Current Biology, Vol. 14, R1048–R1050, December 29, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.cub.2004.11.043

Diatom Genomics: Genetic Acquisitions and Mergers

Dispatch

R. Ellen R. Nisbet, Oliver Kilian and Geoffrey I. McFadden

Diatom algae arose by two-step endosymbiosis. The complete genome of the diatom *Thalassiosira pseudonana* has now been sequenced, allowing us to reconstruct the remarkable intracellular gene transfers that occurred during this convoluted cellular evolution.

Imagine you are running a successful small business converting carbon dioxide into sugar. Suddenly, you are taken over by a bigger company. They commandeer your intellectual property, relocate it to head office, and — to add insult to injury — they ship your own tools back to you and expect you to keep making sugar. Such a business takeover is the perfect analogy for the endosymbiotic origin of plastids. The small business is a photosynthetic cyanobacterium, the aggressive takeover merchant is the eukaryotic host, and the intellectual property is the cohort of genes encoding the machinery for photosynthesis, most of which have now been relocated from the endosymbiont's genome to the host nucleus [1].

Now imagine a second takeover. You and your parent company are swallowed up by an even bigger company, which then raids what is left of the intellectual property in your safe and also pillages your parent company's safe, taking almost everything of value to the new head office. This double takeover scenario describes the sequential or secondary endosymbiosis that created major lineages of marine phytoplankton such as diatoms. A recent paper [2] reports the complete genome sequence of the diatom *Thalassiosira pseudonana*, allowing us to reconstruct the sometimes complicated pilgrimages of photosynthesis genes from endosymbiont to primary host and then onwards to the secondary host — the diatom nucleus.

Diatoms are unicellular eukaryotic algae with golden brown pigments and an ornate silica casing. Fossil diatoms are known as early as the Cretaceous, 144–65 million years ago [3], but their evolutionary history may extend as far back as 550 million years [4]. Carbon fixation and primary productivity by the enormously abundant marine diatoms is a major component of global photosynthesis, and diatoms have diversified into a spectacular array of forms with many thousands of species [5]. Diatoms arose by secondary endosymbiosis when a eukaryote engulfed a red alga to spawn a new lineage known as the chromalveolates, which also includes haptophytes, brown seaweeds, dinoflagellates and plastid-containing parasites such as malaria [4]. The engulfed red alga had previously been created by a primary endosymbiosis between a cyanobacterium and a phagotrophic eukaryote [4].

Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Victoria 3010, Australia.

The complete genome sequence of a diatom species [2] now provides the last missing piece of a puzzle that allows us to determine the fate of the plastid protein genes at all key stages of this extraordinary process. We now have the genomes of several cyanobacteria, for example *Synechocystis* spp. [6], a red alga, *Cyanidioschyzon merolae* [7], and an intermediate form of a secondary endosymbiont, namely the cryptomonad *Guillardia theta*, which probably diverged before the origin of diatoms and still retains a vestige of the red algal nucleus [8]. Figure 1 shows the fate of selected genes along the evolutionary pathway to diatoms, and sketches out the fascinating range of possible outcomes.

Primary endosymbiosis involved the transfer of many hundreds of genes from the cyanobacterial endosymbiont to the host nucleus [1,9]. For instance, genes such as *hcf136*, *psbO* and *tic22* — which encode proteins involved in photosystem II assembly, photosystem II function and protein import, respectively — were transferred from the genome of the cyanobacterial endosymbiont to the primary host nucleus soon after the primary endosymbiosis event [9], and today we find that these cyanobacterial genes persist in the red algal nucleus [7,9] (Figure 1).

At the outset of the secondary endosymbiotic event that created chromalveolates, these genes would have been located in the nucleus of the eukaryotic endosymbiont, but the diatom genome sequence [2] now reveals that *hcf136*, *psbO* and *tic22* (amongst many others) were relocated into the secondary host nucleus. These genes have thus undergone two relocations: one from the cyanobacterial genome to the primary host nucleus, and a second transfer from the endosymbiont nucleus to the secondary host nucleus (Figure 1).

