Helicosporidia: a genomic snapshot of an early transition to parasitism

Yukun Sun, Jean-François Pombert*
Department of Biological and Chemical Sciences, Illinois Institute of Technology, Chicago, IL 60616, USA

Abstract

Helicosporidia are gut parasites of invertebrates. These achlorophyllous, non-photosynthetic green algae are the first reported to infect insects. Helicosporidia are members of the green algal class Trebouxiophyceae and are further related to the photosynthetic and non-photosynthetic genera Auxenochlorella and Prototheca, respectively, the latter of which can also turn to parasitism under opportunistic conditions. Molecular analyses suggest that Helicosporidia diverged from other photosynthetic trebouxiophytes less than 200 million years ago and that its adaptation to parasitism is therefore recent. In this minireview, we summarize the current knowledge of helicosporidian genomics. Unlike many well-known parasitic lineages, the Helicosporidium sp. organelle and nuclear genomes have lost surprisingly little in terms of coding content aside from photosynthesis-related genes. While the small size of its nuclear genome compared to other sequenced trebouxiophycean representatives suggests that Helicosporidium is going through a streamlining process, this scenario cannot be ascertained at this stage. Genome expansions and contractions have occurred independently multiple times in the green algae, and the small size of the Helicosporidium genome may reflect a lack of expansion from a lean ancestor state rather than a tendency towards reduction.

Keywords: Helicosporidium; green algae; comparative genomics; entomopathogen

Introduction

Parasitism comes in many guises, transient or obligate, extracellular or intracellular. Transient parasites are free-living or symbiotic organisms that are turning to parasitism upon a set of favorable opportunistic conditions. These organisms do not rely exclusively on parasitism for their survival nor are they always well-equipped to the task. Obligate parasites, on the other hand, cannot thrive without hosts and, by necessity, came up over the course of their evolution with the strategies and the tools to facilitate infection [1]. While the modes of action (extracellular or intracellular) and levels of specialization differ between various parasites, the goal remains the same: to benefit at the expense of the host.

The road to parasitism is littered with acquisitions and losses [2–6]. Intuitively, functions and strategies that can help avoid, disrupt or even disable the hosts’ defenses are valuable targets for acquisition by parasites, even more so when their survival depends on it. What is not developed intrinsically over time can be acquired sometimes quickly by horizontal transfer of foreign genetic material. Parasites can gain new functions by capturing useful genes from their hosts (e.g. [7–10]), from co-pathogens or co-symbionts (e.g. [11,12]), from other species during transit (e.g. [13]), or from other various sources (e.g. [14–17]). Conversely, functions that are no longer necessary are often discarded. This is especially true for intracellular parasites, whose switch from external to internal environments shifted many functions from necessary to accessory. Given their internalization, intracellular parasites are constantly pressured to adapt to the responses of their host(s), which in turn may lead to accelerated rates of evolution compared to their free-living relatives [18–20].

Due to their reduced and/or highly adapted nature, the physiological features of parasites are often cryptic, which hinders taxonomical classification. Two well-known examples are the malaria-causing agent Plasmodium and the once enigmatic Microsporidia, which have undergone several rounds of taxonomical revisions [21,22]. Helicosporidia, the subject of this minireview, is another protist lineage whose position within the tree of life has long remained elusive. First characterized in 1921 [23], Helicosporidians were in consequence over time classified as sporozoan, myxosporean, fungal and protozoan parasites before their position within the green algae was first hinted at nearly 80 years later based on morphological similarities with Prototheca wickerhamii, a non-photosynthetic trebouxiophyte [24]. This affiliation was soon corroborated by molecular phylogenetic inferences based on nuclear-encoded actin and rRNA genes [25] and further confirmed with additional plastid [26,27], mitochondrial [28] and nuclear data [29–31].

