020-7519(00)00124-7).
- Tomavo S**, Slomianny C, Meissner M, Carruthers VB. 2013. Protein trafficking through the endosomal system prepares intracellular parasites for a home invasion. *PLoS Pathogens* **9**:e1003629. doi: [10.1371/journal.ppat.1003629](https://doi.org/10.1371/journal.ppat.1003629).
- Trapnell C**, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**:562–578. doi: [10.1038/nprot.2012.016](https://doi.org/10.1038/nprot.2012.016).
- Trecek M**, Sanders JL, Elias JE, Boothroyd JC. 2011. The phosphoproteomes of *Plasmodium falciparum* and *Toxoplasma gondii* reveal unusual adaptations within and beyond the parasites' boundaries. *Cell Host & Microbe* **10**:410–419. doi: [10.1016/j.chom.2011.09.004](https://doi.org/10.1016/j.chom.2011.09.004).
- Tsai IJ**, Otto TD, Berriman M. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biology* **11**:R41. doi: [10.1186/gb-2010-11-4-r41](https://doi.org/10.1186/gb-2010-11-4-r41).
- Van de Peer Y**, Frickey T, Taylor J, Meyer A. 2002. Dealing with saturation at the amino acid level: a case study based on anciently duplicated zebrafish genes. *Gene* **295**:205–211. doi: [10.1016/S0378-1119\(02\)00689-3](https://doi.org/10.1016/S0378-1119(02)00689-3).
- van Dooren GG**, Kennedy AT, McFadden GI. 2012. The use and abuse of heme in apicomplexan parasites. *Antioxidants & Redox Signaling* **17**:634–656. doi: [10.1089/ars.2012.4539](https://doi.org/10.1089/ars.2012.4539).
- Wilson D**, Charoensawan V, Kummerfeld SK, Teichmann SA. 2008. DBD—taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Research* **36**:D88–D92.
- Wong W**, Webb AI, Olshina MA, Infusini G, Tan YH, Hanssen E, Catimel B, Suarez C, Condrón M, Angrisano F, Nebi T, Kovar DR, Baum J. 2014. A mechanism for actin filament severing by malaria parasite actin depolymerizing factor 1 via a low affinity binding interface. *The Journal of Biological Chemistry* **289**:4043–4054. doi: [10.1074/jbc.M113.523365](https://doi.org/10.1074/jbc.M113.523365).
- Woo Y**, Affourtit J, Daigle S, Viale A, Johnson K, Naggert J, Churchill G. 2004. A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms. *Journal of Biomolecular Techniques* **15**:276–284.
- Woo YH**, Li WH. 2011. Gene clustering pattern, promoter architecture, and gene expression stability in eukaryotic genomes. *Proceedings of the National Academy of Sciences of USA* **108**:3306–3311. doi: [10.1073/pnas.1100210108](https://doi.org/10.1073/pnas.1100210108).
- Wu TD**, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**:873–881. doi: [10.1093/bioinformatics/btq057](https://doi.org/10.1093/bioinformatics/btq057).
- Xu P**, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, Puiu D, Manque P, Akiyoshi D, Mackey AJ, Pearson WR, Dear PH, Bankier AT, Peterson DL, Abrahamsen MS, Kapur V, Tzipori S, Buck GA. 2004. The genome of *Cryptosporidium hominis*. *Nature* **431**:1107–1112. doi: [10.1038/nature02977](https://doi.org/10.1038/nature02977).
- Xu Z**, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* **35**:W265–W268. doi: [10.1093/nar/gkm286](https://doi.org/10.1093/nar/gkm286).
- Zahradnickova H**, Tomcala A, Berkova P, Schneedorferova I, Okrouhlik J, Simek P, Hodkova M. 2014. Cost effective, robust, and reliable coupled separation techniques for the identification and quantification of phospholipids in complex biological matrices: application to insects. *Journal of Separation Science* **37**:2062–2068.
- Zdobnov EM**, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**:847–848. doi: [10.1093/bioinformatics/17.9.847](https://doi.org/10.1093/bioinformatics/17.9.847).
- Zerbino DR**, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**:821–829. doi: [10.1101/gr.074492.107](https://doi.org/10.1101/gr.074492.107).

Appendix 1

Genome characteristics.

The statistics of the genome assembly and annotation are shown in **Supplementary file 1**. There was bacterial contamination in 20% and 80% of the sequence reads in *Chromera* and *Vitrella*, respectively. There was a high amount of low-complexity DNA sequence repeats and TEs in *Chromera* (**Supplementary file 1**). By various bioinformatics methods ('Materials and methods'), we generated assemblies containing 5953 and 1064 scaffolds for *Chromera* and *Vitrella*, respectively. The total number of predicted genes differed between *Chromera* and *Vitrella* primarily due to significant differences in TE gene content between the two chromerids but the number of expressed genes was similar (**Supplementary file 1**).

We examined how genomes of the chromerids and other species were organized (**Supplementary file 1**). The median gene length is roughly the same between the two chromerids. The number of introns in a given gene was similar between the chromerids, although the size of introns was larger in *Chromera* than in *Vitrella* (**Supplementary file 1**). Compared to these chromerids, the number of introns in Apicomplexa was drastically less, raising the possibility that introns were compacted and reduced during apicomplexan evolution, which would need to be confirmed with further detailed investigation. For many genes (13,912 and 17,569 respectively for *Chromera* and *Vitrella*), we were able to assign 5' and 3' UTRs, using strand-specific transcriptome (RNA-seq) data sets. The distance between the protein-coding genes in *Vitrella* was short (median 92 base-pairs (bp)), indicating compactness of its genome. On the other hand, such distance was longer in *Chromera* (median 989 bp). Determining whether the common ancestor of chromerids had a compact genome or not would require analysis of genomes from more closely related species. There are three possible orientations by closely spaced neighboring genes can be clustered, that is, those with short intergenic spaces between the gene boundaries: tandem, head-to-head, or tail-to-tail. In both *Chromera* and *Vitrella* genomes, closely spaced (<1000 bp) genes were in head-to-head orientation more often than expected by chance (data not shown). It was previously shown that many neighboring genes in head-to-head clusters showed correlated expressions across various conditions; however, most of the co-expressions were modest; instead, head-to-head clustering is a major mechanism for stabilizing transcription of genes in fundamental cellular processes rather than for co-regulating the two genes (**Woo and Li, 2011; Russell et al., 2013**). Head-to-head clustering probably provided evolutionary and regulatory stability to genes involved in fundamental cellular processes. Other related species had different gene orientations, for example, the dinoflagellate *Symbiodinium microtinum* has tandem clusters driven by tandem gene duplication (**Shoguchi et al., 2013**). Given the dynamic nature of genome organization, we propose that different groups of species evolved different strategies for genome organization (**Woo and Li, 2011**).

Repetitive sequences constitute a significant proportion of eukaryotic genomes (**Fedoroff, 2012**). Thus, they play a significant role in evolution of host genomes. Systematic TE clustering, classification, and annotation were performed on 1064 *Vitrella* scaffolds (72.7 Mb genome—72,700,666 bp) and 5953 *Chromera* scaffolds (193.6 Mb genome—193,664,168 bp) *Chromera*. In both species, Class I elements (**Tempel, 2012**) make up a larger proportion of the genome than Class II elements (**Tempel, 2012**) (**Supplementary file 2**). The RT domain variation shows that *Eimeria tenella* TEs grouped separately and are not related to chromerid TEs (**Supplementary file 2**), suggesting gains of TEs in *E. tenella* (**Reid et al., 2014**) independently from chromerids. *Vitrella* forms a separate clade in the phylogenetic analysis of the RT domains.

Appendix 2

Metabolism.

Materials and methods

Reconstructing global metabolic map based on KEGG

Global metabolic pathways were mapped to KEGG metabolic pathways for the predicted protein-coding gene sets for the 26 species. KEGG ortholog (KO) assignments for the respective proteomes were made using the KO identification tools available on the KEGG website (<http://genome.jp/tools/kaas>) (Moriya *et al.*, 2007; Kanehisa *et al.*, 2014), and then the assigned KO numbers were used to identify and map metabolic pathways using the 'Search and color pathway' tool available on the KEGG site (http://genome.jp/kegg/tool/map_pathway2.html). The output of this mapping exercise was then manually inspected to compile the set of enzymes present in all major metabolic pathways (Figure 2—source data 1).

As KEGG is very strict in mapping orthologs and assigning KO numbers (so as to minimize false positives), we found numerous pathway holes (missing enzymes), many of which were readily apparent as false negatives. In order to resolve this, we then resorted to identifying orthologs from OrthoMCL-DB for all the genomes compared here (Chen *et al.*, 2006). The resulting ortholog assignments were then used to manually verify presence/absence of missing enzymes for filling pathway holes where possible. This curated data, based on both KEGG and OrthoMCL assignments, were used to generate the final mapping of enzymes to pathways, and using this info a metabolic pathway network was drawn to represent all major pathways involved in carbohydrate, energy, fatty acid, lipid, isoprene, steroid, amino acid, nucleotide, cofactor, polyamine, and redox metabolism (Figure 2—figure supplement 3).

Based on the enzymes mapped, we calculated the completeness of metabolic pathways by comparing the fraction of enzymes present for each pathway in each species. The complete set of enzymes mapped to each pathway (originally taken from KEGG and further curated to eliminate non-specific entries) is given in column B of **Supplementary file 3**. The fractional values were then color-coded and the resulting data are shown in **Figure 2B**. In order to visualize the retention, loss or gain of higher level metabolic functions, the fraction of enzymes mapped to these pathways is indicated as a pie chart for hypothetical ancestors of selected apicomplexan groups and chromerids (Figure 2B). We used presence of enzymes across the species and the phylogenetic relationship to infer presence of enzymes in the hypothetical ancestors based on Dollo parsimony (Csuros, 2010). Dollo parsimony is based on an assumption that it is unlikely that the same enzymes were gained multiple times independently in different lineages.

Phylogeny of heme pathway enzymes, the urea pathway CPS and enzymes involved in fatty acid biosynthesis.

Predicted proteins from *Vitrella* (Chromera heme pathway is already published [Kořený and Oborník, 2011; Kořený *et al.*, 2011]) were searched for enzymes involved in the synthesis of tetrapyrroles (aminolevulinic acid [ALA] synthase, ALAS; ALA dehydratase, ALAD; Porphobilinogen deaminase, PBGD; Uroporphyrinogen synthase UROS; Uroporphyrinogen decarboxylase, UROD; Coproporphyrinogen oxidase, CPOX; Protoporphyrinogen oxidase, PPOX; and Ferrochelatase FeCH). All genes identified were aligned to the homologs available in public databases such as NCBI and JGI using Muscle (Edgar, 2004), with the alignments further edited in SeaView (Gouy *et al.*, 2010). The results from these analyses are shown in **Supplementary file 4**. The same procedure but searching in the predicted proteomes of both chromerids was used to construct the alignment of carbamoyl phosphate synthases (CPS). Genes coding for enzymes containing ketoacyl synthase domain were searched using BLAST. Functional domains were searched using InterProScan (Zdobnov and Apweiler, 2001). Phylogenetic trees of all investigated enzymes were constructed using the ML approach

(RAxML [Stamatakis, 2014]), Bayesian inference (PHYLOBAYES [Lartillot and Philippe, 2004]), and a method designed to deal with amino acid saturation (AsaturA [Van de Peer et al., 2002]). ML trees were computed under the gamma corrected LG4X model of evolution as implemented in RAxML 7.4.8a using the rapid-bootstrap optimization algorithm in 1000 replicates. Bayesian phylogeny was inferred using empirical site-heterogeneous model C40 as implemented in Phylobayes 3.2f. Two independent chains were run until they converged (i.e., maximum observed discrepancy was lower than 0.2), and the effective number of model parameters was at least 100 after the first 1/5 generation was omitted from topology and posterior probability inference.

AsaturA trees were computed using a Poisson corrected LG model and the support was assessed from 1000 replicates. Sequences from *Vitrella* (all enzymes under investigation) and *Chromera* (CPS and FAS enzymes) were inspected for the presence of N-terminal leader sequences using SignalP (Bendtsen et al., 2004) and TargetP (Emanuelsson et al., 2007) software respectively, suggesting targeting to either mitochondrion (with mitochondrial transit peptide) or plastid (with bipartite leader composed of ER signal peptide and transit peptide).

Fatty acid synthesis pathway

C. velia cells were grown in the f2 medium. Cultures were kept in 25 cm² flasks under artificial light with photoperiod 12/12, light exposure between 70 and 120 μmol/m²/s and temperature of 26°C. 1 ml of *C. velia* stationary culture was added to each flask with 20 ml of f2 solution. The cultures were grown for one month to reach a high density of cells. Since triclosan is not soluble in water, dimethyl sulfoxid (DMSO) was used as a soluble mediator. Four experimental groups were established: control, control with DMSO, *Chromera* treated with triclosan in concentrations of 1 mM and 0.5 mM, respectively. After 16 days of incubation, cultures were harvested via centrifugation, and pellets were stored in –20°C for subsequent lipid extraction. Homogenization of algal sample was achieved by Mini-beadbeater (Biospec Products). Homogenates were dried and weighted. Lipids were extracted using on chloroform and methanol, as described before (Folch et al., 1957). An aliquot of 100 μl volume was subjected to HPLC ESI/MS. The technique was performed on an ion trap LTQ mass spectrometer coupled to Allegro ternary HPLC system equipped with Accela autosampler with the thermostat chamber (all by Thermo, San Jose, CA, USA). 5 μl of sample was injected into a Gemini column 250 × 2 mm i.d. 3 μm (Phenomenex, Torrance, CA, USA). The mobile phase consisted of (A) 5 mmol/l ammonium acetate in methanol, (B) water, and (C) 2-propanol. The analysis was completed within 80 min with a flow rate of 200 μl/min by following gradient of 92% A and 8% B in 0–5 min, then 100% A till the 12th minute, subsequently increasing the phase C to 60% till 50 min and holding for 15 min and then in at the 65th minute returning back to the 92:8% A:B mixture and 10 min to column conditioning. The column temperature was maintained at 30°C. The mass spectrometer was operated in the positive and negative ion detection modes at +4 kV and –4 kV with capillary temperature at 220°C. Nitrogen was employed as shielding and auxiliary gas for both polarities. Mass range of 140–1400 Da was scanned every 0.5 s to obtain the full scan ESI mass spectra of lipids. For investigation of the lipid molecules structures the collisionally induced decomposition multi-stage ion trap tandem mass spectra MS² in both polarity settings were simultaneously recorded with a 3 Da isolation window. Maximum ion injection time was 100 ms, and normalized collision energy was 35%. Major phospholipids, galactolipids, and neutral lipids molecular species that are detected were separated by reversed-phase HPLC. The structure of each entity was identified by MS² experiments in positive or negative mode. Peak areas for each detected lipid component were summarized and their relative contents estimated to sum of all obtained peaks.

Raw extracted lipids have to be transformed to methylesters of fatty acids (FAMES) to enable application of the GC technique. For this purpose sodium methoxide was employed as a transesterification reagent, as previously described (Zahradnickova et al., 2014). FAMES were then analyzed by GC/FID. Hydrocarbon with 26-carbon chain was chosen as an internal standard. The chromatography was performed using gas chromatograph GC-2014 (Shimadzu) equipped by with column BPX70 (SGE)—0.22 mm ID; 0.25 μm film; 30 m length. μl of derivatized

sample was injected via autosampler and injector AOC—20i (Shimadzu) to the column in split mode (split ratio 10). The temperature of the injector was 220°C. The starting temperature of the column was 120°C holding for 4 min. Then, the temperature increased to the 180°C in at the rate of 10°C per minute, and after that 7°C per minute to 230°C. Temperature of the flame ionization detector was 260°C. The whole analysis takes took approximately 20 min. H₂ was used as a carrier gas. For the identification of particular FAs, a mixture of 37 standards purchased from Supelco Inc. was used.

Results and discussion

Global metabolic map

Metabolic annotations based on ortholog assignments with KEGG and OrthoMCL database showed that chromerids contain all major primary metabolic pathways typically found in free-living unicellular eukaryotes (**Figure 2—figure supplement 3**). 2918 *Chromera* proteins and 2985 *Vitrella* proteins were assigned KO numbers, from which 432 *Chromera* (1.3% of proteome) and 425 *Vitrella* proteins (1.8% of proteome), respectively, were identified as enzymes with a metabolic function based on EC number association.

In support of their autotrophic lifestyle, the chromerids appear to be capable of generating de novo all primary carbon metabolites such as the various sugars and other reduced carbon compounds (presumably via photosynthesis and associated carbon fixation pathways), amino acids, nucleotides, fatty acids and lipids, isoprene and steroid derivatives, and most vitamins (except biotin and vitamin B12). These organisms are also capable of assimilating both nitrate and sulfite and can generate energy from photosynthesis as well as mitochondrial respiration. A full complement of enzymes involved in sugar and sugar derivative metabolism, such as glycolysis, Krebs's cycle, pentose phosphate pathway, inositol mono- and poly-phosphate formation, polysaccharide formation, and amino- and nucleotide-sugar formation, is encoded by the chromerids. Chromerids are also capable of synthesizing sulfoquinovosyl-diacyl-glycerol lipids, which are found associated with the chloroplast in photosynthetic organisms.

Figure 2—figure supplement 3 illustrates a complete representation of all major pathways mapped to chromerids in comparison to selected apicomplexan lineages.

Generally, *Chromera* and *Vitrella* have similar sets of metabolic enzymes. Enzymes for the oxidative arm of pentose phosphate pathway, conversion of diacyl-glycerol to phosphatidyl ethanolamine, phosphatidyl ethanolamine to phosphatidyl serine, and XppppX to XTP are absent in *Chromera*, while, on the other hand, the enzymes involved in conversion of glucose-1P to UDP-glucose, and cytidine to uridine are missing in *Vitrella*. One major difference between the two chromerids is that the complex III of the mitochondrial respiratory chain (cytochrome c reductase) is missing in *Chromera*, but present in *Vitrella* (**Flegontov et al., 2015**). This feature of the *Chromera* mitochondrion, that is, absence of complex III but presence of complex IV, makes it unique among all mitochondria and mitochondria-derived organelles.

The crucial enzyme for the urea pathway, mitochondrially targeted carbamoyl phosphate synthase (CPS) (**Allen et al., 2011**), is absent from both chromerids. However, while *Chromera* contains single CPS involved in pyrimidine biosynthesis, *Vitrella* genome encodes two CPSs. But both these genes are closely related suggesting they are recent duplicates (ML tree is shown in **Figure 2—figure supplement 2A**) and they lack a mitochondrial leader sequence at the N-terminus (data not shown). This means that *Vitrella* duplicated CPSs are not likely to be involved in urea cycle. In contrast to *Vitrella*, *Chromera* lacks the gene encoding argininosuccinate lyase (ASL), an enzyme of the ornithin cycle.

Plastid-related metabolic pathways

Chromerid photosystems have a reduced set of genes similarly to that of other algae with a complex plastid. The patterns of reduction were lineage-specific, even between the two known chromerids. We found *psbM*, *petN*, *psaJ*, *Psb27*, *Ycf4*, and *Ycf44* genes in *Vitrella* but absent in *Chromera*; vice versa, *Ycf39* and *Ycf54* are absent in *Vitrella* but present in *Chromera* (**Supplementary file 3**). This demonstrates that the plastid of *Chromera* is more diverse and reduced when compared to *Vitrella* with respect to both the composition of photosystems and the number of genes encoded in the plastid genome (**Janoušková et al., 2010**). In spite of substantial reduction of photosystems, photosynthesis in *Chromera* is highly efficient (**Quigg et al., 2012**).

The heme pathway in *Vitrella* is homologous to that found in *Chromera* (**Kořený et al., 2011**) for most of the involved enzymes (ALAS, ALAD, PBGD, UROS, PPOX), however, *Vitrella* and *Chromera* do not constitute sister groups in the CPOXs and FeCH trees (trees not shown). Some enzymes are not present in single copies (UROD) in *Vitrella*, in contrast to *Chromera*, where three orthologs originating from cyanobacteria, endosymbiont nucleus, and exosymbiont nucleus are present (**Kořený et al., 2011**). For some investigated enzymes (UROD, CPOX), only the endoplasmic reticulum signal peptide was found with transit peptide missing from the sequence, suggesting their possible location in the endoplasmic reticulum or periplastidal space.

