

## MICROBIOLOGY

# Seeing Green and Red in Diatom Genomes

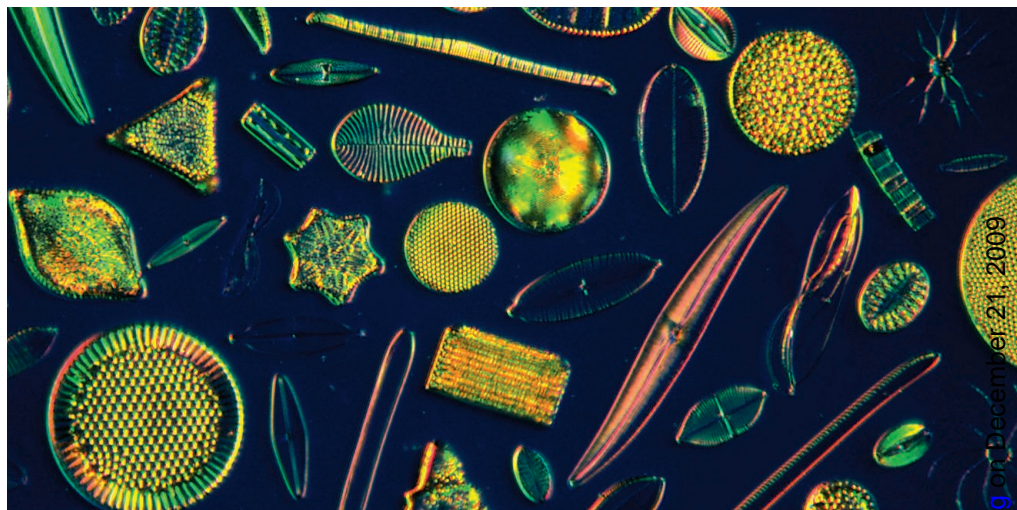
Tal Dagan and William Martin

The genomes of eukaryotes, particularly algae, are providing more and more evidence for the workings of endosymbiosis, an evolutionary source of complex cell organization where one cell (the symbiont) comes to live within another (the host). Some of that evidence is expected, but other evolutionary findings emerging from genomes are unanticipated. On page 1724 of this issue, Moustafa *et al.* (1) uncover such an evolutionary surprise from diatom genomes. The results are likely to be controversial.

Endosymbiotic theory predicts that a substantial fraction of the plant genome was likely acquired from the endosymbiont that made plants what they are—photosynthetic. Early in eukaryote evolution, one lineage acquired a cyanobacterial endosymbiont that became stably integrated into its host cell and eventually turned into a plastid (the photosynthetic compartment of plant cells that transforms solar energy, carbon dioxide, and water into sugar). Plastids still have their own DNA as testimony to their endosymbiotic origin, but that DNA contains only ~1 to 3% as many genes as found in cyanobacteria; yet plastids typically contain about as many proteins as their cyanobacterial cousins. This observation led to the expectation that the plant genome should contain many genes of cyanobacterial origin that were transferred during the course of evolution from the endosymbiont genome to the host genome through a type of lateral gene transfer (the spread of genetic information across species boundaries) called endosymbiotic gene transfer (the transfer of genes from endosymbiont to host).

Moustafa *et al.* (1) set out to look for evidence for gene transfer from organelles to the nucleus in the evolutionary history of diatom genomes. To do so they compared the thousands of genes in two completely sequenced diatom genomes to all genes from hundreds of other sequenced genomes (the search set), and created individual phylogenetic trees for all these genes. They then determined, for each diatom gene, which gene among the search set appears as the sister to the diatom gene in the phylogenetic tree, and hence is

Institut für Botanik III, Heinrich-Heine Universität Duesseldorf, Universitätsstrasse 1, 40225 Duesseldorf, Germany. E-mail: w.martin@uni-duesseldorf.de



**Diatom diversity.** Based on a comparative study of diatom genomes, Moustafa *et al.* provide evidence for a previously unknown endosymbiotic event early in the evolution of diatoms.

most likely to be the progenitor of the diatom nuclear gene.