We can be quite sure about this remarkable stepwise gene pilgrimage because Nature has provided us with a snapshot of this transfer process in the form of a little-known group of organisms called cryptomonads. According to the chromalveolate hypothesis, cryptomonads diverged after the secondary endosymbiosis of a red alga but before the divergence of diatoms [4], and in cryptomonads it appears that nucleus-to-nucleus transfer process has not proceeded to completion because about 30 of the all important plastid protein genes — including *hcf136* and *tic22*, but not *psbO* — are still located in a tiny relict red algal nucleus known as the nucleomorph [8]. Thus, it appears that, in cryptomonads, these genes are trapped in a kind of genetic limbo, awaiting relocation to the new head office: only then will it be possible to dispense with the endosymbiont nucleus altogether, as has occurred in diatoms (Figure 1).

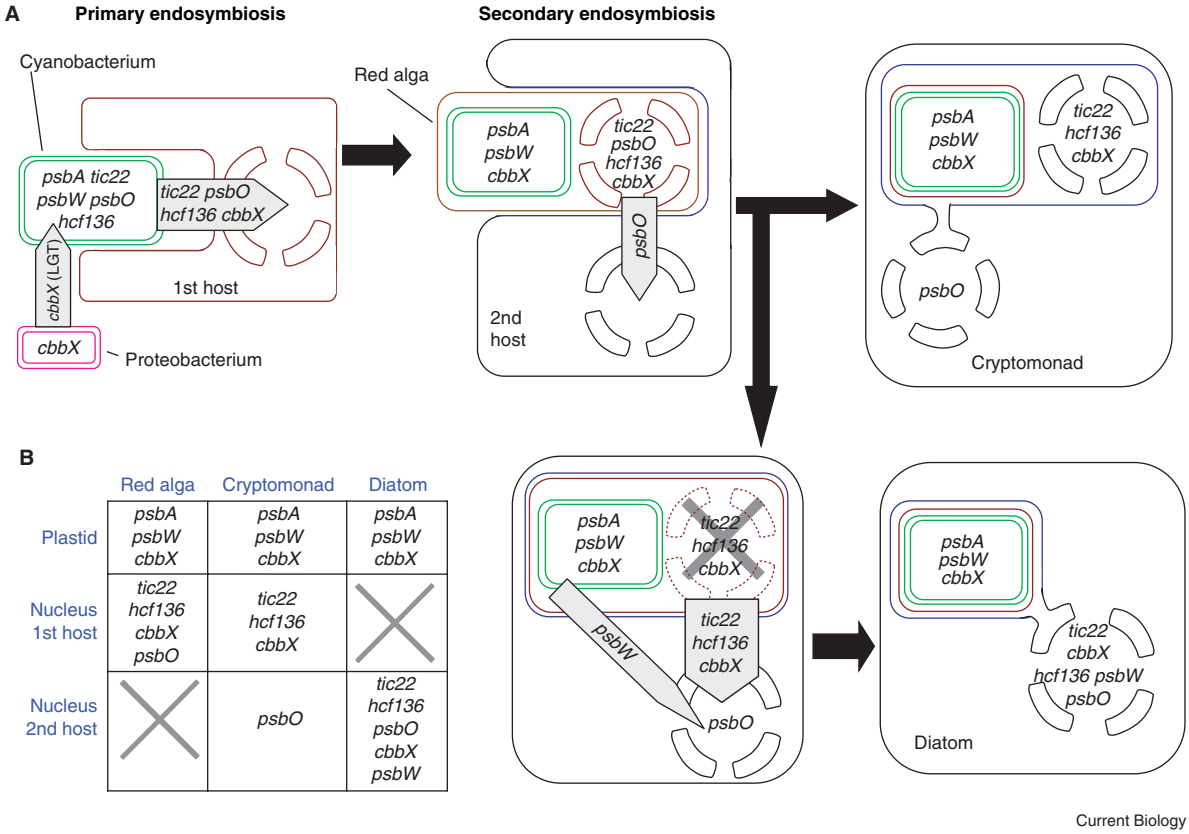

Figure 1. The evolutionary origin of diatoms.

