Minireview

Diatom genomes come of age
Assaf Vardi, Kimberlee Thamatrakoln, Kay D Bidle and Paul G Falkowski

Address: Environmental Biophysics and Molecular Ecology Program, Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ 08901, USA.

Correspondence: Paul G Falkowski. Email: falko@marine.rutgers.edu

Published: 2 January 2009
Genome Biology 2008, 9:245 (doi:10.1186/gb-2008-9-12-245)
The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2008/9/12/245
© 2008 BioMed Central Ltd

Abstract

The results of two published genome sequences from marine diatoms provide basic insights into how these remarkable organisms evolved to become one of the most successful groups of eukaryotic algae in the contemporary ocean.

Diatoms are one of the most successful clades of eukaryotic, single-celled photosynthetic organisms in the contemporary ocean [1]. Their hallmark feature is an ornate, siliceous cell wall (Figure 1). Diatoms often form extensive blooms in temperate and boreal seas. Their productivity supports most of the world's fisheries and their fossilized remains are the major source of petroleum.

Diatoms are secondary symbionts, derived from the engulfment by a heterotrophic eukaryote host cell of a red alga, which then became integrated as a plastid [2]. Although their chromalveolate ancestor probably arose over a billion years ago [3], long before evidence of animal life, the first diatoms do not appear in the fossil record until about 146 million years ago and rose to ecological prominence only about 35 million years ago. Two major clades of diatoms are distinguished by 'body' plans: a radially symmetrical 'centric' form (Figure 2), which is ancestral to a bilaterally symmetrical 'pennate' form (Figure 1). Together, these two groups comprise about 20,000 morphological species [4], although it is believed, on the basis of molecular genetic analyses, that there are over 100,000 cryptic species [5]. In an effort to elucidate how diatoms evolved and rose to ecological prominence, the genomes of two species have been completely sequenced at the Joint Genome Institute: Thalassiosira pseudonana (Figure 2), a centric species [6], and Phaeodactylum tricornutum (Figure 1), a distantly related, recently evolved pennate species [7]. Although these two species diverged over 90 million years ago, about 60% of their genome is shared. Here we briefly review what the genomic analyses have revealed so far. Several other diatom genome sequences are in the pipeline; these include the psychrophilic diatom Fragilariaopsis cylindrus, which is common in polar seas and sea ice, and Pseudo-nitzschia multiseries, which produces the neurotoxin domoic acid.

Basic genome structure and modes of evolution
The vegetative cells of diatoms are diploid, and the genomes are relatively large, containing approximately 30 megabases with 10,000-12,000 predicted genes. Approximately 95% of the DNA is non-coding (Table 1). Diatoms are one of the most rapidly evolving eukaryotic taxa on Earth [8]. The rapid tempo of evolution is suggested to be due to a high proportion of long terminal repeat (LTR) retrotransposons and other transposable elements as well as insertion/deletion mutation (indels). The prevalence of transcripts from LTR retrotransposons in several diatom expressed sequence tag (EST) libraries [9] is hypothesized to be related to their possible role in adaptation to stress conditions, especially nutrient limitation (Maumus F, Allen AE, Jabbari K, Vardi A, Bowler C, unpublished observations). The P. tricornutum genome contains over 50% of the introns found in T. pseudonana, whereas the latter shares less than 10% of conserved intron positions present in the chromalveolate ancestor. Moreover, the evolution of indels in T. pseudonana appears to be extremely rapid and follows a logistic rate that is proportional to genome size [8,10]. Unlike in multicellular plants, however [11], large scale duplication events do not seem to have a pivotal role in the evolution of diatom
genomes, as shown by the similar numbers of genes in the two species (Table 1).

A second, more surprising source of genetic variability is horizontal gene transfer (HGT). Phylogenetic analysis of *P. tricornutum* suggests that about 5% of the genome (587 genes) is derived from bacterial orthologs; more than half of these are shared with *T. pseudonana*, implying that they were acquired by diatoms early in their evolutionary history and perform essential functions [6,7]. In particular, several genes of prokaryotic origin seem to have been recruited for metabolism of organic carbon and nitrogen, including genes involved in a urea cycle that probably evolved in the primordial heterotrophic host cell before acquisition of the secondary symbiont. The mechanism of HGT in diatoms is not understood. Viral infection is one obvious pathway; indeed, several viruses, including single-stranded RNA and single- and double-stranded DNA types, have been isolated that target specific diatoms [12]. Virally mediated HGT can be inferred from the gene encoding a putative photoreceptor, phytochrome, that is clustered in the *P. tricornutum* genome with two viruses that infect brown algae [13]. Other mechanisms proposed to facilitate acquisition of bacterial genes by HGT include phagotrophy and association with organelles or with intracellular endosymbionts or parasites. Furthermore, 22 genes in the diatoms are of chlamydial origin [14]; these genes were hypothesized to be derived from an ancient endosymbiosis event between chlamydiae and the ancestor of primary photosynthetic eukaryotes [15].

