

# CORRELATED EVOLUTION OF GENOME SIZE AND CELL VOLUME IN DIATOMS (BACILLARIOPHYCEAE)

*Jessica A. Connolly*

*Matthew J. Oliver*

*Jeremy M. Beaulieu, Charles A. Knight*

*Lars Tomanek and Mark A. Moline*

**A correlation between genome size and cell volume has been observed across diverse assemblages of eukaryotes. We examined this relationship in diatoms (Bacillariophyceae), a phylum in which cell volume is of critical ecological and biogeochemical importance. In addition to testing whether there is a predictive relationship across extant species, we tested whether evolutionary divergences in genome size were correlated with evolutionary divergences in cell size (using independent contrasts). We estimated total DNA content for 16 diatom species using a flow cytometer and estimated cell volumes using critical dimensions with scaling equations. Our independent contrast analyses indicated a significant correlated evolution between genome size and cell volume. We then explored the evolutionary and ecological implications of this evolutionary relationship. Diatom cell volume is an important component of the global carbon cycle; therefore, understanding the mechanisms that drive diatom genome evolution has both evolutionary and ecological importance.**

**Key index words:** Bacillariophyceae; cellular volume; comparative methods; correlated evolution; diatom evolution; genome size variation

**Abbreviations:** CEN, chicken erythrocyte nuclei; CRBC, chicken red blood cells; SMA, standardized major axis

The correlation between cell volume and genome size has been observed in many lineages of eukaryotes, including animals (Mirsky and Ris 1951, Gregory 2001), plants (Bennett 1971), and a variety of

unicells (Shuter et al. 1983, Veldhuis et al. 1997). Proposed mechanistic theories for existence of this relationship are varied and invoke a variety of linking factors, ranging from simple space constraints to potential osmolyte function of nucleotides and optimal transport of RNA from the nucleus to the cell (Cavalier-Smith 1978). The existence of this relationship in diatoms is of particular interest because of the prominent role cell volume plays in diatom ecology. In marine diatoms, cell volume is related to metabolic rate (Finkel 2001, Raven and Kubler 2002), intrinsic growth rate (Tang 1995), photosynthetic capacity (Finkel 2001), respiratory rate (Tang and Peters 1995, Sarthou et al. 2005), and sexual reproduction (Round et al. 1990). Additionally, cell volume is important in nutrient acquisition (Irwin et al. 2006), light absorption (Agusti 1991), and primary production in the marine environment (Joint and Pomroy 1988, Sarthou et al. 2005). Diatom cell volume also has a significant effect on the efficiency of grazing by herbivores (Smetacek et al. 2004). On global biogeochemical scales, a significant portion of annual phytoplankton organic carbon is produced by diatoms, which account for 45% of total new primary production in the oceans (Field et al. 1998). Their requirement for silica in the construction of their frustules marks them as major players in the global biogeochemical silica cycle (Treguer et al. 1995, Yool and Tyrrell 2003). Export of diatom biomass to the deep ocean, which is driven by both their cell volume and their siliceous frustules, accounts for 50% of the organic carbon that is exported to the ocean interior (Falkowski et al. 1998, 2003). Therefore, understanding the forces and selection pressure that govern cell volume in diatoms has ramifications for how we understand their ecological and biogeochemical importance (Finkel et al. 2005).

Given the variety of physiological and ecological factors that are correlated with diatom cell volume, it seems almost certain that cell volume of diatoms is determined by multiple sources of selection, ranging from grazing (Smetacek et al. 2004) to nutrient status of the upper ocean (Finkel et al. 2005). To explain the variety of trait correlations to cell volume, it is often implied that these correlations are a result of the correlated evolution of some trait to cell volume. Here we examine the possibility that diatom genome size is positively associated with diatom cell volume, and we explore the implications of this relationship in an organism of such global importance.

Most cross-species bivariate correlation and regression tests explicitly assume that each measurement of a given trait, such as cell volume or genome size, is independent. However, this is rarely the case. Any two species in a phylogeny share some degree of evolutionary history, thus violating the assumption of independence (Felsenstein 1985). For example, sister species are likely to have similar traits because they have diverged from a common ancestor. Therefore, it is possible that trait-to-trait correlations, such as the cell volume–genome size correlation, could be an artifact of the phylogenetic history of the species in the analysis and not correlated evolution.

In this study, our approach was threefold. First, we reexamined the relationship between cell volume and genome size using modern techniques to determine the present-day trait association. Second, we generated a diatom phylogeny, inferring evolutionary relationships from nuclear 18S rDNA sequences. Third, using this phylogeny to identify evolutionary relationships between 16 species, we tested for correlated evolution between genome size and cell volume using phylogenetically independent contrasts (Felsenstein 1985). Briefly, we determined that cell volume and genome size clearly show an evolutionary correlation in diatoms, and we discuss the implications of this relationship with respect to the evolution and ecology of diatoms.