Given that diatoms are known to have acquired their plastids from red algae via a process known as secondary endosymbiosis (2–4), one would have expected their search to uncover many nuclear genes with a red algal origin. Yet, the authors find that most symbiont-derived genes in the diatom genome have a green algal origin; only a small minority derive from red algae. To account for this unexpected find, the authors consider several possibilities. They favor the following scenario: The diatom lineage once possessed a green algal endosymbiont that donated green genes to the nucleus. This green plastid was later replaced by an endosymbiotic event that gave rise to the contemporary red plastid. The molecular evidence for this green plastid disappearing act lies in the genes that it left behind in the nucleus.

The outright replacement of well-established plastids by fresh ones of different origin through endosymbiosis might sound unlikely to some readers, but such things do happen during algal evolution, for example, among the dinoflagellates (2–4). Like lateral gene transfer among prokaryotes, endosymbiosis among algae is a normal and widespread mechanism of natural variation, but one that does not strictly adhere to Darwin's metaphor of a universal "tree of life" (5). Algal

Diatom genomes contain remnants of both green and red algal genomes, hinting at a complex evolutionary past.

genomes are chimaeras whose constituent parts have different origins, and approaches to unraveling their genome evolution must take that into account. The approach of looking at all gene histories individually does just that.

The report raises many questions. First and foremost, did the authors make some serious error in their analysis? If they did, then everyone else is making the same error. If we read the fine print of both diatom genome sequence papers (6, 7), both teams evidently saw that there was something suspiciously green about the diatom genome data, which on all counts was supposed to be purely red. Even the genome of the oomycete *Phytophthora infestans*—supposed to be descended from the same red algal symbiosis as the diatoms—contains genes of inexplicably green origin (8). Moustafa *et al.*'s suggestion of a lost green endosymbiont would account for these green evolutionary traces, but at the same time would date the shared green endosymbiotic event back to their common ancestor.

Another question is whether the small size of the red algal genome used for comparison, that of *Cyanidioschyzon merolae* (9), might bias the results. With 5300 protein coding genes, it harbors only about one-fifth as many genes as most genomes from the green lineage do. Hence, one possibility is that this is biasing our view of things. Moustafa *et al.* (1)

looked into that by checking the cases where genes from both the green and red lineages were present in the same tree and found that even there, the green signal came through; thus, it cannot readily be explained away by the tininess of the red algal genome. More data from red algae would soothe all concerns; if anyone is searching for a reason to sequence larger red algal genomes, there it is.

Another question is whether the green signal is spurious. When we are dealing with thousands of trees, as in the present study, some trees will give erroneous results purely by chance, because that is the nature of phylogenetic inference (10). So, might the green signal just be phylogenetic noise?

The ciliate genome highlights the problem of distinguishing signal from noise in gene evolution studies. Ciliates are a group of eukaryotes that lack plastids but, like the oomycetes, are relatives of the diatoms (2–4). The ciliates may have had plastids in the past, the same plastids as diatoms, but later lost them. If so, then ciliate genomes should harbor genes reflecting that photosynthetic past. In the first ciliate genome sequences, Eisen *et al.* (11) uncovered a few dozen genes among

27,000 (~0.1% of the genome) harboring potential evidence for a photosynthetic past. According to Eisen *et al.*, that signal does not rise above background noise, but Moustafa *et al.* found a few dozen of their green diatom genes in the ciliate genome.

Are a few dozen positive results among tens of thousands of cases more than one would expect by chance? That issue is not fully resolved to everyone's satisfaction so far. Thus, what constitutes "evidence" in the analysis of thousands of gene trees remains subjective. Moreover, trees are made from alignments, and alignments themselves can be a burgeoning source of phylogenetic error. Tools to help separate phylogenetic signal from noise at the level of alignments (12) are only slowly coming into use.

On the reassuring side, Moustafa *et al.* (1) are not talking about a dozen genes: They are talking about ~1000 green genes, or ~16% of the diatom genome. A phylogenetic signal of that magnitude surely tells us something important about algal evolution. Like much other recent data from genomes (13, 14), the present findings do not fit comfortably into current theories for algal evo-

lution (2–4). Recent advances in eukaryote phylogeny (15) are bringing order to chaos among the protists that lack plastids, where the evolutionary process is mostly tree-like in nature. But among algae, the lines of descent are becoming more tangled all the time.