Helicosporidia are the first reported green algae that infect insects [23,24]. These entomopathogens are found...
in the outside environment as cysts surrounded by a thin pellicle and composed of three ovoid cells wrapped within a helicoidal filamentous one [24]. The relatively small helicosporidian cysts (sometimes referred as spores in the older literature), up to 6 µm wide, are ingested per os and dehisc in the host’s gut at the start of the invasive stage to liberate the four cells within [32]. The long filamentous cell is then uncoiled and attaches itself to the host epithelium with the help of the barbed protuberances located at its extremities. Upon successful breach of the epithelial barrier, the ovoid cells invade the hemolymph wherein they will replicate during the vegetative stage [24]. Despite their mode of replication within insect hosts, helicosporidian vegetative cells can also be cultured in vitro on limited nutrients media suggesting that, unlike many parasites, they have retained the metabolic pathways required for saprobitic growth. For more information on the morphology and lifecycle of Helicosporidia, we refer to the readers to the educated description by Boucias et al. [24] and the thorough review by Tartar [32]. In this minireview, we will focus on the genomic features of Helicosporidia.

Helicosporidians are of particular interest for our understanding of the genetic changes that occur during the gradual transition to parasitism. This lineage emerged recently and is part of a green algal lineage that encompasses free-living, symbiotic and parasitic organisms [31]. Helicosporidia are members of the trebouxiiophycean order Chlorellales, whose best-known genera include Nannochloris, the type species Chlorella and the achlorophyllous Prototheca [33]. The latter has been reported to sometimes infect humans and other mammals under opportunistic conditions [34]. The Chlorellales order originated circa 350 million years ago (mya), within which the Prototheca and Chlorella genera diverged somewhere between 200 to 350 mya [35]. Helicosporidians are closely related to the genus Prototheca, but as can be seen in Fig. 1, their surprisingly fast rate of evolution renders molecular clock inferences based on 18S rDNA somewhat unreliable. Nevertheless, Helicosporidia are bound to have arisen after the late Paleozoic/early Mesozoic era, likely within the last 200 million years or so, making them one of the youngest parasitic lineages known. In comparison, the well-known parasitic lineages from Giardia, Microsporidia and Plasmodium arose 2.2, 1.2 and 1 bya, respectively [36,37]. The recent transition from free-living to insect pathogen in Helicosporidia thus offers the opportunity to shed some light onto the early stages of parasitism in this lineage.

The Helicosporidium organelle genomes are present and conserved

Organelles are often the target of major modifications in parasites. Severe alterations to organelles can force organisms towards alternate lifestyles while changes in lifecycles can impact organelles to the point that these can become barely recognizable over time. Following internalization, the chloroplast is often victim of severe atrophy, for the capability to harness sunlight and convert it to energy is limited, if not absent within the confines of the host. Keeping a functional photosynthetic apparatus in this environment is rarely beneficial if not deleterious. In the apicomplexan Plasmodium, the relict organelle corresponding to the plastid is so derived that it took decades before it was finally identified as a remnant of a photosynthetic organelle [38–40]. The functions encoded within have been greatly diminished and the apicoplast genome has retained a limited set of genes allowing for its replication and maintenance [41,42]. It is unclear, however, if the Plasmodium plastid was already reduced prior to its adaptation to parasitism or if its reduction is a consequence of its pathogenic lifestyle. The blood vessels carrying the erythrocytes that these parasites infect are iron-rich environments favorable to heme synthesis but are almost systematically shrouded in darkness, such that a photosynthetic Plasmodium species would have little reason to exist.

Not surprisingly, the entomopathogenic helicosporidians are also non-photosynthetic. Losing photosynthetic capability however, by itself, does not define a parasite. Many free-living lineages never had this ability and the loss of atmospheric CO₂ fixation capability occurred independently and recurrently across many photosynthetic lineages, reverting the organisms to heterotrophy. The closest known relatives of Helicosporidia from the Prototheca genus (Fig. 1) are also non-photosynthetic and, while sometimes parasitic, they can do so only under opportunistic conditions [43]. The Helicosporidium sp. plastid genome has been streamlined to a greater extent than that of Prototheca wickerhamii (Junbiao Dai, personal communication) but not as much as that of the apicoplast from Plasmodium falciparum [27,38]. As one would expect, the Helicosporidium plastid gene losses mainly affect those associated with the fixation of atmospheric carbon dioxide, i.e. none of the genes coding for products involved in it has been retained, but it also lacks all genes coding for ATP synthases that are present in Prototheca ([27,44]; Fig. 2) and four ribosomal proteins (Rpl23, Rps2, Rps9 and Rps18) that are found in other green algae. The Helicosporidium plastid genome is compact and arrayed with a single replication origin (inferred by GC-skew analyses [27]), suggesting an optimization towards efficiency and a reduction in the amount of energy spent per replication, but it is unknown if its streamlining is still actively ongoing or if it has reached a certain equilibrium.

