Algal Genomics: Exploring the Imprint of Endosymbiosis

The nuclear genomes of photosynthetic eukaryotes are littered with genes derived from the cyanobacterial progenitor of modern-day plastids. A genomic analysis of *Cyanophora paradoxa* — a deeply diverged unicellular alga — suggests that the abundance and functional diversity of nucleus-encoded genes of cyanobacterial origin differs in plants and algae.

John M. Archibald

In the billion odd years since their cyanobacterial predecessors first inhabited the cytoplasm of a non-photosynthetic eukaryote [1], the plastids (chloroplasts) of plants and algae have been radically transformed. The bulk of this transformation has come in the form of genomic 'downsizing' — the elimination of superfluous genes and the transfer of essential genes to the host nuclear genome. As reported recently in *Current Biology*, Reyes-Prieto *et al.* [2] have estimated the cyanobacterial contribution to the nuclear genome of the 'coelacanth' of the algal world, the freshwater unicell *Cyanophora paradoxa*. They conclude that, in contrast to what is seen in the genome of the land plant *Arabidopsis* [3], the *Cyanophora* genome harbors a modest set of cyanobacteria-derived genes, the functions of which appear tightly linked to the plastid. Diverse plants and algae seem to differ in the extent to which they have taken advantage of the cyanobacterial gene pool present in their common ancestor.

The fate of cyanobacterial genes in photosynthetic eukaryotes has been debated for more than 20 years. Long before the era of genome sequencing, Weeden [4] proposed that plant nuclei must possess large numbers of genes derived from cyanobacteria. He based this proposition on the fact that plastid genomes are small and cannot possibly encode all the proteins necessary for plastid function. Weeden's insight was spot-on: plastid genomes are indeed extremely limited in their coding potential (rarely >200 protein genes [5]; Figure 1) and, as predicted, plant nuclear genomes harbor thousands of cyanobacterial genes whose proteins are synthesized in the cytoplasm and targeted back to the plastid [6–8]. The reality is even more complex than Weeden could have imagined. We now know that, in addition to producing proteins that service their compartment of origin, nucleus-encoded cyanobacterial genes have the potential to acquire all manner of functions in plant and algal cells (Figure 1).

The full extent to which the function of cyanobacterial genes can be uncoupled from their prokaryotic ancestry was revealed in 2002. Building on earlier studies indicating that, in plants, metabolic pathways are often cobbled together from enzymes with diverse evolutionary histories (reviewed in [9]), Martin *et al.* [3] scrutinized the complete sequence of the *Arabidopsis* nuclear genome [10] with the goal of answering the following questions: How many cyanobacterial proteins does it encode? And of these, how many are plastid-targeted? The answer was unexpected: of the 24,990 analyzed *Arabidopsis* proteins, a whopping 18% (4,500 in total) were inferred to have come from cyanobacteria — as many or more than the number of proteins encoded in most cyanobacterial genomes [3]. Even more surprising was the fact that the predicted functions of >50% of the analyzable proteins have nothing to do with the plastid or photosynthesis, and fall into a wide range of functional categories, including metabolism, cell division and intracellular transport [3]. The cyanobacterial contribution to the nuclear genome of *Arabidopsis* appears to extend well beyond the