#### MATERIALS AND METHODS

**Total DNA content measurements.** We selected a total of 16 diatom species representing a broad range of taxa for total DNA content analysis ( $\text{pg DNA} \cdot \text{cell}^{-1}$ ) based on their taxonomic diversity and culture availability. Cultures were obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP, Boothbay Harbor, ME, USA) and maintained in f/2 media (Guillard and Ryther 1962, Guillard 1975) at a 12:12 light:dark (L:D) at the recommended growth temperature. Fifteen milliliters of each diatom strain was harvested in their log growth phase to maintain an accurate comparison of total DNA content. Nuclear DNA content of selected diatom species was measured using flow cytometry. A CyFlow SL (Partec Headquarters, Münster, Germany) was used to measure fluorescence at 535 nm, from base-pair-specific staining of the double stranded nuclear DNA molecule using SYTOX orange. SYTOX orange (S11366; Invitrogen Corp., Carlsbad, CA, USA) is a nucleic-acid-specific dye characterized by excitation and emission spectra of 547 nm and 570 nm,

respectively, and was chosen to minimize interference with the emission spectra of chl. We developed a working solution by diluting the stock solution (5 mM) 100-fold with Milli-Q water (Millipore, Billerica, MA, USA, A10); this yielded a 50  $\mu\text{M}$  concentration (Veldhuis et al. 1997), which we stored at  $-20^\circ\text{C}$  for a maximum of 3 weeks. To ensure that chl autofluorescence would not interfere with the emission signal detected by the flow cytometer, cultures were centrifuged (1,000g for 4 min) and extracted in 10 mL of ethanol for 48 h at  $4^\circ\text{C}$  before staining (LaJeunesse et al. 2005). Samples were then washed three times and resuspended in a 1 $\times$  PBS buffer solution immediately before analysis. To minimize variation in the nuclear DNA signal, cells were treated with RNase (0.2  $\mu\text{g} \cdot \text{mL}^{-1}$ , R-6513; Sigma Inc., St. Louis, MO, USA) at  $37^\circ\text{C}$  for 30 min (Veldhuis et al. 1997) and stained with 100  $\mu\text{M}$  of a working solution of SYTOX orange for a minimum of 30 min. Chicken erythrocyte nuclei (CEN; Biosure, Grass Valley, CA, USA) with single nuclei and doublet and triplet peaks were used as a performance standard to establish consistent linearity of the flow cytometer. Chicken red blood cells (CRBC; Biosure) were used as an internal DNA standard and were subjected to the same staining protocol as the diatom cultures. Cellular DNA content of whole cells (G1/0 peak) was compared to the internal standard (CRBC) value of 3.0  $\text{pg DNA} \cdot \text{cell}^{-1}$  (Johnston et al. 1999) and expressed as CRBC units. Nuclear DNA content was inferred from a linear relationship detected by the instrument and calculated relative to the CRBC measurement (Fig. 1). Samples were run in triplicate, and average DNA content was used to measure genome size.

Stained diatom cells were visually inspected to confirm that staining was restricted to nuclear DNA. We took a two-dimensional image capture of *Chaetoceros muelleri* (see Table 1 for taxonomic authors) cells that were chl extracted and subjected to the staining protocol described above. Images were acquired using an Olympus IX81 microscope (Center Valley, PA, USA) and a digital camera (Orca ER AG, Hamamatsu, Shizuoka, Japan). Cell fluorescence was detected using a Tritc filter and Slidebook imaging software (Olympus). These images validated our staining protocol and confirmed that detected fluorescence localized in the nucleus and was a result of nucleic-acid-specific staining of nuclear DNA (Fig. 2, a and b). Additionally, the known genome sizes of *Thalassiosira pseudonana* (Armbrust et al. 2004) and *Phaeodactylum tricoratum* (Scala et al. 2002, Joint Genome Institute, U.S. Department of Energy 2006) were included in the data set. All estimates of diatom genome size are direct measurements of whole-cell DNA content, and use of the term genome size throughout this article refers to total DNA content of the diatom cell.

**Cell volume measurements.** Classically, phytoplankton cell volume has been calculated as a cylinder or as assumed ellipsoid volume. However, many diatom genera deviate significantly from these simple shapes. We accounted for the extensive variation in cell shape by calculating final cell volume specific to each diatom genera. Biovolume ( $\mu\text{m}^3$ ) for each species was estimated using geometric calculations specific to the shape of the genus according to Hillebrand et al. 1999. All cell volume measurements include the vacuole and were estimated from pure cultures using light microscopy and a stage micrometer. These measurements were supplemented with reported length and width measurements from CCMP. Multiple estimates were taken for each cell axis, and means were used to calculate cell biovolume.

**Constructing a diatom phylogeny.** We obtained 148 diatom nuclear 18S rDNA sequences from the National Center for Biotechnology Information (NCBI; Bethesda MD, USA). We aligned the sequences using a ClustalW multiple alignment (Thompson et al. 1994). The parameters used for the pair-wise

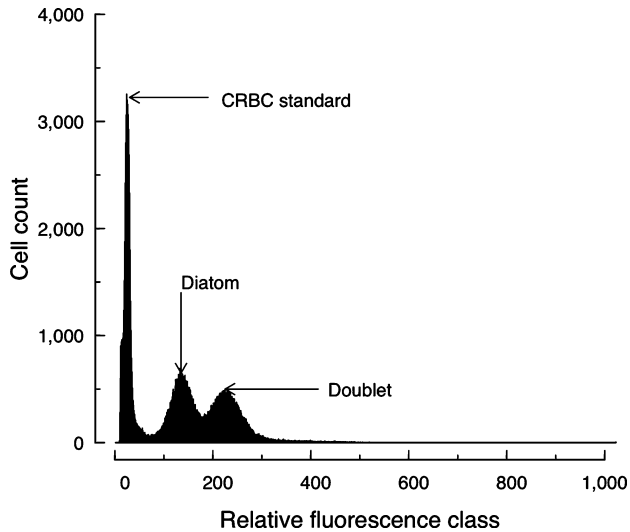


FIG. 1. Cyflow SL output displaying *Thalassiosira weissflogii* and the internal DNA standard, chicken red blood cells (CRBC). Peaks represent nuclear DNA content detected by relative fluorescence of the base-pair-specific stain. The doublet represents dividing diatom cells.

alignment were a gap opening penalty of 10.00 and a gap extension penalty of 0.10. The parameters for the multiple alignment were a gap opening penalty of 10.00 and a gap extension penalty of 0.20 (Hall 2004). Sequences were then aligned by hand, eliminating truncated sequences and nonhomologous regions (Hall 2004).

