Minireview

Genomics and chloroplast evolution: what did cyanobacteria do for plants?

John A Raven* and John F Allen†

Addresses: *Division of Environmental and Applied Biology, University of Dundee, Dundee DD1 4HN, UK. †Department of Plant Biochemistry, Center for Chemistry and Chemical Engineering, Box 124, Lund University, SE-221 00 Lund, Sweden.

Correspondence: John A Raven. E-mail: j.a.raven@dundee.ac.uk

Abstract

The complete genome sequences of cyanobacteria and of the higher plant *Arabidopsis thaliana* leave no doubt that the plant chloroplast originated, through endosymbiosis, from a cyanobacterium. But the genomic legacy of cyanobacterial ancestry extends far beyond the chloroplast itself, and persists in organisms that have lost chloroplasts completely.

The endosymbiont hypothesis is mainstream

Chloroplasts, the sites of photosynthesis within plant cells, comprise a prominent and well-known class of plastids, subcellular organelles with diverse, specialist functions in plant and algal cells. Mereschkowsky [1,2] is widely recognized as having written the first clear exposition of the hypothesis that plastids are derived from endosymbiotic cyanobacteria, then known as blue-green algae. Initially greeted with skepticism or even derision, Mereschkowsky's 1905 hypothesis gained support from electron microscopical and biochemical studies which showed that plastids contain DNA, RNA and ribosomes, supplying a structural and biochemical basis for non-Mendelian, cytoplasmic inheritance of plastid-related characters [3]. Subsequent molecular genetic studies have demonstrated the ubiquity of plastid genomes and confirmed that their replication, transcription and translation closely resemble those of (eu)bacteria.

Molecular phylogenetic studies now make it abundantly clear that the closest bacterial homologs of plastids are indeed cyanobacteria [4], supporting earlier conclusions from the comparative biochemistry of photosynthesis. Only cyanobacteria and chloroplasts have two photosystems and split water, to make oxygen, as a source of reducing power. But it has long been clear that many of the proteins needed for plastid functions, including photosynthesis, are now encoded in the nuclear genome and arrived there during evolution by the wholesale uptake of cyanobacteria, including their genomes, followed by gene transfer into the nucleus [5]. Recent advances in genomics have greatly enhanced our understanding of the evolution of plastids, allowing us to address specific questions such as which genes were moved or retained, and why. It also becomes possible to see clearly the algal ancestry of cells that have vestigial and otherwise unrecognizable plastids, and even to discern the unmistakable genomic footprint of plastids long lost from organisms one might never imagine to have descended from plants.

Molecular genetic studies of plastid genomes show that they encode only 60-200 proteins, while perhaps as many as 5,000 nuclear-coded gene products are targeted to plastids [6]. From complete sequences it is known that each cyanobacterial genome codes for at least 1,500 proteins, and they are therefore at least an order of magnitude larger than plastid genomes. It is perhaps surprising that the size of the proteome of a free-living cyanobacterium is not greatly different from that of a subcellular organelle. Genomic studies have been very important in showing the evolutionary fate of the cyanobacterial genes that originated from the endosymbiotic pre-plastids. The genes in pre-plastids were either retained, lost, or transferred to the nucleus. The process of transfer of genes to the nucleus would have involved duplication of each

plastid gene, and a nuclear copy of the gene becoming able to produce a functional product in the cytosol or, with appropriate targeting sequences, in other compartments.

The fates of endosymbiont genes

An important recent analysis by Martin *et al.* [6] has put limits on the number of genes in the nucleus of *Arabidopsis thaliana* that derive from the plastid ancestor. Previous analyses of limited portions of the *A. thaliana* nuclear genome suggested that 800-2000 genes from the plastid ancestor were transferred to the nucleus. The analysis by Martin *et al.* [6] was based on comparison of the whole nuclear genome of *A. thaliana* with whole genomes of three cyanobacteria (*Nostoc punctiforme*, *Prochlorococcus marinus* and *Synechocystis* sp. PCC 6803), 16 other prokaryotes, and *Saccharomyces cerevisiae* (yeast). The analysis was restricted to the 9,368 *A. thaliana* gene products that are sufficiently conserved for the comparison of primary sequences. Of these, the greatest number of similarities were detected with the yeast nuclear genome; these common genes were presumably inherited by *Arabidopsis* from the host cell that acquired the plastid(s) [6]. The second most numerous class of genes in the *Arabidopsis* nuclear genome are those directly homologous to cyanobacterial genes. A decreasing number of similarities is found in Gram-positive bacteria, non-proteobacterial Gram-negative bacteria, proteobacteria, and least of all in archaebacteria [6]. Extrapolating the data from the 9,368 conserved proteins to the total of 24,990 non-redundant nuclear genes of *Arabidopsis* gives a total of some 4,500 genes, or 18% of the nuclear genes, that came from the cyanobacterial ancestor of the plastids. More than half of these are not targeted back to the plastids but to other cell compartments (including the secretory pathway) [6]. The protein products of many nuclear genes that were not acquired from the plastid ancestor are now targeted to the plastid. The genes within the nuclear genomes that originated from the plastid ancestor cover all of the functional categories defined by The *Arabidopsis* Genome Initiative [7].