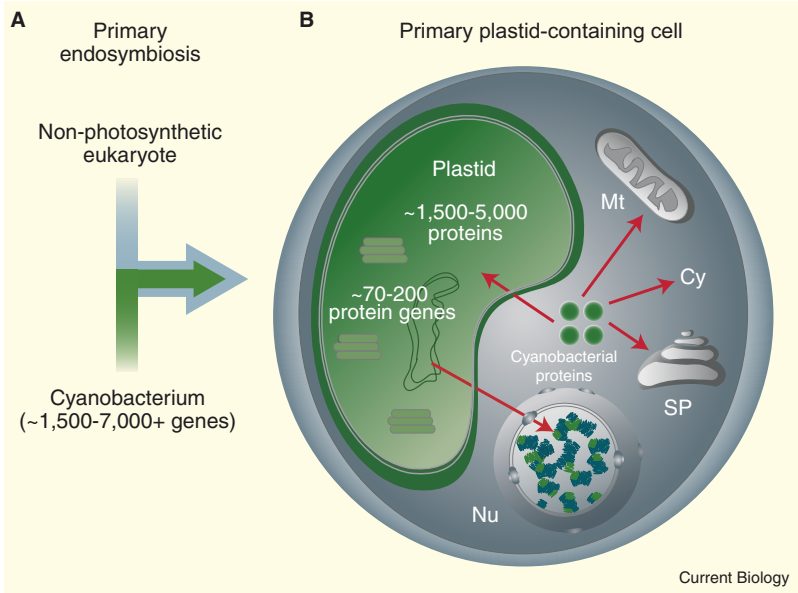

Figure 1. Endosymbiotic origin of plastids.

(A) Primary endosymbiosis between a non-photosynthetic eukaryote and a cyanobacterium, with the predicted number of cyanobacterial genes based on the completely sequenced genomes of modern-day species. (B) Schematic diagram of a primary plastid-containing cell. The number of plastid genes is based on the gene content of sequenced red, green and glaucophyte plastid genomes. Arrows show the known or predicted flow of genes and proteins. Gene transfers involving the mitochondrion are omitted for simplicity, as is the targeting of non-cyanobacterial proteins to the plastid and other cellular compartments. Abbreviations: Nu, nucleus; Mt, mitochondrion; Cy, cytoplasm; SP, secretory pathway.

Figure 2. Light micrograph of *Cyanophora*.

Image by D. Patterson, provided with permission by http://microscope.mbl.edu.

provision of essential genes for plastid function (Figure 1). Is this true of other plants and algae?

In order to grasp the significance of this question it is important to recognize that despite their current terrestrial dominance, land plants (or embryophytes) had a humble beginning. They are essentially the multicellular cousins of unicellular green algae [11], and an increasingly robust body of evidence indicates that all plastids — including those of land plants — can be traced back (directly or indirectly) to a single 'primary' endosymbiosis that occurred in the common ancestor of green algae, red algae and an enigmatic unicellular lineage called glaucophytes ([1,12] and references therein). From this perspective, one wonders how much of the creative usage of cyanobacterial genes in *Arabidopsis* occurred relatively recently during the evolution of land plants, and how much of it happened earlier, in the common ancestor they share with green, red and glaucophyte algae?

Reyes-Prieto *et al.* [2] set out to answer this question by surveying the nuclear genome of the glaucophyte *Cyanophora paradoxa* (Figure 2). Glaucophyte plastids (or cyanelles) bear a striking resemblance to free-living cyanobacteria — they were once assigned their own genus and species [13] — and have several cyanobacterial features that are absent in canonical plastids, most notably a peptidoglycan layer between their two cell membranes [13,14]. While molecular data indicate that the *Cyanophora* plastid is not quite the 'missing link' its morphology would suggest — its genome is similar in size and coding capacity to those of other plastids [15,16] — the glaucophytes do appear to be the earliest diverging of the three primary plastid-containing lineages [3,16]. *Cyanophora* is therefore as evolutionarily distant from *Arabidopsis* as any photosynthetic eukaryote known and thus an important lineage in which to test the extent of ancient cyanobacterial gene recruitment.

From a dataset of 11,176 *Cyanophora* expressed sequence tags (ESTs), Reyes-Prieto *et al.* [2] identified 3,576 unique genes and 1,226 of these produced significant database hits to known or conserved proteins. Consideration of the functional distribution of these genes revealed no obvious sampling bias, and the authors conclude that their dataset is a random sample corresponding to 24–30% of the *Cyanophora* genome [2]. To search for cyanobacterial genes in their ESTs, a two-tiered 'similarity' and 'phylogenomic' search strategy was employed, yielding a set of 132 proteins of putative cyanobacterial origin. This corresponds to 3.7% of analyzable genes and ~10.8% of the 1,226 genes with clear homologs in other organisms. In other words, ~1,500 of the estimated 12,000–15,000 genes in the *Cyanophora* genome are cyanobacterial in origin [2]. Intriguingly, >90% of the putative cyanobacterial proteins in the *Cyanophora* dataset (12 of 132) are predicted to have plastid functions. These numbers provide an interesting contrast to those obtained for *Arabidopsis*, where 4,500 cyanobacterial proteins were inferred, fewer than 50% of which were predicted to be plastid-localized [3]. What are we to make of these differences?