Genes containing the ketoacyl synthase domain and thus likely involved in fatty acid or polyketide synthesis were searched in the genomes of chromerids. We found that both algae possess multi-modular enzymes responsible for fatty acid synthesis type I (FASI), similar to some apicomplexan parasites, such as *Cryptosporidium* spp. and *T. gondii*. The longest multi-modular enzyme found in *Chromera* contains five multi-domain modules, reaching over 11,000 amino acids in length (**Figure 2—figure supplement 2B**).

Evolution of metabolic pathways

Apicomplexan parasites differ drastically from each other in their metabolic functions, and have a significantly reduced metabolic capability in comparison to the chromerids. Apicomplexans are non-photosynthetic and therefore lack all associated metabolic activities including photosynthetic carbon fixation. Interestingly, however, all plastid-bearing parasites have retained only the ferredoxin-NADP⁺ reductase (FNR)/ferredoxin redox system of the photosynthetic electron transport (**Lim and McFadden, 2010**). In photosynthetic organisms, this redox system mediates the transfer of electrons originating from water to NADP⁺, resulting in the formation of NADPH (cofactor for fatty acid biosynthesis and Calvin cycle), and it is likely that this role is conserved in chromerids. In apicomplexans, the source of electrons for generating reduced ferredoxin is not clear, but it is evident that reduced ferredoxin is required for generating reducing equivalents and is a cofactor for several plastid-associated enzymes, including those involved in isoprene synthesis (**Lim and McFadden, 2010**). Other notable pathways missing in apicomplexans but present in the chromerids include the following: glyoxalate pathway; steroid biosynthesis; synthesis of aromatic and branched-chain amino acids; purine synthesis; and synthesis of cofactors such as thiamine, riboflavin, and nicotinate ribonucleotide.

Despite the reduced metabolic capabilities, certain core metabolic functions are conserved in chromerids as well as in all apicomplexan parasites, and many of these are likely to be essential. These pathways include: glycolysis; synthesis of ubiquinone, inositol-P derivatives, GPI-anchor, mono-, di- and tri-acyl glycerol, isoprene derivatives, and N-glycans; a subset of scavenge reactions for purine and pyrimidine bases and their conversion to nucleotides; glycine-serine inter-conversion; one-carbon folate cycle and S-adenosyl-methionine formation. There are many metabolic pathways that are retained in specific apicomplexan lineages but shared with the chromerids (see **Figure 2B** and **Figure 2—figure supplement 3**). The following are notable examples of pathways shared between chromerids and an apicomplexan lineage: with *Plasmodium*, polyamine synthesis; with *Cryptosporidium*, conversion of serine to tryptophan; with *Toxoplasma*, branched-chain amino acid degradation, synthesis of aspartate, lysine, and methionine; synthesis of molybdopterin, bipterin, pyridoxal-phosphate, and pantothenate

cofactors. Surprisingly, with respect to the chromerids and other apicomplexans, *Cryptosporidium* appears to be the only apicomplexan lineage to have gained a metabolic function of conversion of thymidine to dTMP by thymidine kinase. In addition, the type I and II pathways for fatty acids biosynthesis show lineage-specific distribution in apicomplexans and chromerids (**Figure 2—figure supplement 3**).

We can also find example of pathways that have been lost in a lineage-specific manner. For example, the ability to synthesize the di-saccharide trehalose is missing only in *Plasmodium*. However, the most dramatic loss of metabolic function in a single lineage can be found in cryptosporidia. These parasites are devoid of all plastid- and mitochondria-associated metabolic functions and other pathways involved in the synthesis of ribose-P, pyrimidine, most amino acids, heme, fatty acids (de novo), and isoprene units. It seems that the lack of mitochondrial oxidative pathways in cryptosporidia led to loss of the ability to generate flavin nucleotide (FMN/FAD) and lipoic acid cofactors.

In order to cope with loss of metabolic pathways, parasites have evolved various mechanisms for scavenging the required nutrients and metabolites from their respective hosts. For example, metabolites such as heme, fatty acids, steroids (specifically cholesterol), and sphingolipids are known to be scavenged by various apicomplexans as needed from their respective hosts.

According to the metabolic pathway maps, certain metabolic functions, which are coupled to each other, have been either retained or lost concomitantly in various species (**Figure 2—figure supplement 3**). For example, piroplasms and cryptosporidia lack de novo fatty acid biosynthesis along with the pyruvate dehydrogenase enzyme complex (plastid-associated in apicomplexans), which is known to supply acetyl CoA units for fatty acid synthesis. On the other hand, these parasites have retained the ability to convert pantothenate to co-enzyme A, which is required for the activation of fatty acids scavenged from the hosts (**Leonardi et al., 2005**). Similarly, activities of the serine hydroxyl methyl transferase and thymidylate synthase enzymes are coupled to each other and to one-carbon folate metabolism. Therefore, these three metabolic functions are retained in all parasites.

Appendix 3

Endomembrane trafficking system.

Materials and methods

The predicted proteomes of the 26 species in **Figure 1** have been searched for endomembrane trafficking components. Initial homology searching was carried out using BLAST (**Altschul et al., 1990**). Known sequences from human (*Homo sapiens*) and yeast (*Saccharomyces cerevisiae*) were used to search the proteomes of each organism including *Chromera* and *Vitrella* to identify potential homologs of proteins implicated in endomembrane trafficking. Any sequences scoring an initial E value of 0.05 or lower were subjected to confirmation by reciprocal BLASTP. This involved the use of candidate homologous sequences as queries against the relevant *H. sapiens* or *S. cerevisiae* genome. Sequences that retrieved the query sequence, or named homologs/paralogs/isoforms thereof, first with an E value of 0.05 or lower were considered true homologs.

Additional searches were carried out using HMMER (**Finn et al., 2011**). The HMMs for the initial queries were built and used to search each proteome. Top hits based on BLASTP results with E values less than 0.05 were considered confirmed homologs, and not subjected to further analysis. Subunits with significant HMMER hits were further investigated by the reciprocal BLASTP as described above. Further HMMER searches were carried out with the addition of homologous sequences from *Bigelowiella natans*, *Phytophthora infestans*, and *T. gondii* to the original HMMs. Results were analyzed identically to the first round. All identified endomembrane components are listed in **Figure 2—source data 2**.

To identify homologous proteins not predicted by the gene prediction software, we used TBLASTN with the homologous protein from the closest related organism in our data set against scaffolds and contigs; E value cut-off was identical to BLASTP analysis. We utilized BLASTP to search either genome with an identified homolog from the other, if it was present. The final results are summarized in **Figure 2—figure supplement 4** using the Coulson Plot Generator software (**Field et al., 2013**).

Results and discussion

Apicomplexa possess numerous unusual features in their membrane trafficking systems. Non-canonical membranous inclusions such as the invasion organelles, the micronemes, rhoptries, and dense granules are present (**Baum et al., 2008**). Though canonical, stacked, Golgi bodies are present in *T. gondii* (**Pelletier et al., 2002**), other apicomplexan species possess Golgi bodies with aberrant morphology and unusual characteristics (**Struck et al., 2008**). Combined with other organelle destinations such as mitochondria, digestive vacuoles involved in hemoglobin catabolism in *P. falciparum*, and plant-like lytic vacuoles in *T. gondii* (**Miranda et al., 2010**), specificity of protein and lipid components of these various organelles suggest a need for unique trafficking pathways mediated by distinct protein machinery.

Interestingly, previous studies demonstrated the loss of trafficking machinery in Apicomplexa, including three key sets of proteins in the ESCRT machinery (**Leung et al., 2008**), adaptor protein complex (AP) families (**Nevin and Dacks, 2009; Hirst et al., 2011**), and multi-subunit tethering complexes (MTCs) (**Koumandou et al., 2007; Klinger et al., 2013a**) have been published. Several of the aforementioned families are involved in trafficking within the late endosomal system in opisthokont models and so may be associated with the evolution of the rhoptries and micronemes within the apicomplexan or myzozoan lineage. Consistent with this

idea, some cases of reduction were not limited to Apicomplexa, and could be observed in the sister phyla of the ciliates and dinoflagellates.

This pattern of loss raises the question of what losses correlate with the transition to parasitism and which are pre-adaptive, arising more deeply in the lineage. The unique phylogenetic position of chromerids (*Janouškovec et al., 2010, 2012; Oborník et al., 2012*) allows finer dissection of the patterns of retention/loss observed previously. Hence, we chose to focus on detailed characterization of the three previously studied sets of membrane trafficking machinery in the predicted proteomes of *Chromera* and *Vitrella*, together with 24 closely related organisms for comparison.

ESCRT machinery

The ESCRT machinery is a set of five sub-complexes involved in recognition of ubiquitylated proteins and recruitment to the multi-vesicular body (MVB)/late endosome for degradation (*Leung et al., 2008*). Most eukaryotes, including *Chlamydomonas reinhardtii* and our representative stramenopile taxa (*Thalassiosira pseudonanna*, *Phaeodactylum tricornutum*, *Ectocarpus siliculosus*, and *Pythium ultimum*), have a complete set of the ESCRT machinery, suggesting that the ancestor of alveolates, and indeed the Last Eukaryotic Common Ancestor (LECA), likely had it. Though this ancestral complement appears to have been reduced in ciliates in the ESCRTI and III complexes, and a few components are missing from dinoflagellate taxa, numerous gene duplications have occurred as well, suggesting sculpting of the machinery. By comparison, apicomplexan parasites exhibit significant reductions in their ESCRT machinery (*Leung et al., 2008*). Cryptosporidia, coccidia, and plasmodia appear to lack any subunits of the ESCRTI and II complexes. ESCRTIII conservation is better, though no apicomplexan encodes Vps24, and multiple taxa have lost Vps20 as well. A similar pattern is seen for the ESCRTIII-a machinery, with piroplasmids encoding only Vps46 and Vps4. Coccidia additionally encode Vps31, and cryptosporidia Vps60, whereas plasmodia encode all subunits (rodent parasites like *Plasmodium chabaudi*), or lack Vps31 (human or simian parasites like *P. falciparum*). *Chromera* and *Vitrella* possess all ESCRT subunits except for the ESCRT-III component CHMP7, which is rarely found outside the opisthokont supergroup (*Leung et al., 2008*). This observation suggests two conclusions regarding the evolution of the ESCRT machinery within alveolates: massive gene loss within the Apicomplexa occurred recently, after the split from the proto-apicomplexan ancestor, and some losses of machinery shared between apicomplexans and other alveolates are due to independent losses. An excellent example of this latter case is that of Vps37, which is present only in chromerids, but in no other alveolate included in the current study, suggesting its function was dispensable in a large number of lineages.

APs

The APs are heterotetrameric complexes that select cargo for inclusion into transport vesicles at organelles of the late secretory system and endocytic system. AP1 and AP3 are involved in the transport between the trans-Golgi network (TGN) and endosomes. AP2 is involved in the transport from the cell surface. AP4 is involved in TGN transport to either endosomes or the cell surface, while the recently described AP5 complex is involved in the transport from late endosomes back to early endosomes. All five complexes are ancient, having likely been present in the LECA (*Nevin and Dacks, 2009; Hirst et al., 2011*). However, the complexes have also been secondarily lost on multiple occasions as well. Outgroup taxa in our data set possess AP1-4 complexes, with the exception of *C. reinhardtii* lacking AP3, but only *Symbiodinium minutum* possesses an AP5 complex.

Apicomplexa display higher variability in AP complex retention. With the exception of AP2M in cryptosporidia, all taxa retain full AP1, 2, and 4 complexes. Piroplasmids lack all subunits of the AP3 complex, and together with *P. falciparum* and *Plasmodium reichenowi*, lack AP5 as well. Other plasmodia possess all AP5 subunits with the exception of the mu subunit. This result was unexpected, based on the usual patterns of conservation seen across *Plasmodium* species.

Presence of AP5 in the majority of these organisms suggests the exciting possibility of a novel trafficking pathway absent from the comparatively well-studied human parasite *P. falciparum*. Additionally, our increased taxon sampling has suggested that AP5 may be well conserved across Myzozoa, a result otherwise indeterminable from previous studies of this protein family (**Hirst et al., 2011**). Cryptosporidia also lack AP3, but unlike piroplasmids, they possess almost a complete AP5 complex, missing only the sigma subunit. Coccidia are the exception, possessing all five AP complexes in their entirety. Excitingly, *Chromera* and *Vitrella*, like coccidia, possess a complete complement of adaptin subunits, suggestive of a more complete set of trafficking pathways to endosomal organelles in these organisms.

MTCs

The MTCs are an assembly of heteromeric protein complexes involved in the first stage of vesicle fusion and delivery of contents from a transport vesicle to a destination organelle. Each one is specific to an organelle or transport pathway and all eight complexes have been deduced as present in the LECA, with some interesting cases of secondary loss. While *C. reinhardtii* and the stramenopiles encode a complete set of MTC machinery, several of these MTCs have interesting patterns of conservation, specifically in the Apicomplexa (**Klinger et al., 2013a**).

The conservation of the TRAPP I–II complexes is unclear through eukaryotes and clear patterns are difficult to draw. However, the apparent absence of the entire TRAPP II complex in *Vitrella* may be due to gaps/biases/absences in sequencing, protein prediction, or analysis, but has interesting ramifications if proven to be a real biological phenomenon.

Exocyst is involved in diverse processes, all of which involve polarized exocytosis (**Liu and Guo, 2012**). *Tetrahymena* appears to encode only four of the Exocyst subunits. None of the eight subunits were identifiable in *Chromera*, *Vitrella*, nor in any of the Apicomplexa or dinoflagellates. This confirms, and extends, a previous result suggesting the absence of this complex within the Myzozoa, suggesting a bona fide ancestral loss concurrent with the acquisition of an apical complex that could have served an analogous tethering function for secretory organelles.

COG is an octameric complex involved in tethering at the Golgi body (**Tomavo et al., 2013; Klinger et al., 2013b**). The COG complex is poorly conserved in Apicomplexa and a ciliate *Tetrahymena thermophila* only encodes half of the COG subunits. In contrast, all eight COG subunits are present in *Chromera* and *Vitrella*. The retention of a complete COG complex in both *Chromera* and *Vitrella* contrasts with the substantial loss of subunits in Apicomplexa, especially outside the coccidians (**Klinger et al., 2013b**) (**Figure 2—figure supplement 4**). Notably, this conservation is consistent with the presence of robust, stacked Golgi bodies in *Chromera* (**Oborník et al., 2011**) and *T. gondii* (**Pelletier et al., 2002**), compared to aberrant morphology in other Apicomplexa.

The complexes of CORVET and HOPS mediate tethering at the early and late endosomes (**Tomavo et al., 2013; Klinger et al., 2013b**). They share a core of four subunits with complex-specific proteins (Vps3 and 8 for CORVET and Vps39/41 for HOPS). Though all taxa encode the complete VpsC core of the HOPS/CORVET complex, all taxa except for *T. gondii* only appear to encode the CORVET-specific interactor Vps3. *Chromera* and *Vitrella*, like Apicomplexa, possess the entire VpsC core complements as well as the HOPS component Vps41 and both CORVET components.

Chromerids exhibit complex life cycles, from immotile vegetative cells to multi-cellular sporangia, and occasionally motile flagellated cells. Both lineages contain numerous potential locales for intracellular trafficking including mitochondria, plastid, starch granules, flagella, micronemes, and, in *Chromera*, the chromerosome. Additionally, vesicular traffic to the sporangial/cyst wall has been visualized in both lineages (**Oborník et al., 2012**). Our results indicate that chromerids possess an appropriately complex complement of membrane trafficking machinery to achieve these requirements.

Though MVBs have not been explicitly imaged or characterized in either lineage to date, both *Chromera* and *Vitrella* encode a complete set of ESCRT machinery, suggestive of the presence of functional MVBs. These may play a key role in modulating surface protein expression in various life cycle stages. Importantly, the close evolutionary position of *Chromera* and *Vitrella* to Apicomplexa suggests that the extensive decrease in ESCRT subunit conservation in Apicomplexa occurred in the immediate ancestor and is not an ancestral feature of a more inclusive group (Leung et al., 2008) (Figure 2—figure supplement 4). Particularly, the lack of some ESCRT subunits such as Vps37 in ciliates and dinoflagellates is most parsimoniously attributed to multiple independent losses. Further evidence for a complete set of ESCRT machinery in the last common alveolate ancestor comes from the conservation of all subunits to the exclusion of CHMP7 in the outgroup stramenopile taxa and in *C. reinhardtii*. The absence of CHMP7 in all taxa is not unusual, as it is lost in numerous taxa across eukaryotes (Leung et al., 2008).

Conservation of adaptin subunits is striking, particularly the complete retention of AP5 in chromerids. In an initial study of seven organisms from the SAR supergroup (the group in which chromerids belong to), only two (*B. natans* and *T. gondii*) were found to encode the complex; conservation across eukaryotes was similarly sparse (Hirst et al., 2011). The presence of a complete AP5 complex in chromerids and coccidians may be indicative of a conserved function in both lineages. Likewise, the retention of an almost complete AP5 in cryptosporidia and plasmodia may have functional significance or may simply represent a reductive evolutionary process that has not yet reached completion. The complete lack of AP5 in *P. falciparum* and *P. reichenowi* supports the latter view. As with the ESCRT complexes, the presence of AP1-5 in chromerids suggests the loss of AP3 and AP5 observed in some Apicomplexa is secondary, as well as the loss of AP5 in *Perkinsus marinus*, and in both ciliate lineages.

Presence of a complete VpsC core along with an additional CORVET subunit Vps3 in the majority of apicomplexan genomes suggests the potential for a modified HOPS/CORVET complex that interacts with Rab5 to direct tethering at the micronemes/rhoptries. This is in keeping with the view of rhoptries/micronemes as divergent endolysosomal organelles (Klinger et al., 2013b). However, chromerids do not appear to possess rhoptries, although chromerids possess cellular components analogous to micronemes (Obornik et al., 2011, 2012). More HOPS/CORVET subunits were found to be conserved in *T. gondii*, which are the only apicomplexan to date to be described as possessing a canonical lysosome-like compartment⁵, suggesting that complete complexes are retained in these lineages because they are required for trafficking to canonical lysosome-related organelles as well. Additionally, *Chromera* possesses the chromerosome, which often displays intraluminal vesicles similar to MVBs, suggesting it may also be derived from endosome-like organelles (Obornik et al., 2011).

In conclusion, apicomplexans possess unusual endomembrane compartments including atypical Golgi and endosome-derived invasion organelles such as micronemes and rhoptries (Klinger et al., 2013b). Modifications in the complement of membrane trafficking machinery, including the loss of key protein complexes found in most eukaryotes, have been observed in the apicomplexan lineage, potentially associated with the specialization of the endomembrane system. The absence of some components (Exocyst, Vps39, Trs120, Tip20) within *Chromera* and *Vitrella* suggests pre-adaptation to parasitism deeper in the apicomplexan lineage. By contrast, the presence of near complete complements of key machinery (AP1-5, ESCRTs, COG) absent in many apicomplexans, pinpoints the timing of the losses at the colpodellid/apicomplexan transition.

Appendix 4

Apical complex and cytoskeleton.

Motility is an essential feature of many living organisms. Some organisms utilize microtubule-based specialized structures such as flagella and cilia for locomotion. Some use actin-based structures like filopodia, lamellipodia, and pseudopodia (**Frenal and Soldati-Favre, 2009**), which are exploited in the amoeboid crawling (**Pollard and Borisy, 2003**) or bacterial and viral movement into and between cells (**Stevens et al., 2006**). Apicomplexan parasites use an unconventional actin-based mode of locomotion known as gliding motility (**Morrisette and Sibley, 2002**). This mechanism allows the parasites to move fast in the absence of canonical microtubular and actin-based structures. Gliding motility is mediated by the apical complex, which is a cellular structure common to all apicomplexan parasites. In the apical complex, proteins secreted from specialized secretory organelles, microneme and rhoptries, mediate adhesion to the cell substrate during motility and invasion or formation of a PV (**Baum et al., 2006**).

