Endosymbiotic and horizontal gene transfer in microbial eukaryotes
Impacts on cell evolution and the tree of life

Cheong Xin Chan,1 Debashish Bhattacharya2,* and Adrian Reyes-Prieto3
1The University of Queensland; Institute for Molecular Bioscience and ARC Centre of Excellence in Bioinformatics; Brisbane, Queensland Australia; 2Department of Ecology and Evolution and Institute for Marine and Coastal Sciences; Rutgers University; New Brunswick, NJ USA; 3Canadian Institute for Advanced Research and Department of Biology; University of New Brunswick; Fredericton, New Brunswick Canada

Keywords: horizontal gene transfer, endosymbiotic gene transfer, diatoms, membrane transporters, eukaryote evolution, tree of life
Submitted: 02/09/12
Accepted: 03/22/12
http://dx.doi.org/10.4161/mge.20110
*Correspondence to: Debashish Bhattacharya; Email: bhattacharya@aesop.rutgers.edu

The evolution of microbial eukaryotes, in particular of photosynthetic lineages, is complicated by multiple instances of endosymbiotic and horizontal gene transfer (E/HGT) resulting from plastid origin(s). Our recent analysis of diatom membrane transporters provides evidence of red and/or green algal origins of 172 of the genes encoding these proteins (ca. 25% of the examined phylogenies), with the majority putatively derived from green algae. These data suggest that E/HGT has been an important driver of evolutionary innovation among diatoms (and likely other stramenopiles), and lend further support to the hypothesis of an ancient, cryptic green algal endosymbiosis in "chromalveolate" lineages. Here, we discuss the implications of our findings on the understanding of eukaryote evolution and inference of the tree of life.

Singl-cell photosynthetic eukaryotes are significant contributors to total biomass production in oceanic and freshwater environments.1,2 The origin of photoautotrophic metabolism in eukaryotes is explained by a primary endosymbiotic event with captured cyanobacteria that evolved into the plastids in the common ancestor of the Plantae (i.e., red algae, green algae and plants, and glaucophytes3-5). The plastids have subsequently spread into a multitude of other eukaryote lineages such as euglenids, chlorarachniophytes, and the diverse chlorophyll c-containing algae (often referred to as "chromalveolates") through secondary and tertiary eukaryote-eukaryote endosymbiosis.6,7 The complex evolutionary history of algae with secondary and tertiary plastids also includes enigmatic non-photosynthetic related groups that presumably have lost photosynthetic capacity several times independently and, in other cases, replaced their original plastid via a subsequent endosymbiosis.8 Analysis of the recently sequenced genome of the glauco-phyte Cyanophora paradoxa substantiates Plantae monophyly and the single origin of primary plastids.4 Importantly, the evidence for this long-sought after result comes not from multigene trees but instead from the more convincing analysis of groups of genes that are involved in complex processes such as fermentation, plastid solute transport, and plastid protein transloca- tion.4 In contrast, even with the availability of genome data from diverse "chromalveolates," the phylogenetic relationships between these groups (e.g., the ubiquitous diatoms, haptophytes, and cryptophytes) remain murky and multigene trees again do not provide unambiguous evidence for their union as a supergroup. These types of analyses are either too difficult to interpret or suggest more complex evolutionary scenarios to explain "chromalveolate" interrelationships.9-11 The persistent problem of resolving the phylogenetic relationships of "chromalveolates" is likely not due to the failure of phylogenetic approaches in general to resolve evolutionary history (although confounding issues such as long-branch attraction abound). Rather, these taxa share a feature that may grossly mislead phylogenetic inference: extensive horizontal gene transfer (HGT). Early genome-wide studies of "chromalveolates"12 and chlorarachniophytes13 revealed a substantial contribution...
of HGT to their nuclear genomes and, importantly, hundreds of transferred genes that were recruiting via endosymbiotic gene transfer (EGT). In “chromalveolates,” EGT is a specific case of HGT resulting from secondary and tertiary algal endosymbiosis. The potential phylogenetic confusion spawned by E/HGT is magnified with each round of secondary endosymbiosis that has occurred, with each event contributing novel sequences to the host genome that add to or replace existing genes.10,12