First, the estimates obtained for *Cyanophora* are based on partial genomic data and will need to be revisited when a complete genome sequence becomes available, and with organelle targeting prediction algorithms trained on experimentally verified *Cyanophora* proteins. The recent funding of a project by the National Science Foundation to generate a complete genome sequence for *Cyanophora* (D. Bhattacharya, personal communication) promises to provide these data. Second, Reyes-Prieto *et al.* [2] note that, at the time of the Martin *et al.* *Arabidopsis* study [3], only one other eukaryotic genome (that of the budding yeast *Saccharomyces cerevisiae*) was available for comparison, a situation that has improved dramatically in recent years. It is possible that differences in datasets and search strategies are at least partly responsible for the variance in the estimated proportions and cellular functions of cyanobacterial proteins in *Arabidopsis* and *Cyanophora*.

Methodological issues aside, an important factor to consider when comparing the *Cyanophora* and *Arabidopsis* numbers is the huge amount of time that has transpired since glaucophytes and green algae diverged from a common ancestor. The apparent differences in the abundance and functional diversity of cyanobacterial genes in the two genomes could in part be the result of 'amplification' of cyanobacteria-derived genes during the evolution of plants via gene and genome duplication. Both processes have figured prominently in the evolution of plant genomes [10,17], and while there is still much to learn about the structure of glaucophyte genomes, very few genes in the *Cyanophora* EST dataset appear to belong to multigene families [2].

Although genome duplication would not increase the percentage of cyanobacterial genes in the *Arabidopsis* genome, it could certainly increase the total number of cyanobacterial homologs available for functional 'reassignment'. Once a core set of ancient cyanobacterial genes has been identified in red, green and glaucophyte algae, it should be possible to determine the extent to which gene and genome duplication has played a role in the functional diversification of cyanobacterial proteins by comparing land plant genomes to those of their closest algal relatives, the charophytes [11]. Indeed, a recent comparison of the

Arabidopsis and rice genomes [18] suggests that lineage-specific tinkering with the composition of the plastid proteome can occur even over short evolutionary timescales.

In the mean-time, multiple algal genome sequences are now publicly available — for example, those of the green algae *Chlamydomonas* and *Ostreococcus* (http://genome.jgi-psf.org/euk_cur1.html) and the red alga *Cyanidioschyzon* (http://merolae.biol.s.u-tokyo.ac.jp) — and it is likely that ten or more plant and algal nuclear genome sequences will be available for even more thorough and systematic analyses within the next few years. An increasing number of genomes from 'secondary' plastid-containing algae such as diatoms, cryptophytes and chlorarachniophytes will also be completely sequenced (http://www.jgi.doe.gov/). These organisms acquired their plastids through the engulfment of red or green algal endosymbionts [19] and the molecular dynamics accompanying the process of secondary endosymbiosis, in which gene transfers between evolutionarily distinct nuclear genomes are also a possibility, adds another layer of complexity to an already complicated picture [20]. Overall, it is sobering to consider how little we know about the nuts and bolts of endosymbiosis and the full scope of its role in the diversification of eukaryotic cells.

