

Testing a hypothesis about timing of intracellular “enslavement” of cells of two algal groups

Jotun Hein and Tom Cavalier-Smith 3.2.11

Motivation and Background

Chromista is a major eukaryotic kingdom comprising algae and former protozoa that is evolutionarily entirely distinct from the kingdoms Plantae and Protozoa (Cavalier-Smith 1981, 2007). Chromist chloroplasts were acquired secondarily by enslavement of a red alga, itself a member of kingdom Plantae, and have a unique membrane topology (Cavalier-Smith 2002, 2003, 2004). Chromista originally included only three predominantly algal groups: Heterokonta (e.g. brown seaweeds, diatoms, chrysophytes), Haptophyta, Cryptomonada. Unlike plants these have chlorophyll c-containing plastid(s) lying within an extra (periplastid) membrane (the relic of the red algal plasma membrane) inside the lumen of the rough endoplasmic reticulum (RER typically within the perinuclear cisterna); chromist chloroplasts were acquired secondarily by enslavement of a red alga, itself a member of kingdom Plantae, and their unique membrane topology is a consequence of this enslavement and the preservation of both host and slave membranes (Cavalier-Smith 2003). In one case (cryptomonads) even the nucleus of the enslaved red algae remains and has the smallest known eukaryote genome (Douglas 2003). A related group of former protozoa, Alveolata, includes the often-photosynthetic chlorophyll-c containing dinoflagellates (which have lost the periplastid membrane), the mostly non-photosynthetic and mostly parasitic apicomplexans (e.g. the malaria parasite) which often have relict colourless plastids (some rare free-living relatives, *Chromera*, are still photosynthetic). Alveolates are now widely believed to have acquired their plastids by the same red algal enslavement event as the chromists (Cavalier-Smith 1999) and as they nest phylogenetically within the Chromista they are now also placed in the kingdom Chromista (Cavalier-Smith 2010; formerly they and the classical chromists were collectively informally called chromalveolates). Another huge group of former protozoa, the Rhizaria (Cercozoa, Foraminifera, Radiozoa), which do not have chloroplasts of red algal origin, has been shown to be phylogenetically nested within Chromista (Burki et al. 2009); therefore Rhizaria also were moved into the Chromista (Cavalier-Smith 2010), which therefore now includes several groups with no visible trace of the ancestral chromist red algal chloroplast (notably ciliates, Rhizaria, and the oomycetes and other Pseudofungi), which was presumably lost early in their evolution. Some of these colourless groups, notably the oomycetes, have been shown still to retain relict algal genes.

The problem

Moustafa et al. (2009) provided evidence that perhaps over 400 genes in the diatoms (two with full genomes available) came not from a red alga but instead from a green alga (also Plantae) and proposed that there may therefore have been a second cryptic and previously unrecognized temporary enslavement of a green algae as well as a red alga early in chromist evolution. Cavalier-Smith (2010) proposed that the green chloroplast and relict green algal nucleus that is still present in one subgroup of Cercozoa, the chlorarachnean algae, may therefore be an evolutionary relic of the very same cryptic symbiosis/enslavement postulated by Moustafa et al. This can be tested by determining whether essentially the same set of genes is present in the nuclear genome of the chlorarachniophyte *Bigeloviella* and whether they are phylogenetically closer to those of diatoms than to green algal outgroups. A second test of the theory of a second (green) algal enslavement in early chromist evolution would be to determine whether many of these genes are also present in the nuclear genome of the cryptomonad *Guillardia theta*, which according to multigene trees diverged even earlier than diatoms and Cercozoa (Burki et al. 2009). This idea of two separate incorporations of different eukaryotic cells into the ancestral chromist is evolutionarily so important that it ought to be tested more thoroughly than it has been. A key question is whether the single-gene trees used by Moustafa et al. (only two of which have been published) are really adequate for distinguishing the phylogenetic origin in green algae (and even a specific subgroup) versus red algae.

The project

The *Bigeloviella* and *Guillardia theta* genomes were recently completely sequenced by JGI; the former will almost certainly be publically available before the project starts and the latter might also be. Similar automated methods to those of Moustafa et al. should be applied to these two genomes and to any other chromistan algal genomes that become public by then to identify genes of likely red or green algal origin in

order to obtain a much more comprehensive data set and construct large numbers of trees semiautomatically.