Modifications to the mitochondrion can also have a drastic, much more potent impact on parasites. As the principal energy factory of the cell, any evolutionary trend that leads to the loss of ATP production will severely affect the parasite and increase its host dependency. In many parasites, the mitochondrion has been severely overhauled, losing its genome in the process. Those highly reduced organelles, referred to as mitosomes or hydrogenosomes, are involved in only a few of the original mitochondrial pathways [45,46]. Mitosomes are involved principally in iron–sulfur cluster assembly [47] and are incapable of oxygenic respiration, such that mitosome-bearing parasites like Microsporidia or Giardia have to rely either on the alternate glycolysis and pentose phosphate pathways for ATP production or on the energy that they hitchhike from their hosts to power all of their biological functions [45,48]. Hydrogenosomes, like those found in Trichomonas, are similarly reduced but are further distinguished by their ability to generate molecular hydrogen [49].

© The Author(s) 2014 Published by Polish Botanical Society Acta Soc Bot Pol 83(4):377–385 378
Fig. 1  Phylogenetic niche of Helicosporidium within the Trebouxiophyceae. The best 18S rDNA maximum likelihood tree shown here was computed with PhyML 3.1 [62] under the general time reversible model of nucleotide substitution, as implemented in SeaView 4 [63]. The tree dataset is derived from de Wever et al. [35], to which Helicosporidium and Prototheca sequences were added. Analyses were run with and without sequences from the Ulvophyceae and Chlorophyceae (not shown), with Prasinophyceae (in purple) and the charophycean algae Mesostigma and Chaetosphaeridium (in gray) as outgroups. The Trebouxiophyceae are shown in shades of green. The grouping of Helicosporidium with Prototheca was recovered in both analyses. For brevity, the tree shown here does not include the Ulvophyceae and Chlorophyceae. Bootstrap values for the Prototheca/Helicosporidium clades above 50% are indicated over the corresponding nodes; asterisks indicate nodes recovered in all bootstrap replicates. Branch lengths are drawn to scale.
This is not the case for Helicosporidia. Helicosporidians can use oxygen as a terminal electron acceptor, as indicated by their successful growth in oxygenic conditions on nutrients-limited petri dishes [24], and have retained a mitochondrion with a full genome similar to other free-living, early-diverging green algae [28]. In fact, if we make abstraction of the Chlorophyceae, a later-diverging lineage with mid- to heavily-modified mitochondrial genomes [50], one could describe the circular, 49.3 kbp-long Helicosporidium mitochondrial chromosome as a typical green algal mitogenome. Perhaps the most salient feature in the otherwise somewhat unremarkable Helicosporidium mitochondrial genome is the presence of a trans-spliced group I intron. Trans-spliced group I introns are rare (e.g. [51] and references within) but have little to do with parasitism.

The Helicosporidium nuclear genome is compact

We often think about parasite genomes in terms of reduction, for the evolutionary pressures applied on them to stay small is intuitive, whether to facilitate replication or minimize energetic expenses. A caveat of that way of thinking is that we sometimes make involuntary shortcuts when comparing with free-living organisms. Quite often, when the genome of the parasite is smaller, we say that it has been reduced. This is true for many instances, but is it always the case? In the green algae, judging by size alone is not enough. Most genome sizes were estimated from C-values (Fig. 3), which are not always accurate. For example, the Ostreococcus genome was thus once estimated at around 100 Mbp [52], or one order of magnitude larger than its real size [53]! Even if the sizes are accurate, many free-living algae harbor genomes that are bloated with repeated elements (e.g. [54,55]) that can propagate rapidly upon the right conditions and which have little effect on the overall metabolic profile of the species. Comparisons based on the total number of genes are therefore preferable but not necessarily accurate either. Automated gene prediction algorithms can struggle with their proper detection, sometimes grossly under- or overestimating their presence by the thousands [55,56], and we lack the manpower to manually curate all of the incoming onslaught of genomic data. Furthermore, a greater gene count is not necessarily synonymous with a greater metabolic potential, as duplications have occurred many times throughout evolution. Without looking at the genes themselves, inferences about parasitic reduction, free-living expansion or a mixture of both processes cannot be made without prior knowledge of the state that was present in the common ancestor.