The extent of EGT can be relatively easily discerned in the case of primary endosymbiosis (i.e., Plantae lineages), whereby the transferred genes are of prokaryotic provenance (i.e., cyanobacterial and α-proteobacterial origin, respectively, for the plastid and mitochondrion) and differ greatly in sequence from the resident eukaryotic nuclear genes. Therefore, the finding of hundreds of nuclear genes in algae and other eukaryotes with a prokaryotic phylogenetic signature is considered unambiguous evidence for EGT associated with the primary endosymbiotic origin of organelles.1,6,13-15 This level of resolution may not occur, however, in the case of serial, eukaryote-eukaryote endosymbioses. In these instances, the closer the phylogenetic affinity between the donor (the endosymbiont) and recipient (the host) lineages the less likely that standard phylogenetic methods can distinguish bona fide EGT events from diverged host genes. This expectation of highly reticulate relationships driven by serial eukaryotic endosymbioses involving closely related lineages has been used to explain the great difficulties in establishing a robust “chromalveolate” tree of life using multigene phylogenetic inference.6,11

Demonstrable evidence for the existence of serial eukaryotic endosymbioses was recently provided with the analysis of complete genome data from diatoms. Moustafa et al.18 showed the presence of hundreds of genes of green algal origin in Thalassiosira pseudonana and Phaeodactylum tricornutum, two species that possess a plastid of unambiguous red algal origin. The “green genes” reside alongside genes of red algal origin in these diatom nuclear genomes and are postulated to derive from an ancient, cryptic green algal secondary endosymbiosis (with associated EGT) that predated the capture of the red algal plastid. This initial “controversial” finding has now been supported with analysis of complete genome data from other stramenopiles (e.g., Ectocarpus siliculosus19). Given the growing evidence suggesting multiple algal endosymbiosis in some lineages, the key questions become: what specific functions do these green genes have and with a putative residence time of hundreds of millions of years, how have they contributed to the large-scale evolution of the host lineage?

Evolution of the Diatom Permeome and the “Green-Algal” Footprints

We recently completed a phylogenomic analysis of genes encoding membrane transporters in the nuclear genomes of diatoms T. pseudonana and P. tricornutum.20 In addition to examples of long-term vertical inheritance of membrane transporters in these and other eukaryotes (Fig. 1A), we also found putative red and green algal origins (Fig. 1B) of many of these genes. Specifically, 172 (ca. 25%) of the 697 encoded membrane transporters were found to be of Plantae origin. Although 50 of these show an unresolved affiliation with red and/or green algae, interestingly, the majority (103) are of putative green algal provenance compared with 19 of putative red algal origin. To explain their long-term retention, we hypothesized that these genes, presumably derived from different algal secondary endosymbionts, contributed important membrane transport functions to diatoms.20

The fluctuations of redox-sensitive transition metals over evolutionary timescales driven most notably by the rise of oxygen ca. 2.5 BYA, would have introduced immense selective pressure for cell survival in aquatic environments. Such environmental pressure is particularly relevant to the trafficking of molecules across cellular membrane, whether to acquire useful molecular compounds or to expel molecules that cause harm to the cell (i.e., cell detoxification). Free-living unicellular algae are known to possess efficient transport systems, particularly of metal cations such as sodium and potassium, to maintain cell homeostasis with respect to the inter- and extra-cellular environments.21 Therefore, the acquisition of red and green algal genes via E/HGT could have been a crucial selective force for the survival of ancient, and hence extant diatom lineages. Our observations are also in general agreement with the hypothesis of a cryptic green algal endosymbiosis in diatoms,18 and likely more generally in “chromalveolates.”22,23

And finally, these data20 suggest that the cryptic endosymbiont contributes gene functions that extend beyond photosynthetic capacity to roles related to cell adaptation to the external environment.

