Communicative & Integrative Biology 4:3, 361-363; May/June 2011; ©2011 Landes Bioscience

Non-random sharing of Plantae genes

Cheong Xin Chan and Debashish Bhattacharya*

Department of Ecology, Evolution and Natural Resources; and Institute of Marine and Coastal Sciences; Rutgers University; New Brunswick, NJ USA

The power of eukaryote genomics relies strongly on taxon sampling. This point was underlined in a recent analysis of red algal genome evolution in which we tested the Plantae hypothesis that posits the monophyly of red, green (including plants) and glaucophyte algae. The inclusion of novel genome data from two mesophilic red algae enabled us to robustly demonstrate the sisterhood of red and green algae in the tree of life. Perhaps more exciting was the finding that >1,800 putative genes in the unicellular red alga *Porphyridium cruentum* showed evidence of gene-sharing with diverse lineages of eukaryotes and prokaryotes. Here we assessed the correlation between the putative functions of these shared genes and their susceptibility to transfer. It turns out that genes involved in complex interactive networks such as biological regulation and transcription/translation are less susceptible to endosymbiotic or horizontal gene transfer, when compared to genes with metabolic and transporter functions.

Key words: red algae, transcriptome, expressed sequence tag, phylogenomics, horizontal gene transfer, endosymbiosis, plantae monophyly, porphyridium

Submitted: 04/01/11

Accepted: 04/01/11

DOI: 10.4161/cib.4.3.15700

*Correspondence to: Debashish Bhattacharya; Email: bhattacharya@aesop.rutgers.edu

Addendum to: Chan CX, Yang EC, Banerjee T, Yoon HS, Martone PT, Estevez JM, et al. Red and green algal monophyly and extensive gene sharing found in a rich repertoire of red algal genes. Curr Biol 2011; 21:328–33; PMID: 21315598; DOI: 10.1016/j.cub.2011.01.037.

The Plantae hypothesis posits that Rhodophyta (red algae; source of nori, agar and carrageenan), Viridiplantae (green algae and plants) and Glaucophyta (a small, poorly studied algal group that includes *Cyanophora paradoxa*) share a common ancestral origin.[1,2] These taxa have played a pivotal role in the field of endosymbiosis because the Plantae share many traits associated with plastid function and photosynthesis,[3-6] suggesting their ancestor captured the plastid (e.g., chloroplast) that has spread throughout the tree of life. Recent phylogenomic analyses of eukaryotes however, recover little[7,8] or no[9,10] support for Plantae monophyly. This may be explained by the fact that investigators have until now relied primarily on data from the highly reduced and specialized genome of the red algal extremophile *Cyanidioschyzon merolae*[11] to represent this species-rich and ancient lineage comprised primarily of mesophilic taxa. Using phylogenomic analysis of >60,000 novel red algal genes from two mesophilic species, *Porphyridium cruentum* and *Calliarthron tuberculosum*, we recently found an emerging signal of red + green (RG) monophyly that was supported by ~50% of the examined protein phylogenies.[12] Interestingly, we also recovered strong evidence of reticulate evolution between the RG lineage and other phyla of eukaryotes and prokaryotes involving 1,808 protein phylogenies (non-parametric bootstrap ≥90%; terminal taxa ≥30, number of distinct phyla ≥3). This surprising result is likely explained by endosymbiotic or horizontal gene transfer (E/HGT) involving the red and/or green algal lineages,[13,14] particularly within the context of plastid evolution.[15,16] Little is known, however, about the functional bias (if any) of this substantial level of non-lineal transfer between eukaryotes. In comparison, the correlation between gene function and susceptibility to HGT has recently been described in prokaryotes.[17]

Even though the function of many *P. cruentum* genes is unknown (62.3% do not have a hit to publicly available gene data[12]), here we asked the question if functional bias exists among the 1,808 genes implicated in E/HGT, when compared to the total number (8,082) of annotated genes in the dataset. Adopting the approach of Chan et al.[17] we postulate that putative functions over-represented in the E/HGT set are associated