By glancing quickly at sizes and gene numbers alone (Fig. 4), the Helicosporidium nuclear genome (12.4 Mbp assembled; 17 Mbp estimated [31]) appears to have shrunken down to almost a third of that of the free-living Coccomyxa (49 Mbp; [57]) and the symbiotic Chlorella (46 Mbp; [58]) and encodes a gene set that is about 40% smaller than that of its relatives. However, when looking at their respective metabolic profiles (Fig. 4), the picture that emerges is quite different. The core green algal metabolic pathways are all present in Helicosporidium and the only clear reduction has been the loss of its photosynthetic ability (Fig. 4). But even here the reduction is not complete and Helicosporidium has retained all of the genes involved in the storage of sugars as starchy polymers. The lower gene count in Helicosporidium can be attributed at least in part to the presence of fewer paralogous gene copies and to the absence of non-essential alternate metabolic branches synthesizing compounds that are either accessory or produced via other pathways. Other factors that contribute to the small size of the Helicosporidium genome are the limited number of introns, about on par...
Fig. 3  Estimated green algal genome sizes in the Streptophyta and Chlorophyta. Genomes that were sequenced are lettered in black, with *Helicosporidium* lettered in red. Genomes whose sizes were estimated based on their 1C-values are lettered in gray. Estimated 1C-values for algae from the Streptophyta (Charophyceae) were inferred from Kapraun ([52]; only the smallest representatives were selected) whereas others from the Chlorophyta (Prasinophyceae, Trebouxiophyceae, Chlorophyceae, Ulvophyceae) were retrieved from the RBG Kew Algae DNA C-values database (http://data.kew.org/cvalues). Dark and light shades in the histogram bars correspond to lower and upper 1C-value size estimates. The histogram is color-coded according to the underlying cladogram depicting the branching order between these lineages.
Fig. 4  Distribution of KEGG metabolic pathways among sequenced green algal nuclear genomes. a The concentric rings, from the outside to the inside, are: *Chlamydomonas reinhardtii*, *Volvox carteri f. nagariensis*, *Helicosporidium* sp., *Chlorella variabilis* NC64A, *Coccomyxa subellipsoidea* C-169, *Micromonas* sp. RCC299, *Micromonas pusilla*, *Ostreococcus tauri*, *Ostreococcus lucimarinus*. Each bar in the concentric rings corresponds to a unique protein in the overarching pathways. KEGG orthology identifiers (KO) indicated at the start and end of each arc correspond to the order in Tab. S1. Color gradients indicate the respective number of each protein-coding gene in the corresponding organisms. The gradient is a heat map standardized per pathway for visual balance. The absence of a protein-coding gene is indicated by a white color, whereas the pink to purple gradient (pink, red, brown, gold, green, cyan, blue, purple) indicate the relative number of gene copies per pathway. The KEGG metabolic pathways are: Ami: amino acids; Car: carbohydrates; Cof: cofactors and vitamins; Ene: energy; Fol: folding, sorting and degradation; Gly: glycans; Lip: lipids; Nuc: nucleotides; Rep: replication and repair; Sec: secondary metabolites; Sig: signal transduction; Ter: terpenoids and polyketides; Tsc: transcription; Tsl; translation; Tsp; transport. The cladogram in the center schematizes the relationships between the prasinophytes (in purple), trebouxiophytes (in green) and chlorophycean algae (in orange). Assembled genome sizes (in Mbp) are indicated above the corresponding branches; values for picoprasinophytes are from *Ostreococcus tauri* and *Micromonas* sp. RCC299. b Zoom-in of the major *Helicosporidium* losses in the KEGG energy-related pathways. The corresponding KEGG orthology identifiers (KO) indicated under the corresponding blocks are further described in Tab. S1.
with picoprasinophytes and roughly three times less than in the other Trebouxiophyceae and Chlorophyceae, and the dearth of transposable elements [31]. Helicosporidium lacks the Dicer and Argonaute proteins hypothesized to suppress the propagation of transposable elements (TEs) [59], and it is unclear if the RNA interference machinery was lost because its function was accessory.