Actin-based gliding motility is essential for apicomplexan invasion (**Skillman et al., 2011**). Apicomplexan gliding motility undergoes actin polymerization/depolymerization for their directional motility with other associated protein classes such as actin-like proteins (ALP), actin-related protein (ARP), capping protein (CP), formin, profilin and cofilin/ADF. Actins elongate in the form of filaments and push the membrane forward. Arp2/3 complex (one of the ARPs) mediates the initiation of new branches on pre-existing filaments. After some growth, CP terminates the elongation of the filaments. Cofilin/ADF promotes de-branching and depolymerization. Profilin mediates the refilling of ATP-actin monomer pools, which are used for elongation through catalyzing ADP-ATP exchange (**Baum et al., 2006; Foth et al., 2006**).

We identified and compared genes encoding actins and other related components in the 26 species according to a method described by a previous study (**Baum et al., 2006**). Chromerids share homology with Apicomplexa for most of the actin, ALP and ARP classes. For example, both chromerids possess actin 1 (ACT1), actin-related (ARP), and actin-like (ALP) homologs. There were fewer actin genes in apicomplexans than in chromerids, indicating losses during apicomplexan evolution. The patterns of losses were the same for closely related species, suggesting non-random, lineage-specific losses (**Figure 2—figure supplement 5C**).

Arp2/3 complex, a nucleator of actins (**Machesky et al., 1994**), consists of seven subunits that regulate actin polymerization (**Mullins et al., 1997; Fehrenbacher et al., 2003**). Initially identified in *Acanthamoeba* (**Machesky et al., 1994**), Arp2/3 complex is conserved in most eukaryotes (**Gordon and Sibley, 2005**). We could not identify genes encoding subunits of Arp2/3 complex in both chromerids (**Figure 2—figure supplement 5C**). Also, all subunits were not found in apicomplexan species, consistent with a previous study (**Gordon and Sibley, 2005**). Individual subunits are important, as subunit ARPC4/p20 was shown to be essential for a complete, functional Arp2/3 complex (**Gournier et al., 2001**). Different subunits were identified in different phyla (**Figure 2—figure supplement 5C**). Within Apicomplexa, ARPC1 and ARPC4 were present in *Cryptosporidium hominis* and *Cryptosporidium parvum*, and ARPC1 and ARPC2 in *Plasmodium* spp. *S. minutum*, a dinoflagellate, contains genes for most of the subunits except ARPC1. This suggests that the common ancestor of Myzozoa (chromerids, apicomplexans, and dinoflagellates) had all the subunits, and they were lost in different lineages. Genes encoding ALP1, hypothesized to function as Arp2/3-like nucleator (**Gordon and Sibley, 2005**), were found in apicomplexans and also in *Vitrella* (Vbra_266.t1). FH2-domain (Pfam-PF02181) containing formins are members of another actin nucleator gene family. They produce unbranched filaments unlike Arp2/3 complex, which induces branched filaments. Both chromerids possess formin1 (FRM1) and formin 2 (FRM2) homologs, which are conserved in all the other studied species as well. Although *Plasmodium* spp. maintained a 1-1 orthology for both FRM1 and FRM2, we found a coccidian-specific FRM3 (TGME49_213370), suggesting a lineage-specific expansion. Maintenance of some formins across chromerids and

apicomplexans and lack of Arp2/3 complex suggest their importance, perhaps reflecting a switch from Arp2/3 complex to formins for actin nucleation during the evolution of Apicomplexa. Taken together, it seems that an Arp2/3 independent actin nucleation mechanism had already evolved before Apicomplexa and chromerids, and losses of ARP have probably begun too, although inferring the exact timing and sequence of losses will require studying more closely related species such as *Colpodella*.

We analyzed coronins, a major conserved gene family with a multifunctional role in actin regulation and vesicular transport (**Rybakin and Clemen, 2005**). These are WD40-repeat containing proteins, which represent the only candidate for actin bundling in apicomplexan parasites. Coronins inhibit the nucleating activity of Arp2/3, unlike other known Arp2/3-binding proteins. We observed absence of coronins in both chromerids, which is consistent with the notion that they, functionally linked to the Arp2/3 complex, were lost (**Figure 2—figure supplement 5C**). Although parasite homologs do not have the microtubule-binding domain of canonical coronins, but essential amino acid residues are conserved (**Gandhi et al., 2010**). Thus, coronin could be playing a role in stabilizing F-actin scaffolds or having an alternative role in vesicular transport in apicomplexan parasites.

Profilins are actin-binding proteins that supply pools of readily polymerizable actin monomers (**Baum et al., 2006**). Genes encoding profilins were found in all 26 species studied except for an oomycete (*P. ultimum*) and diatoms (*P. tricornutum*, *Thalassiosira pseudonana*). Apicomplexa-specific profilins have β mini1 and β mini2 domains, which provide an extended interface with actin and formed the structural basis of their actin-binding function in *Toxoplasma* (**Kucera et al., 2010**) and *Plasmodium* (**Kursula et al., 2008**). These domains are not found in other eukaryotes. Sequence alignment of these profilins reveals an intriguing observation that *Vitrella* (Vbra_7301.t1) had these β -domains previously thought to be specific to Apicomplexa, with partial conservation in *Chromera* (Cvel_18957.t1) and in dinoflagellate *P. marinus* (XP_002774080). The β -domains were not detected in other non-apicomplexan species. All species studied has had only one profilin gene except for chromerids where we observed 2 in *Chromera* and 3 in *Vitrella*, including an one-to-one ortholog of the apicomplexan profilin in both chromerids.

Cyclase-associated proteins (CAPs) are evolutionary conserved G-actin-binding proteins, which participate in filament turnover regulation by acting on actin monomers (**Chaudhry et al., 2010**). CAP proteins are made up of three significant regions: N-terminal adenylate cyclase binding domain (CAP_N, linked to the cAMP-RAS signaling), a central proline-rich segment, and a C-terminal actin-binding domain (CAP_C). Apicomplexans do not possess the N-terminal (CAP_N) domain altogether with few genes in stramenopiles and in *Vitrella* (Vbra_7026.t1) also showing a similar pattern of loss. However, the *Chromera* gene Cvel_8488.t1 possesses both domains. This suggests the dispensable nature of CAP_N domain (**Figure 2—figure supplement 5C**), and we speculate that in parasites CAP functions are reduced to actin sequestration only.

The F-actin capping, CapZ duplex, a dimer of α - and β -CPs, prevents polymerization from the 'barbed' (plus) end. It is conserved across Apicomplexa except for in piroplasmids. It is also conserved in most of the species studied including stramenopiles, dinoflagellates, and both chromerids. In Apicomplexa, several gelsolin domain-containing proteins were found but they are unlikely to be functionally related and are speculated to be Sec23/Sec24-like proteins, which function in vesicular transport (**Baum et al., 2006**).

Cofilin/ADF genes promote de-branching of actin filaments and are well conserved among species studied. However, plasmodia differ from the rest of the Apicomplexa species in having an additional copy of the ADF gene. Phylogenetic analysis shows that ADF in plasmodia has duplicated and diverged with respect to the rest of the Apicomplexa, and recent structural studies explain the mechanism of action of *Plasmodium* ADFs (**Singh et al., 2011; Wong et al., 2014**). This represents a clear example of additional innovations of actin regulation in certain apicomplexan clades.

In addition, we identified myosin families in the 26 species using a myosin HMM model (Foth et al., 2006) (Figure 2—figure supplement 5C). Members of piroplasmids such as *Theileria annulata* and *Theileria parva* have the fewest genes among the apicomplexan species examined, likely because piroplasms do not require motility for intracellular invasion. On the other hand, we detected the most complete myosin family repertoire in *Chromera* and *Vitrella*. We detected certain myosin families in some apicomplexan genera, but not among non-apicomplexan species, indicating lineage-specific gains (data not shown). In summary, combinations of lineage-specific losses and gains have led to streamlined, unique repertoires of actins and myosins in various apicomplexan species.

Appendix 5

Extracellular proteins in chromerids.

We curated the chromerid genomes for genes with extracellular domains and domain architectures like similar to those of apicomplexans (**Figure 3—figure supplement 4; Supplementary file 5**). Both chromerids possess mucin-like proteins having long stretches of threonine and serine residues with predicted O-linked glycosylation, as well as the enzyme pathways involved in O-linked glycosylation (**Templeton et al., 2004a; Anantharaman et al., 2007**). Proteins encoding combinations of von Willibrand factor A (vWA) and thrombospondin 1 (TSP1) were also observed, although none with apparent orthologous relationship to the vWA and TSP1 domain proteins (TRAP) that serve as receptors mediating gliding motility in apicomplexans. The chromerid genomes possess numerous secreted proteins with domains predicted to participate in binding of sugar moieties (**Figure 3—figure supplement 4**). Chromerids share FRINGE domains with *Cryptosporidium*, and HINT domains with *Cryptosporidium* and *Gregarina*, to the exclusion of other apicomplexans, in support of early divergence of these genera within the Apicomplexa (**Figure 3—figure supplement 4B**). *Vitrella* genome contains multiple copies of proteins, which have arrays of the cysteine-rich oocyst wall protein (OWP) domain found in *Cryptosporidium* and coccidians, which are associated with forming environmentally durable walls of oocysts (**Templeton et al., 2004b**).

Several EC domain architectures thought to be distributed apicomplexan-wide have homologs in the chromerids; for example, the LCCL domain-containing proteins, CCp1 and CCp2/3, as well as the CPW-WPC domain proteins (**Figure 3—figure supplement 4C**). Ultrastructures reminiscent of micronemes have been observed in both chromerids (**Obornik et al., 2012**); consistent with this, we identified EC proteins having domains and architectures typical of *Toxoplasma* and *Plasmodium* micronemal secretory proteins. Examples include expansions of proteins containing SUSHI, EGF, TSP1, and vWA domains (data not shown). Chromerids possess unique architectures of proteins containing the macrophage perforin (MacPerf) domain (**Figure 3—figure supplement 4E**), which, previously found in apicomplexans and ciliates (as large expansions), are thought to function in apicomplexans to mediate membrane lysis during host cell egress and tissue traversal (**Roiko and Carruthers, 2009**). The *Chromera* MacPerf domain proteins also contain arrays of a domain, WSC, thus far not found in other alveolates, as well as a unique C-terminal DERM domain. *Chromera* possesses four MacPerf domain proteins with various domain architectures, whereas *Vitrella* a single MacPerf protein with a stand alone MacPerf domain (Vbra_18070.t1).

The ciliate genomes possess highly amplified and antigenically diverse repertoires of GPI-linked proteins termed 'immobilization antigens' (**Caron and Meyer, 1989**). We did not see amplifications of GPI-linked gene families in either chromerid species. Lineage-specific gene amplifications include a predicted secreted protein in *Vitrella*, which contains an arenylsulfonase domain (**Figure 3—figure supplement 4E**). The chromerids possess highly amplified gene family, annotated as 'CAST multi-domain protein' in the ciliate, *Oxytricha* (e.g., UniProt ID: OXYTRI_15408), and which comprises conserved cysteine-rich domains in the extracellular region, a single transmembrane domain, and a conserved predicted coiled-coil region in the cytoplasmic domain (e.g., Cvel_3066.t1). Representatives of this protein are found in the ciliate *Oxytricha*, but not in *Tetrahymena* and *Paramecium*; in stramenopiles, choanoflagellates and coccidians, but are absent from other apicomplexans such as *Cryptosporidium*, *Theileria*, *Babesia*, and *Plasmodium*.

Appendix 6

ApiAP2 proteins.

We examined the abundance of apicomplexan-specific AP2 (apiAP2) genes, transcription factors that play regulatory roles in key aspects of apicomplexan biology (**Campbell et al., 2010; Flueck et al., 2010; Radke et al., 2013; Kafsack et al., 2014; Sinha et al., 2014**). We scanned the protein-coding gene sets of the 26 species using the apicomplexan-specific apiAP2 HMM, which was constructed with the AP2 domain sequences from apicomplexan species. We found that apiAP2 genes were abundant in both chromerids and in all apicomplexans. ApiAP2 genes were moderately abundant in the two dinoflagellates and rare or absent in ciliates and stramenopiles, respectively (**Figure 3—figure supplement 1D**). There were very few apiAP2 genes that were shared between apicomplexans and non-apicomplexan species; most were shared between closely related species, that is, from the same clade (**Figure 3—figure supplement 1B**). These lineage-specific apiAP2 genes in the present-day species could have arisen from de novo gene birth or modification of the full-length sequences of existing genes beyond recognition. In the former case, the proto-apicomplexan ancestor had a small set of apiAP2 genes. In the latter case, the common ancestor already had a large set of apiAP2 genes, which continued to change, giving the appearance of ‘new’ clade-specific genes. The latter case, the turnover scenario, is more parsimonious because, according to the de novo gene birth scenario, apiAP2 genes must have expanded independently in every descending lineage from the proto-apicomplexan ancestor. In summary, our data support the notion that massive apiAP2 expansion occurred in the common ancestor before Apicomplexa and chromerids split, and the apiAP2s continued to change as the common ancestor split into chromerids and apicomplexans, which continued to radiate and adapt to their host niches and life cycle strategies.

We sought to determine if gene duplication and divergence was a significant mechanism for the expansion and the turnover of apiAP2 genes. The number of apiAP2 genes that have other homologous apiAP2 genes within the species based on OrthoMCL clustering, which are likely mediated by paralogous expansions, was high (93 out of 136) in chromerids (**Figure 3—figure supplement 1D**). In *Vitrella*, we identified one cluster of 50 homologous apiAP2 genes. This means that gene duplication played an important role in expanding apiAP2 gene repertoire in chromerids. The number was significantly less (5 out of 13) in dinoflagellates than in chromerids (**Figure 3—figure supplement 1D**). We suspect that gene duplication and diversifications drove expansion of apiAP2 genes significantly after the split of dinoflagellates. In apicomplexan species, evidence for recent duplications was sparse, as only 4 out of 409 apiAP2 genes had homologous copies in the same species. This does not necessarily mean that apiAP2 genes do not duplicate readily in apicomplexans, but rather that redundant copies of apiAP2 are quickly lost or diversified beyond recognition in part by selective pressure to reduce gene repertoires and genome sizes (**Katinka et al., 2001**) and due to higher rate of sequence divergence in parasites (**Hafner et al., 1994**).

Previous studies have shown that plant genomes contain a large repertoire of AP2 genes, and that plant AP2 domains evolved from an endonuclease domain in a cyanobacteria (**Magnani et al., 2004**). According to our phylogenetic analysis, AP domains among bacteria are many and diverse, with both plant-like and apicomplexan-like AP2s (data not shown). We did not find significant homology with bacterial AP2 genes at the full gene length level. It is not clear if the originally transferred AP2 gene has evolved beyond recognition or if only the domain has been transferred to these eukaryotes. The exact genetic events that led to acquisition of AP2s in apicomplexans are not clear. However, what is the most probable scenario is that AP2 domains in alveolates came from bacteria and have expanded in myzozoans, independent of those in plants. Both functional studies and more taxon sampling would be required for elucidating how AP2s in alveolates were acquired in the first place.

Paper III

Fatty Acid Biosynthesis in Chromerids

Article

Fatty Acid Biosynthesis in Chromerids

Aleš Tomčala^{1,2,†}, Jan Michálek^{1,3,†}, Ivana Schneedorferová^{1,3}, Zoltán Füßy¹, Ansgar Gruber¹, Marie Vancová¹ and Miroslav Oborník^{1,3*}

¹ Biology Centre CAS, Institute of Parasitology, Branišovská 31, 370 05 České Budějovice, Czech Republic; a.tomcala@centrum.cz (A.T.); jan.michalek@entu.cas.cz (J.M.); Ivana.Schneedorferova@email.cz (I.S.); zoltan.fussy@gmail.com (Z.F.); ansgar.gruber@paru.cas.cz (A.G.); vancova@paru.cas.cz (M.V.)

² Faculty of Fisheries and Protection of Waters, CENAKVA, Institute of Aquaculture and Protection of Waters, University of South Bohemia, Husova 458/102, 370 05 České Budějovice, Czech Republic

³ Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic

* Correspondence: obornik@paru.cas.cz; Tel.: +420-38777-5464

† These authors equally contributed to this work.

Received: 14 May 2020; Accepted: 15 July 2020; Published: 24 July 2020

Abstract: Fatty acids are essential components of biological membranes, important for the maintenance of cellular structures, especially in organisms with complex life cycles like protozoan parasites. Apicomplexans are obligate parasites responsible for various deadly diseases of humans and livestock. We analyzed the fatty acids produced by the closest phototrophic relatives of parasitic apicomplexans, the chromerids *Chromera velia* and *Vitrella brassicaformis*, and investigated the genes coding for enzymes involved in fatty acids biosynthesis in chromerids, in comparison to their parasitic relatives. Based on evidence from genomic and metabolomic data, we propose a model of fatty acid synthesis in chromerids: the plastid-localized FAS-II pathway is responsible for the de novo synthesis of fatty acids reaching the maximum length of 18 carbon units. Short saturated fatty acids (C14:0–C18:0) originate from the plastid are then elongated and desaturated in the cytosol and the endoplasmic reticulum. We identified giant FAS I-like multi-modular enzymes in both chromerids, which seem to be involved in polyketide synthesis and fatty acid elongation. This full-scale description of the biosynthesis of fatty acids and their derivatives provides important insights into the reductive evolutionary transition of a phototrophic algal ancestor to obligate parasites.

Keywords: *Chromera velia*; *Vitrella brassicaformis*; fatty acids; de novo biosynthesis; evolution; elongation; desaturation

1. Introduction

The phototrophic alveolates *Chromera velia* [1] and *Vitrella brassicaformis* [2] (referred to as chromerids) were isolated from stony corals in Australia using methods for the isolation of intracellular symbionts. Both chromerids can live as free-living algae without hosts in laboratory cultures; however, they are associated with corals in the wild [1–4]. Although they were found in the larvae of several coral species [5], they have never been detected in adult corals [6]. It was speculated that chromerids live as mutualists like symbiotic dinoflagellates of the genus *Symbiodinium* [1,7]. However, the latest transcriptomic survey of coral larvae experimentally infected by *C. velia* opens the possibility that the alga can be their facultative or even accidental parasite [8]. The environmental detection of uncultured and undescribed chromerids and related organisms (ARL-V) also implies their probable commensal or parasitic relationship to corals [9]. Such finding is congruent with the accumulating evidence supporting the common ancestry of phototrophic chromerids and parasitic apicomplexans [3,4,10–14]. Analyses of plastid [11] and mitochondrial [15] and nuclear genomes [10,13], together with the determination of ultrastructural characters such as cortical alveoli, four-membrane plastid envelopes [1], the presence of a pseudoconoid in *C. velia* [16], and zoosporegenesis

by budding in *V. brassicaformis* [17], support the unique phylogenetic position of chromerids at the root of the parasitic Apicomplexa. Further molecular phylogenetic and phylogenomic analyses have placed the predatory colpodellids within the polyphyletic chromerids (Figure 1), constituting a group named Apicomonada [18] or chrompodellids [10].