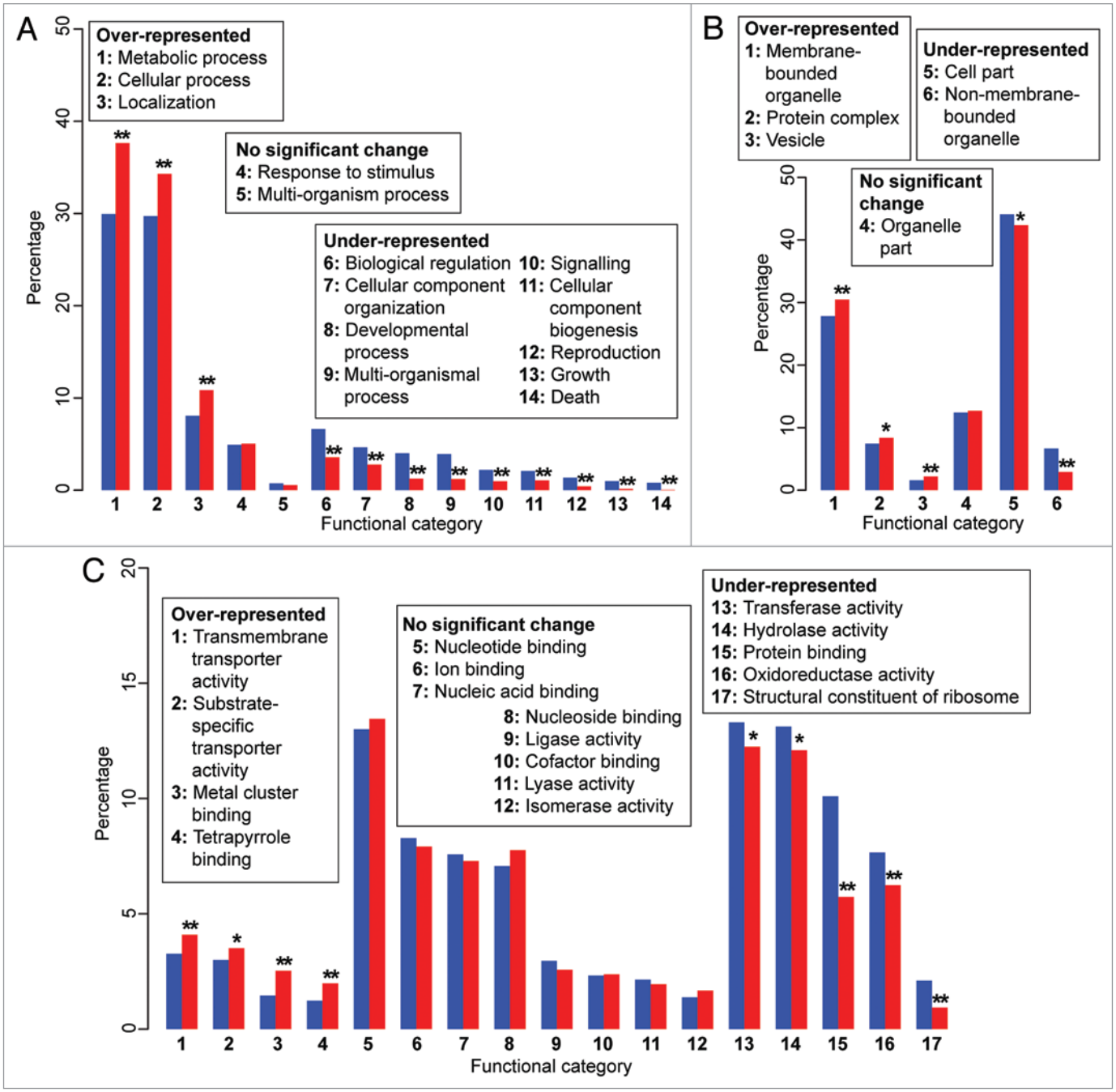

Figure 1. Annotated functional categories of proteins encoded by *Porphyridium cruentum*. The charts are organized based on three principal categories in the Gene Ontology database: (A) biological process, (B) cellular component and (C) molecular function. The blue bars show the proportional representation of the functional categories (in percentage) across the overall dataset (8,082 annotated genes), whereas the red bars show the proportional representation of the same functional categories (in percentage) among the 1,808 genes implicated in E/HGT. The categories are numbered independently for each panel, as shown in the legends. Significance of over- or under-representation is represented by single ($p \leq 0.05$) and double asterisks ($p \leq 0.01$), as inferred based on the approach of Chan et al.[17]

with genes that are more susceptible to gene sharing/transfer when compared to functions distributed across all annotated genes; i.e., the null model postulates no significant difference between the distribution of annotated functions in the E/HGT set and that across all genes. **Figure 1** shows the comparative representation of functional categories between these two groups of genes for three principal groups of "biological process," "cellular component," and "molecular function" as assigned based on sequence similarity[18] and information extracted from the Gene Ontology database (www.geneontology.org/).

Within the principal group of "biological process" (**Fig. 1A**), the categories metabolic process, cellular process and localization ($p \leq 0.01$) are significantly over-represented in genes showing

non-lineal sharing, whereas categories such as biological regulation, cellular component organization and developmental process are under-represented ($p \leq 0.01$). This finding suggests that genes involved in anabolic or catabolic pathways could have been more susceptible to non-lineal transfer than the others, and that genes involved in the intricate network of cell development and organization are less likely to have undergone such transfer. Similarly within the principal group of "cellular component" (**Fig. 1B**), protein functions related to membrane-bounded organelle (e.g., plastid and mitochondrion), vesicle ($p \leq 0.01$, respectively), and protein complex ($p \leq 0.05$) are more susceptible to E/HGT than others, in contrast to those related to ribosomes, the cytoskeleton, and chromosomes (non-membrane-bounded organelle; $p \leq 0.01$) or to structural constituents of a cell (cell part; $p \leq 0.05$).