The cyanobacterial ancestor of the plastids was, relative to the three cyanobacteria with completed genome sequences that were examined by Martin *et al.* [6], closer to *N. punctiforme* than to *P. marinus* or *Synechocystis* sp. Although three genomes is not a large sample size, it is of interest that *N. punctiforme* is a diazotroph, so the plastid ancestor could also have been a nitrogen-fixer. Were early plastids perhaps also able to fix atmospheric nitrogen?

Losing chloroplasts but keeping cyanobacterial genes

The work of Brinkman *et al.* [8] re-examines the processes that have led to the high proportion of proteins of a bacterial human pathogen, *Chlamydia*, that are similar to those of plants. This similarity was formerly attributed to horizontal gene transfer from plants, or plant-like host organisms, to the bacterium. Brinkman *et al.* [8] point out that such gene transfer is unlikely since all extant Chlamydiaceae are obligate intracellular parasites of animals. Instead, the analysis by Brinkman *et al.* [8] shows that the majority of the plant-like genes in *Chlamydia* are, in plant cells, targeted to the chloroplast. But the conclusion that this targeting of proteins to chloroplasts is necessarily a function of their origin from a plastid ancestor is not always sound. Furthermore, Martin *et al.* [6] did not find a much similarity between *Chlamydia* and *Arabidopsis* (see Figure 1 in [6]). Clearly, further investigation is needed.

Figure 1 illustrates the various endosymbiotic events described here. Amongst eukaryotes, the apicomplexan parasitic pathogens *Toxoplasma* and *Plasmodium* have curious cytoplasmic organelles bounded by three membranes, namely 'apicoplasts', which genome sequencing has established as *bona fide* plastids complete with a characteristic inverted repeat within the plastid genome [9]. The presence of three membranes, as is found around chloroplasts of dinoflagellates and euglenoids, betrays an ancestry from a secondary symbiosis, as does the presence of four membranes surrounding the plastids of, for example, photosynthetic heterokonts (a diverse group, some of which are algae) such as diatoms and brown algae. The function of the apicoplast is not clearly understood, but one suggestion is that it is indispensable for the synthesis of iron-sulfur proteins. The function of the residual plastid genome is even less clear, and it provides a test case for any theory for the function of organellar genes. Although *Plasmodium* has a plastid genome that some think is on the way out, trypanosomes, which are also non-photosynthetic, have no plastid or plastid genome at all, but are now clearly seen to be former euglenoids because of the remaining genes for a variety of plant-like enzymes, including sedoheptulose-1,7-bisphosphatase (otherwise found only in the Benson-Calvin cycle) [10,11].

