Genome complexity in a lean, mean photosynthetic machine

John M. Archibald*

Canadian Institute for Advanced Research, Program in Evolutionary Biology, and Department of Biochemistry and Molecular Biology, Dalhousie University, Sir Charles Tupper Medical Building, 5850 College Street, Halifax, NS, Canada B3H 1X5

Photosynthetic organisms come in all shapes and sizes. From a human perspective, the trees and plants of dry land are the most conspicuous examples, but the next time you admire a colorful tulip or marvel at the girth of a giant Sequoia, consider the following: Approximately half of the oxygen we breathe is generated by single-celled photosynthesizers, phytoplankton, adrift in the world's oceans, invisible to the naked eye and unfathomably large in number, quietly harnessing solar energy, fixing carbon dioxide, and producing oxygen by the bucket-load. In this issue of PNAS, Derelle *et al.* (1) present the complete genome sequence of the smallest of the small eukaryotic (nucleus-containing) phytoplankton, *Ostreococcus tauri*. This organism is best known for its diminutive cell size, about that of a typical bacterium. Its genome is equally remarkable for its small size and extreme compactness. However, the *O. tauri* genome is also unexpectedly complex and provides a fascinating glimpse into the genetic makeup and metabolic potential of the smallest known eukaryote at the base of the marine food chain.

Oxygenic photosynthesis first evolved in the ancestors of modern-day cyanobacteria. In terms of sheer numbers, these organisms dominate the ocean (2), but from the perspective of primary productivity, eukaryotic algae are considered more significant. Marine diatoms, for example, produce up to 40% of the organic carbon generated in the ocean each year (3) and represent just one of the abundant and well studied algal lineages in the sea. Least understood of all eukaryotic phytoplankton are those with a diameter of <2–3 μm, the so-called "picoeukaryotes." The first descriptions of bacterial-sized eukaryotes date back more than 40 years (e.g., ref. 4), but it is only with the application of flow cytometry (2) and molecular approaches (5) to the study of marine microbes that we have begun to grasp the extent of their abundance and diversity.

O. tauri is perhaps the most famous of all picoeukaryotes and, together with its close relatives, has become the focus of concerted efforts to understand the global distribution and ecological significance of eukaryotic picoplankton (e.g., refs. 6–8). *O. tauri* was first discovered in 1994 in France's Thau lagoon, a shallow offshoot of the Mediterranean Sea known for its oyster farming. Barely 1 μm in diameter and practically invisible under the light microscope, *O. tauri* was detected by flow cytometry and hailed as the "smallest eukaryotic organism" (9). It also proved to be shockingly simple in its ultrastructure: *O. tauri* cells lack flagella and a cell wall and contain one mitochondrion, one chloroplast, a single Golgi apparatus, and a nucleus containing a single nuclear pore (Fig. 1) (10). Molecular data (11, 12) indicate that *O. tauri* belongs to a group of green algae called prasinophytes, a lineage thought to be of key importance in elucidating the earliest events in the evolution of chlorophyll *b*-containing organisms. *O. tauri* appears to be ubiquitous in coastal waters and in the open ocean (e.g., refs. 6, 8, and 12), and its minimal cell structure and high growth rate have made it a promising model picoeukaryote.

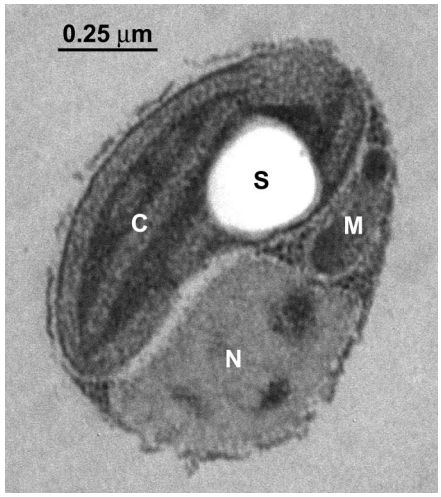

Fig. 1. Transmission electron micrograph of *O. tauri* strain OTH95, modified and reproduced with permission from Hervé Moreau (Université Paris, Paris, France). C, chloroplast; S, starch granule; M, mitochondrion; N, nucleus.