References

1. Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. *21*, 809–818.
2. Reyes-Prieto, A., Hackett, J.D., Soares, M.B., Bonaldo, M.F., and Bhattacharya, D. (2006). Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. Curr. Biol. *16*, 2320–2325.
3. Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., and Penny, D. (2002). Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. USA *99*, 12246–12251.
4. Weeden, N.F. (1981). Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. J. Mol. Evol. *17*, 133–139.
5. Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. *5*, 123–135.
6. Bruce, B.D. (2000). Chloroplast transit peptides: structure, function and evolution. Trends Cell Biol. *10*, 440–447.
7. McFadden, G.I. (1999). Plastids and protein targeting. J. Eukaryot. Microbiol. *46*, 339–346.
8. Soll, J., and Schleiff, E. (2004). Protein import into chloroplasts. Nat. Rev. Mol. Cell Biol. *5*, 198–208.
9. Martin, W., and Schnarrenberger, C. (1997). The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Curr. Genet. *32*, 1–18.
10. The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature *408*, 796–815.
11. Karol, K.G., McCourt, R.M., Cimino, M.T., and Delwiche, C.F. (2001). The closest living relatives of land plants. Science *294*, 2351–2353.
12. Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Löffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B.F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr. Biol. *15*, 1325–1330.
13. Hall, W.T., and Claus, G. (1963). Ultrastructural studies on the blue-green algal symbiont in *Cyanophora Paradoxa* Korschikoff. J. Cell Biol. *19*, 551–563.
14. Steiner, J.M., Berghöfer, J., Yusa, F., Pompe, J.A., Klösgen, R.B., and Löffelhardt, W. (2005). Conservative sorting in a primitive plastid. The cyanelle of *Cyanophora paradoxa*. FEBS J. *272*, 987–998.
15. Löffelhardt, W., Bohnert, H.J., and Bryant, D.A. (1997). The complete sequence of the *Cyanophora paradoxa* cyanelle genome. In Origins of Algae and Their Plastids, D. Bhattacharya, ed. (Wien: Springer), pp. 149–162.
16. Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M., and Kowallik, K.V. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. Nature *393*, 162–165.
17. De Bodt, S., Maere, S., and Van de Peer, Y. (2005). Genome duplication and the origin of angiosperms. Trends Ecol. Evol. *20*, 591–597.
18. Richly, E., and Leister, D. (2004). An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of *Arabidopsis* and rice. Gene *329*, 11–16.
19. Bhattacharya, D., Yoon, H.S., and Hackett, J.D. (2003). Photosynthetic eukaryotes unite: endosymbiosis connects the dots. Bioessays *26*, 50–60.
20. Archibald, J.M. (2005). Jumping genes and shrinking genomes—probing the evolution of eukaryotic photosynthesis with genomics. IUBMB Life *57*, 539–547.

The Canadian Institute for Advanced Research, Program in Evolutionary Biology, Department of Biochemistry and Molecular Biology, Dalhousie University, Sir Charles Tupper Medical Building, 5850 College Street, Halifax, Nova Scotia, B3H 1X5, Canada.
E-mail: jmarchib@dal.ca

DOI: 10.1016/j.cub.2006.11.008

Quality Control: Linking Retrotranslocation and Degradation

Misfolded proteins in the ER require the p97 AAA ATPase for dislocation across the membrane prior to degradation by the cytosolic proteasome. The mechanism by which dislocated proteins are delivered to the proteasome from p97 is unclear, but recent studies suggest an important regulatory role for the protein ataxin-3.

Colin J. Stirling[1] and J. Michael Lord[2]

The fatal neurodegenerative Machado-Joseph disease, also known as spinocerebellar ataxia type 3, is caused by mutations of the polyglutamine-containing protein ataxin-3 [1]. The biochemical properties of ataxin-3 are well known, but its physiological role has been elusive. Two recent studies [2,3] now implicate ataxin-3 in the process known as endoplasmic-reticulum-associated degradation (ERAD), suggesting a link between endoplasmic reticulum (ER) stress and the neuropathology associated with disease.

After targeting to the ER, proteins are screened by a quality control system to prevent misfolded forms from progressing through the secretory pathway. Rather than accumulating within the cell, these aberrant proteins are disposed of by ERAD. This requires the dislocation of such proteins across the ER membrane and their subsequent degradation