A primary endosymbiosis event creates a red alga with two plastid membranes; this is followed by a secondary endosymbiosis event that creates the cryptomonads and diatoms, both of which have four-membraned plastids. The fate of selected photosynthesis genes is tracked (grey arrows) and the derivation of the plastid membranes is colour coded. The *cbbX* gene originated in proteobacteria and underwent lateral gene transfer (LGT) into cyanobacteria and then relocation into the primary host nucleus followed by another relocation into the secondary host nucleus. Only when all the essential genes in the primary host nucleus were transferred to the secondary host nucleus did this genome disappear. Other genes like *psbW* originated in cyanobacteria and remained in the plastid until after secondary endosymbiosis then transferred directly into the secondary host nucleus. Select genes such as *psbA* have never left the photosynthetic plastid.

Other genes have apparently traced a different course to the diatom nucleus, bypassing the primary host nucleus altogether. Consider *psbW*, which encodes another protein of photosystem II and entered the primary endosymbiosis as part of the cyanobacterial genome [6], but has apparently resisted relocation to the nucleus in red algae [7]. In diatoms, however, we find *psbW* in the nucleus and the gene product is apparently targeted back to the diatom plastid. Intriguingly, this plastid-to-secondary nucleus transfer appears to have been relatively recent, as the diatom's plastid copy of *psbW* has not yet been lost and appears functional, so it would seem that both copies are competing for evolutionary survival. Transfer of *psbW* to the nucleus has probably not occurred in the cryptomonad, where *psbW* is in the plastid genome but absent from the nucleomorph [8], although it could be in the unsequenced nucleus. The genome sequence from a second diatom, *Phaeodactylum tricornutum* [10], will be available soon and should provide more detail about the timing of this transfer as it will allow us to trace *psbW*'s movements within the diatom radiation.

The gene with the most extraordinary history must be *cbbX*, which encodes a protein necessary for photoautotrophic growth [11]. Sometime prior to the primary endosymbiotic event, a cyanobacterium-like ancestor apparently acquired a transmissible plasmid with a small operon containing three photosynthesis-related genes, including *cbbX*, from the purple (proteo)-bacteria [12]. This suite of genes, acquired through bacterium-to-bacterium transfer, became part of the cyanobacterial genome and then, after the primary endosymbiosis event a component of the plastid genome (the operon structure is retained in the red algal plastid).

Meanwhile, a copy of the *cbbX* gene was integrated into the red algal nuclear genome [7]. After secondary endosymbiosis, the *cbbX* gene was transferred to the diatom nuclear genome (from the captured red algal nucleus). In the cryptomonad algae, however, the *cbbX* gene remained in the nucleomorph (derived from the red algal nucleus) and was not transferred to the secondary host nucleus [12]. In both cases, the chloroplast copy is retained. Thus *cbbX* has traipsed from bacterium to bacterium, then

to the primary host nucleus and then finally to the secondary host nucleus of the diatom.

Although the great majority of plastid protein genes are peripatetic, wandering from genome to genome, a quorum of genes — *psbA*, which encodes a photosystem II protein, is a diatom example — have not left the plastid [9]. What keeps these homebodies anchored in place when their compatriots so readily relocate? To answer this question we first need to examine why genes move at all, and why transfer is unidirectional. Clearly DNA moves from plastids to nuclei frequently, so the opportunity for transfer is high [13]. Endosymbionts, either primary or secondary, are genetically isolated, doomed to a clonal existence with no opportunity for recombination. Endosymbiont genes are thus particularly prone to error accumulation. Add to this a high rate of mutation in plastids, due to free radicals generated by photosynthesis, and we have a dangerous situation for precious genetic resources. Relocation to a sexual milieu remote from redox damage would thus have been favoured by selection and this might explain the dominant trend to relocate [14].