We constructed a phylogenetic tree including 148 diatom species from aligned sequences using a general-time-reversal (GTR) model of evolution implemented into a maximum-likelihood (ML) analysis (Medlin and Kaczmarek 2004) using PHYML (Guindon and Gascuel 2003). The proportion of invariable sites and gamma distribution parameters were estimated by the program, and eight substitution rate categories were implemented. We obtained bootstrap values from 1,000 replicates of the data set.

**Statistical analyses.** The statistical framework we applied consisted of both conventional regression and independent contrasts analyses (see below). In all cases, trait values were  $\log_{10}$  transformed to fit the assumption of normality.

**Regression:** We determined the cross-species relationship between genome size and cell volume using conventional regression statistics. However, because we provide first estimates of both genome size and cell volume, there may be inherent error in both the independent and dependent variables, respectively. This error may influence the regression slope. To deal with this problem, we fitted a standardized major axis (SMA, or model II regression) slope to the cross-species data (nonphylogenetically corrected data, see below). SMA is useful because it minimizes the residual variance in both dependent and independent variables rather than just the dependent variable, which is customary when using ordinary least-squares regression models (Warton et al. 2006). We used the freeware (S)MATR (Falster et al. 2006) to obtain the coefficient of determination ( $R^2$ ) and an SMA slope estimate for the cross-species data.

**Independent contrasts:** We converted cross-species data into statistically independent data points using phylogenetic independent contrasts. First, we calculated the difference (termed here a divergence) in trait values among sister species all the way to the root of the tree. Subsequently, we calculated the difference in the internal node averages among sister species. In the presence of a fully bifurcating tree, the method results in  $N-1$  contrasts, where  $N$  is the number of input species values. Assuming Brownian motion evolution, the magnitude of trait differences among sister species should be proportional to time (Garland 1992), where more distantly related species will tend to have larger contrasts. Therefore, each trait divergence is divided by the square root of the sum of the branch lengths (the standard deviation) arising from the most common hypothetical ancestor to effectively “standardize” each contrast. This calculation also ensures statistically independent data, drawn from a normal distribution with equal variances, which can be analyzed using conventional statistics (Felsenstein 1985, Garland et al. 1992).

For the 16 species with both genome size and cell volume estimates, we derived a subset tree from our complete diatom phylogeny using a tree-pruning algorithm in Phylocom (Webb et al. 2006). We then used the analysis of traits module (AOT, Ackerly 2006) in Phylocom to calculate contrasts from extant species and internal node averages standardized by branch length information (see above). First, we performed a standardization test (sensu Garland et al. 1992) to assess whether our branch lengths derived from the nuclear 18S rDNA sequences adequately standardized the contrasts. A significant correlation between the absolute

TABLE 1. Total DNA content (pg) and cell biovolume ( $\mu\text{m}^3$ ) for the 16 species used to compute phylogenetically independent contrasts.

Species	Culture source	pg DNA	Biovolume ( $\mu\text{m}^3$ )
<i>Achnanthes brevipes</i> C. Agardh	CCMP100	6.000	290.990
<i>Amphipora paludosa</i> VanHeurck	CCMP125	1.695	176.715
<i>Chaetoceros muelleri</i> Lemmerm.	CCMP1316	0.475	117.800
<i>Coscinodiscus</i> sp.	CCMP1814	50.000	45410.740
<i>Cyclotella meneghiniana</i> Kütz.	CCMP338	12.860	678.580
<i>Ditylum brightwellii</i> (T. West) Grunow	CCMP358	12.900	3750.000
<i>Lithodesmium undulatum</i> Ehrenb.	CCMP472	43.100	54161.250
<i>Minutocellus polymorphus</i> (Hargraves et Guillard) Hasle, Von Stosch et Syversten	CCMP497	1.900	106.029
<i>Navicula pelliculosa</i> (Bréb. et Kütz.) Hilse	CCMP543	0.071	19.635
<i>Nitzschia frustulum</i> (Kütz.) Grunow	CCMP558	0.770	62.500
<i>Pauliella taeniata</i> (Grunow) Round et Basson	CCMP1115	6.730	593.760
<i>Phaeodactylum tricorutum</i> Bohlin	CCMP1327	0.030	61.261
<i>Pseudo-nitzschia multiseriis</i> (Halse) Hasle	T. Peterson	16.490	468.000
<i>Skeletonema japonicum</i> (Grev.) Cleve	CCMP1281	0.336	137.440
<i>Thalassiosira pseudonana</i> (Hustedt) Hasle et Heimdal	CCMP1335	0.035	88.357
<i>Thalassiosira weissflogii</i> (Grunow) Fryxell et Hasle	CCMP1049	17.250	4555.309

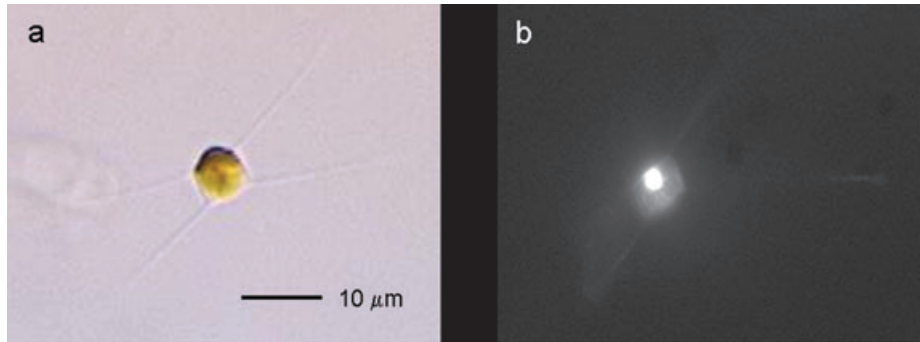


FIG. 2. (a) Live *Chaetoceros muelleri* cell imaged at Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP1316). (b) Two-dimensional image capture of a *C. muelleri* cell after staining with SYTOX orange nucleic acid specific stain.

