The tiny enslaved genome of a rhizarian alga

Thomas Cavalier-Smith*
Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom

At least twice in the history of life a predatory nonphotosynthetic protozoan cell ate a eukaryotic algal cell and enslaved it internally instead of digesting it, thereby becoming a chimeric photophagotrophic cell with two distinct nuclei and remarkably complex membrane topology—far surpassing that of animals or plants. In their descendants, the host nucleus became dominant, whereas that of the algal slave shrank by transfer of most of its genes into the main nucleus and retargeting the proteins they encode back into the enslaved chloroplast. Some descendants managed to transfer all essential genes and lose the enslaved nuclei altogether while retaining the algal chloroplast and plasma membrane (e.g., brown seaweeds), but in two groups of algae, cryptomonads (1) and chlorarachneans (2–4), the enslaved nucleus remains in the chimeric cell many millions of years afterward, raising fascinating questions for cell and evolutionary biology. How are functions shared between two evolutionarily unrelated nuclei, and how do their proteins integrate into one harmonious cell? The first inklings of answers came with the first nucleomorph genome sequence from a cryptomonad (1). In this issue of PNAS, Gilson et al. (5) report the first genome sequence of a chlorarachnean nucleomorph, the tiniest nucleus in nature.

Fig. 1 shows how these two unusual algal groups fit on the eukaryote evolutionary tree (6, 7). Cryptomonads belong to a vast branch, the chlorovolales, formed by a single enslavement of a red alga (8, 9) and containing eight phyla: two mainly photosynthetic (Ochrophyta (including brown algae, diatoms, and eight other algal classes) and Haptophyta; two with one algal class each (dinoflagellates and cryptomonads) but many heterotrophs; and four entirely nonphotosynthetic (e.g., ciliate protozoa, and Pseudofungi) (10). By contrast, chlorarachneans belong to Rhizaria, a typically amoeboid group characterized by long, thin-branching pseudopods that often anastomose as a net. Rhizaria have two phyla: Cercozoa, wherein chlorarachneans are the only algae, and Retaria, including the largely marine Foraminifera and Radiolaria, which often cultivate algae temporarily in their cells, much as corals do, but never permanently enslave them as a true chloroplast with its own protein-import machinery like in chlorarachneans. Chlorarachnea is Greek for green spider, referring to the web-like body of the archetypal genus Chlorarachnion, in which green-chloroplast-containing cells form a multicellular network by temporary fusions of thread-like pseudopods that trap diverse prey for phagocytic engulfment and digestion.

Compared with other algae, chlorarachneans are little known partly because they live largely in subtropical waters. Japanese researchers found the biodiversity of chlorarachneans to be greater than previously recognized (11) and that they include classical amoeboid forms, others with cell walls, and flagellates like Bigelovia (12) with potential as a laboratory model, being easier to grow; many display all three growth forms at different life-cycle stages.

Thus, chlorarachneans and cryptomonads represent entirely independent natural experiments in nuclear genome miniaturization that are most interesting to compare (13). What do such comparisons tell us? As expected, Bigelovia nucleomorphs have far fewer genes than those of cryptomonads; only 331 protein-coding genes compared with 464 in the cryptomonad Guillardia (1) and 512 in the archetypal genus Chlorarachnion (14). As in Guillardia, most genes are for housekeeping, merely to maintain the nucleomorph itself and the ribosomes that make its proteins. The only end-product functions of obvious direct value to the host are provided by 17 genes that encode proteins imported into the former green algal chloroplast. The 20-fold more numerous housekeeping genes are kept merely to allow expression of these 17 and are an evolutionary load, reflecting the sometimes bizarre imperfections of evolution by mutation and selection.

Organisms are not optimized, let alone designed, because selection is powerless without the requisite chance mutations. Thus, the nucleomorph of Bigelovia was retained solely because 17 genes are essential for the chloroplast, just as were the cryptomonad nucleomorphs because of the 30 chloroplast proteins they encode (1). If, by chance, copies of all of these few genes had moved into the nucleus and their products successfully retargeted into the chloroplasts through the four surrounding membranes, the nucleomorph genome would have disappeared. Such loss occurred in chromalveolates, both in the chromobiote sisters of cryptomonads and in algae (e.g., dinoflagellates and sporozoa like the malaria parasite) (10), but never in chlorarachneans.

However, euglenoid algae (members of Excavata) also have green plastids acquired from a green alga long ago. Their ancestor must once have had a nucleomorph and transferred all essential chloroplast protein genes into its own nucleus, retargeted their proteins to the chloroplast, and then lost the nucleomorph entirely. It is possible that retargeting was easier for euglenoids, in which proteins have to cross only three membranes and the targeting machinery via Golgi vesicles that fuse with the outermost membrane copes with proteins temporarily stuck in vesicle membranes by hydrophobic sequences (14). It is widely assumed that chlorarachneans

Conflict of interest statement: No conflicts declared.