Arabidopsis is not the only plant

The article by Martin *et al.* [6] uses chloroplast genomics to infer plastid phylogeny, as well as gene loss and gene transfer, for 16 sequenced plastid genomes. An important conclusion from this analysis is that two secondary endosymbiotic events involving a red alga are needed to explain the occurrence of plastids in cryptophytes (algae with phycobilin pigments in the thylakoid lumen rather than in particles on the thylakoid membrane as in cyanobacterial and red algae; an example is *Guillardia*) and heterokonts (the diatom *Odontella*). This contrasts with the arguments of Cavalier-Smith (recently set out in [12]) for a single endosymbiotic event, based on evidence such as the replacement of the glyceraldehyde-3-phosphate dehydrogenase gene derived from the red algal plastid with one of host origin in both cases.
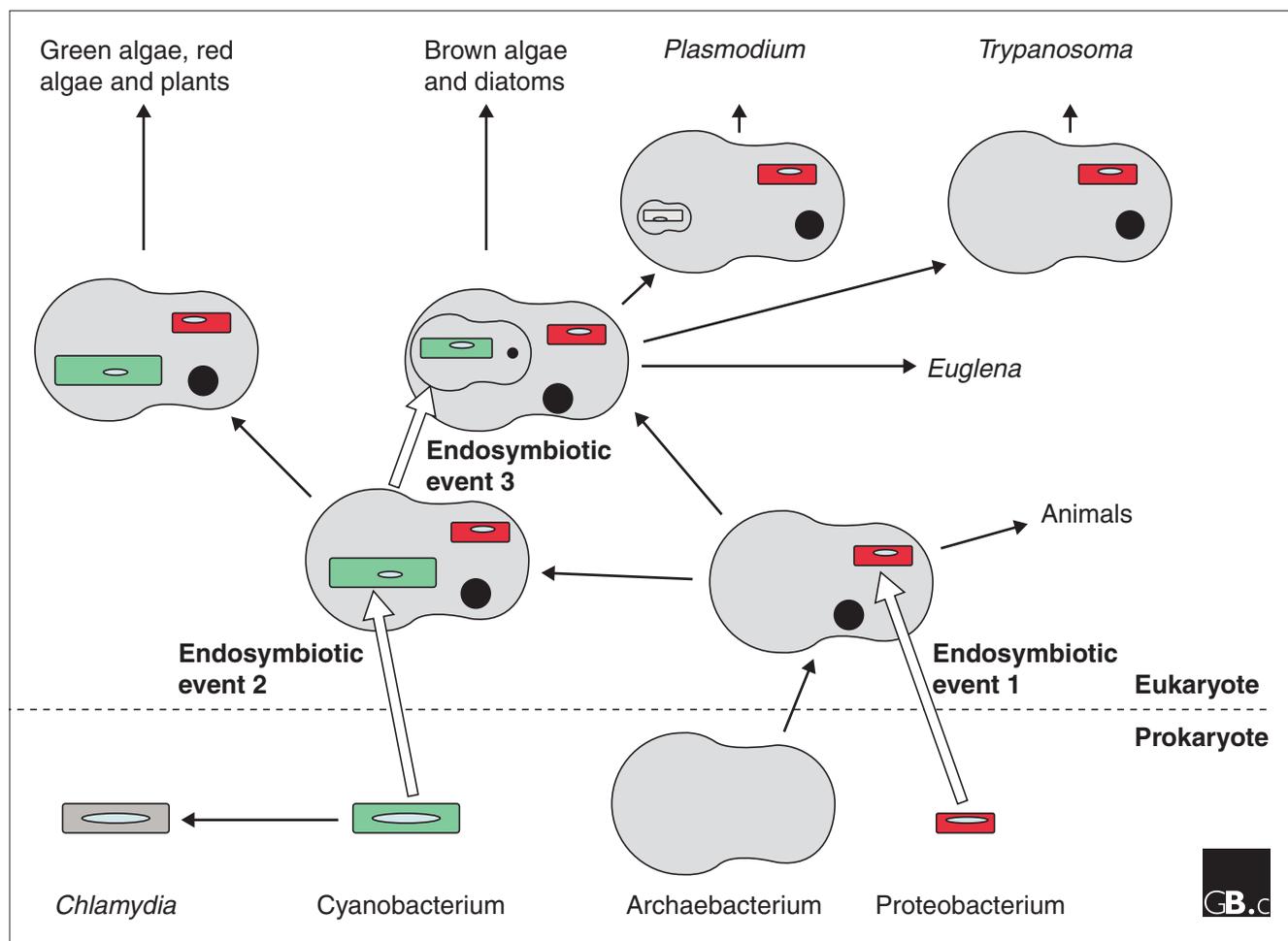
Figure 1
A schematic outline of the acquisition, reduction, and loss of genomes and compartments during evolution. Black arrows indicate evolutionary pathways; white arrows indicate endosymbiotic events in the host cell. Endosymbiotic event 1 occurred at the origin of eukaryotes. The proteobacterial endosymbiont gave rise to mitochondria (the smaller organelles in the bottom part of the diagram). Endosymbiotic event 2 occurred at the origin of plastid-containing cells. Endosymbiotic event 3 represents the secondary and higher-order endosymbioses giving rise to numerous algal phyla, as well as apicomplexans (such as *Plasmodium*) which have residual plastids, and to trypanosomes, which have no plastid at all. Black, filled circles indicate nuclei or nucleomorphs; ellipses within organelles indicate bacterially derived genomes, which may be reduced or lost completely. More than one kind of host cell and of endosymbiont is involved in the secondary, and in the higher-order, symbioses. The genome of the Archaebacterium is not represented in the diagram.