value of the standardized contrasts and the standard deviation (derived from the branch lengths, see above) indicates that the contrasts are not properly standardized and different branch length transformations may be needed (Garland et al. 1992). It should be noted, however, that results of independent contrasts are generally robust even when deviations from a Brownian motion model of evolution occur (Diaz-Uriate and Garland 1996).

A significant relationship between the contrasts would indicate correlated evolution between the two traits, where divergences in genome size are associated with divergences in cell volume. Since the direction of subtraction in an independent contrast analysis is arbitrary, reversing the direction of subtraction would result in a contrast of the opposite sign. This property gives the mean value of zero to all contrasts. Therefore, we performed a regression analysis forced through the origin (Garland et al. 1992). Because the contrasts were derived from variables that contained errors (see the “Regression” section), we applied a model II regression forced through the origin in (S)MATR (Falster et al. 2006) to obtain the SMA slope estimate and  $R^2$ . We also analyzed the independent contrast relationship using a sign test to determine the proportion of contrasts with a positive sign (i.e., positive divergence in genome size associated with a positive divergence in cell volume). Sign-test significance was determined using critical values of the binomial distribution (Ackerly 2000).

## RESULTS

The ML sequence analysis resulted in a phylogenetic tree including 148 diatom species (accession numbers and nexus output file are available from the author by request). This larger phylogenetic tree was pruned using the “sample prune” function in Phylocom (Webb et al. 2006), resulting in a tree with conserved topology and reoptimized genetic distances for the 16 species for which both genome size and cell volume were measured (Fig. 3). This pruned phylogeny was compared to the recently published diatom phylogeny by Alverson et al. (2006), and a similar topology was observed for the 16 species. This phylogeny was then used to analyze the evolutionary relationship between genome size and cell volume using phylogenetically independent contrasts.

We first considered the present-day cross-species relationship between genome size and cell volume

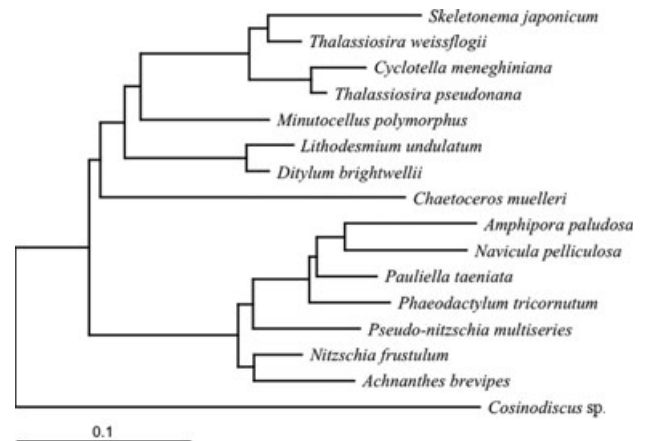


FIG. 3. 18S rDNA gene phylogeny of 16 diatom species used to test hypotheses of correlated evolution between genome size and cell volume. This tree was pruned from a larger tree of 148 diatom species inferred using the maximum-likelihood algorithm.

among 16 species of diatoms. We determined that genome size was significant and positively associated with cell volume across the 16 species sampled (Fig. 4a;  $R^2 = 0.688$ , slope = 0.951,  $P < 0.001$ ).

Results from our independent contrast analysis, which incorporated the phylogeny (Fig. 3), indicated that divergences in genome size were significant and positively associated with divergences in cell volume (Fig. 4b;  $R^2 = 0.673$ , SMA slope = 0.676,  $P < 0.001$ ). In addition, 14 out of the 15 contrasts were positively related. The greater proportion of positively related contrasts was significant when compared with the binomial expectation (Fig. 4b;  $P < 0.001$ ). For both variables, the branch lengths derived from the 18S rDNA sequences adequately standardized the contrasts (genome size:  $r = -0.491$ ,  $P = 0.063$ ; cell volume:  $r = -0.319$ ,  $P = 0.245$ ).

## DISCUSSION

We observed a significant positive relationship between genome size and cell volume across extant diatom species (Fig. 4a). These data