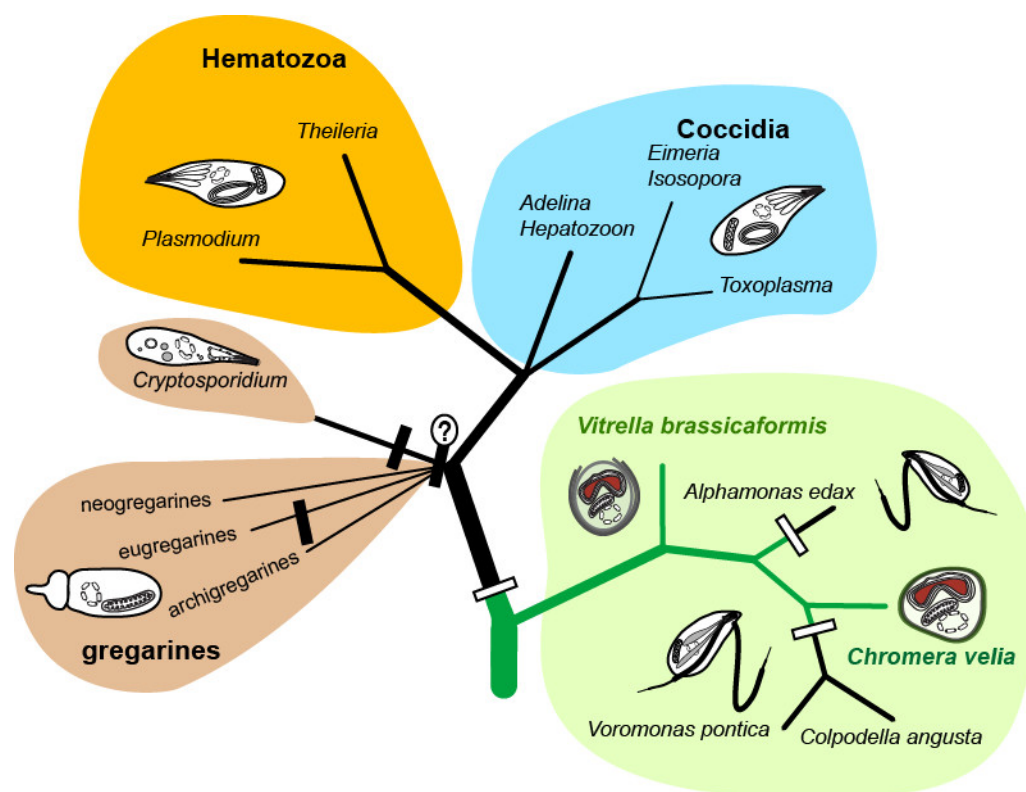


Figure 1. Phylogenetic tree of chromerid algae and apicomplexan parasites. The white rectangles indicate the loss of photosynthesis, the black ones indicate the loss of plastid.

Fatty acids (FAs), carboxylic acids with aliphatic chains, represent a fundamental component of biological lipids and can be divided into several categories according to their structure [19]. Here, we focus on the most abundant subclass—straight-chained saturated and unsaturated FAs with a terminal carboxylic acid group. Generally, FAs are synthesized by repeated addition of two-carbon units to a growing carboxylic acid chain attached to the acyl carrier protein (ACP) [20]. In cellular organisms, the enzymes responsible for the de novo synthesis of FAs are either the type I FA synthases (FAS-I) typically found in heterotrophic eukaryotes or the type II FA synthases (FAS-II) usually referred to as bacterial-type FAS-II. Both pathways contain identical reaction steps but they substantially differ in the architecture of the responsible enzymes. In eukaryotes, cytosolic FAS-I is a huge multi-modular enzyme [21], while FAS-II consists of individual separate enzymes located in the plastid [22–24]. Modules in the FAS-I multi-modular enzymes have the full set of domains needed for the attachment of FA carbon units. If such a multi-modular enzyme contains an incomplete module lacking some of the domains (often acyltransferase) or has some additional domains (e.g., methyltransferase), it usually has a different function. Such giant multi-modular enzymes known as polyketide synthases (PKSs) are responsible for the synthesis of antimicrobial compounds called polyketides (PK) [25]. It was proposed that FAS-I derived from FAS-II via a gene fusion [25]. Plants and algae primarily use bacterial FAS-II located in the plastid, with the involved enzymes encoded in the nuclear genome [24,26]. Many algae also possess FAS-I/PKS-like enzymatic equipment. Their function is still enigmatic; however, they likely do not act as FA synthases [25]. Various lineages of apicomplexan parasites differ in the presence of the particular FAS system. Most apicomplexan parasites possess the apicoplast-located FAS-II for de novo production of FAs [20,27,28], with Haemosporidia such as *Plasmodium* utilizing only this prokaryotic type of the synthesis. The coccidians *Toxoplasma gondii* or *Eimeria tenella* were described to use the apicoplast FAS-II system and

the cytosolic FAS-I in parallel, with life stage-specific FA metabolism [20,29–31]. However, the apicomplexan minimalists *Theileria parva* and *Babesia bigemina* completely lost all the enzymes for FA synthesis and scavenge all FAs from their hosts [20]. Current findings suggest independent multiple FAS-II losses in apicomplexans and their relatives [32].

Last but not least, intestinal parasites of the genus *Cryptosporidium* do not possess the apicoplast at all and lack any de novo FA synthesis. The huge FAS-I-like multi-modular enzyme coded in the *Cryptosporidium parvum* genome is responsible for the elongation of palmitic acid acquired from the host [13,20,33]. The high divergence of the apicomplexan sources of fatty acids reflects different stages of reductive evolution driven by adaptations to diverse hosts [13]. Genomic analysis performed by John and co-workers [34] revealed that some representatives among apicomplexans such as *C. parvum*, *E. tenella*, and *T. gondii* possess PKS genes. However, the presence of PKS genes is not universal among apicomplexans. For instance, in the genomes of *Plasmodium falciparum* and *T. parva*, these sequences are lacking [34].

Here, we studied FAs and their synthesis in the chromerids *C. velia* and *V. brassicaformis* by the combination of analytical chemistry and genomics methods. We compared chromerid and apicomplexan FA biosynthesis and propose a scenario of their evolution.

2. Materials and Methods

2.1. Growth Experiments

Cultures of *C. velia* and *V. brassicaformis* were grown in f/2 medium in 25 cm² flasks under artificial light, with a photoperiod of 12 h light/12 h dark and light intensities of 30–50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ at a temperature of 26 °C. As an inoculum, 1 mL of *C. velia* (1×10^6 cells/mL) or 5 mL of *V. brassicaformis* (0.3×10^6 cells/mL) stationary culture was added to each flask with 40 mL of f/2 medium. The cultures were grown for 50 days and harvested via centrifugation. Pellets for lipid extraction were stored at –20 °C.

The cultures of *C. velia* were treated with Triclosan (TCI Europe N. V., Zwijndrecht, Belgium) [35] in the same cultivations conditions as mentioned above. Triclosan was diluted in dimethyl sulfoxide (DMSO, Sigma Aldrich, St. Louis, Missouri, USA). The inhibitor was added to cultures after 14 days, at six concentrations in the media (0.02, 0.04, 0.08, 0.133, 0.333, and 0.667 μM). The densities of the cultures were periodically measured by a spectrophotometer (Infinite[®] 200 PRO, Tecan, Männedorf, Switzerland) over the course of 36 days. Fifty-day-old cultures were harvested by centrifugation in 50 mL flacons (Hettich Micro 22R centrifuge with rotor radius 100 mm, set to 6000 rpm at the of temperature 4 °C). The obtained pellets were split; the first part was stored at –20 °C for fatty acid analyses; the second part was subjected to transmission electron microscopy (TEM).

To investigate the effects of various levels of inorganic nitrogen in the media, *C. velia* cultures were grown in standard conditions, as mentioned above (light, temperature, cultivation flasks). Flasks containing 120 mL of f/2 solution with four different concentrations of NaNO₃ (from 0.000 to 0.062 g of pure nitrogen per 1 L of f/2 medium) were inoculated using 1 mL of stationary-stage culture of *C. velia* (1×10^6 cells/mL). The highest concentration of nitrate in this experiment was five times higher than in the regular f/2 medium. The cultures were harvested by centrifugation after 30 days. The dry pellets were weighed and investigated using gas chromatography coupled with a flame ionization detector (GC/FID).

2.2. Genomic Search

BLAST search and InterProScan in the publicly available dataset of predicted proteins [13] (CryptoDB; <https://cryptodb.org/cryptodb/>) were used to search for the enzymes of the FA biosynthesis in *C. velia* and *V. brassicaformis*. The obtained sequences were complemented with homologs found in other public databases (NCBI, KEGG, JGI) and aligned using MAFFT [36]. The alignment was manually edited by excluding gaps and ambiguously aligned positions using Geneious [37]. Phylogenetic trees were computed using the maximum likelihood method (RAxML; [38] and Bayesian inference (MrBayes 3.2.4.; [39,40]). The subcellular locations of the enzymes were

predicted using TargetP [41,42], SignalP [43], and ASAFind [44]. Localizations were established with the support of multiple predictors together; previously published data were also considered. For phylogenetic analyses, we took advantage of computational resources freely provided by the CIPRES Science Gateway [45].

2.3. Fatty Acids—Gas Chromatography Coupled with a Flame Ionization Detector (GC/FID)

Lipids were extracted following the chloroform–methanol solution method [46] modified by Košťál and Šimek [47], with additions described by Tomčala et al. [48]. Briefly, homogenized algal samples were dried, weighed, and extracted. The raw extracts in chloroform were dried and resolved in 500 µl of chloroform. Then, 100 µl aliquots were transformed into methyl esters of fatty acids (FAMES) to allow for GC analyses. For this purpose, sodium methoxide was employed as a transesterification reagent, as described previously [49]. FAMES were then analyzed by GC/FID using a gas chromatograph GC-2014 (Shimadzu, Shimadzu Scientific Instruments Inc., Kyoto, Japan) equipped with the column BPX70 (SGE, SciTech, Prague, Czech Republic) and an autosampler with the injector AOC—20i (Shimadzu), with the settings according to a previous work published by Tomčala et al. [48]. For the identification of FAs, a mixture of 37 standards purchased from Supelco Inc. (Bellefonte, Pennsylvania, USA) was used.

2.4. Transmission Electron Microscopy

Pellets were transferred to HPF (High Pressure Freeze) carriers (Leica Microsyst. Vienna, Austria), then f/2 medium with 20% (*w/v*) bovine serum albumin (BSA) was added to fill the carriers completely. The samples were immediately frozen using the high-pressure freezer Leica EM PACT2. The frozen samples were incubated with 2% osmium diluted in 100% acetone in a freeze substitution unit, with the following substitution program: −90 °C for 96 h; then the temperature was increased to −20 °C (with a slope of 5 °C/h); after 24 h, the temperature was raised (3 °C/h) to the final temperature of 4 °C. After 18 h of incubation, the samples were washed with acetone and infiltrated by 25%, 50%, 75% EMBed 812 resin (EMS)/acetone solutions, 1 h at each step, at room temperature. Finally, fresh resin was added, and the specimens were placed overnight in a vacuum desiccator. The resin was allowed to polymerize at 60 °C for 48 h. Images were obtained by a transmission electron microscope (JEM 1010, JEOL Inc., USA) with an acceleration voltage of 100 kV [16].

3. Results

3.1. De novo Fatty Acid Synthesis

The presence of a FAS-II in *C. velia* was first described by Dahmen et al. [50] and Woo et al. [13], who indicated the presence of FAS-I and FAS-II in both chromerids. We searched in the chromerid genomic and transcriptomic data (CryptoDB) for all the enzymes possibly involved in the synthesis of FAs and their derivatives (Tables 1–5). In addition to the canonical FAS-I and FAS-II enzymes, the genome contains a large number of genes likely encoding ketoacyl synthases (36 genes), short-chain dehydrogenases (73), or acyltransferase domains (29), present in various gene arrangements, often in a form of multi-modular enzymes with unknown putative function in FA or PK synthesis (Tables 1–5). We performed a phylogenetic analysis based on ketoacyl synthase (KS) domains to distinguish FAS-I and FAS-II enzymes of chromerids and further calculated detailed phylogenies of FAS-I- and FAS-II-type enzymes (Figures 2–4). Both chromerids possess numerous giant multi-modular enzymes in sizes ranging from 1000 to 11,000 amino acids composed of 1–5 modules and classified as putative FAS-I/PKS-I [13]. In general, KSs of type II FAS show two main lineages, i.e., α -proteobacterial and cyanobacterial, with the latter genes involved in plastid-located KSs in chromerids (Figure 3). Ketoacyl synthases from FAS-I/PKS-I enzymes of chromerids, apicomplexans, and dinoflagellates form a sister group to PKSs responsible for the production of polyketide-based algal toxins found in dinoflagellates, chlorophytes, haptophytes, and stramenopiles (Figure 4). Since the evolutionary conversion between FAS and PKS function can be caused by a single amino acid substitution affecting the functionality of the participating domain [51], it is hard to identify them

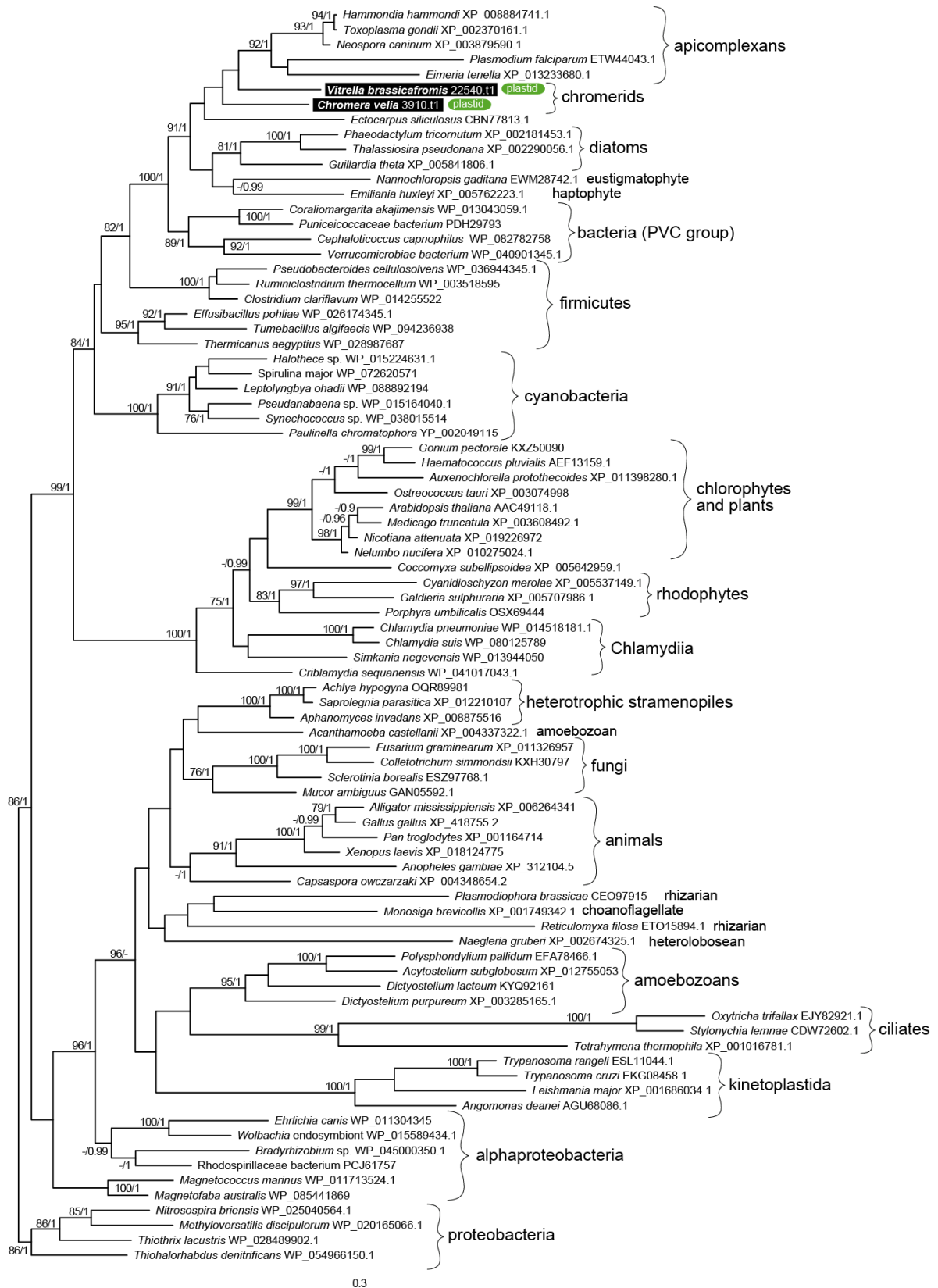


Figure 3. The detailed maximum likelihood phylogeny of ketoacyl synthase domains of type II fatty acid synthases. Bootstrap for 1000 repetitions and Bayesian pp values (BS1000/1M, CAT/LG models) are displayed above the branches. Predicted subcellular locations are shown.

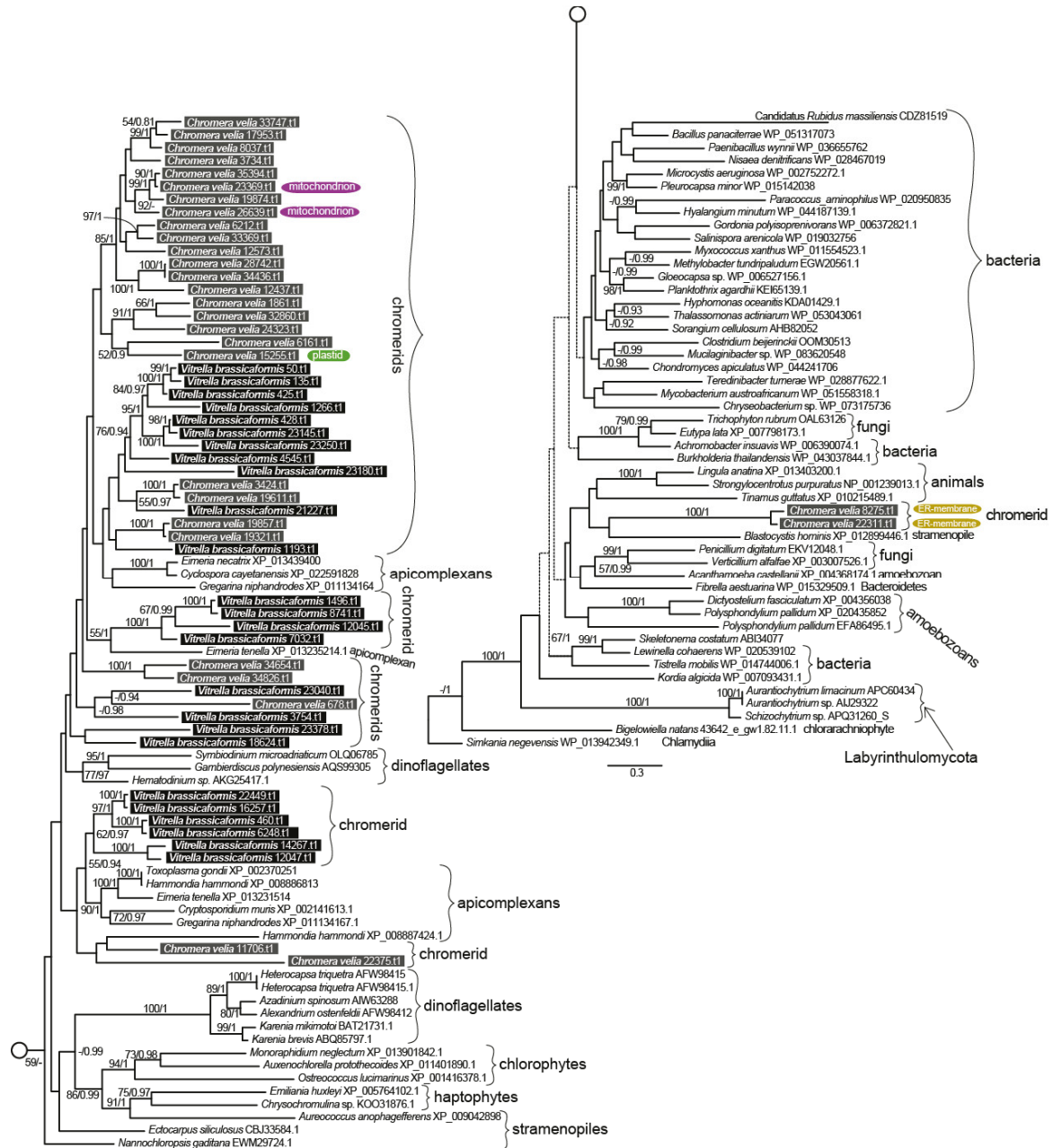


Figure 4. The detailed phylogeny of ketoacyl synthase domains of type I fatty acid synthases and polyketide synthases. Bootstrap for 1000 repetitions and Bayesian pp values (BS1000/1M, CAT/LG models) are displayed above the branches. Predicted subcellular locations other than cytosolic are shown.