At least three hypotheses have been proposed [15–17] to rationalise the persistence of those few genes still in the plastid. One is that the gene products are too hydrophobic to undertake the journey across the membranes bounding a plastid. A more compelling reason for retaining certain genes in a plastid lies in the fine-tuning of photosynthesis and the need to maintain redox poise. This argument holds that key redox proteins with high turnover rates require subtle management on-site [16,18], but this does not explain the retention of a cyanobacterium-style genome in the non-photosynthetic plastids of malaria parasites [19]. A third possibility is that there simply has not yet been enough time to relocate all the genes, and that today's gene distribution is just a snapshot of an ongoing, as yet incomplete, process. According to this view, all endosymbiont genes might ultimately be relocated to the host nucleus, at which point the giant corporation of the eukaryotic cell will have won the battle to control essential intellectual property. Genomes are now giving us the information to audit the gene transfer process and refine our understanding of this remarkable process.

References

1. 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.
2. Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.A., Martinez, D., Putnam, N.H., Zhou, S., Allen, A.E., Apt, K.E., Bechner, M., et al. (2004). The genome of the diatom *Thalassiosira pseudonona*: ecology, evolution and metabolism. Science *306*, 79-86.
3. Falkowski, P.G., Katz, M.E., Knoll, A.H., Quigg, A., Raven, J.A., Schofield, O., and Taylor, F.J. (2004). The evolution of modern eukaryotic phytoplankton. Science *305*, 354-360.
4. Cavalier-Smith, T. (2002). Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). Phil. Trans. R. Soc. London B *358*, 109-134.
5. Round, F., Crawford, R., and Mann, D. (1990). The Diatoms (Cambridge: Cambridge University Press).
6. Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., *et al.* (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. *3*, 109-136.
7. Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S.Y., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., *et al.* (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature *428*, 653-657.
8. Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L.-T., X., W., Reith, M., Cavalier-Smith, T., and Maier, U.-G. (2001). The highly reduced genome of an enslaved algal nucleus. Nature *410*, 1091-1096.
9. Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M., and Kowallik, K. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. Nature *393*, 162-165.
10. Scala, S., Carels, N., Falciatore, A., Chiusano, M.L., and Bowler, C. (2002). Genome properties of the diatom *Phaeodactylum tricornutum*. Plant Physiol. *129*, 993-1002.
11. Gibson, J.L., and Tabita, F.R. (1997). Analysis of the *cbbXYZ* operon in *Rhodobacter sphaeroides*. J. Bacteriol. *179*, 663-669.
12. Maier, U.G., Fraunholz, M., Zauner, S., Penny, S., and Douglas, S. (2000). A nucleomorph-encoded CbbX and the phylogeny of RuBisCo regulators. Mol. Biol. Evol. *174*, 576-583.
13. Huang, C.Y., Ayliffe, M.A., and Timmis, J.N. (2003). Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature *422*, 72-76.
14. Martin, W. (2003). Gene transfer from organelles to the nucleus: frequent and in big chunks. Proc. Natl Acad. Sci. USA *100*, 8612-8614.
15. Palmer, J. (1997). Organelle genomes: going, going, gone! Science *275*, 790-791.
16. Allen, A.E. (2002). The function of genomes in bioenergetic organelles. Phil. Trans. R. Soc. London B *358*, 19-38.
17. Martin, W., and Herrmann, R.G. (1998). Gene transfer from organelles to the nucleus: how much, what happens, and why? Plant Physiol. *118*, 9-17.
18. Allen, J.F., and Raven, J.A. (1996). Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. J. Mol. Evol. *42*, 482-492.
19. Wilson, R.J.M., Denny, P.W., Preiser, P.R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D.J., Moore, P.W., *et al.* (1996). Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. J. Mol. Biol. *261*, 155-172.