Interestingly, for the principal group of "molecular function" (**Fig. 1C**), categories related to transporter activities, i.e., transmembrane ($p \leq 0.01$) and substrate-specific proteins ($p \leq 0.05$), as well as proteins binding to metal cluster and tetrapyrrole ($p \leq 0.01$, respectively) are over-represented in E/HGT-implicated genes compared to the overall dataset. These proteins are crucial for the transport of secondary metabolites and photosynthesis. Biomolecules that bind to metal cluster and tetrapyrrole include heme (resulting in hemoproteins; e.g., cytochrome) and chlorophyll, which are key components of the photosynthetic reaction complex. This observation suggests that algal genes encoding photosynthetic functions or transport among different cellular compartments are more likely to have a non-lineal gene history than expected under the null model.

Our results underline the notion that endosymbiosis has a significant impact both on the quantity (extensive E/HGT) and quality (function) of gene transfer between algal lineages. We found functional categories related to the transcription/translation apparatus (e.g., protein binding) and structural constituents of the ribosome ($p \leq 0.01$ respectively), among other enzymatic activities, are under-represented in the 1,808 genes compared to the overall dataset. The complexity hypothesis[19,20] postulates that genes encoding "informational" proteins that are involved in complex interactive networks implicating multi-protein complexes (e.g., processes related to transcription and translation) are less likely to be horizontally transferred than genes encoding "operational" proteins. A recent study[21] suggests that the connectivity of protein-protein interaction networks rather than biological function better explains the susceptibility of a gene to undergo HGT. This aspect remains to be rigorously tested, because the complexity of protein-protein interactions is necessarily related to overall biological function. Our results do not validate nor did we intend to test the complexity hypothesis in this study, but our findings are consistent with this working hypothesis. In conclusion, we demonstrate potential functional biases in non-lineal gene sharing in Plantae lineages that are associated with endosymbiosis, e.g., towards functions related to the establishment of cell organelles. The addition of genome data from understudied groups such as mesophilic red algae promises to further advance understanding of eukaryote evolution in general.