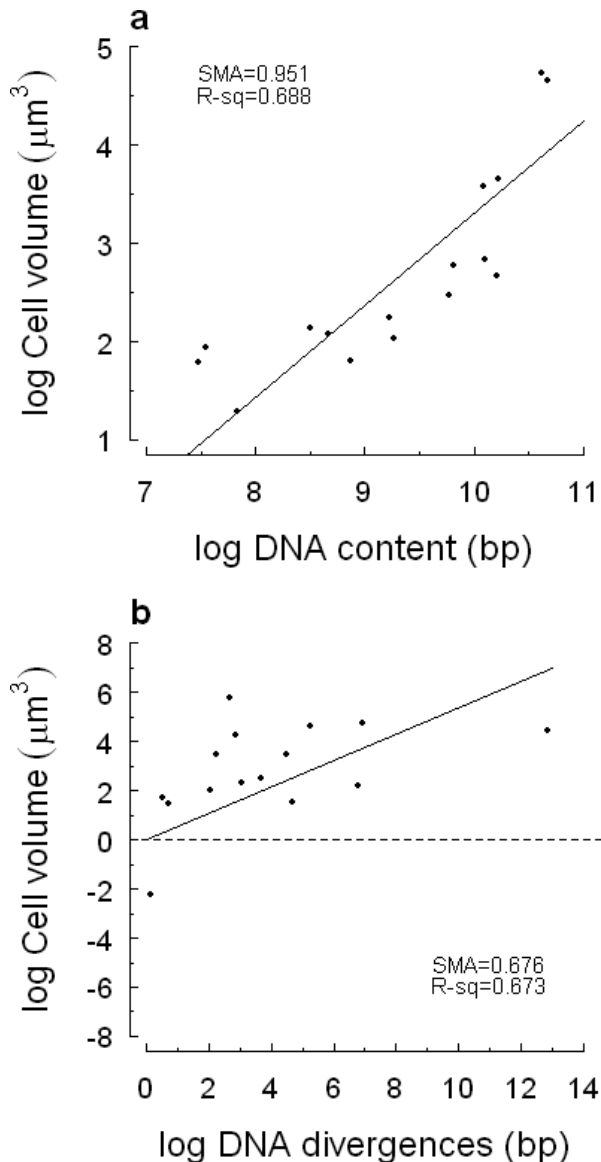


FIG. 4. (a) Standardized major axis (SMA) regression of log cell volume ( $\mu\text{m}^3$ ) and log DNA content (bp) displaying a positive linear relationship across 16 extant diatom species. (b) SMA regression analysis of independent contrasts forced through the origin, showing correlated evolution of log cell volume ( $\mu\text{m}^3$ ) and log DNA content (bp). Independent contrasts account for the phylogenetic signal by rendering each data point independent. Resulting values represent contrasts between sister taxa or nodes embedded within the phylogeny. This method of trait comparison incorporates evolutionary relationships between species, resulting in a test for correlated divergences between traits. R-sq, coefficient of determination ( $R^2$ ).

represent the present-day association of these two traits. The scaling exponent close to 1 confirms the predictive relationship between changes in genome size and changes in cell volume. These results are consistent with earlier surveys of unicellular eukaryotes, which also reported a positive relationship between total DNA content and cell volume (Shuter et al. 1983).

However, our initial regression tests did not account for the evolutionary relationships between species and therefore do not accurately encapsulate the evolutionary relationship between cell volume and genome size. Traditionally, trait correlations across species assume statistical independence. However, closely related species may resemble each other more than distantly related species because of their shared ancestry; thus, these trait values cannot be considered statistically independent in cross-species analyses. For this reason, Felsenstein (1985) developed the method of independent contrasts that transforms cross-species data with phylogenetic information, so that species data more closely approximate independence. These independent contrast results give insight into the evolutionary relationship between traits and can be thought of as a test for correlated divergences between traits. The regression results can be directly compared with results of the independent contrasts and provide a more complete view of the evolutionary relationship between two traits. For example, significant cross-species relationships may have nonsignificant independent contrast results if deeper divergences (i.e., higher taxonomic level divergences) weight the overall relationship (Felsenstein 1985, Moles et al. 2005, Beaulieu et al. 2007).

After incorporating the phylogeny of our diatom species using the independent contrast method, we determined that divergences in genome size have been significant and positively associated with divergences in cell volume (Fig. 4b). Moreover, in 13 out of 14 instances where genome size increased, cell volume also increased. Only in one instance did cell volume decrease with increasing genome size. These results show that divergences in genome size and cell volume are coupled, which strengthens the hypothesis that there is a mechanistic link between these traits in diatoms. Our data do indicate the direction of causality, whether genome size causes cell volume to increase or whether increases in cell volume allow for further expansion of the genome. However, in the remainder of this article, we explore both possibilities and draw on examples and theories that attempt to explain this relationship and discuss the possible ecological and evolutionary consequences of this clear evolutionary link in diatoms.

*Implications of diatom genome size driving changes in cell volume.* There are a few examples of genome size changing cell volume in eukaryotes. One of these examples is in the development of cells in the flowering plant *Arabidopsis thaliana* (Brassicaceae), where the rapid increase in DNA within a cell through the process of endoreduplication immediately leads to a manifold increase in cell volume. While the molecular mechanism behind this rapid growth in cell volume is not well understood, it appears that *Arabidopsis* can control cell size simply by increasing genome size (Melaragno et al. 1993).

Cavalier-Smith (1985) produced a more general theory of cell volume control by genome size with his nucleoskeletal theory. This theory postulates that genome size has a function beyond encoding genetic information; it also determines the size of the nuclear envelope, which is set mechanistically and is dependent on how much DNA it surrounds (Cavalier-Smith 2005). Cavalier-Smith (1985) postulates that all eukaryotes select for a karyoplasmic ratio (the ratio between nuclear envelope volume and cell volume) that allows for optimal transfer of RNA from the nucleus to the surrounding cytoplasm. Therefore, understanding the factors that control genome size may ultimately expose one of the many factors controlling cell volume.

Genome size reflects the balance of DNA insertions and deletions in a genome, including random errors from DNA polymerase (Albertini et al. 1982), polyploidy (Soltis and Soltis 1999), and the activity of mobile genetic elements, such as retrotransposons (Kazazian 2004). It is well established that genome size variation in plants is commonly due to episodes of polyploidy (Soltis and Soltis 1999). In addition, recent evidence suggests that across a taxonomically broad data set of angiosperms, there is a strong positively correlated evolution between genome size and cell size. Furthermore, it was shown that polyploidy inflated the minimum rate of genome size evolution, and this was exhibited by cellular traits (J. Beaulieu and C. Knight, unpublished data). Currently, little is known about the formation of polyploids in diatoms. However, polyploidy has been reported in some species (Mann 1994, Chepurinov et al. 2002), which suggests that this may be an important process in diatom genome evolution as well.