In addition more extensive phylogenetic analyses and statistical tests on the robustness of the trees and reliability of inferring source genomes would be done on a large sample of chromist (a) red alga-like and (b) green-alga like genes, including both those targeting to the chloroplast and those that are not. The latter would be searched for in a wide variety of non-photosynthetic chromists that putatively had algal ancestors to see how many can still be traced, as well as in more distant eukaryotic outgroups as controls, and semi-automated phylogenetic analysis done to identify genes that would repay a more thorough phylogenetic and statistical analysis of their phylogeny.

The methodologies needed for this project is phylogeny inference (Felsenstein, 2003), sequence/genome alignment, database search and detection of horizontal transfers. Much of this can be done with available packages, but there is also scope for new models and implementations.

The Data – When Moustafa *et al.* did their study they started by Blasting the genomes of two diatoms against a local database composed of the latest release of RefSeq (release 34), the genome of the red alga *Cyanidioschyzon merolae*, additional heterokonts, and green algal genomes available from JGI, and partial EST data for organism without complete genomes such as dinoflagellates and cryptophytes. Our project would use the latest RefSeq (currently release 45) and update the local database by substituting complete genomes (e.g. for the haptophyte *Emiliania*) for EST data, and use not just two but 6 or 7 chromist genomes for the initial BLAST for selecting genes for making trees (by adding a third diatom *Fragillariopsis*, the haptophyte *Emiliania huxleyi*, the heterokont *Aureococcus anophagefferens*, and the cercozoan chromist *Bigelowiella natans* and if publically available then the cryptomonad *Guillardia theta*). If the red algal *Porphyra* genome, completed but not yet available for download, becomes available in time, it will be very important to add it also to the local data base.

Project Plan – This would initially follow the same procedure as in Moustafa *et al.* The major differences would be threefold: (1) We would use more chromist query genomes and more genomes in the target database and thus address the problems with several major chromist groups, not just diatoms, and get taxonomically more comprehensive and thus potentially more reliable and meaningful trees. (2) We would examine individual trees more critically and apply statistical tests to determine for how many of the genes do they really have the resolution to show whether the genes in question came (a) from red or from green algae and (b) if the latter whether they came specifically from prasinophytes. (3) Look more critically at the results to see if we can work out whether all the genes of purported green origins are likely to have entered chromists by a single symbiogenetic event, and especially whether such genes in chromists with red algal chloroplasts have the same origin as those of *Bigelowiella* with a green algal chloroplast.

Expertise needed – the ideal constellation clearly is a biologist, statistician and computer scientist. These three expertises should be present in the three students collaborating on this project. The project is motivated in a biological question, so there is also scope exploring interests in organismal biology. The project scales well into a DPhil at many levels as the increased availability of genomes will allow many classic questions to be investigated and the many genomes will pose many and interesting methodological challenges.

References

- Burki F, Inagaki Y, Bråte J, Archibald JM, Keeling PJ, Cavalier-Smith T, Sakaguchi M, Hashimoto T, Horak A, Kumar S, Klaveness D, Jakobsen KS, Pawlowski J, Shalchian-Tabrizi K (2009) Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, Telonemia and Centroheliozoa, are related to photosynthetic chromalveolates. *Genome Biol Evol* 1: 231-238
- Cavalier-Smith T (1981) Eukaryote kingdoms: seven or nine? *BioSystems* 14: 461-481
- Cavalier-Smith T (1999) Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryotic family tree. *J Euk Microbiol* 46: 347-366
- Cavalier-Smith T (2002) Chloroplast evolution: secondary symbiogenesis and multiple losses. *Curr Biol* 12: R62-4
- Cavalier-Smith T (2003) Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). *Phil Trans Roy Soc Lond B* 358: 109-134
- Cavalier-Smith T (2004) Chromalveolate diversity and cell megaevolution: interplay of membranes, genomes and cytoskeleton. In: Hirt RP, Horner DS (eds) *Organelles, genomes and eukaryote phylogeny*. CRC Press, London, pp 75-108
- Cavalier-Smith T (2007) Evolution and relationships of algae: major branches of the tree of life. In: Brodie J, Lewis J (eds) *Unravelling the Algae*. CRC Press, Boca Raton, pp 21-55
- Cavalier-Smith T (2010) Kingdoms Protozoa and Chromista and the eozoan root of eukaryotes. *Biol Lett* 6: 342-345
- Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu X, Reith M, Cavalier-Smith T, Maier UG (2001) The highly reduced genome of an enslaved algal nucleus. *Nature* 410: 1091-1096

Felsenstein, J (2003) *Inferring Phylogenies* Sinauer

Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324: 1724-1726