Implications for Inferring the Tree of Life

A well-sampled phylogenetic tree of eukaryotes (i.e., the eukaryote tree of life) is essential to understand the tempo and mode of endosymbiotic events that have occurred during algal evolution, and to test different hypotheses about the relationships between the major algal groups. The recent findings of extensive EGT from putative serial endosymbioses,4,18,20,23 in combination with HGT events from diverse sources,24 stresses the significant impact that gene transfer has had on algal and eukaryote evolution in general. Gene transfer events are not only of significance to cell evolution, but may also mislead inference of the tree of life. Some of our previous studies3,25 were aimed at identifying genes that do not “follow” the presumed vertical (i.e., host lineage) history of algal groups and their non-photosynthetic relatives (e.g., Plantae, Alveolata and stramenopiles) to study the endosymbiotic origin of plastids, and, importantly, the impact of E/HGT on algal genome evolution. It has now become clear that the identification of bona fide cases of EGT or HGT is often saddled by well-recognized stochastic errors (e.g., insufficient data) and systematic biases (e.g., base composition, high mutation rates) inherent to molecular phylogenetics that can lead to equivocal support for phylogenetic hypotheses.26,27

In spite of these difficulties, independent analyses using multiple single-locus phylogenetic trees provide support for the existence of EGT during the evolution of
eukaryotes, such as the cyanobacterial contribution to Plantae genomes,\textsuperscript{4,25,28} the presence of red algal derived genes in diatoms and haptophytes,\textsuperscript{29} green algal genes in chlorarachniophytes,\textsuperscript{13} and cyanobacterial sequences in the nuclear genome of the photosynthetic filose amoeba \textit{Paulinella chromatophora}.\textsuperscript{30,31} Once the existence of large-scale EGT from putative serial endosymbioses and sporadic HGT events are recognized in diverse taxa, it is then fair to ask whether it remains feasible to reconstruct genealogical history of the host lineages that comprise major algal groups. The major challenge in this task is the successful exclusion of the vast collection of eukaryotic genes recruited via E/HGT that depict evolutionary histories that conflict with the vertical evolution of the host lineages. In principle, the presence of genes acquired by HGT should not impede reliable inference of genealogical history of eukaryotic lineages using molecular data. Genes implicated in E/HGT could be excluded from phylogenetic analyses (this is critical when using multi-gene concatenated alignments) and, in some cases, used as a marker to unite lineages that share a common E/HGT event,\textsuperscript{32} such as a defined genetic exchange community.\textsuperscript{33}

If appropriate phylogenetic methods and phylogenetic hypothesis-testing strategies\textsuperscript{20,27} are used to ameliorate or exclude the impact of E/HGT when inferring organismal evolution, then the representation of eukaryote evolution as a tree-like process is legitimate. As a complement to the widely used concatenated multi-gene analyses,\textsuperscript{5,10} we consider the exhaustive phylogenetic analyses of core biochemical pathways; e.g., the Calvin Cycle\textsuperscript{34} and the permeome\textsuperscript{20} to provide useful molecular markers to infer phylogenetic relationships between algal groups. At the same time, phylogenetic studies of single proteins involved in key biochemical processes, such as the plastid ADP/ATP translocator,\textsuperscript{36} carotenoid,\textsuperscript{37} and starch biosynthesis\textsuperscript{38} provide clear examples of genealogical relationships between major algal groups that are based on a history of shared E/HGTs. These genes are of fundamental importance for understanding the evolution of key cellular functions but do not serve as markers of vertical evolution.