See companion article on page 9566.

*E-mail: tom.cavalier-smith@zoo.ox.ac.uk.

© 2006 by The National Academy of Sciences of the USA
enslaved a green alga independently of euglenoids (4), but the possibility that this occurred once only in a hypothetical common ancestor of Rhizaria and Excavata and see heteroxenous members of both groups all lost chloroplasts (6, 8) is not yet decisively ruled out.

Deciding which history is true is important for estimating the age of chlorarachnean nucleomorphs. Judging from sequence trees (15), the last common ancestor of chlorarachnians may be only a quarter of the age of Cercozoa, but, because their lineage branches early, the first chlorarachnean could have been 70% of that age. Thus, if a single green alga were enslaved by an excavate/rhizarian ancestor, chlorarachnean nucleomorphs must be marginally older than Cercozoa, which probably go back 540 million years (My) (16). If, instead, chlorarachnians acquired plastids independently of euglenoids, their nucleomorphs are probably 135–380 My old. Cryptomonad nucleomorphs are as old as chromalveolates, probably ~570 My (16). Thus, chlorarachnean nucleomorphs are no older than those of cryptomonads and are possibly much younger, yet they did shrink more. Thus, degree of compaction is not simply related to elapsed evolutionary time, as also shown by chromalveolates that lost nucleomorphs and were no older than cryptomonads that did not.

Nucleomorph genome reduction involves three processes: transfer of gene copies to the nucleus, their acquisition of bipartite targeting signals for protein import across four membranes (17), and deletion of the original nucleomorph copy. This gene-transfer process has both chance and deterministic aspects. The role of historical accident is shown by the same as any of the 30 retained genes of *Guillardia*. Given originally ~1,000 chloroplast protein genes in the enslaved algal nuclei, shared retention of these two is probably just a coincidence. There appears no rhyme or reason why certain genes were successfully transferred and others were not (5). It is probably through historical accidents that chlorarachneans and cryptomonads failed to transfer all of their genes to the nucleus, whereas chromobiotics, alveolates, and euglenoids independently succeeded. Retention of Toc75/Tic20 but not Tic22 genes for chloroplast protein import, unlike *Guillardia*, also is likely accidental.

Determinism is evident in the way intergenic spacers have been pared down to as little as 1 nt or less: some genes overlap by up to 110 nt. These very short spacers and overlapping genes must reflect systematically greater reproductive success of deletions compared with insertions. Elimination of intergenic noncoding DNA, pseudogenes, and gene duplicates attest to strong selection for small genomes (18), but chlorarachnean nucleomorph genes riddled with introns [852 in all, 160 independently confirmed as spliced by parallel sequencing of DNA complementary to mRNAs (5)] show that such selection is mutationally limited: precise deletion is probably difficult.

Gilton et al. (5) show that intron density is essentially the same as in free-living green plants; some occur in the same position (four are even shared with *Guillardia* and must predate the red/green algal divergence), implying that 135–570 My of selection for small size failed to streamline the genome as much as a designer could have. Yet intron size was dramatically reduced to 18–21 nt (mostly 19 nt), suggesting a novel, possibly length-dependent mechanism of splicing.

Unlike in cryptomonads, the nucleomorph lost genes for tubulins and proteasomes, raising the question whether proteasome proteins are imported into the periplastid space or are dispensed with. Previously, divergent nuclear-coded tubulins were found with leaders, suggestive of targeting signals (19); thus, it is likely that they are imported and that the nucleomorph divides by a relict mitotic spindle as in *Guillardia* (1).

What remains for the future? With unpublished sequences of *Bigelowiella* chloroplast and mitochondrial genomes completed in Canada and Japan, we greatly need nuclear genome sequences for *Bigelowiella* and *Guillardia* to understand how four genomes cooperate to make such complex chimeric cells. Rhizaria are the only major group in the living world with no complete nuclear genome sequences. For fuller understanding of the cell biology and lifestyles of these unusual organisms, which may be the most abundant predators on Earth as Cercozoa dominate the soil and Retaria dominate marine sediments and the oligotrophic ocean vastnesses, we need genome sequences from divergent heterotrophic Rhizaria, especially Cercozoa (far easier to cultivate). A final answer to the conundrum whether chlorarachnean and euglenoid plastids were enslaved in a common ancestor or not may depend on elucidating the chlorarachnean protein-targeting systems.

**Degree of compaction is not simply related to elapsed evolutionary time.**