Table 1. Ketoacyl synthases (KSs), elongases, and desaturases in chromerids and their predicted subcellular location. ER – endoplasmic reticulum.

Ketoacyl Synthases	Accession	Length (AA)	SignalP 3.0	TargetP 1.1	TargetP (trimmed SP)	ASAFind	Pred. Localization
Type I FAS	Cvel_12437.t1	11,656	-	-	-	Cytosol/ER	Cytosol
	Cvel_19857.t1	6729	-	-	-	Cytosol/ER	Cytosol
	Vbra_3754.t1	7413	-	-	-	Cytosol/ER	Cytosol
	Vbra_4545.t1	9001	-	-	-	Cytosol/ER	Cytosol
	Vbra_18624.t1	8182	-	-	-	Cytosol/ER	Cytosol
Type II FAS	Cvel_3910.t1	463	SP (1.000, L19)	M (0.579, R5, L38)	M (0.625, R5, L19)	Plastid/mitoch	Plastid
	Vbra_22540.t1	504	SP (0.920, L33)	SP (0.968, R1, L33)	M (0.653, R4, L19)	Plastid	Plastid
ACR-B coupled	Cvel_8275.t1	4523	Anch (0.716, L26)	M (0.778, R3, L26)	-	Cytosol/ER	ER/ Microsome memb.
	Cvel_22311.t1	4384	Anch (0.868, L46)	M (0.706, R4, L15)	-	Cytosol/ER	ER/ Microsome memb.
Other KSs	Cvel_11706.t1	2033	-	M (0.489, R5, L21)	-	Cytosol/ER	Cytosol
	Cvel_12573.t1	2656	-	-	-	Cytosol/ER	Cytosol
	Cvel_15255.t1	2854	SP (0.877, L29)	M (0.333, R5, L17)	M (0.669, R3, L7)	Cytosol/ER	Plastid
	Cvel_17953.t1	1403	-	-	-	Cytosol/ER	Cytosol
	Cvel_18613.t1	1930	-	-	-	Cytosol/ER	Cytosol
	Cvel_19321.t1	3353	-	-	-	Cytosol/ER	Cytosol
	Cvel_19611.t1	6404	-	-	-	Cytosol/ER	Cytosol
	Cvel_19874.t1	528	-	-	-	Cytosol/ER	Cytosol
	Cvel_22375.t1	2562	-	-	-	Cytosol/ER	Cytosol
	Cvel_23369.t1	2489	-	M (0.494, R5, L50)	-	Mitochondrion	Mitochondrion
	Cvel_24220.t1	1093	-	-	-	Cytosol/ER	Cytosol
	Cvel_24323.t1	782	-	-	-	Cytosol/ER	Cytosol
	Cvel_26639.t1	2832	-	M (0.683, R4, L11)	-	Cytosol/ER	Cytosol/Mitochondrion
	Cvel_28742.t1	2860	-	SP (0.546, R5, L17)	-	Cytosol/ER	Cytosol
	Cvel_32860.t1	565	-	-	-	Cytosol/ER	Cytosol
Cvel_33368.t1	1048	-	SP (0.447, R5, L24)	-	Cytosol/ER	Cytosol	

Table 2. Ketoacyl synthases, elongases, and desaturases in chromerids and their predicted subcellular location.

Ketoacyl Synthases	Accession	Length (AA)	SignalP 3.0	TargetP 1.1	TargetP (trimmed SP)	ASAFind	Pred. Localization	
Other KSs	Cvel_33369.t1	720	-	-	-	Cytosol/ER	Cytosol	
	Cvel_33747.t1	1598	-	-	-	Cytosol/ER	Cytosol	
	Cvel_3424.t1	4498	-	-	-	Cytosol/ER	Cytosol	
	Cvel_34436.t1	1211	-	-	-	Cytosol/ER	Cytosol	
	Cvel_34654.t1	1165	-	-	-	Cytosol/ER	Cytosol	
	Cvel_34826.t1	1174	-	-	-	Cytosol/ER	Cytosol	
	Cvel_35138.t1	1013	-	-	-	Cytosol/ER	Cytosol	
	Cvel_35394.t1	1007	-	-	-	Cytosol/ER	Cytosol	
	Cvel_35394.t2	968	-	-	-	Cytosol/ER	Cytosol	
	Cvel_36416.t1	675	-	-	-	Cytosol/ER	Cytosol	
	Cvel_36618.t1	349	-	-	-	Cytosol/ER	Cytosol	
	Cvel_3734.t1	4498	-	-	-	Cytosol/ER	Cytosol	
	Cvel_6161.t1	6550	-	-	-	Cytosol/ER	Cytosol	
	Cvel_6212.t1	2635	-	-	-	Cytosol/ER	Cytosol	
	Cvel_678.t1	2469	-	-	-	Cytosol/ER	Cytosol	
	Vbra_1266.t1	6691	-	-	-	Cytosol/ER	Cytosol	
	Vbra_1193.t1	1857	-	-	-	Cytosol/ER	Cytosol	
	Vbra_12045.t1	3268	-	-	-	Cytosol/ER	Cytosol	
	Vbra_12047.t1	1204	-		SP (0.777, R4, L19)	-	Cytosol/ER	Cytosol
	Vbra_135.t1	1603	-	-	-	-	Cytosol/ER	Cytosol
	Vbra_14267.t1	1337	-	-	-	-	Cytosol/ER	Cytosol
	Vbra_1496.t1	3575	-	-	-	-	Cytosol/ER	Cytosol
	Vbra_16257.t1	1220	-	-	-	-	Cytosol/ER	Cytosol
Vbra_21227.t1	6651	-	-	-	-	Cytosol/ER	Cytosol	
Vbra_2242.t1	995		SP (0.959, L16)	M (0.442, R5, L12)	-	Cytosol/ER	ER/Microsome lumen	

Table 3. Ketoacyl synthases, elongases, and desaturases in chromerids and their predicted subcellular location.

Ketoacyl Synthases	Accession	Length (AA)	SignalP 3.0	TargetP 1.1	TargetP (trimmed SP)	ASAFind	Pred. Localization
Other KSs	Vbra_22449.t1	1668	-	-	-	Cytosol/ER	Cytosol
	Vbra_23040.t1	4167	-	-	-	Cytosol/ER	Cytosol
	Vbra_23145.t1	480	-	-	-	Cytosol/ER	Cytosol
	Vbra_23180.t1	2233	-	-	-	Cytosol/ER	Cytosol
	Vbra_23216.t1	301	-	-	-	Cytosol/ER	Cytosol
	Vbra_23250.t1	1732	-	-	-	Cytosol/ER	Cytosol
	Vbra_23378.t1	1092	-	-	-	Cytosol/ER	Cytosol
	Vbra_424.t1	1469	-	-	-	Cytosol/ER	Cytosol
	Vbra_425.t1	1098	-	-	-	Cytosol/ER	Cytosol
	Vbra_428.t1	3041	-	-	-	Cytosol/ER	Cytosol
	Vbra_460.t1	3015	-	-	-	Cytosol/ER	Cytosol
	Vbra_483.t1	2820	-	SP (0.479, R4, L43)	-	Cytosol/ER	Cytosol
	Vbra_50.t1	526	-	-	-	Cytosol/ER	Cytosol
	Vbra_502.t1	710	SP (0.871, L20)	SP (0.479, R4, L20)	-	Plastid	Plastid/ER
	Vbra_51.t1	3829	-	-	-	Cytosol/ER	Cytosol
	Vbra_6248.t1	1297	-	-	-	Cytosol/ER	Cytosol
	Vbra_7032.t1	2189	-	-	-	Cytosol/ER	Cytosol
Vbra_7414.t1	461	-	-	-	Cytosol/ER	Cytosol	
Vbra_8741.t1	1101	-	-	-	Cytosol/ER	Cytosol	
FASII (remaining) parts)	Accession	Length (AA)	SignalP 3.0	TargetP 1.1	TargetP (trimmed SP)	ASAFind	Pred. Localization
Acyl transferase	Cvel_4616.t1	349	SP (0.983, L20)	SP (0.729, L19, R3)	M 0.526 (R5, L19)	Plastid	Plastid
	Vbra_21812.t1	362	SP (0.996, L21)	SP (0.694, L20, R3)	M 0.635 (R4, L15)	Plastid	Plastid
Enoyl reductase	Cvel_5563.t1	355	SP (0.958, L20)	SP (0.785, L19, R3)	M 0.912 (R2, L23)	Plastid	Plastid
	Vbra_11747.t1	412	SP (0.996, L22)	SP (0.982, L33, R1)	M 0.919 (R1, L20)	Cytosol/ER	Plastid
Ketoacyl reductase (KR)	Cvel_3619.t1	392	SP (0.912, L19)	SP (0.915, L18, R1)	M 0.700 (R3, L17)	Plastid	Plastid
	Vbra_710.t1	338	SP (0.996, L22)	SP (0.943, L21, R1)	M 0.766 (R3, L10)	Plastid	Plastid
Dehydrogenase	Cvel_14912.t1	213	SP (0.849, L18)	SP (0.884, L17, R2)	M 0.526 (R5, L19)	Plastid	Plastid
	Vbra_19455.t1	225	SP (0.993, L25)	SP (0.800, L22, R2)	M 0.827 (R2, L43)	Plastid	Plastid

Table 4. Ketoacyl synthases, elongases, and desaturases in chromerids and their predicted subcellular location. S/MUFAs, saturated and monounsaturated FAs, PUFA, poly-unsaturated FAs.

Elongases								
S/MUFA	Cvel_6334.t1	193	SP (0.960, L16)	SP (0.861, R2, L16)	M (0.689, R3, L5)	Cytosol/ER	Plastid	
	Cvel_13090.t1	525	Anch (0.838, L41)	-	-	Cytosol/ER	ER/Microsome memb.	
	Cvel_14249.t1	884	Anch (0.791, L39)	-	-	Cytosol/ER	ER/Microsome memb.	
	Cvel_15641.t1	2131	Anch (0.606, L50)	-	-	Cytosol/ER	ER/Microsome memb.	
	Cvel_12461.t1	415	-	-	-	Cytosol/ER	Cytosol	
	Vbra_3929.t1	307	-	-	-	Cytosol/ER	Cytosol	
	Vbra_9666.t1	885	SP (0.675, L24)	SP (0.622, R3, L40)	S (0.950, R1, L16)	Cytosol/ER	ER/Microsome lumen	
	Vbra_11085.t1	281	Anch (0.560, L16)	-	-	Cytosol/ER	ER/Microsome memb.	
	Vbra_12050.t1	330	Anch (0.600, L41)	-	-	Cytosol/ER	ER/Microsome memb.	
	Vbra_13163.t1	363	-	-	-	Cytosol/ER	Cytosol	
PUFA	Cvel_1604.t1	297	-	-	-	Cytosol/ER	Cytosol	
	Vbra_3441.t1	363	Anch (0.984, L55)	-	-	Cytosol/ER	ER/Microsome memb.	
	Vbra_11843.t1	306	Anch (0.986, L63)	-	-	Cytosol/ER	ER/Microsome memb.	
	Vbra_16961.t1	291	Anch (0.638, L62)	-	-	Cytosol/ER	ER/Microsome memb.	

Table 5. Ketoacyl synthases, elongases, and desaturases in chromerids and their predicted subcellular location.

Desaturases	Accession	Length (AA)	SignalP 3.0	TargetP 1.1	TargetP (trimmed SP)	ASAFind	Pred. Localization
Omega	Cvel_2615.t1	461	-	-	-	Cytosol/ER	Cytosol
	Cvel_21003.t1	440	-	-	-	Cytosol/ER	Cytosol
	Cvel_22707.t1	425	SP (0.999, L17)	M/SP (0.631/0.771,L17, R5)	M (0.852, R2, L71)	Plastid	Plastid
	Vbra_7407.t1	445	-	M (0.564, L12, R5)	-	Cytosol/ER	Cytosol
	Vbra_15192.t1	465	-	-	-	Cytosol/ER	Cytosol
	Vbra_20615.t1	447	-	-	-	Cytosol/ER	Cytosol
Delta 9	Cvel_14249.t1	884	Anch (0.791, L39)	-	-	Cytosol/ER	ER/Microsome memb.
	Cvel_21149.t1	395	SP (0.910, L37)	SP (0.979, L37, R1)	M 0.814 (R2, L6)	Cytosol/ER	Plastid
	Vbra_9666.t1	885	SP (0.675, L40)	SP (0.622, L40, R3)	-	Cytosol/ER	ER/Microsome lumen
	Vbra_15445.t1	565	-	M (0.549, L19, R5)	-	Cytosol/ER	Cytosol
Front-end	Cvel_8966.t1	507	-	-	-	Cytosol/ER	Cytosol
	Cvel_17413.t1	543	-	-	-	Cytosol/ER	Cytosol
	Vbra_17659.t1	439	-	-	-	Cytosol/ER	Cytosol
	Vbra_20473.t1	437	-	M (0.738, L17, R3)	-	Mitochondrion	Mitochondrion
Acetyl-CoA Carboxylases	Accession	Length (AA)	SignalP 3.0	TargetP 1.1	TargetP (trimmed SP)	ASAFind	Pred. Localization
	Cvel_530.t2	2097	-	M (0.478, L18, R5)	-	Cytosol/ER	Cytosol
	Cvel_25292.t1	1651	-	-	-	Cytosol/ER	Cytosol
	Vbra_9562.t1	2702	-	-	-	Cytosol/ER	Cytosol
	Vbra_15163.t1	2146	SP (0.999, L19)	SP (0.828, L19, R2)	M 0.833 (R2, L20)	Cytosol/ER	Plastid

3.2. Elongation

The repertoire of FA elongases (Tables 1–5) differs between the two chromerids, even though their detected FA spectra are highly similar (Figure 5). The class-specific sequence motifs [52] and phylogenetic signal in desaturase domains allowed us to recognize two major functional classes (Figure 6):

i) Elongases specific for the saturated and monounsaturated fatty acids (S/MUFAs), forming four different orthologous groups. The total number of genes in each species is four, but their phylogenetic distribution within the tree and their predicted intracellular locations appeared to be species-specific. The first group predicted as cytosolic in location and counting one gene per species resulted to branch with stramenopiles. The second distant clade, putatively localized to the ER membrane, was found to be composed of *V. brassicaformis* (three genes) and *C. velia* (two genes). The remaining two S/MUFA elongases of *C. velia* and *V. brassicaformis* share origin with chlorophytes and plants at the very base of the S/MUFA group. A single distinct elongase of *C. velia* was found to form a sister long branch to haptophytes.

ii) Elongases specific for poly-unsaturated FAs (PUFA). A single gene coding for a cytosolic elongase (Cvel_1604.t1) of PUFA was found in *C. velia*, whereas *V. brassicaformis* appeared to encode three paralogous enzymes putatively located in the ER membrane (Tables 1–5). The single cytosolic PUFA desaturase of *C. velia* clusters with stramenopiles, while the three ER-located elongases from *V. brassicaformis* branch together with apicomplexans (Figure 6).

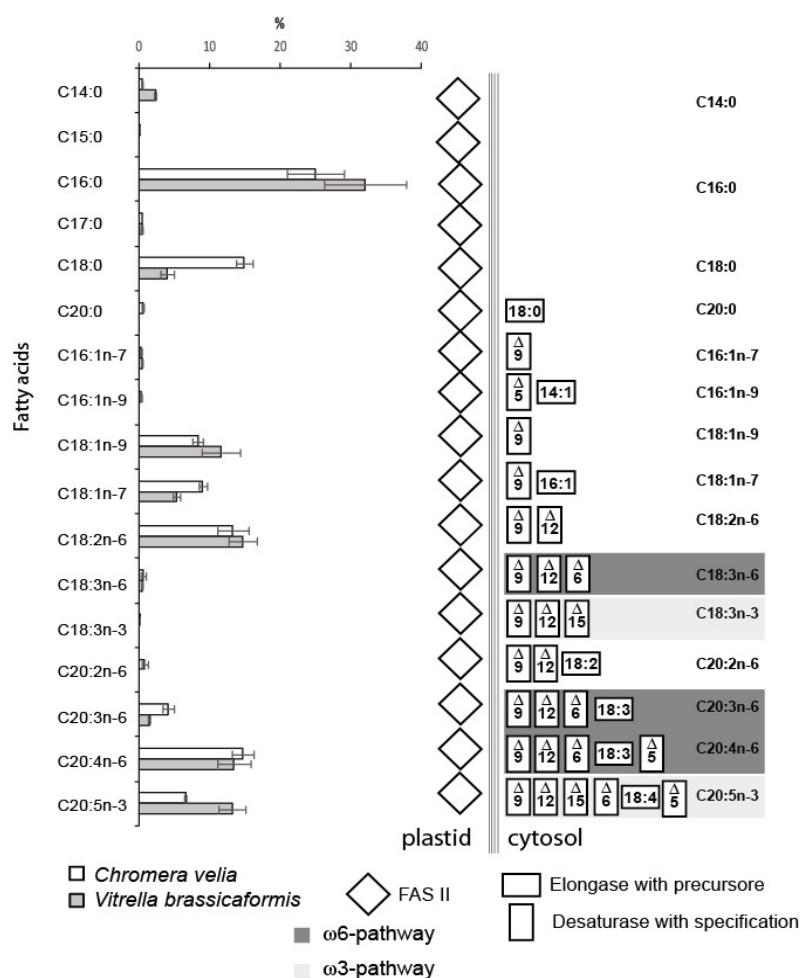


Figure 5. The fatty acid composition of both algae obtained by GC/FID (*Chromera velia* 350 ± 38 mg/g and *Vitrella brassicaformis* 4 ± 0.8 mg/g dry mass) and enzymatic repertoire for particular fatty acid synthesis and modification ($n = 5$). The concentrations of FAs are displayed in percentages with standard deviations. *C. velia* tends to accumulate storage lipids and contains approximately 100-fold

higher FAs concentrations than *V. brassicaformis*. The absolute concentrations of individual fatty acids can be calculated from absolute FAs concentrations values.

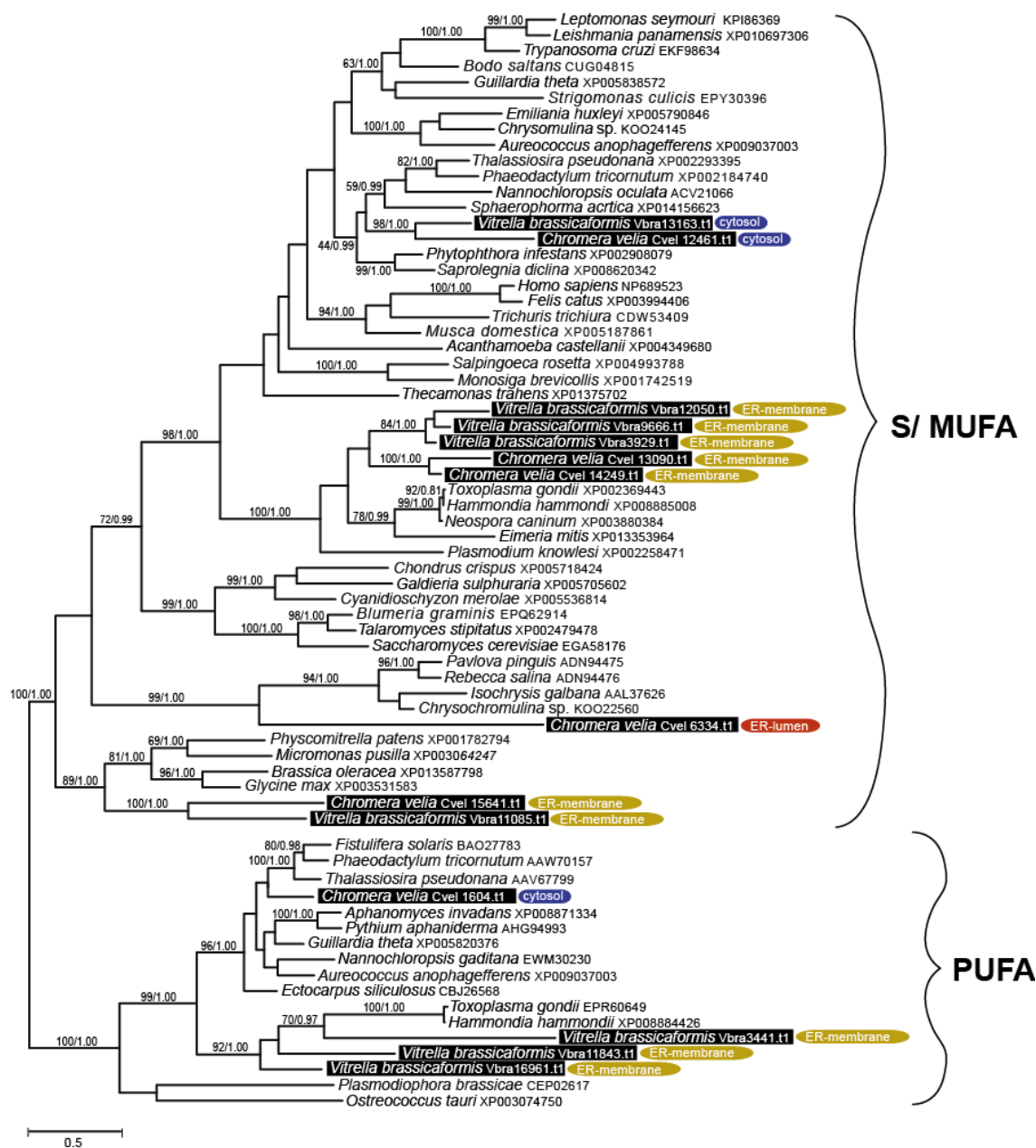


Figure 6. Bayesian phylogenetic tree inferred from elongase domains (70 taxa, 188 amino acid residues). Numbers above the branches display ML bootstrap/Bayesian PP support (BS1000/3M, CAT/LG). Predicted subcellular locations are displayed.