#### References

1. A. Moustafa, *et al.*, *Science* **324**, 1724 (2009).
2. S. B. Gould, R. F. Waller, G. I. McFadden, *Annu. Rev. Plant Biol.* **59**, 491 (2008).
3. T. Kleine, U.-G. Maier, D. Leister, *Annu. Rev. Plant Biol.* **60**, 115 (2009).
4. J. M. Archibald, *Curr. Biol.* **19**, R81 (2009).
5. W. F. Doolittle, E. Bapteste, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2043 (2007).
6. E. V. Armbrust, *et al.*, *Science* **306**, 79 (2004).
7. C. Bowler, *et al.*, *Nature* **456**, 239 (2008).
8. B. M. Tyler, *et al.*, *Science* **313**, 1261 (2006).
9. M. Matsuzaki, *et al.*, *Nature* **428**, 653 (2004).
10. O. Deusch, *et al.*, *Mol. Biol. Evol.* **25**, 748 (2008).
11. J. A. Eisen, *et al.*, *PLoS Biol.* **4**, e286 (2006).
12. G. Landan, D. Graur, *Mol. Biol. Evol.* **24**, 1380 (2007).
13. M. V. Sanchez-Puerta, C. F. Delwiche, *J. Phycol.* **44**, 1097 (2008).
14. I. B. Rogozin, M. K. Basu, M. Csürös, E. V. Koonin, *Genome Biol. Evol.* doi:10.1093/gbe/evp011 (2009).
15. A. J. Roger, A. G. B. Simpson, *Curr. Biol.* **19**, R167 (2008).

10.1126/science.1175765

## EVOLUTION

# Auxin at the Evo-Devo Intersection

William E. Friedman

“Alles ist Blatt.” With this simple but profound assertion (“All is leaf”), Johann Wolfgang von Goethe launched the modern age of comparative biology (1–3). By articulating the concept that plants can be broken down into modular and iterative variants of an archetypal structure (the leaf, in the form of bud scales, spines, petals, stamens and so forth), Goethe propelled the analysis of plant and animal structure (4) for the next two centuries and beyond. The idea that variant forms of a basic organ “type” are homologous (morphologically equivalent) within an organism and between organisms has emerged as a central conceptual (and testable) framework in the burgeoning field of comparative molecular analyses of development. On page 1684 of this issue, a study by Pagnussat *et al.* (5) brings together a remarkable set of experiments that bear on the developmental biology and modular construction

of the microscopic egg-producing structure (female gametophyte or embryo sac) buried deep within a flowering plant's reproductive tissues. The findings have great importance for understanding and further examining the evolutionary developmental history of flowering plants.

Pagnussat *et al.* demonstrate in the flowering plant *Arabidopsis thaliana* that the phytohormone auxin is a key determinant of cell fates within the angiosperm female gametophyte. The embryo sac contains seven cells and eight nuclei: an egg cell, two synergids (one of which will receive the pollen tube bearing two sperm cells), a binucleate central cell, and three sterile antipodal cells at the opposite pole from the egg (see the figure). The egg cell and central cell serve as female gametes and, upon receipt of the two sperm from a pollen tube during the process of double fertilization, will yield a diploid zygote and a triploid endosperm, the embryo-nourishing tissue within the seed. Pagnussat *et al.* show that the distribution of auxin within the develop-

The female gametophyte of flowering plants may have evolved through iteration of a modular structure and a gradient of the hormone auxin.

ing embryo sac is polarized, and they propose that gradient-based variation in the concentration of auxin determines the identity that cells will assume during the transition from a single-celled syncytium to a seven-celled, eight-nucleate mature structure. This finding is seminal, as auxin's centrality to patterning and differentiation in the angiosperm female gametophyte—or, for that matter, any land plant gametophyte—had not been anticipated.

Following the recent discovery that the earliest flowering plants likely produced a female gametophyte with only four cells and four nuclei, it was proposed that the angiosperm female gametophyte is a modular and iterative structure involving quartets of nuclei (6–9) (see the figure). Development of a basic angiosperm female gametophyte module was hypothesized to involve three ontogenetic stages: positioning of a single nucleus within a developmentally autonomous cytoplasmic domain of the female gametophyte; two nuclear division events to yield four nuclei within that domain; and par-

Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA. E-mail: ned@colorado.edu