In plants, it has also been shown that increases in genome size are accompanied by a dramatic proliferation of transposable elements (Kidwell 2002). The discovery of such mobile genetic elements in the maize genome by McClintock (1950) challenged the view that genomes were stable entities. The self-replicating activity and mobility of retrotransposons fit well with the “selfish gene” hypothesis (Orgel and Crick 1980). The inference of this theory is that over evolutionary time, retrotransposons have “selfishly” reproduced themselves and slowly inundated genomes. Because of their self-replicating activity, retrotransposons can act as a population, thus having the potential to modulate significantly an organism’s genomic structure. For this reason, they are considered major drivers of genome evolution (Kazazian 2004) and may represent one possible mode of genome size change in organisms.

Retrotransposons have been detected in the diatom *T. pseudonanna* and, based on their sequence structure, appear to be active (Armbrust et al. 2004). Additionally, one of the attributes of many retrotransposon families is that they are induced by environmental stress (Wilke et al. 1992, Hirochika

et al. 1996, Grandbastien 1998, Kalendar et al. 2000, Kimura et al. 2001). There is good evidence that these elements are also activated by nutrient stress in diatoms (A. Allen and C. Bowler, unpublished data; M. J. Oliver, unpublished data). Given these studies, there is mounting evidence that genome restructuring through the activities of retrotransposons in response to stress is not a rare occurrence, but an active evolutionary response to local environments that may also be regulating genome size and thus cell volume in diatoms through currently unknown pathways.

As significant contributors to global primary production, increases in diatom cell volume as a result of DNA accumulation could have ecosystem consequences as large cells are responsible for the majority of exported carbon (Tremblay et al. 1997, Laws et al. 2000, Le Borgne et al. 2002). Although strategies to minimize sinking have proved effective in sustaining phytoplankton populations across evolutionary timescales, the settling of growing and senescent cells remains a contributing factor to the annual export of organic carbon to the ocean interior (Field et al. 1998). An increase in sinking of cells with larger cell volumes because of their larger genome sizes would suggest an increase in total cell carbon export as well (Veldhuis et al. 1997). In this way, random DNA accumulation processes within a cell, such as retrotransposition or formation of polyploids, may have ecosystem consequences.

*Implications of diatom cell volume driving changes in genome size.* As mentioned earlier, the volume of a cell is clearly related to many essential physiological parameters (see introduction), and changes in diatom cell volume are of particular importance because of the significant role diatoms play in global biogeochemical cycles. Because of the tight link between diatom cell volume and their ecology, environmental factors, such as nutrient acquisition (Irwin et al. 2006), light absorption (Agusti 1991), primary production (Joint and Pomroy 1988, Sarthou et al. 2005), and grazing by herbivores (Smetacek et al. 2004) may all play a role in governing cell volume. If cell volume primarily drives genome size, then any environmental factors that influence cell volume would also affect the evolution of genome size.

One of the major findings of the eukaryotic genome sequencing projects is that despite large variations in genome size, eukaryotes basically all have  $10^4$  protein coding genes (Walbot and Petrov 2001). The remainder of the genome is made up of introns, repeats, and mobile elements (transposons and retrotransposons). The function of this remainder or “junk” DNA is still in question (Gregory 2001 and references therein). However, what is clear from these sequencing efforts is that the quantity of introns and mobile elements is not an unpredictable, taxa-specific quantity but is strongly correlated to genome size (Lynch and Conery 2003). Therefore, one potential

explanation for changes in diatom genome size due to changes in cell volume could be the activity and abundance of mobile elements within the diatom genome. Environmental forces that select for cell volumes may then also incorporate the potential deleterious (Doolittle and Sapienza 1980, Orgel and Crick 1980) or beneficial (McClintock 1984, Takeda et al. 1999) effects of mobile elements, offering diatoms genetic solutions to a changing environment.

In summary, this study is the first analysis of correlated evolution between genome size and cell volume using phylogenetically independent contrasts in diatoms. We demonstrated an evolutionary association between cell volume and genome size in diatoms and explored some potential ecological and evolutionary implications of this relationship. Because of the prominent role cell volume plays in diatom ecology and in biogeochemistry, the evolutionary relationship between genome size and cellular volume suggests that changes in cell volume could impact the genomic size in diatoms and indirectly diatom function in ocean ecosystems.

We thank Marcel Veldhuis, Francis Villablanca, Jason Morgan, David Jackson, and French Morgan for very helpful comments and contributions regarding methodology. We would also like to thank three anonymous reviewers for their time and dedication in reviewing this manuscript. We also thank the California Polytechnic State University student fee committee for partial funding of this project.

Ackerly, D. D. 2000. Taxon sampling, correlated evolution, and independent contrasts. *Evolution* 54:1480–92.

Ackerly, D. D. 2006. *Analysis of Traits (AOT): A Module of Phylocom*. Version 3.1. <http://www.phylodiversity.net/phylocom> (accessed on January 5, 2006).

Agustí, S. 1991. Allometric scaling of light absorption and scattering by phytoplankton cells. *Can. J. Fish. Aquat. Sci.* 48:763–7.

Albertini, A. M., Hofer, M., Calos, M. P. & Miller, J. H. 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell* 29:319–28.

Alverson, A. J., Cannone, J. J., Gutell, R. R. & Theriot, E. C. 2006. The evolution of elongate shape in diatoms. *J. Phycol.* 42:655–68.