3.3. Desaturation

The desaturation of FAs is accomplished by desaturases of four different major classes defined by the specific products of their action. In both chromerid genomes, we identified desaturases of all functional classes. Most of the genes appeared to have no detectable N-terminal-targeting pre-sequence (Tables 1–5), with omega desaturases making an exception (Tables 1–5). Therefore, we presume that most desaturation processes likely take place in the cytosol. To distinguish the main functional enzyme groups, we constructed a maximum likelihood phylogeny (Figure 7) based on the dataset of desaturase domains from various organisms previously published by Gostincar et al. [53]. Using this initial phylogeny-based sorting, we built more detailed phylogenies focused on particular desaturase classes (Figures S1, S2, S3).

Sphingolipid desaturases were not an object of this study. However, they were included in the phylogeny (Figure 7) to test the completeness of the dataset. Regarding the evolution of desaturases, Δ -9 desaturases are generally considered as the ancestral group. The Δ -9 class is represented by two genes in both *C. velia* and *V. brassicaformis*. The likelihood tree (Figure S1) suggested their origin within the SAR clade (Stramenopila + Alveolata + Rhizaria). The highly divergent gene Cvel_21149.t1 forms the long branch far distant from the SAR group or any sequence data deposited in databases. The prediction suggests a highly possible plastid targeting of Cvel_21149.t1 (Tables 1–5).

The Δ -5/6 (front-end) enzyme class was found to be represented by two enzymes in each chromerid genome. The phylogenetic position of Cvel_17413.t1 is not fully resolved, likely due to high sequence divergence. The remaining genes from *C. velia* (1) and *V. brassicaformis* (2) coding for Δ -5/6 (front-end) desaturases constitute a monophyletic sister group to their chlorophyte counterparts, branching in the proximity of perkinsids, rhizarian, haptophyte, kinetoplastids, stramenopiles, and choanoflagellate in the well-supported eukaryotic clade. TargetP predicted Vbra_20473.t1 as mitochondrially located (Figure 8, Figure S2).

The Δ 12-desaturase group found in *C. velia* forms two distinct clades: the first clade consists of two genes from *C. velia*, branching with plants and cyanobacteria, suggesting its plastid origin. The second clade is composed of one desaturase from *C. velia* together with three genes from *V. brassicaformis* and branches between stramenopiles and ciliates, within a major clade composed mainly of SAR group members, haptophytes, and cryptophytes (Figure S3). The N-terminal pre-sequence detected in the desaturase Cvel_22707.t1 sharing its origin with plants displays a bi-partite architecture characteristic for targeting complex plastids (Tables 1–5). The desaturase Vbra_7407.t1 shows only a hint of mitochondrial transit peptide, with low probability and reliability for the prediction. Thus, we considered this structure to be an artifact and this whole clade of chromerid desaturases as cytosolic.

Fatty acid	<i>C. velia</i>	<i>V. brassicaformis</i>	<i>P. falciparum</i> ¹	<i>T. gondii</i> ²	<i>C. parvum</i> ³	<i>Symbiodinium</i> sp. ⁴	<i>P. tricornutum</i> ⁵	<i>C. reinhardtii</i> ⁶
C14:0	●	●	●	●	●	●	●	●
C16:0	●	●	●	●	●	●	●	●
C18:0	●	●	●	●	●	●	●	●
C20:0	●	×	●	●	●	×	×	●
C22:0	×	×	●	●	●	×	●	●
C24:0	×	×	●	●	●	×	×	●
C16:1	●	●	●	●	●	●	●	●
C18:1	●	●	●	●	●	●	●	●
C20:1	×	×	●	●	●	×	×	×
C22:1	×	×	●	●	×	●	×	×
C24:1	×	×	●	●	×	×	×	×
C26:1	×	×	×	●	×	×	×	×
C16:2	×	×	×	×	×	×	●	●
C18:2	●	●	●	×	●	●	●	●
C20:2	●	×	×	×	×	×	×	×
C16:3	×	×	×	×	×	×	●	●
C18:3	●	●	×	×	●	×	●	●
C20:3	●	●	●	×	×	×	●	×
C16:4	×	×	×	×	×	×	●	●
C18:4	×	×	×	×	×	×	●	●
C20:4	●	●	●	×	●	×	●	×
C20:5	●	●	×	×	×	●	●	×
C22:6	×	×	●	×	×	●	●	×

× absent or not detected ● synthesized ● obtained from host ○ found in traces ● nitrogen depletion ● increased under nitrogen depletion

Figure 8. Comparison of the fatty acid profile of particular organisms and the origin of particular fatty acids. Liver-stage of *Plasmodium*, intracellular stages of *Toxoplasma gondii* 1) Botté et al., 2013 [54], 2) Ramakrishnan et al., 2012, 2015 [52,55], 3) Mitschler et al., 1994 [56], 4) Weng et al., 2014 [57], 5) Siron et al., 1989; Popko et al., 2016 [58,59], 6) James et al., 2011; Puzanskiy et al., 2015 [60,61]. The position of the last double bond was not specified in some mentioned publications; therefore, it is not included in the table.

3.4. Fatty Acid Profile

The list of all FAs detected in chromerids and their abundances are shown in Figure 5. The outcome of the described complex enzymatic machinery is a broad spectrum of FAs that are variable in abundance, number of carbons, number and position of double bonds. *C. velia* and *V. brassicaformis* have a high content of palmitic acid (C16:0), covering 25% and 33% of all FAs in the two algae. The FA profiles of chromerids are very similar to each other, with two exceptions: the occurrence of stearic acid (C18:0) in *C. velia* is three times higher than in *V. brassicaformis*, and *V. brassicaformis* contains two times more eicosapentaenoic acid (C20:5n-3) (EPA) than *C. velia*. The detected FAs (Figure 8, Figure 5) indicate the presence of both PUFA pathways ω 6 (animals) and ω 3 (algae), resulting in a chain of 20 carbon units with five double bonds (EPA). However, the last enzyme of the ω 6 pathway, Δ -17desaturase, is missing in both algae in correlation with their high content of arachidonic acid (C20:4n-6). It seems that *V. brassicaformis* engages the ω 3 pathway preferentially, unlike *C. velia*, which employs both pathways at a similar level (Figure 5, Figure S4a,b). The precursors of both pathways occur at very low levels or below the limit of detection (e.g., C18:3n-6, C18:3n-3, and C18:4n-3). The presence of C20:2n-6 among FAs of *C. velia* indicates an alternative ω 6 pathway (Figure 3, Figure S4a). In contrast, C20:2n-6 is not present in *V. brassicaformis*, with no obvious alternative ω 6 pathway (Figure S4b). The high variability of FAs detected in chromerids is also highlighted by the presence of oleic FA (C18:1n-9) and cis-vaccenic acid (C18:1n-7). These two FAs are based on palmitic acid, modified by only two enzymes, Δ -9 desaturase and elongase. C18:1n-9 is synthesized by elongation of palmitic acid (C16:0) to stearic acid (C18:0) and subsequent desaturation at the 9th carbon position,

in contrast to the synthesis of cis-vaccenic acid, which begins with the desaturation of palmitic acid at the 9th carbon position followed by elongation, resulting in a shift of the double bond to the 7th position (for details, see Figure S4a,b).

After 20 days of triclosan treatment, all cultures treated with a concentration higher than 40 μM stopped growing (Figure S5), bleached, and died. Cells treated with solvent control (DMSO) were not affected (data not shown). At the concentration of 40 μM , triclosan drastically inhibited the production of unsaturated FAs and slightly affected the abundance of saturated FAs. When the inhibitor concentration was increased up to 330 μM , we observed the vanishing of unsaturated FAs and a dramatic downregulation of C16:0. No such drastic decrease was observed in the production of C18:0 (Figure S6). Even a slightly higher concentration of triclosan in the medium caused a fatal collapse of the whole membrane system, as seen by transmission electron microscopy (Figure S7a,b).

Our experiment with nitrogen repletion and depletion showed changes in the lipid profiles of *C. velia*: FAs up to C18 increased in abundance by approximately 14% during nitrogen depletion, while FAs longer than C18 were more abundant during nitrogen repletion (Figure S8). The modulation of the FA profile by environmental conditions was also described [62].

4. Discussion

Both types of FAS have been detected in chromerids, *T. gondii*, and other coccidians (Figure 9). Type I FAS enzymes are preserved mainly in organisms that have lost their plastids, such as *Cryptosporidium*, or in those that never had plastids. It is possible that in plastid-bearing organisms, the rarely present cytosolic FAS-I is employed in the synthesis of some special FAs; however, it is more likely that FAS-I multi-modular enzymes in chromerids and coccidia are involved in the elongation of FAs rather than in de novo synthesis, as suggested for FAS-I enzyme in *C. parvum*. It was nicely shown that the multi-modular FAS-I enzyme from this plastid-lacking intestinal apicomplexan parasite is responsible for the elongation of short fatty acids scavenged from the host [33]. Since the acetyl-CoA (coenzyme A) carboxylase (ACC) from *C. parvum*, which is used for elongation and not FA de novo synthesis branches together with the cytosolic ACCs from *T. gondii* and chromerids (Figure S9), they can hypothetically do the same job. Although chromerids, coccidians, and *Cryptosporidium* are known to encode numerous FAS-I-like multi-modular enzymes in their genomes, only some of them are believed to be involved in FAS metabolism. Others more likely represent evolutionary related genes coding for type I polyketide synthases (PKS-I) responsible for the synthesis of polyketides, secondary metabolites known to act as antimicrobial compounds and toxins, such as cyanobacterial toxins, dinoflagellate red-tide toxins (e.g., brevetoxin), mycotoxins (e.g., aflatoxin), and even some compounds used as antibiotics (geldanamycin, doxycycline, and others) [51]. All these enzymes contain at least one incomplete module lacking an acyltransferase (AT) domain, with the absence of this domain being considered as a hallmark of most algal type I PKSs [25]. Furthermore, one and two multi-modular FAS-like enzymes in *C. velia* and *V. brassicaformis*, respectively, contain additional methyltransferase (MT) domains, which is also typical for PKSs [63].

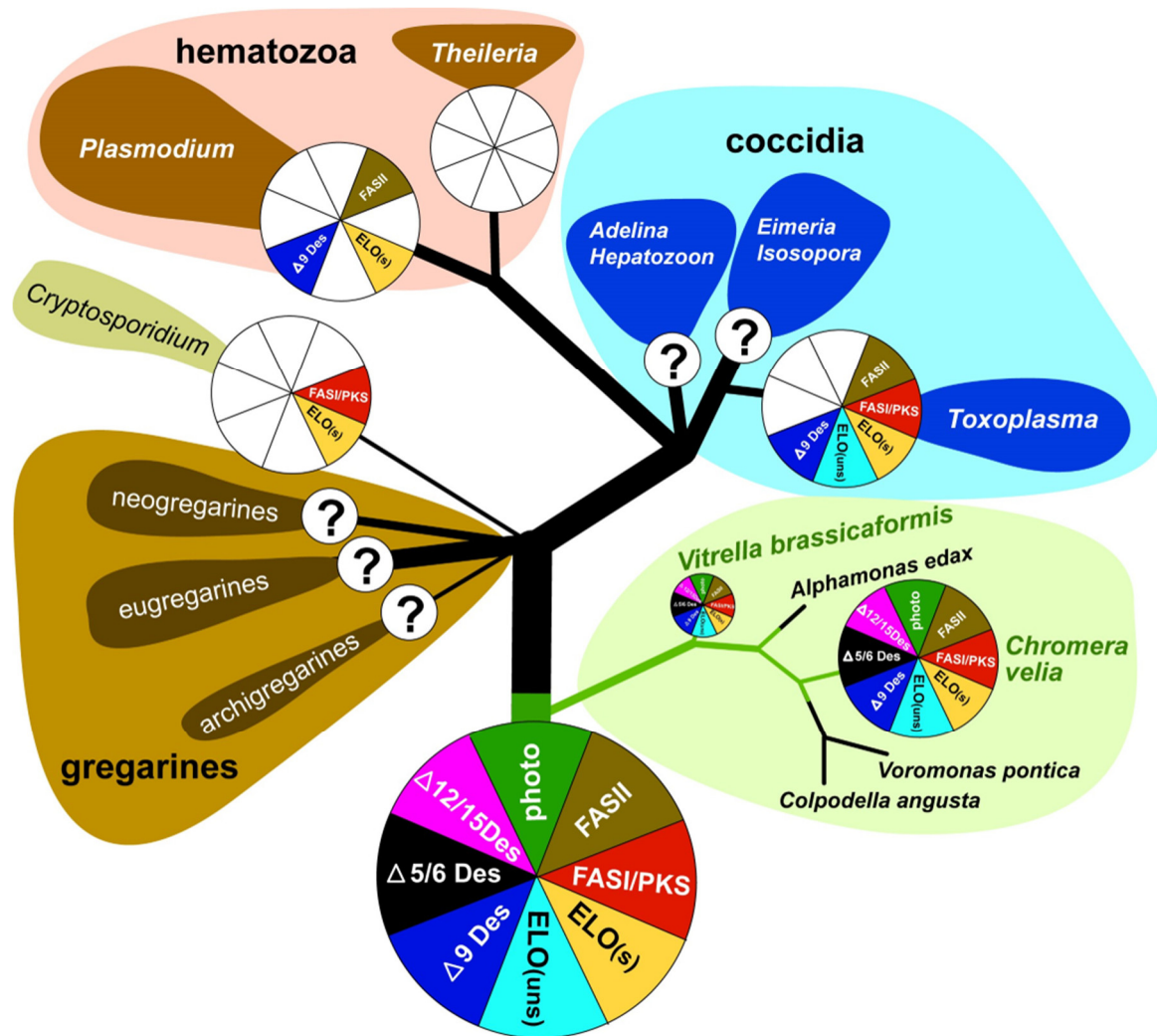


Figure 9. Overview of FAS evolution in chromerids and apicomplexan parasites.

The coccidian *T. gondii* contains at least one FAS-I enzyme with four complete modules (9940 amino acids) [57] and PKS-I enzymes [22]. *C. parvum* contains, in addition to the FAS-I multi-modular gene with three modules, also a PKS gene, which consists of seven modules containing 13,411 amino acids. The exclusive feature of all FAS-I and PKS-I in apicomplexans is likely the formation of highly resilient oocysts (*Cryptosporidium*, *Eimeria*, *Toxoplasma*) [30]. Since these enzymes have affinity to similar substrates, it is speculated that FAS-I, together with multi-domain PKSs, collaborate to produce special lipids constituting the oocyst membrane. This cooperation between FAS-I and PKS-I has also been reported in certain dinoflagellates, fungi, and bacteria [22]. A durable cell wall is also characteristic of certain life stages of *C. velia* and *V. brassicaformis* [1,2]. Although most multi-domain FAS-I/PKS-I enzymes in chromerids show sequence structures typical for PKS rather than FAS-I, both algae have two different ACCases (Figure S9, S10). The first one is related to the plastid originated and located ACCases, similar to those of other apicomplexans, dinoflagellates, cryptophytes, haptophytes, and chlorophytes. The second one, which is likely cytosolic, branches with that of *T. gondii* and *C. parvum*. Since it was proposed that cytosolic acetyl CoA carboxylase could be a source of malonyl-CoA for the elongation of fatty acids [22,64], FAS-I in apicomplexans may be involved in elongation. As canonical elongases are also present in chromerid genomes, FAS-I multi-modular enzymes can be involved in FA elongation in specific life stages, like in *T. gondii* [65] and *C. parvum* [33].

As mentioned previously [13,58], chromerids are equipped with cytosolic FAS-I multi-modular enzymes originating from the eukaryotic heterotrophic host (exosymbiont) and with plastid-located FAS-II, which was acquired during the complex endosymbiotic event from the endosymbiont (Tables 1–5). This pathway arrangement correlates with the elevated expression of plastidial ACCase during nitrogen depletion, which has been described in *C. velia*, and in the haptophyte *Isochrysis galbana* [64]. The tendency of chromerids to switch metabolism towards storing energy in the form of lipids, which is enhanced in nitrogen-poor media, is shared with other algae like the symbiotic dinoflagellate *Symbiodinium* spp. [59], the eustigmatophyte *Nannochloropsis gaditana* [66], the diatom *Phaeodactyllum tricornutum* [60,61,67], the rhodophyte *Cyanidioschyzon merolae*, and the chlorophytes *Chlorella vulgaris* [68] and *Chlamydomonas reinhardtii* [55,69,70]. The complete FAS-II apicoplast-located system was recorded in *T. gondii*, *P. falciparum*, and *E. tenella* [22,29,54,71–75], and all these apicomplexans use it for the de novo synthesis of saturated FAs [20,29,57,76]. The apicomplexan minimalist *Cryptosporidium* does not possess the apicoplast, consequently lacks the FAS-II pathway, and fully depends on FAs scavenged from the host and in-house elongation [33]. The blood parasites *Babesia* and *Theileria* completely lack FAS and modification enzymes, despite the presence of the apicoplast in their cells [20,22,33] (Figure 9). Importantly, the FAS-II pathway was found in all organisms bearing red algal-derived plastids [23]. Based on experiments with the reportedly selective inhibitor of FAS-II triclosan [77], it was initially suggested that it is the FAS-I system which is responsible for FA de novo synthesis in *C. velia*, while FAS-II is involved in elongation and desaturation of primary FAs to final PUFAs arising from both ω -3 and ω -6 pathways [13]. A similar FA biosynthetic and modulatory system was initially proposed for *T. gondii*, again partly based on experiments with triclosan and other selective inhibitors. Mazumdar and Striepen [30] deduced that *T. gondii* employs FAS-I in the de novo synthesis of FAs and FAS-II in lipoic acid synthesis. However, experiments performed in the absence of selective inhibitors finally suggested that the de novo synthesis of FAs in *T. gondii* is maintained exclusively by the apicoplast-located type II FAS pathway [54,57]. After several of our experiments with triclosan and considering the above-mentioned published results, we doubt its inhibitory specificity. The target of triclosan is supposed to be the enoyl reductase domain, which plays an indispensable role in both FAS-I and FAS-II systems. Thus, the inhibition should affect both types of synthesis [78]. Since the enzymatic steps involved in the elongation process are like those of FAS-I and II, but the growing chain is held by CoA instead of ACP [30], we hypothesize that lower concentrations of triclosan mainly affect cytosolic enoyl reductases, while the plastidial ones are likely protected by multiple plastid membranes. It seems that plastid membrane protection is not efficient at triclosan concentrations of 80 μ M and higher. A fatal collapse of the whole membrane system, which is apparent in the TEM pictures (Figure S6), could be induced by the inhibition of plastidial enoyl reductase. We assume that the specificity of triclosan is at least questionable and that this compound non-specifically inhibits elongases and desaturases in the cytosol and endoplasmic reticulum rather than FA synthesis in the plastid. Therefore, we propose that the de novo synthesis of FAs occurs in the plastid using a prokaryotic type of fatty acid synthesis, like in other phototrophic eukaryotes. The FAS-II pathway was found responsible for FA synthesis in all organisms bearing red algal-derived plastids [23]. It also appears reasonable that one of the most energetically expensive processes, for which also reduction equivalents are needed, is associated with the plastid, the main ‘power station’ and producer of reduction equivalents of the phototrophic cell. Taking all the above-mentioned facts into account, we propose that chromerids synthesize at least the major pool of fatty acids in the plastid using the FAS-II system.