So, did the Helicosporidium genome shrink over time, was it the other trebouxiophycean genomes that have expanded or perhaps a bit of both? Almost concurrently with the publication of the Helicosporidium genome, Gao and co-authors reported the genome of another free-living Trebouxiophyceae, Chlorella protothecoides, assembled at 22.9 Mbp with a maximum estimated size of 27.6 Mbp [60]. The C. protothecoides genome is also quite compact and displays features that are intermediates between the Helicosporidium and C. variabilis/Coccomyxa genomes. In particular, like Helicosporidium, multi-copy genes are fewer in C. protothecoides than in the other trebouxiophycean genomes, suggesting that expansion by duplication may have occurred in this lineage [60]. The variation in size observed in the Trebouxiophyceae is not limited to this lineage and all green algal classes feature a mixture of species with small, medium or large genomes (Fig. 3). The problem is that we actually don’t know what the genomic state of the green algal ancestor was, but here we infer that it was likely small. The genomes of prasinophytes occupying the most basal branches of the Chlorophyta phylum tend to be on the smaller scale and, while charophycean genomes in the Streptophyta have been somewhat stigmatized as humongous (many of the early ones investigated were indeed quite large [61]), species with genomes in the 100 Mbp range also exists [52] and, again, tend to branch at the base of the clade. We cannot rule out that the ancestral genome of the green algae may have been bloated and then reduced repeatedly and independently over time, but the opposite scenario is more parsimonious and thus, we find it more likely. Considering the above, we think that the Helicosporidium genome did not really experience much reduction, aside from a few clear photosynthetic and other gene losses, and that it has retained a genome whose features are probably closer to the ancestral ones than that of Coccomyxa and C. variabilis.

Our picture of the evolution of Heliosporidia is far from complete however. The Helicosporidium genome reported corresponds to the one with the shortest branch in our 18S rDNA tree (Fig. 1) and may thus represent one of the less derived helicosporidians. Looking at the other species may reveal greater levels of specialization and/or host dependency and help identify the components that are important for their infection strategies. Helicosporidium sp. was found to harbor numerous chitinases that can help breach its invertebrate host gut epithelium but also featured over 850 unique predicted proteins that were found to be expressed and for which no homology was found [31]. It is unclear if these transcripts are functional or the result of transcriptional noise, but their unicity in some cases may reflect novel or derived-beyond-recognition components potentially involved in pathogenicity. In that regard, sequencing the genomes of the closest known relatives of Helicosporidia from the genus Protothea would most likely provide interesting insights, considering that these related species can also parasitize various hosts upon favorable conditions, potentially with the help of shared genetic components.

Acknowledgments
This work was supported by funds from the Illinois Institute of Technology.

Authors’ contributions
The following declarations about authors’ contributions to the research have been made: performed the phylogenetic analysis: YS; wrote the mini-review: YS, JFP.

Competing interests
No competing interests have been declared.

Supplementary material
The following supplementary material for this article is available online at http://pbsociety.org.pl/journals/index.php/asbp/rt/suppFiles/asbp.2014.039/0:
1. Tab S1: distribution of KEGG metabolic pathways among sequenced green algal genomes.

References
12. Weigl LM, Clewell DB, Gill SR, Clark NC, McDougal LK, Flannigan
Pombert JF, Blouin NA, Lane C, Boucias D, Keeling PJ. A lack of


© The Author(s) 2014 Published by Polish Botanical Society Acta Soc Bot Pol 83(4):377–385

384