Armbrust, V. E., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S., et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306:79–86.

Beaulieu, J. M., Moles, A. T., Leitch, I. J., Bennett, M. D., Dickie, J. B. & Knight, C. A. 2007. Correlated evolution of genome size and seed mass. *New Phytol.* 173:422–37.

Bennett, M. D. 1971. The duration of meiosis. *Proc. R. Soc. Lond Ser. B Biol. Sci.* 178:277–99.

Cavalier-Smith, T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate and the solution of the DNA C-value paradox. *J. Cell Sci.* 34:247–78.

Cavalier-Smith, T. 1985. Eukaryote gene numbers, non-coding DNA and genome size. In Cavalier-Smith, T. [Ed.] *The Evolution of Genome Size*. Wiley, Chichester, UK, pp. 253–76.

Cavalier-Smith, T. 2005. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Ann. Bot.* 95:147–75.

Chepurinov, V. A., Mann, D. G., Vyverman, W., Sabbe, K. & Danielidis, D. 2002. Sexual reproduction, mating system, and

protoplast dynamics of *Seminavis* (Bacillariophyceae). *J. Phycol.* 38:1004–19.

Diaz-Uriate, R. & Garland, T., Jr. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Syst. Biol.* 45:27–47.

Doolittle, W. F. & Sapienza, C. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284:601–3.

Falkowski, P., Barber, R. & Smetacek, V. 1998. Biogeochemical controls and feedbacks on ocean primary production. *Science* 281:200–5.

Falkowski, P. G., Laws, E. A., Barbar, R. T. & Murray, J. W. 2003. Phytoplankton and their role in primary, new and export production. In Fasham, M. J. R. [Ed.] *Ocean Biochemistry. The Role of the Ocean Carbon Cycle in Global Change*. Springer-Verlag, Berlin, pp. 99–121.

Falster, D. S., Warton, D. I. & Wright, I. J. 2006. *SMATR: Standardised Major Axis Tests and Routines*. Version 2.0. Available at: <http://www.bio.mq.edu.au/ecology/SMATR> (accessed January 2, 2007).

Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.

Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281:237–40.

Finkel, Z. V. 2001. Light absorption and size scaling of the light-limited metabolism in marine diatoms. *Limnol. Oceanogr.* 46:86–94.

Finkel, Z. V., Katz, M. E., Wright, J. D., Schofield, O. M. E. & Falkowski, P. G. 2005. Climatically driven macroevolutionary patterns in the size of marine diatoms over the Cenozoic. *Proc. Natl. Acad. Sci. U. S. A.* 102:8927–32.

Garland, T., Jr. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am. Nat.* 140:509–19.

Garland, T., Jr., Harvey, P. H. & Ives, A. R. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* 42:265–92.

Grandbastien, M. 1998. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* 3:181–7.

Gregory, T. R. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol. Rev. Camb. Philos. Soc.* 76:65–101.

Guillard, R. R. L. 1975. Culture of phytoplankton for feeding marine invertebrates. In Smith, W. L. & Chanley, M. H. [Eds.] *Culture of Marine Invertebrate Animals*. Plenum Press, New York, pp. 26–60.

Guillard, R. R. L. & Ryther, J. H. 1962. Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt and *Detonula confervacea* Cleve. *Can. J. Microbiol.* 8:229–39.

Guindon, S. & Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696–704.

Hall, B. G. 2004. *Phylogenetic Trees Made Easy: A How-To Manual*. 3rd ed. Sinauer Associates Inc., Sunderland, Massachusetts, 244 pp.

Hillebrand, H., Dürselen, C., Kirschtel, D., Pollinger, U. & Zohary, T. 1999. Biovolume calculation for pelagic and benthic microalgae. *J. Phycol.* 35:403–24.

Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H. & Kanda, M. 1996. Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl. Acad. Sci. U. S. A.* 93:7783–8.

Irwin, A. J., Finkel, Z. V., Schofield, O. M. E. & Falkowski, P. G. 2006. Scaling-up from nutrient physiology to the size-structure of phytoplankton communities. *J. Plankton Res.* 28:459–71.

Johnston, J. S., Bennett, M. D., Rayburn, A. L., Galbraith, D. W. & Price, J. H. 1999. Reference standards for the determination of DNA content of plant nuclei. *Am. J. Bot.* 86:609–13.

Joint Genome Institute, U.S. Department of Energy. 2006. *Phaeodactylum Tricornutum Finished Genome Sequence v2.0*. Available at: <http://genome.jgi-psf.org/Phatr2/Phatr2.home.html> (accessed January 8, 2007).