Sequence identification of the enzymes involved in FAS and their localization results (Tables 1–5) together with GC/FID analysis of FAs (Figure 5) provided enough information for proposing a model for FA biosynthesis, elongation, and desaturation in chromerids (Figure S4 a, b). The primary sources of energy and carbons for the synthesis of FAs in chromerids are photosynthesis and the Calvin cycle, as described in plants [56,79]. Based on our findings, we suggest that the plastid-located FAS-II system de novo synthesizes short saturated FAs—mainly C16:0, followed by C18:0 and C14:0 acids—and exports them to the cytosol in the form of FAs or associated with ACP. FA-ACP complexes are then modified by the set of desaturases and elongases. Most enzymes participating in

the final steps of FA synthesis are either located in the cytosol or associated with the membrane of the endoplasmic reticulum, with the exception of one plastid-targeted $\Delta 9$ desaturase (Cvel_21149), one plastid-targeted ω -desaturase (Cvel_22707), and one front-end desaturase (Vbra 20473), likely located in the mitochondrion (Tables 1–5). Plastid-targeted $\Delta 9$ and ω -desaturases were also described in plants [24] and are the only desaturases retained in *Plasmodium* lacking any traces of FAS-I system (Figure 9). It should also be pointed out that chromerids do not produce C16 fatty acids with more than one double bond. Such poly-unsaturated C16 fatty acids are common in a broad spectrum of non-chromerid algae; for example, the diatom *P. tricornutum* and the green alga *C. reinhardtii* contain these FAs even with four double bonds in the 16 carbons chain (Figure 8). The presence of an alternative pathway to produce EPA in *C. velia* is very common and has been previously reported in the chlorophyte *Pariechloris incista* [80]. Furthermore, the diatom *P. tricornutum* has three [81], and *Euglena gracilis* has four different pathways to produce EPA [82]. The absence of any alternative EPA synthesis pathway in *V. brassicaformis* is even more interesting (Figure S4b). Chromerids lack long FAs (longer than C20), like other mentioned eukaryotic phototrophs, except for *P. tricornutum*, containing polyunsaturated FAs (22:6) (Figure 8).

Chromerids encode a broad spectrum of enzymes involved in the synthesis of short-chain saturated FAs and their derivatives. The evolutionary diagram in Figure 9 displays differences between chromerids and apicomplexans, which have completely lost $\Delta 5/6$ and $\Delta 12/15$ desaturases essential for the biosynthesis of PUFAs. However, previous works dealing with fatty acid determination showed the occurrence of PUFA in *P. falciparum* [76] and *C. parvum* [83] (for details, see Figure 8.). The presence of PUFA in these two parasites can be explained by their salvage from the host. The only known apicomplexans completely lacking the apicoplast are neogregarines (like *Gregarina niphandrodes*) [84] and members of the genus *Cryptosporidium* [85,86]. Due to the loss of the complete plastidial FAS II pathway and all desaturases and elongases, *C. parvum* displays the most reduced enzymatic equipment for FA synthesis and modification among apicomplexans. Like *C. velia*, *T. gondii* and *E. tenella* have retained the majority of the enzymatic equipment, except for $\Delta 5/6$ and $\Delta 12/15$ desaturases. While *T. gondii* uses three elongases and only a single $\Delta 9$ desaturase [54], *C. velia* and *V. brassicaformis* encode six and eight elongases, respectively, and two $\Delta 9$ desaturases. The loss of photosynthesis and switch to parasitism have been accompanied by losses of enzymes involved in FA synthesis in a lineage-specific pattern. The ancestral state was most likely equipped as the typical photosynthetic primary producer.

Cytosolic S/MUFA elongases branch along with stramenopiles (Figure 4), suggesting the possible transfer from a hypothetical tertiary stramenopile endosymbiont linked to the origin of the chromerid plastid [2,4,87,88]. The group of S/MUFA elongases branching along with apicomplexans reflects the common origin of the chromerids and apicomplexans in the frame of the myzozoan lineage. A single gene of *C. velia* (Cvel_6334.t1) constitutes a long branch, sister to haptophytes. Similarly, a single PUFA elongase found in *C. velia* (Cvel_1604.t1) is related to stramenopiles, in contrast to three genes from *V. brassicaformis*, with apicomplexan relatives. Long branching and orphan phylogenetic positions are common phenomena for genes of *C. velia*. It appears that the alga retained genes of ancient origin, acting as a collector of genetic fossils.

5. Conclusions

Sequence identification of particular enzymes involved in FA synthesis and their predicted localization (Tables 1–5) together with GC analysis of particular FAs (Figure 5) provided enough information for proposing a system of fatty acid biosynthesis, elongation, and desaturation in chromerids (Figure S4). The primary sources of energy and carbons for the synthesis of fatty acids in chromerids are photosynthesis and the Calvin cycle, as described in plants [56,79]. We propose that the FAS II system localized in the plastid produces de novo short saturated fatty acids—mainly C16:0 (palmitic acid), followed by C18:0 (stearic acid) and C14:0 (myristic acid)—and exports them to the cytosol as fatty acids or in association with ACP. Fatty acids with ACP are then modified by the set of desaturases and elongases. Most enzymes participating in the final steps of fatty acid synthesis are either localized in the cytosol or associated with the membrane of the endoplasmic reticulum, with

the exclusion of one $\Delta 9$ desaturase (Cvel21149), one ω -desaturase (Cvel22707), which are putatively localized in the plastid (Tables 1–5), as described in plants [24], and one front-end desaturase (Vbra 20473) probably localized in the mitochondrion.

Supplementary Materials: The following are available online at www.mdpi.com/2218-273X/10/8/1102/s1: Figure S1. The $\Delta 9$ desaturases Likelihood/Bayesian tree. Figure S2. The $\Delta 5/6$ (front-end) desaturase Likelihood/Bayesian tree. Figure S3. The omega ($\Delta 12/15$) desaturase Likelihood/Bayesian tree. Figure S4. The model of fatty acid biosynthesis in *Chromera velia* (A) and in *Vitrella brassicaformis* (B). Figure S5. Growing curves of *Chromera velia* treated by the various concentration of triclosan recorded by the Tecan instrument ($n = 5$). Figure S6. Spectra of fatty acid composition changes after triclosan treatment obtained by GC/FID. Figure S7. Transmission electron microscopy of *C. velia* grown in standard f/2 medium (A) and grown in f/2 medium with 333 μ M triclosan (B), where the cell is deformed and the inner structure is collapsed. LD – lipid droplets; N – nucleus; P – plastid. Figure S8. The ratio of fatty acids shorter than C18 and C18 (blue) and longer than C18 (orange) in *C. velia* during nitrogen deprivation and repletion ($n = 5$). Figure S9. Maximum-likelihood phylogenetic tree of acetyl-CoA decarboxylases. Figure S10. The alignment of ACCases of chromerids, showing the first 408 amino acid residues, with the N-terminal motif similarity shared between plastid-targeted gene variants.

Author Contributions: M.O. conceived the study. A.T. performed experiments, interpreted GC/FID data, proposed the model of fatty acid biosynthesis. J.M. performed genomic searches and phylogenetic analyses and interpreted the results. I.S. extracted lipids, prepared samples for derivatization, and performed the GC/FID analysis. M.V. prepared the T.E.M. samples. A.T., J.M., M.O., Z.F. and A.G. wrote the manuscript. MO obtained funding. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Czech Science Foundation (project 18-13458S), Czech Biolmaging (MEYS CR, LM2015062), Ministry of Education, Youth, and Sports of the Czech Republic—the CENAKVA project (LM2018099), CENAKVA Center Development (CZ.1.05/2.1.00/19.0380), and ERDF/ESF Centre for research of pathogenicity and virulence of parasites (No.CZ.02.1.01/0.0/0.0/16_019/0000759).

Acknowledgments: We thank Geoffrey I. McFadden for helpful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Moore, R.B.; Oborník, M.; Janouškovec, J.; Chrudimský, T.; Vancová, M.; Green, D.H.; Wright, S.W.; Davies, N.W.; Bolch, C.J.S.; Heimann, K.; et al. A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **2008**, *451*, 959–963, doi:10.1038/nature06635.
2. Oborník, M.; Modrý, D.; Lukeš, M.; Černotiková-Stříbrná, E.; Cihlář, J.; Tesařová, M.; Kotabová, E.; Vancová, M.; Prášil, O.; Lukeš, J. Morphology, Ultrastructure and Life Cycle of *Vitrella brassicaformis* n. sp., n. gen., a Novel Chromerid from the Great Barrier Reef. *Protist* **2012**, *163*, 306–323, doi:10.1016/j.protis.2011.09.001.
3. Oborník, M.; Janouškovec, J.; Chrudimský, T.; Lukeš, J. Evolution of the apicoplast and its hosts: From heterotrophy to autotrophy and back again. *Int. J. Parasitol.* **2009**, *39*, 1–12, doi:10.1016/j.ijpara.2008.07.010.
4. Füssy, Z.; Oborník, M. Chromerids and Their Plastids. In *Secondary Endosymbioses*; Hirakawa, Y., Ed.; Academic Press Ltd-Elsevier Science Ltd.: London, UK, 2017; Volume 84, pp. 187–218.
5. Cumbo, V.R.; Baird, A.H.; Moore, R.B.; Negri, A.P.; Neilan, B.A.; Salih, A.; van Oppen, M.J.H.; Wang, Y.; Marquis, C.P. *Chromera velia* is Endosymbiotic in Larvae of the Reef Corals *Acropora digitifera* and *A. tenuis*. *Protist* **2013**, *164*, 237–244, doi:10.1016/j.protis.2012.08.003.
6. Janouškovec, J.; Sobotka, R.; Lai, D.; Flegontov, P.; Koník, P.; Komenda, J.; Ali, S.; Prášil, O.; Pain, A.; Oborník, M.; et al. Split Photosystem Protein, Linear-Mapping Topology, and Growth of Structural Complexity in the Plastid Genome of *Chromera velia*. *Mol. Biol. Evol.* **2013**, *30*, 2447–2462.
7. Okamoto, N.; McFadden, G.I. The mother of all parasites. *Future Microbiol.* **2008**, *3*, 391–395, doi:10.2217/17460913.3.4.391.
8. Mohamed, A.R.; Cumbo, V.R.; Harii, S.; Shinzato, C.; Chan, C.X.; Ragan, M.A.; Satoh, N.; Ball, E.E.; Miller, D.J. Deciphering the nature of the coral-Chromera association. *ISME J.* **2018**, *12*, 776–790, doi:10.1038/s41396-017-0005-9.
9. Mathur, V.; del Campo, J.; Kolisko, M.; Keeling, P.J. Global diversity and distribution of close relatives of apicomplexan parasites. *Environ. Microbiol.* **2018**, *20*, 2824–2833, doi:10.1111/1462-2920.14134.

10. Janouškovec, J.; Tikhonenkov, D.V.; Burki, F.; Howe, A.T.; Kolisko, M.; Mylnikov, A.P.; Keeling, P.J. Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10200–10207, doi:10.1073/pnas.1423790112.
11. Janouškovec, J.; Horák, A.; Oborník, M.; Lukeš, J.; Keeling, P.J. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10949–10954, doi:10.1073/pnas.1003335107.
12. Kořený, L.; Sobotka, R.; Janouškovec, J.; Keeling, P.J.; Oborník, M. Tetrapyrrole Synthesis of Photosynthetic Chromerids Is Likely Homologous to the Unusual Pathway of Apicomplexan Parasites. *Plant Cell* **2011**, *23*, 3454–3462, doi:10.1105/tpc.111.089102.
13. Woo, Y.H.; Ansari, H.; Otto, T.D.; Klinger, C.M.; Kolisko, M.; Michálek, J.; Saxena, A.; Shanmugam, D.; Tayyrov, A.; Veluchamy, A.; et al. Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *Elife* **2015**, *4*, doi:10.7554/eLife.06974.
14. Oborník, M. Endosymbiotic Evolution of Algae, Secondary Heterotrophy and Parasitism. *Biomolecules* **2019**, *9*, doi:10.3390/biom9070266.
15. Flegontov, P.; Michálek, J.; Janouškovec, J.; Lai, D.H.; Jirků, M.; Hajdůšková, E.; Tomčala, A.; Otto, T.D.; Keeling, P.J.; Pain, A.; et al. Divergent Mitochondrial Respiratory Chains in Phototrophic Relatives of Apicomplexan Parasites. *Mol. Biol. Evol.* **2015**, *32*, 1115–1131, doi:10.1093/molbev/msv021.
16. Oborník, M.; Vancová, M.; Lai, D.-H.; Janouškovec, J.; Keeling, P.J.; Lukeš, J. Morphology and Ultrastructure of Multiple Life Cycle Stages of the Photosynthetic Relative of Apicomplexa, *Chromera velia*. *Protist* **2011**, *162*, 115–130, doi:10.1016/j.protis.2010.02.004.
17. Füßy, Z.; Masařova, P.; Kručinská, J.; Esson, H.J.; Oborník, M. Budding of the Alveolate Alga *Vitrella brassicaformis* Resembles Sexual and Asexual Processes in Apicomplexan Parasites. *Protist* **2017**, *168*, 80–91, doi:10.1016/j.protis.2016.12.001.
18. Cavalier-Smith, T. Kingdom Chromista and its eight phyla: A new synthesis emphasising periplastid protein targeting, cyto-skeletal and periplastid evolution, and ancient divergences. *Protoplasma* **2018**, *255*, 297–357, doi:10.1007/s00709-017-1147-3.
19. Fahy, E.; Subramaniam, S.; Brown, H.A.; Glass, C.K.; Merrill, A.H.; Murphy, R.C.; Raetz, C.R.H.; Russell, D.W.; Seyama, Y.; Shaw, W.; et al. A comprehensive classification system for lipids. *J. Lipid Res.* **2005**, *46*, 839–861, doi:10.1194/jlr.E400004-JLR200.
20. Goodman, C.D.; McFadden, G.I. Fatty acid biosynthesis as a drug target in apicomplexan parasites. *Curr. Drug Targets* **2007**, *8*, 15–30, doi:10.2174/138945007779315579.
21. Chirala, S.S.; Wakil, S.J. Structure and function of animal fatty acid synthase. *Lipids* **2004**, *39*, 1045–1053, doi:10.1007/s11745-004-1329-9.
22. Ramakrishnan, S.; Serricchio, M.; Striepen, B.; Butikofer, P. Lipid synthesis in protozoan parasites: A comparison between kinetoplastids and apicomplexans. *Prog. Lipid Res.* **2013**, *52*, 488–512, doi:10.1016/j.plipres.2013.06.003.
23. Ryall, K.; Harper, J.T.; Keeling, P.J. Plastid-derived Type II fatty acid biosynthetic enzymes in chromists. *Gene* **2003**, *313*, 139–148, doi:10.1016/s0378-1119(03)00671-1.
24. Schmid, K.M. *Biochemistry of Lipids, Lipoproteins and Membranes*. Elsevier B.V.: 2016; pp. 113–147.
25. Shelest, E.; Heimerl, N.; Fichtner, M.; Sasso, S. Multimodular type I polyketide synthases in algae evolve by module duplications and displacement of AT domains in trans. *BMC Genom.* **2015**, *16*, doi:10.1186/s12864-015-2222-9.
26. McFadden, G.I. Plastids and protein targeting. *J. Eukaryot. Microbiol.* **1999**, *46*, 339–346, doi:10.1111/j.1550-7408.1999.tb04613.x.
27. Wilson, R.J.M. Plastid functions in the Apicomplexa. *Protist* **2004**, *155*, 11–12, doi:10.1078/1434461000158.
28. Zhu, G. Current progress in the fatty acid metabolism in *Cryptosporidium parvum*. *J. Eukaryot. Microbiol.* **2004**, *51*, 381–388, doi:10.1111/j.1550-7408.2004.tb00384.x.
29. Lu, J.Z.; Muench, S.P.; Allary, M.; Campbell, S.; Roberts, C.W.; Mui, E.; McLeod, R.L.; Rice, D.W.; Prigge, S.T. Type I and type II fatty acid biosynthesis in *Eimeria tenella*: Enoyl reductase activity and structure. *Parasitology* **2007**, *134*, 1949–1962, doi:10.1017/s0031182007003319.
30. Mazumdar, J.; Striepen, B. Make it or take it: Fatty acid metabolism of Apicomplexan parasites. *Eukaryot. Cell* **2007**, *6*, 1727–1735, doi:10.1128/ec.00255-07.
31. Sonda, S.; Hehl, A.B. Lipid biology of Apicomplexa: Perspectives for new drug targets, particularly for *Toxoplasma gondii*. *Trends Parasitol.* **2006**, *22*, 41–47, doi:10.1016/j.pt.2005.11.001.