**Core metabolic pathways**

In the ocean, essential nutrients such as nitrate, phosphate and silicate are brought up to the surface from the interior by wind-driven mixing (for example, storms) or deep convection. Diatoms assimilate these nutrients very rapidly in excess of their immediate growth demands, storing the
Table 1

| Comparison of the genome properties of *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* genomes* |
|--------------------------------------------------|
| **Thalassiosira pseudonana** | **Phaeodactylum tricornutum** |
| Genome size (Mb) | 32.4 | 27.4 |
| Predicted genes | 11,776 | 10,402 |
| Introns | 17,880 | 8,169 |
| Number of chromosomes | 24 | 33 |
| G+C content | About 48% | About 47% |
| Percentage of genome that is non-coding | About 97% | About 94% |
| ESTs in GenBank | 61,913 | 133,871 |

*Data from [6,7].

The primary mode of nutrition in diatoms is oxygenic photosynthesis. Although the core machinery for this process is highly conserved, it has been known since the mid-1970s that the affinity of diatoms for inorganic carbon is considerably higher than that of their primary carbon-fixing enzyme, Rubisco, for CO$_2$, suggesting that diatoms must concentrate inorganic carbon in their cells [18]. The genome sequences [6,7] have revealed the unique nature of nitrogen cycling in diatoms: a catabolic urea cycle has been identified involving ornithine and citruline and potentially yielding urea and subsequently ammonia from hydrolysis of the substrate by urease. However, diatoms do not excrete urea and subsequently ammonia from hydrolysis of the oval morphotype contains a lightly silicified valve [24]. Genome sequencing [7] revealed genes for four silicon transporters in *P. tricornutum*, all with strong support from ESTs, but only one silaffin-like protein.

The formation of the silicate-based cell wall in diatoms is one of the most interesting areas of research. Silicic acid is translocated across the plasma membrane via specific transporters and is subsequently conveyed to a silica deposition vesicle, a slightly acidic environment in which the new cell wall is completely formed before it is exported by exocytosis. Silaffins and long-chain polyamines have a role in the polymerization of silica, but the mechanism of pattern formation remains unknown. Genome analysis reveals that *T. pseudonana* contains three silicon transporters [22] and three silaffin genes [23]. *P. tricornutum*, however, is an atypical diatom in that it does not have an obligate requirement of silicon for growth and exists as three distinct morphotypes: oval, triradiate and fusiform (Figure 1a). Only the oval morphotype contains a lightly silicified valve [24] and is the only diatom reported to take up the anionic form of silicon (silicate, or Si(OH)$_4$), rather than the more commonly transported form, orthosilicic acid (Si(OH)$_4$) [25]. Genome sequencing [7] revealed genes for four silicon transporters in *P. tricornutum*, all with strong support from ESTs, but only one silaffin-like protein.

Iron limits primary production in three major areas of the ocean: the eastern equatorial Pacific, the subarctic Pacific and the Southern Ocean. So far, 11 iron enrichment experiments have been conducted in the open ocean, covering all three environments; these involve adding iron to the sea in order to stimulate growth of phytoplankton. In all experiments, the first major group of organisms to grow following fertilization was pennate diatoms. One major factor in the ability of diatoms to take advantage of the nutrient enrichment is the vacuole, a sort of 'food pantry', which does not, as yet, have a clear genetic marker. However, analysis of the *T. pseudonana* and *P. tricornutum* genomes has revealed the presence of several Fe acquisition and storage genes in *P. tricornutum* that are absent from *T. pseudonana*. Iron acquisition in *T. pseudonana* seems to work through a ferroxidase/permease pathway for Fe(II) uptake. In contrast, *P. tricornutum* may acquire iron through a cell-surface reductase. A recent discovery of iron storage ferritin in bloom-forming pennate diatoms contributes to their success in chronically low-iron oceanic regions [26]. More data are required before we can be sure that this strategy...
can explain the success of pennate diatoms specifically in low Fe environments.