Another recent article [13] deals with genome-based phylogenies of plastids; 19 complete chloroplast genomes are studied using a new computational method, and broadly similar conclusions are reached to those of Martin and coworkers [6]. This work also allows novel functional assignments to a number of chloroplast open reading frames. The functional implications of chloroplast genomics, with special reference to experimental opportunities and 'directional genetics' in *Arabidopsis thaliana*, have recently been reviewed by Leister [14].

An important question relating to the evolution of plastid genomes in higher plants is the timing of the changes in the plastid genome in the streptophyte clade (made up of charophytes, a group of green algae or chlorophytes, plus embryophytes, or higher plants), which evolved more than 500 million years ago. From the unicellular flagellate *Mesostigma*, which is either a basal chlorophyte or lies at the split between Chlorophyta and Streptophyta, to the embryophytes, of which the liverwort *Marchantia* is the most basal to have been sequenced, the changes are gene losses, including transfer to the nucleus, scrambling of gene order, and intron insertion [15].

An important contribution to bridging the evolutionary gap between *Mesostigma* and *Marchantia* is the work of Turmel *et al.* [15] on a member of the charophytes *sensu stricto* (that is, excluding *Mesostigma*) *Chaetosphaeridium globosum*. Before the work of Turmel *et al.* [15] only fragmentary data addressed the issue of gene content and organization of the

eight charophytes *sensu stricto*. The complete plastid genome sequence of *Chaetosphaeridium globosum* [15] shows that most of the embryophyte characteristics were present in the charophyte alga, so that the major changes had occurred between the branch to *Mesostigma* and that to *Chaetosphaeridium*. The common features shared by the plastid DNA of *Chaetosphaeridium* and of embryophytes include the gene content, the intron composition, and the gene order. Thus, the *Chaetosphaeridium* chloroplast genome has 124 genes (compared to 136 in *Mesostigma* and 110-120 in embryophytes), one Group I intron (there are none in *Mesostigma* and one in embryophytes), 16 *cis*-spliced Group II introns (none in *Mesostigma* and 18-19 in embryophytes) and one *trans*-spliced Group II intron (none in *Mesostigma*, one in embryophytes). Genome size (118-155 kilobases) is relatively constant among *Mesostigma*, *Chaetosphaeridium* and higher plant plastids. By contrast, the mitochondrial genome of *Chaetosphaeridium* is closely similar to that of *Mesostigma* in terms of size (57 kb and 42 kb, respectively), gene content and, perhaps, intron content. *Chaetosphaeridium* has a much smaller genome size than the obese mitochondrial genomes of *Marchantia* (187 kb) or *Arabidopsis* (367 kb), and many more *cis*-spliced Group II introns (18-25 rather than two). The apparently different tempo of evolution in mitochondria and plastids of the charophytes deserves further investigation. An important point about the functional genomics of the plastid is the determinant of which genes essential for plastid function are retained in the plastid genome. Higher plant plastid genomes have slightly fewer genes than in the plastids of the charophytes *sensu lato* (that is, the charophytes *sensu stricto* plus *Mesostigma*).

Cells within cells

One requirement of the endosymbiont hypothesis is whole-scale gene transfer from the chloroplast to the nucleus. Long thought to be either impossible or, at best, highly problematical, its difficulties are often thought to relate to the failure of some genes to move at all. Gene transfer from chloroplast to nucleus is now estimated to occur naturally in tobacco at a frequency of one transposition in 16,000 pollen grains [16]. In natural populations and over evolutionary time, this frequency represents a massive informational onslaught and highlights the urgency of the question of why chloroplasts have genomes at all. There must be some crucial, over-riding, selective advantage in retaining certain genes in chloroplasts but not others. Evidence is now accruing for the ten-year-old proposal that genes are maintained in the chloroplast and are thus able to permit regulation of their expression so as to be regulated by the activities of their gene products in photosynthesis and electron transport [17,18].