32. Salomaki, E.D.; Kolisko, M. There Is Treasure Everywhere: Reductive Plastid Evolution in Apicomplexa in Light of Their Close Relatives. *Biomolecules* **2019**, *9*, doi:10.3390/biom9080378.
33. Zhu, G.; Li, Y.N.; Cai, X.M.; Millership, J.J.; Marchewka, M.J.; Keithly, J.S. Expression and functional characterization of a giant Type I fatty acid synthase (CpFAS1) gene from *Cryptosporidium parvum*. *Mol. Biochem. Parasitol.* **2004**, *134*, 127–135, doi:10.1016/j.molbiopara.2003.11.011.
34. John, U.; Beszteri, B.; Derelle, E.; de Peer, Y.V.; Read, B.; Moreau, H.; Cembella, A. Novel insights into evolution of protistan polyketide synthases through phylogenomic analysis. *Protist* **2008**, *159*, 21–30, doi:10.1016/j.protis.2007.08.001.
35. Heath, R.J.; Rubin, J.R.; Holland, D.R.; Zhang, E.L.; Snow, M.E.; Rock, C.O. Mechanism of triclosan inhibition of bacterial fatty acid synthesis. *J. Biol. Chem.* **1999**, *274*, 11110–11114, doi:10.1074/jbc.274.16.11110.
36. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066, doi:10.1093/nar/gkf436.
37. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649, doi:10.1093/bioinformatics/bts199.
38. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313, doi:10.1093/bioinformatics/btu033.
39. Huelsenbeck, J.P.; Bollback, J.P. Empirical and hierarchical Bayesian estimation of ancestral states. *Syst. Biol.* **2001**, *50*, 351–366, doi:10.1080/106351501300317978.
40. Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572–1574, doi:10.1093/bioinformatics/btg180.
41. Emanuelsson, O.; Nielsen, H.; Brunak, S.; von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **2000**, *300*, 1005–1016, doi:10.1006/jmbi.2000.3903.
42. Nielsen, H.; Engelbrecht, J.; Brunak, S.; von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **1997**, *10*, 1–6, doi:10.1093/protein/10.1.1.
43. Bendtsen, J.D.; Nielsen, H.; von Heijne, G.; Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **2004**, *340*, 783–795, doi:10.1016/j.jmb.2004.05.028.
44. Gruber, A.; Roco, G.; Kroth, P.G.; Armbrust, E.V.; Mock, T. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J.* **2015**, *81*, 519–528, doi:10.1111/tpj.12734.
45. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the 2010 Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010; pp. 1–8.
46. Folch, J.; Lees, M.; Stanley, G.H.S. A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* **1957**, *226*, 497–509.
47. Košťál, V.; Šimek, P. Changes in fatty acid composition of phospholipids and triacylglycerols after cold-acclimation of an aestivating insect prepupa. *J. Comp. Physiol. B Biochem. Syst. Environ. Physiol.* **1998**, *168*, 453–460, doi:10.1007/s003600050165.
48. Tomčala, A.; Kyselová, V.; Schneedorferová, I.; Opekarová, I.; Moos, M.; Urajová, P.; Kručinská, J.; Oborník, M. Separation and identification of lipids in the photosynthetic cousins of Apicomplexa *Chromera velia* and *Vitrella brassicaformis*. *J. Sep. Sci.* **2017**, *40*, 3402–3413, doi:10.1002/jssc.201700171.
49. Zahradníčková, H.; Tomčala, A.; Berková, P.; Schneedorferová, I.; Okrouhlík, J.; Šimek, P.; Hodková, M. Cost effective, robust, and reliable coupled separation techniques for the identification and quantification of phospholipids in complex biological matrices: Application to insects. *J. Sep. Sci.* **2014**, *37*, 2062–2068, doi:10.1002/jssc.201400113.
50. Dahmen, J.L.; Khadka, M.; Dodson, V.J.; Leblond, J.D. Mono- and digalactosyldiacylglycerol composition of dinoflagellates. VI. Biochemical and genomic comparison of galactolipid biosynthesis between *Chromera velia* (Chromerida), a photosynthetic alveolate with red algal plastid ancestry, and the dinoflagellate, *Lingulodinium polyedrum*. *Eur. J. Phycol.* **2013**, *48*, 268–277, doi:10.1080/09670262.2013.809610.
51. Jenke-Kodama, H.; Sandmann, A.; Muller, R.; Dittmann, E. Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol.* **2005**, *22*, 2027–2039, doi:10.1093/molbev/msi193.
52. Hashimoto, K.; Yoshizawa, A.C.; Okuda, S.; Kuma, K.; Goto, S.; Kanehisa, M. The repertoire of desaturases and elongases reveals fatty acid variations in 56 eukaryotic genomes. *J. Lipid Res.* **2008**, *49*, 183–191, doi:10.1194/jlr.M700377-JLR200.

53. Gostincar, C.; Turk, M.; Gunde-Cimerman, N. The Evolution of Fatty Acid Desaturases and Cytochrome b5 in Eukaryotes. *J. Membr. Biol.* **2010**, *233*, 63–72, doi:10.1007/s00232-010-9225-x.
54. Ramakrishnan, S.; Docampo, M.D.; MacRae, J.I.; Pujol, F.M.; Brooks, C.F.; van Dooren, G.G.; Hiltunen, J.K.; Kastaniotis, A.J.; McConville, M.J.; Striepen, B. Apicoplast and Endoplasmic Reticulum Cooperate in Fatty Acid Biosynthesis in Apicomplexan Parasite *Toxoplasma gondii*. *J. Biol. Chem.* **2012**, *287*, 4957–4971, doi:10.1074/jbc.M111.310144.
55. James, G.O.; Hocart, C.H.; Hillier, W.; Chen, H.C.; Kordbacheh, F.; Price, G.D.; Djordjevic, M.A. Fatty acid profiling of *Chlamydomonas reinhardtii* under nitrogen deprivation. *Bioresour. Technol.* **2011**, *102*, 3343–3351, doi:10.1016/j.biortech.2010.11.051.
56. Rawsthorne, S. Carbon flux and fatty acid synthesis in plants. *Prog. Lipid Res.* **2002**, *41*, 182–196, doi:10.1016/s0163-7827(01)00023-6.
57. Ramakrishnan, S.; Docampo, M.D.; MacRae, J.I.; Ralton, J.E.; Rupasinghe, T.; McConville, M.J.; Striepen, B. The intracellular parasite *Toxoplasma gondii* depends on the synthesis of long-chain and very long-chain unsaturated fatty acids not supplied by the host cell. *Mol. Microbiol.* **2015**, *97*, 64–76, doi:10.1111/mmi.13010.
58. Füßy, Z.; Faitova, T.; Oborník, M. Subcellular Compartments Interplay for Carbon and Nitrogen Allocation in *Chromera velia* and *Vitrella brassicaformis*. *Genome Biol. Evol.* **2019**, *11*, 1765–1779, doi:10.1093/gbe/evz123.
59. Weng, L.C.; Pasaribu, B.; Lin, I.P.; Tsai, C.H.; Chen, C.S.; Jiang, P.L. Nitrogen Deprivation Induces Lipid Droplet Accumulation and Alters Fatty Acid Metabolism in Symbiotic Dinoflagellates Isolated from *Aiptasia pulchella*. *Sci. Rep.* **2014**, *4*, doi:10.1038/srep05777.
60. Siron, R.; Giusti, G.; Berland, B. Changes in the fatty-acid composition of *Phaeodactylum tricornutum* and *Dunaliella tertiolecta* during growth and under phosphorus deficiency. *Mar. Ecol. Prog. Ser.* **1989**, *55*, 95–100, doi:10.3354/meps055095.
61. Popko, J.; Herrfurth, C.; Feussner, K.; Ischebeck, T.; Iven, T.; Haslam, R.; Hamilton, M.; Sayanova, O.; Napier, J.; Khozin-Goldberg, I.; et al. Metabolome Analysis Reveals Betaine Lipids as Major Source for Triglyceride Formation, and the Accumulation of Sedoheptulose during Nitrogen-Starvation of *Phaeodactylum tricornutum*. *PLoS ONE* **2016**, *11*, doi:10.1371/journal.pone.0164673.
62. Lukeš, M.; Giordano, M.; Prasil, O. The effect of environmental factors on fatty acid composition of *Chromera velia* (Chromeridae). *J. Appl. Phycol.* **2017**, *29*, 1791–1799, doi:10.1007/s10811-017-1114-6.
63. Skiba, M.A.; Sikkema, A.P.; Fiers, W.D.; Gerwick, W.H.; Sherman, D.H.; Aldrich, C.C.; Smith, J.L. Domain Organization and Active Site Architecture of a Polyketide Synthase C-methyltransferase. *ACS Chem. Biol.* **2016**, *11*, 3319–3327, doi:10.1021/acscchembio.6b00759.
64. Huerlimann, R.; Steinig, E.J.; Loxton, H.; Zenger, K.R.; Jerry, D.R.; Heimann, K. The effect of nitrogen limitation on acetyl-CoA carboxylase expression and fatty acid content in *Chromera velia* and *Isochrysis aff. galbana* (TISO). *Gene* **2014**, *543*, 204–211, doi:10.1016/j.gene.2014.04.022.
65. Dubois, D.; Fernandes, F.; Amiar, S.; Dass, S.; Katris, N.J.; Botté, C.Y.; Yamaro-Botté, Y. *Toxoplasma gondii* acetyl-CoA synthetase is involved in fatty acid elongation (of long fatty acid chains) during tachyzoite life stages. *J. Lipid Res.* **2018**, *59*, 994–1004, doi:10.1194/jlr.M082891.
66. Simionato, D.; Block, M.A.; La Rocca, N.; Jouhet, J.; Marechal, E.; Finazzi, G.; Morosinotto, T. The Response of *Nannochloropsis gaditana* to Nitrogen Starvation Includes De Novo Biosynthesis of Triacylglycerols, a Decrease of Chloroplast Galactolipids, and Reorganization of the Photosynthetic Apparatus. *Eukaryot. Cell* **2013**, *12*, 665–676, doi:10.1128/ec.00363-12.
67. Burrows, E.H.; Bennette, N.B.; Carrieri, D.; Dixon, J.L.; Brinker, A.; Frada, M.; Baldassano, S.N.; Falkowski, P.G.; Dismukes, G.C. Dynamics of Lipid Biosynthesis and Redistribution in the Marine Diatom *Phaeodactylum tricornutum* Under Nitrate Deprivation. *Bioenergy Res.* **2012**, *5*, 876–885, doi:10.1007/s12155-012-9201-7.
68. Stephenson, A.L.; Dennis, J.S.; Howe, C.J.; Scott, S.A.; Smith, A.G. Influence of nitrogen-limitation regime on the production by *Chlorella vulgaris* of lipids for biodiesel feedstocks. *Biofuels* **2010**, *1*, 47–58, doi:10.4155/bfs.09.1.
69. Yang, D.W.; Song, D.H.; Kind, T.; Ma, Y.; Hoefkens, J.; Fiehn, O. Lipidomic Analysis of *Chlamydomonas reinhardtii* under Nitrogen and Sulfur Deprivation. *PLoS ONE* **2015**, *10*, doi:10.1371/journal.pone.0137948.
70. Puzanskiy, R.K.; Shavarda, A.L.; Tarakhovskaya, E.R.; Shishova, M.F. Analysis of Metabolic Profile of *Chlamydomonas reinhardtii* Cultivated under Autotrophic Conditions. *Appl. Biochem. Microbiol.* **2015**, *51*, 83–94, doi:10.1134/s0003683815010135.
71. Bisanz, C.; Bastien, O.; Grando, D.; Jouhet, J.; Marechal, E.; Cesbron-Delauw, M.F. *Toxoplasma gondii* acyl-lipid metabolism: De novo synthesis from apicoplast-generated fatty acids versus scavenging of host cell precursors. *Biochem. J.* **2006**, *394*, 197–205, doi:10.1042/bj20050609.

72. Lack, G.; Homberger-Zizzari, E.; Folkers, G.; Scapozza, L.; Perozzo, R. Recombinant expression and biochemical characterization of the unique elongating beta-ketoacyl-acyl carrier protein synthase involved in fatty acid biosynthesis of *Plasmodium falciparum* using natural and artificial substrates. *J. Biol. Chem.* **2006**, *281*, 9538–9546, doi:10.1074/jbc.M509119200.
73. Sharma, S.K.; Kapoor, M.; Ramya, T.N.C.; Kumar, S.; Kumar, G.; Modak, R.; Sharma, S.; Surolia, N.; Surolia, A. Identification, characterization, and inhibition of *Plasmodium falciparum* beta-hydroxyacyl-acyl carrier protein dehydratase (FabZ). *J. Biol. Chem.* **2003**, *278*, 45661–45671, doi:10.1074/jbc.M304283200.
74. Surolia, N.; Surolia, A. Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum* (vol 7, pg 167, 2000). *Nat. Med.* **2001**, *7*, 636.
75. Waller, R.F.; Keeling, P.J.; Donald, R.G.K.; Striepen, B.; Handman, E.; Lang-Unnasch, N.; Cowman, A.F.; Besra, G.S.; Roos, D.S.; McFadden, G.I. Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12352–12357, doi:10.1073/pnas.95.21.12352.
76. Botte, C.Y.; Yamaryo-Botte, Y.; Rupasinghe, T.W.T.; Mullin, K.A.; MacRae, J.I.; Spurck, T.P.; Kalanon, M.; Shears, M.J.; Coppel, R.L.; Crellin, P.K.; et al. Atypical lipid composition in the purified relict plastid (apicoplast) of malaria parasites. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 7506–7511, doi:10.1073/pnas.1301251110.
77. Goodman, C.D.; McFadden, G.I. Fatty acid synthesis in protozoan parasites: Unusual pathways and novel drug targets. *Curr. Pharm. Des.* **2008**, *14*, 901–916, doi:10.2174/138161208784041088.
78. Liu, B.Q.; Wang, Y.Q.; Fillgrove, K.L.; Anderson, V.E. Triclosan inhibits enoyl-reductase of type I fatty acid synthase in vitro and is cytotoxic to MCF-7 and SKBr-3 breast cancer cells. *Cancer Chemother. Pharmacol.* **2002**, *49*, 187–193, doi:10.1007/s00280-001-0399-x.
79. Kruger, N.J.; von Schaewen, A. The oxidative pentose phosphate pathway: Structure and organisation. *Curr. Opin. Plant Biol.* **2003**, *6*, 236–246, doi:10.1016/s1369-5266(03)00039-6.
80. Bigogno, C.; Khozin-Goldberg, I.; Cohen, Z. Accumulation of arachidonic acid-rich triacylglycerols in the microalga *Parietochloris incisa* (Trebuxiophyceae, Chlorophyta). *Phytochemistry* **2002**, *60*, 135–143, doi:10.1016/s0031-9422(02)00037-7.
81. Guschina, I.A.; Harwood, J.L. Mechanisms of temperature adaptation in poikilotherms. *FEBS Lett.* **2006**, *580*, 5477–5483, doi:10.1016/j.febslet.2006.06.066.
82. Khozin-Goldberg, I.; Didi-Cohen, S.; Shayakhmetova, I.; Cohen, Z. Biosynthesis of eicosapentaenoic acid (EPA) in the freshwater euglenoid *Monodus subterraneus* (Euglenophyceae). *J. Phycol.* **2002**, *38*, 745–756, doi:10.1046/j.1529-8817.2002.02006.x.
83. Mitschler, R.R.; Welti, R.; Upton, S.J. A comparative-study of lipid compositions of *Cryptosporidium parvum* (Apicomplexa) and madin-darby bovine kidney-cells. *J. Eukaryot. Microbiol.* **1994**, *41*, 8–12, doi:10.1111/j.1550-7408.1994.tb05927.x.
84. Toso, M.A.; Omoto, C.K. *Gregarina niphandrodes* may lack both a plastid genome and organelle. *J. Eukaryot. Microbiol.* **2007**, *54*, 66–72, doi:10.1111/j.1550-7408.2006.00229.x.
85. Zhu, G.; Marchewka, M.J.; Keithly, J.S. *Cryptosporidium parvum* appears to lack a plastid genome. *Microbiology* **2000**, *146*, 315–321, doi:10.1099/00221287-146-2-315.
86. Keithly, J.S.; Langreth, S.G.; Buttle, K.F.; Mannella, C.A. Electron tomographic and ultrastructural analysis of the *Cryptosporidium parvum* relict mitochondrion, its associated membranes, and Organelles. *J. Eukaryot. Microbiol.* **2005**, *52*, 132–140, doi:10.1111/j.1550-7408.2005.04-3317.x.
87. Ševčíková, T.; Horák, A.; Klimeš, V.; Zbránková, V.; Demir-Hilton, E.; Sudek, S.; Jenkins, J.; Schmutz, J.; Příbyl, P.; Fousek, J.; et al. Updating algal evolutionary relationships through plastid genome sequencing: Did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci. Rep.* **2015**, *5*, doi:10.1038/srep10134.
88. Sobotka, R.; Esson, H.J.; Koník, P.; Trsková, E.; Moravcová, L.; Horák, A.; Dufková, P.; Oborník, M. Extensive gain and loss of photosystem I subunits in chromerid algae, photosynthetic relatives of apicomplexans. *Sci. Rep.* **2017**, *7*, doi:10.1038/s41598-017-13575-x.



CURRICULUM VITAE

Mgr. Jan Michálek

July 4, 1984, Plzeň, Czech Republic

Address: Nové Hrady 448, 37005; phone: +420 721 589 270; e-mail:

gonzo.dr@seznam.cz, jan.michalek@entu.cas.cz

EDUCATION

2006-2009 University of South Bohemia in České Budějovice, Faculty of Biology
BSc. program: General Biology

BSc. thesis: Diversity analysis of microfungi in phylloplane of pedunculate oak (*Quercus robur*) infected by oak powdery mildew (*Erysiphe alphitoides*) classical versus molecular methods (review).
Supervisor: Ing. Miloslava Kavková PhD.

2009-2013 University of South Bohemia in České Budějovice, Faculty of Science
MSc. program: Experimental biology, Genetics and Genetic Engineering

MSc. thesis: *Erysiphe alphitoides* and a community of micromycetes in phylloplane of pedunculate oaks assessed using DGGE.
Supervisor: Ing. Miloslava Kavková PhD.

2013-pres. University of South Bohemia in České Budějovice, Faculty of Science
Ph.D. program: Molecular and Cell Biology and Genetics

Ph.D. thesis: Genomes of Chromerid Algae.
Supervisor: Prof. Ing. Miroslav Oborník, PhD.

EMPLOYMENT

2012-2016 Graduate student

Laboratory of Evolutionary Protistology, Institute of Parasitology, Biology Centre CAS, v.v.i.

2016-pres. Research assistant

Laboratory of Evolutionary Ecology, Institute of Entomology, Biology Centre CAS, v.v.i.

CONFERENCE CONTRIBUTIONS

44th Jírovec's Protozoological Days, Týn nad Vltavou, Czech Republic (2013)

Do Not Curse the Contamination – An Unexpected Discovery of the Novel Marine Fungus (poster)

44th Jírovec's Protozoological Days, Krásná, Czech Republic (2014) The Fat

Cousin of Apicomplexans: From Genes to Fats. (oral presentation)

- 2nd Chromera Meeting**, České Budějovice, Czech Republic (2014) The Fatty Acid Synthesis in Chromerids. (oral presentation)
- 25th Annual Molecular Parasitology Meeting**, Woodshole MA, USA (2014) Synthesis of Fatty Acids in Chromerid Algae – The Closest Autotrophic Relatives of Apicomplexans. (poster)
- 45th Jirovec's Protozoological Days**, Pelhřimov (2015) The Weird Respiratory Chain of Chromerid *Chromera velia*. (oral presentation)
- The annual meeting of the Society for Molecular Biology and Evolution**, Vienna, Austria (2015) The broken respiratory chain in *Chromera velia* - a phototrophic relative of apicomplexans. (poster)

PUBLICATIONS

- Woo YH, Ansari H, Otto TD, Klinger CM, Kolisko M, Michálek J, Saxena A, Shanmugam D, Tayyrov A, Veluchamy A, Ali S, Bernal A, del Campo J, Cihlář J, Flegontov P, Gornik SG, Hajdušková E, Horák A, Janouškovec J, Katris NJ, Mast FD, Miranda-Saavedra D, Mourier T, Naeem R, Nair M, Panigrahi AK, Rawlings ND, Padron-Regalado E, Ramaprasad A, Samad N, Tomčala A, Wilkes J, Neafsey DE, Doerig C, Bowler C, Keeling PJ, Roos DS, Dacks JB, Templeton TJ, Waller RF, Lukeš J, Oborník M, Pain A. 2015. **Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites**. *Elife* 4:e06974.7
- Flegontov P, Michálek J, Janouškovec J, Lai DH, Jirků M, Hajdušková E, Tomčala A, Otto TD, Keeling PJ, Pain A, Oborník M, Lukeš J. 2015. **Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites**. *Molecular Biology and Evolution* 32(5):1115-31. doi: 10.1093/molbev/msv021.
- Souto-Vilarós D, Macháč A, Michálek J, Darwell C, Sisol M, Kuyaiva T, Isua B, Weiblen G, Novotný V, Segar ST. 2019. **Faster speciation of fig-wasps than their host figs leads to decoupled speciation dynamics: Snapshots across the speciation continuum**. *Molecular Ecology* 28 : 3958-3976. DOI: 10.1111/mec.15190
- Souto-Vilarós D, Proffit M, Buatois B, Rindoš M, Sisol M, Kuyaiva T, Isua B, Michálek J, Darwell C, Hossaert-McKey M, Weiblen G, Novotný V, Segar ST. 2018. **Pollination along an elevational gradient mediated both by floral scent and pollinator compatibility in the fig and fig-wasp mutualism**. *Journal of Ecology* 106 : 2256-2273. DOI: 10.1111/1365-2745.12995
- Souto-Vilarós D, Houadria MYI, Michálek J, Sisol M, Isua B, Kuyaiva T, Weiblen G, Novotný V, Segar ST. 2020. **Contrasting patterns of fig wasp communities along Mt. Wilhelm, Papua New Guinea**. *Biotropica* 52 : 323-334. DOI: 10.1111/btp.12763
- Tomčala A, Michálek J, Schneedorferová I, Füssy Z, Gruber A, Vancová M, Oborník M 2020. **Fatty Acid Biosynthesis in Chromerids**. *Biomolecules* 10: 1102.