Preliminary molecular investigations pegged the *O. tauri* genome at well under 15 megabase pairs (Mbp) (11), and, like most model organisms these days, *O. tauri* quickly became the focus of a genome project (13). The complete genome sequence presented by Derelle *et al.* (1) weighs in at 12.56 Mbp, making it among the smallest, although not the smallest, nuclear genome of a free-living eukaryote characterized thus far [that honor belongs to the 9.2-Mbp genome of the fungus *Ashbya gossypii* (14)]. The *O. tauri* genome is composed of 20 linear chromosomes between 1.07 and 0.16 Mbp (1) and, given its small size, is remarkable in the number of genes it encodes: 8,166 protein-coding genes are predicted (6,265 by similarity to known genes), far more than in the ≈16-Mbp genome of the red alga *Cyanidioschyzon merolae* (5,331 genes) (15) or in the ≈12-Mbp genome of the laboratory yeast *Saccharomyces cerevisiae* (6,563 genes) (16). With a mean intergenic distance of only 197 bp, an average intron size of 103 bp, and multiple gene fusions, the *O. tauri* genome appears to be the product of intense genome compaction. One wonders to what extent the complexities of transcription initiation and termination have been affected.

In terms of structure, the most unusual feature of the *O. tauri* genome is its heterogeneity. The genome as a whole has a G+C content of ≈58%, but chromosome 19 and approximately half of chromosome 2 differ significantly from this average (54% and 52% G+C, respectively) and contain 77% of the 417 transposable elements encoded in the genome (1). Genes encoded in the low G+C portion of chromosome 2 also exhibit a different codon usage pattern than genes elsewhere in the genome, and they possess smaller and more compositionally biased introns. From a phylogenetic perspective, 43% of the genes on chromosome 2 are most similar to green algal homologs, which is a similar proportion to that seen for the other chromosomes (excluding chromosome 19). Therefore, despite its anomalous composition and structure, there is no evidence that the low G+C region of chromosome 2 is of exogenous origin. Derelle *et al.* (1) raise the possibility that it is a sex chromosome, citing the

Conflict of interest statement: No conflicts declared.

See companion article on page 11647.

*E-mail: jmarchib@dal.ca.

fact that bona fide sex chromosomes in other organisms are often similarly riddled with transposable elements (17). Meiosis has never been observed in *O. tauri*, although the presence of a near-complete set of meiotic genes encoded in its genome (1) suggests that sex is at least a possibility.

The composition of chromosome 19 is even more intriguing. More than 60% of the predicted protein genes on this chromosome have no similarity to known genes, and, of those that do, only 18% are demonstrably green algal in origin. The bulk of the remaining genes appear most similar to bacterial homologs (although modestly so), and a significant fraction of the non-green algal genes are predicted to encode surface membrane proteins and proteins involved in glycoconjugate synthesis (1). Based on these observations, the authors hypothesize that chromosome 19 has a different evolutionary history than the rest of the genome (1). Given that *O. tauri* lacks a cell wall and is susceptible to grazing in nature (8), it is tempting to speculate that the cell surface genes on chromosome 19 were acquired by lateral gene transfer and selected for as an adaptation to predation. If this is true, however, how these genes were acquired, and from where, remains a mystery.

What does the *O. tauri* genome reveal about the cell biology and metabolism of this tiny organism? The genome possesses complete or nearly complete gene sets for proteins involved in cell division, starch metabolism, and nitrogen assimilation, as well as a diverse set of transcription factors and proteins with putative kinase- and calcium-binding domains (1). As expected, a complete suite of enzymes essential for carbon fixation and the Calvin cycle are present, as is a complex gene family encoding prasinophyte-specific light-harvesting antenna proteins. Most unexpected is the presence of genes implicated in C_4 photosynthesis. This process has evolved repeatedly in higher plants as an adaptation to environmental stress (e.g., drought and low CO_2 concentrations) and involves modifications to leaf structure and altered biochemistry (reviewed in ref. 18). The existence of bona fide C_4 photosynthesis in phytoplankton is controversial. The morphological transformations that occur in plants are obviously impossible for a microbe, but an intracellular C_4 cycle has been documented

The most unusual feature of the *Ostreococcus tauri* genome is its heterogeneity.

in several plants, including the aquatic monocot *Hydrilla verticillata* (19). *O. tauri* appears to possess the right combination of enzymes in the right cellular locations to drive such a process, including a putatively cytosolic phospho*enol*pyruvate carboxylase and at least one chloroplast-targeted NADP-dependent malic enzyme. Much experimentation will be required to determine whether C_4 photosynthesis actually occurs in the tiny cells of *O. tauri*, but it is significant that its genome does not encode any obvious "carbon-concentrating mechanism" genes comparable with those in *Chlamydomonas* (1). If C_4 photosynthesis does exist, it is not difficult to imagine the competitive advantage it would bestow on *O. tauri* cells under conditions of high cell density and low CO_2 levels.