Acknowledgments

This research was supported by grants from the National Science Foundation (EF 08-27023, DEB 09-36884). We acknowledge the support of our colleagues in the RedToL project (lead PI, Hwan Su Yoon).

References

1. Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, et al. Monophyly of primary photosynthetic eukaryotes: green plants, red algae and glaucophytes. Curr Biol 2005; 15:1325-30.
2. Weber AP, Linka M, Bhattacharya D. Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. Eukaryot Cell 2006; 5:609-12.
3. Colleoni C, Linka M, Deschamps P, Handford MG, Dupree P, Weber APM, et al. Phylogenetic and biochemical evidence supports the recruitment of an ADP-glucose translocator for the export of photosynthate during plastid endosymbiosis. Mol Biol Evol 2010; 27:2691-701.
4. Huang JL, Gogarten JP. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? Genome Biol 2007; 8:99.
5. Reyes-Prieto A, Bhattacharya D. Phylogeny of Calvin cycle enzymes supports Plantae monophyly. Mol Phylogenet Evol 2007; 45:384-91.
6. Tyra HM, Linka M, Weber AP, Bhattacharya D. Host origin of plastid solute transporters in the first photosynthetic eukaryotes. Genome Biol 2007; 8:212.
7. Patron NJ, Inagaki Y, Keeling PJ. Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. Curr Biol 2007; 17:887-91.
8. Burki F, Shalchian-Tabrizi K, Minge M, Skjæveland Å, Nikolaev SI, Jakobsen KS, et al. Phylogenomics reshuffles the eukaryotic supergroups. PLoS ONE 2007; 2:790.
9. Baurain D, Brinkmann H, Petersen J, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V, et al. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes and stramenopiles. Mol Biol Evol 2010; 27:1698-709.
10. Nozaki H, Maruyama S, Matsuzaki M, Nakada T, Kato S, Misawa K. Phylogenetic positions of Glaucophyta, green plants (Archaeplastida) and Haptophyta (Chromalveolata) as deduced from slowly evolving nuclear genes. Mol Phylogenet Evol 2009; 53:872-80.
11. Matsuzaki M, Misumi O, Shin-i T, Maruyama S, Takahara M, Miyagishima SY, et al. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature 2004; 428:653-7.
12. Chan CX, Yang EC, Banerjee T, Yoon HS, Martone PT, Estevez JM, et al. Red and green algal monophyly and extensive gene sharing found in a rich repertoire of red algal genes. Curr Biol 2011; 21:328-33.
13. Li S, Nosenko T, Hackett JD, Bhattacharya D. Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. Mol Biol Evol 2006; 23:663-74.
14. Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. Science 2009; 324:1724-6.
15. Chan CX, Gross J, Yoon HS, Bhattacharya D. Plastid origin and evolution: new models provide insights into old problems. Plant Physiol 2011; 155:1552-60.
16. Keeling PJ. Chromalveolates and the evolution of plastids by secondary endosymbiosis. J Eukaryot Microbiol 2009; 56:1-8.
17. Chan CX, Beiko RG, Darling AE, Ragan MA. Lateral transfer of genes and gene fragments in prokaryotes. Genome Biol Evol 2009; 1:429-38.
18. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 2005; 21:3674-6.
19. Jain R, Rivera MC, Lake JA. Horizontal gene transfer among genomes: the complexity hypothesis. Proc Natl Acad Sci USA 1999; 96:3801-6.
20. Aris-Brosou S. Determinants of adaptive evolution at the molecular level: the extended complexity hypothesis. Mol Biol Evol 2005; 22:200-9.
21. Cohen O, Gophna U, Pupko T. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. Mol Biol Evol 2011; 28:1481-9.