\textbf{Figure 1.} Membrane transporter evolution in algae. (A) Maximum likelihood (ML) tree of a membrane transporter in diatoms and other “chromalveolates” that is widely shared among eukaryotes and appears to be vertically inherited. This protein encodes a member of the mitochondrial carrier family that, although of unknown function in diatoms, is annotated as a Fe\textsuperscript{2+} (or potentially other cations) transporter in yeast (GI: 6322328) that is active under low-iron conditions. (B) ML phylogenetic tree of an endoplasmic reticulum nucleotide sugar transporter that has a history of E/HGT. This pan-eukaryote membrane transporter in a distinct group of “chromalveolates” is apparently of prasinophyte (green algal) origin. RAxML and PhyML bootstrap support values based on 100 pseudoreplicates (\textasciitilde 50\%) are shown (above and below the nodes, respectively). The unit of branch lengths is the number of substitutions per site (see scale bars). Red algae are shown in red text, green algae and plants in green text and “chromalveolates” in brown text. The NCBI GI number for each sequence is shown where available.
In recent years, our understanding of algal evolution and diversification has benefited greatly from the availability of genome data from non-photosynthetic groups such as picobiliphytes\textsuperscript{39} and kata-algal evolution and diversification have been illuminated with the inclusion of these data.\textsuperscript{10} The original idea of a united "chromalveolate" lineage is no longer tenable after the identification of the SAR group\textsuperscript{10} and the finding of the intermingled positions of picobiliphytes,\textsuperscript{39} telonomids, and kata-algal plastids with photosynthetic "chromalveolates."\textsuperscript{7,9,10} Genome-scale data from many of these algal and protist groups is still lacking. Major sequencing efforts of key algal groups (e.g., red algae, glaucophytes, prasinophytes) and non-photosynthetic taxa (e.g., picobiliphytes, telonomids, and katablepharids) are therefore required both for exploring the relationships between these lineages and for elucidating the number and tempo of endosymbiotic events that have generated the astonishing diversity of microbial eukaryotes.

Acknowledgments

We acknowledge generous support from Rutgers University and valuable discussions with colleagues in the Bhattacharya lab.
33. Skippington E, Ragan MA. Lateral genetic transfer and the construction of genetic exchange communities. FEMS Microbiol Rev 2011; 35:707-35; PMID:21223321; http://dx.doi.org/10.1111/j.1574-6976.2010.00261.x
34. Reyes-Prieto A, Bhattacharya D. Phylogeny of Calvin cycle enzymes supports Plantae monophyly. Mol Phylogenet Evol 2007; 45:384-91; PMID:17482838; http://dx.doi.org/10.1016/j.ympev.2007.02.026
35. Gross J, Bhattacharya D. Revaluating the evolution of the Toc and Tic protein translocons. Trends Plant Sci 2009; 14:13-20; PMID:19042148; http://dx.doi.org/10.1016/j.tplants.2008.10.003
36. Linka N, Hurka H, Lang BF, Burger G, Winkler HH, Stamme C, et al. Phylogenetic relationships of non-mitochondrial nucleotide transport proteins in bacteria and eukaryotes. Gene 2003; 306:27-35; PMID:12657464; http://dx.doi.org/10.1016/S0378-1119(03)00429-3
37. Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zouna S, et al. Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. Mol Biol Evol 2008; 25:2653-67; PMID:18799712; http://dx.doi.org/10.1093/molbev/msn206
38. Ball S, Colleoni C, Cenci U, Raj JN, Titaiaux C. The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. J Exp Bot 2011; 62:1775-801; PMID:21220783; http://dx.doi.org/10.1093/jxb/erq411
39. Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH, et al. Single-cell genomics reveals organismal interactions in uncultivated marine protists. Science 2011; 332:714-7; PMID:21551060; http://dx.doi.org/10.1126/science.1203163
40. Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rümmele SE, Bhattacharya D. Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of Rhizaria with chromalveolates. Mol Biol Evol 2007; 24:1702-13; PMID:17488740; http://dx.doi.org/10.1093/molbev/msm089