**Signaling and regulation**

Diatoms use sophisticated mechanisms to monitor and adapt appropriately to changes in environmental stress conditions [27,28]. The mosaic multi-lineage nature of the diatom genomes predicts interesting signaling pathways that are similar to features not only of plants and animals but also of prokaryotes. Both diatom genomes contain a bacterially derived two-component system composed of a novel domain organization of histidine kinase (sensor) and response regulator (transcriptional activators) [6,7]. Calcium and nitric oxide were recently shown to act as important second messengers in diatom perception and transduction of stress conditions. A novel calcium-regulated protein, induced by nitric oxide (NO) and regulating cell death, has also been identified [29]. Furthermore, a diatom alternative oxidase contains a calcium-binding EF-hand domain that is induced under iron starvation [30]. Genetic manipulation of a chloroplast-localized protein PtNOA in diatoms has revealed the interplay between sensing chemicals cues (infochemicals), oxidative stress and cell death through NO-based signaling [31].

One of the more enigmatic aspects of evolution of protists is the emergence of programmed cell death pathways. *T. pseudonana* has homologs to key components of programmed cell death biochemical machinery, including metacaspases, HtrA-family proteases, apoptosis-associated nuclear factors of the E2F and DP1 families, cell death suppressor proteins and a cellular apoptosis susceptibility protein [13]. Diatom genomes contain five to six metacaspases, some of which are constitutively expressed whereas others are induced by nutrient deprivation [6,7,32]. However, *T. pseudonana* lacks homologs of important elements of metazoan apoptotic pathways, such as p53 and the Bcl-2 family of apoptosis regulators, as well as TIR adaptor proteins and AP-ATPases, both of which are abundant in *Arabidopsis thaliana*. These findings raise fundamental questions about whether *T. pseudonana* has a functional programmed cell death pathway in response to iron starvation [32], raising fundamental questions about how it is regulated.

**Functional genomics and biotechnological applications**

Genetic transformation methods have been established for both *T. pseudonana* and *P. tricornutum* [33,34]. Expression vectors for *T. pseudonana* have been developed that allow constitutive and inducible protein expression [34]. Also recently developed was a method for growing cultures of *T. pseudonana* synchronously [35], making it possible to gain insights into cell division and other metabolic processes tightly coupled to the cell cycle, such as silification. A useful tool for reverse functional genomics has also been developed in *P. tricornutum* that allows high-throughput cloning and expression of a target gene [36]. These tools are important advances that will enable insights into the molecular mechanisms of diatom biology that were not possible even 5 years ago. However, the field still lacks classical genetic techniques, such as a method for gene knockout. Perhaps with the newly available genome sequence and the growing interest in diatom genomics, these tools will soon become available.

Diatoms have inspired many biologists and engineers. Silicon-based nanotechnology is a multi-billion-dollar industry, but there is an increasing need for the efficient and cost-effective production of such devices, for example, in solar energy capture, charge separation in battery technologies, or even in separation technologies involving purification of gases or solutes in fluids. Diatoms provide an unparalleled system for studying the basic mechanism of silica nanofabrication because they can make complex, reproducible three-dimensional structures under ambient conditions. In addition, because diatoms have been such an important component of petroleum, potential genetic manipulation may lead to more efficient use of these organisms as biofuel feedstock. Indeed, the development of a model organism such as *P. tricornutum*, combined with system-level approaches for better understanding of how carbon is allocated to specific sinks, may ultimately provide a source of advanced, sustainable biofuels that does not compete with food production.

Diatom genomes are coming of age. These protists, long studied by marine biologists for their complexity and ecological success, are now becoming a source of information not only about the evolutionary history of eukaryotes, but as a potential source of nanodevices and energy for our future. These genome sequences [6,7] are the beginning of a long learning process that will potentially teach us how the complex web of metabolic processes was selected by specific clades and how we can use that information to develop a sustainable world in the coming centuries.

**References**

1. Smietacek V: Diatoms and the ocean carbon cycle. Progist 1999, 150: 25-32.
2. Delwiche C: Tracing the thread of plastid diversity through the tapestry of life. An Nat 1999, 154:164-S177.
3. Cavalier-Smith T: Chloroplasts: secondary symbiosis and multiple losses. Curr Biol 2002, 12:R62-R64.
4. Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJ: The evolution of modern eukaryotic phytoplankton. Science 2004, 305:354-365.
5. Kooistra WHF, Gerssonde R, Medlin LS, Mann DG: The origin and evolution of the diatoms: their adaptation to a planktonic existence. In Evolution of Primary Producers in the Sea. Edited by Falkowski PG, Knoll AH. New York: Academic Press; 2007: 207-249.
6. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Beckner M, Brzezinski MA, Chai BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Heilstein U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WW, Lane TW, Larimer FW, Lippmeier JC, Lucas S: The genome diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 2004, 306:79-86.