- Joint, I. & Pomroy, A. 1988. Allometric estimation of the productivity of phytoplankton assemblages. *Mar. Ecol. Prog. Ser.* 47:161–8.
- Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E. & Schulman, A. H. 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimate divergence. *Proc. Natl. Acad. Sci. U. S. A.* 97:6603–7.
- Kazazian, H. H. 2004. Mobile elements: drivers of genome evolution. *Science* 303:1626–32.
- Kidwell, M. G. 2002. Transposable elements and the evolution of genome sizes in eukaryotes. *Genetica* 115:49–63.
- Kimura, Y., Tosa, Y., Shimada, S., Sogo, R., Kusaba, M., Sunaga, T., Betsuyaku, S., Eto, Y., Nakayashiki, H. & Mayama, S. 2001. *OARE-1*, a *Ty1-copia* retrotransposon in oat activated by abiotic and biotic stresses. *Plant Cell Physiol.* 42:1345–54.
- LaJeunesse, T. C., Lambert, G., Anderson, R. A., Coffroth, M. A. & Galbraith, D. W. 2005. *Symbiodinium* (Pyrrophyta) genome sizes (DNA content) are smallest among dinoflagellates. *J. Phycol.* 41:880–6.
- Laws, E. A., Falkowski, P. G., Smith, W. O. J., Ducklow, H. & McCarthy, J. J. 2000. Temperature effects on export production in the open ocean. *Glob. Biogeochem. Cycles* 14:1231–46.
- Le Borgne, R., Feely, R. A. & Mackey, D. J. 2002. Carbon fluxes in the equatorial Pacific: a synthesis of the JGOFS programme. *Deep-Sea Res. Part II Top. Stud. Oceanogr.* 49:2425–42.
- Lynch, M. & Conery, J. S. 2003. The origins of genome complexity. *Science* 302:1401–4.
- Mann, D. G. 1994. Auxospore formation, reproductive plasticity and cell structure in *Navicula ulvacea* and the resurrection of the genus *Dickieia* (Bacillariophyceae). *Eur. J. Phycol.* 29:141–57.
- McClintock, B. 1950. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U. S. A.* 36:344–55.
- McClintock, B. 1984. The significance of responses of the genome to challenge. *Science* 256:792–801.
- Medlin, L. K. & Kaczmarek, I. 2004. Evolution of the diatoms: V. Morphological and cytological support for the major clades and a taxonomic revision. *Phycologia* 43:245–70.
- Melaragno, J. E., Mehrotra, B. & Coleman, A. W. 1993. Relationship between endopolyploidy and cell size in epidermal tissue of *Arabidopsis*. *Plant Cell* 5:1661–8.
- Mirsky, A. E. & Ris, H. 1951. The deoxyribonucleic acid content of animal cells and its evolutionary significance. *J. Gen. Physiol.* 34:451–62.
- Moles, A. T., Ackerly, D. D., Webb, C. O., Tweddle, J. C., Dickie, J. B., Pitman, A. J. & Westoby, M. 2005. Factors that shape seed mass evolution. *Proc. Natl. Acad. Sci. U. S. A.* 102:10540–4.
- Orgel, L. E. & Crick, F. H. C. 1980. Selfish DNA: the ultimate parasite. *Nature* 284:604–7.
- Raven, J. A. & Kubler, J. E. 2002. New light on the scaling of metabolic rate with the size of algae. *J. Phycol.* 38:11–6.
- Round, F. E., Crawford, R. M. & Mann, D. G. 1990. *The Diatom: Biology and Morphology of the Genera*. Cambridge University Press, Cambridge, UK, 747 pp.
- Sarthou, G., Timmermans, K. R., Blain, S. & Treguer, P. 2005. Growth physiology and fate of diatoms in the ocean: a review. *J. Sea Res.* 53:25–42.
- Scala, S., Carels, N., Falciatore, A., Chiusano, M. L. & Bowler, C. 2002. Genome properties of the diatom *Phaeodactylum tricorutum*. *Plant Physiol.* 129:993–1002.
- Shuter, B. J., Thomas, J. E., Taylor, W. D. & Zimerman, A. M. 1983. Phenotypic correlates of genomic DNA content in unicellular eukaryotes and other cells. *Am. Nat.* 122:26–44.
- Smetacek, V., Assmy, P. & Henjes, J. 2004. The role of grazing in structuring southern ocean pelagic ecosystems and biogeochemical cycles. *Antarc. Sci.* 16:541–58.
- Soltis, D. E. & Soltis, P. S. 1999. Polyploidy: recurrent formation and genome evolution. *Trends Ecol. Evol.* 14:348–52.
- Takeda, S., Sugimoto, K., Otsuki, H. & Hirochika, H. 1999. A 13-bp *cis*-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *Plant J.* 18:383–93.
- Tang, E. P. Y. 1995. The allometry of algal growth rates. *J. Plankton Res.* 17:1325–35.
- Tang, E. P. Y. & Peters, R. H. 1995. The allometry of algal respiration. *J. Plankton Res.* 17:303–15.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–80.
- Treguer, P., Nelson, D. M., Van Bennekom, A. J., DeMaster, D. J., Leynaert, A. & Queguiner, B. 1995. The silica balance in the World Ocean: a reestimate. *Science* 268:375–9.
- Tremblay, J. E., Klein, B., Legendre, L., Rivkin, R. B. & Therriault, J. 1997. Estimation of fratics in oceans based on phytoplankton size structure. *Limnol. Oceanogr.* 42:595–601.
- Veldhuis, M. J., Ciccì, T. L. & Sieracki, M. E. 1997. Cellular DNA content of marine phytoplankton using two new fluorochromes: taxonomic and ecological implications. *J. Phycol.* 33:527–41.
- Walbot, V. & Petrov, D. A. 2001. Gene galaxies in the maize genome. *Proc. Natl. Acad. Sci. U. S. A.* 98:8163–4.
- Warton, D. I., Wright, I. J., Falster, D. S. & Westoby, M. 2006. Bivariate line-fitting methods for allometry. *Biol. Rev.* 81:259–91.
- Webb, C. O., Ackerly, D. D. & Kembel, S. W. 2006. *Phylocom: Software for the Analysis of Community Phylogenetic Structure and Character Evolution*. Version 3.40. Available at: <http://www.phylodiversity.net/phylocom> (accessed January 5, 2006).
- Wilke, C. M., Maimon, E. & Adams, J. 1992. The population biology and evolutionary significance of Ty elements in *Saccharomyces cerevisiae*. *Genetica (Dordr.)* 86:155–73.
- Yool, A. & Tyrrell, T. 2003. Role of diatoms in regulating the ocean's silicon cycle. *Glob. Biogeochem. Cycles* 17:1103.