In summary, given its small size, the *O. tauri* genome packs plenty of surprises. However, as is so often the case in comparative genomics, the biological significance of many of its interesting features will be fully revealed only by comparison to closely related genomes. It is therefore significant that the strain sequenced by Derelle *et al.* (1) (OTH95, the original Thau Lagoon isolate) is the first of a trio of complete *Ostreococcus* genomes soon to be available. The Joint Genome Institute (www.jgi.doe.gov) has already sequenced the genome of a Californian surface-isolated strain (CCE9901; see ref. 8) and is now sequencing a "low-light" strain from the Atlantic Ocean (RCC141; see ref. 7). Recent work by Moreau, Vaulot, and colleagues (7) has revealed that these and other *Ostreococcus* strains constitute different "ecotypes" with distinct growth patterns, karyotypes, and pigment compositions. As was the case for the cyanobacterium *Prochlorococcus* (20), a comparison of the genomic differences between ecologically distinct *Ostreococcus* strains should greatly improve our understanding of the genetic determinants of niche adaptation in oceanic picoplankton communities. Despite its size, the ease with which *Ostreococcus* (and other prasinophytes) can be cultured and studied in the laboratory makes it a promising target for even more ambitious attempts to study the diversity and evolution of eukaryotic picoplankton using the combined strengths of oceanography, microbial ecology, and comparative genomics.

1. Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A. Z., Robbens, S., Partensky, F., Degroeve, S., Echeynié, S., Cooke, R., *et al.* (2006) *Proc. Natl. Acad. Sci. USA* **103,** 11647–11652.
2. Li, W. K. W. (1995) *Mar. Ecol. Prog. Ser.* **122,** 1–8.
3. Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A. & Queguiner, B. (1995) *Global Biogeochem. Cycles* **9,** 359–372.
4. Butcher, R. W. (1952) *J. Mar. Biol. Assoc. U.K.* **31,** 175–191.
5. Moreira, D. & López-García, P. (2002) *Trends Microbiol.* **10,** 31–38.
6. Countway, P. D. & Caron, D. A. (2006) *App. Env. Microbiol.* **72,** 2496–2506.
7. Rodríguez, F., Derelle, E., Guillou, L., Le Gall, F., Vaulot, D. & Moreau, H. (2005) *Environ. Microbiol.* **7,** 853–859.
8. Worden, A. Z., Nolan, J. K. & Palenik, B. (2004) *Limnol. Oceanogr.* **49,** 168–179.
9. Courties, C., Vaquer, A., Troussellier, M. & Lautier, J. (1994) *Nature* **370,** 255.
10. Chrétiennot-Dinet, M.-J., Courties, C., Vaquer, A., Neveux, J., Claustre, H., Lautier, J. & Machado, M. C. (1995) *Phycologia* **34,** 285–292.
11. Courties, C., Perasso, R., Chrétiennot-Dinet, M.-J., Gouy, M., Guillou, L. & Troussellier, M. (1998) *J. Phycol.* **34,** 844–849.
12. Guillou, L., Eikrem, W., Chrétiennot-Dinet, M. J., Le Gall, F., Massana, R., Romari, K., Pedrós-Alió, C. & Vaulot, D. (2004) *Protist* **155,** 193–214.
13. Derelle, E., Ferraz, C., Lagoda, P., Eychenié, S., Cooke, R., Regad, F., Sabau, X., Courties, C., Delseny, M., Demaille, J., *et al.* (2002) *J. Phycol.* **38,** 1150–1156.
14. Dietrich, F. S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S., *et al.* (2004) *Science* **304,** 304–307.
15. 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) *Nature* **428,** 653–657.
16. Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., *et al.* (1996) *Science* **274, 546,** 563–567.
17. Fraser, J. A. & Heitman, J. (2004) *Mol. Microbiol.* **51,** 299–306.
18. Sage, R. F. (2004) *N. Phytol.* **161,** 341–370.
19. Rao, S. K., Magnin, N. C., Reiskind, J. B. & Bowes, G. (2002) *Plant Physiol.* **130,** 876–886.
20. Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., *et al.* (2003) *Nature* **424,** 1042–1047.