It is clear that genomics, in the sense of whole-genome analyses, is making very important contributions to our understanding of the evolution of plastids, and is complementing, and to a significant extent supplanting, 'single gene' phylogenies. Genomics is revolutionizing our understanding of the changes involved in the primary endosymbiosis that produced the plastids of red, green and glaucophyte algae, and in the subsequent genetic changes in green (charophycean) plastids with the evolution of higher plants. Genomics is also indispensable for understanding how red and green algae yielded the plastids derived from secondary endosymbiosis.

The endosymbiont hypothesis took a long time to graduate from wild and untestable speculation to an accepted view of plastid origins and evolution. In contrast, comparative genomics has quickly elevated the kinship of chloroplasts and cyanobacteria to a keystone of our understanding of the most abundant of cells, the primary producers on which life now depends, not to mention some vicious and enterprising pathogens whose exploits are a global burden to human health. The title of this article asks what the cyanobacteria have done for plants. "What have they not done?" is a question perhaps more easily addressed.

References

1. Mereschkowsky C: **Über Natur und Ursprung der Chromatophoren im Pflanzenreiche.** *Biol Centralbl* 1905, **25:**593-604.
2. Martin W, Kowallik KV: **Annotated English translation of Mereschkowsky's 1905 paper Über Natur und Ursprung der Chromatophoren im Pflanzenreiche.** *Eur J Phycol* 1999, **34:**287-295.
3. Margulis L: *Symbiosis in Cell Evolution.* New York: WH Freeman; 1981.
4. Raven JA, Douglas AE: **Genomes at the interface between bacteria and organelles.** *Philos Trans R Soc Lond B Biol Sci* 2003, **358:**5-18.
5. Doolittle WF, Boucher Y, Nesbø CL, Douady CJ, Andersson JO, Roger AJ: **How big is the iceberg of which organellar genes in nuclear genomes are but the tip?** *Philos Trans R Soc Lond B Biol Sci* 2003, **358:**39-58.
6. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D: **Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus.** *Proc Natl Acad Sci USA* 2002, **99:**12246-12251.
7. The *Arabidopsis* Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.*** *Nature* 2000, **408:**796-815.
8. Brinkman FSL, Blanchard JL, Cherkasov A, Av-Gay Y, Brunham RC, Fernandez, RC, Finlay BB, Otto SP, Ouellette BFF, Keeling PJ, *et al*: **Evidence that plant-like genes in *Chlamydia* species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast.** *Genome Res* 2002, **12:**1159-1167.
9. Wilson RJM, Rangachari K, Saldanha JW, Rickman L, Buxton RS, Eccleston JF: **Parasite plastids: maintenance and functions.** *Philos Trans R Soc Lond B Biol Sci* 2003, **358:**155-164.
10. Hannaert V, Saavedra E, Duffieux F, Szikora J-P, Rigden DJ, Michels PAM, Opperdoes FR: **Plant-like traits associated with metabolism of *Trypanosoma* parasites.** *Proc Natl Acad Sci USA* 2003, **100:**1067-1071.
11. Martin W, Borst P: **Secondary loss of chloroplasts in trypanosomes.** *Proc Natl Acad Sci USA* 2003, **100:**765-767.

12. Cavalier-Smith T: **The phagotrophic origin of eukaryotes and phylogenetic classification of protozoa.** *Int J System Evol Microbiol* 2002, **52:**297-354.

13. De las Rivas J, Lozano JJ, Ortiz AR: **Comparative analysis of chloroplast genomes: Functional annotation, genome-based phylogeny, and deduced evolutionary patterns.** *Genome Res* 2002, **12:**567-583.

14. Leister D: **Chloroplast research in the genomic age.** *Trends Genet* 2003, **19:**47-56.

15. Turmel M, Otis C, Lemieux C: **The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants.** *Proc Natl Acad Sci USA* 2002, **99:**11275-11280.

16. Huang CY, Ayliffe MA, Timmis JN: **Direct measurement of the transfer rate of chloroplast DNA into the nucleus.** *Nature* advance online publication 5 February 2003 (doi:0.1038/nature01435).

17. Allen JF: **Why chloroplasts and mitochondria contain genomes.** *Comp Funct Genomics* 2003, **4:**31-36.

18. Allen JF: **The function of genomes in bioenergetic organelles.** *Philos Trans R Soc Lond B Biol Sci* 2003, **358:**19-38.