5-2014

Transcriptomic Insights into the Diplontic Life History of Diatoms

Colton Richard Kessenich
University of Arkansas, Fayetteville

Transcriptomic Insights into the Diplontic Life History of Diatoms

Transcriptomic Insights into the Diplontic Life History of Diatoms

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Biology

by

Colton Kessenich
Taylor University
Bachelor of Science in Biology, 2012

May 2014
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

_____
Andrew J. Alverson, Ph.D.
Thesis Director

_____
Marlis R. Douglas, Ph.D.
Committee Member

_____
Jeffery D. Silberman, Ph.D.
Committee Member

ABSTRACT

An organism's life cycle is the direct result of its evolutionary history and represents a fundamental aspect of its ancestry and ecology. Yet the process of linking alternating life-history stages has proven to be challenging, if not impossible in some cases. Diatoms (Bacillariophyceae) are no exception to this challenge, and their diversity of life stages and reproductive strategies add further challenges. A central focus of diatom research has been to unravel the evolutionary events that led to their extraordinary diversity, a line of inquiry that has been greatly aided by the availability of next-generation sequence data. Yet without proper taxonomic sampling, many fundamental questions have remained unresolved. Furthermore, the incorporation of data from outgroup lineages provides crucial insights into the ancestral state of a taxon. However the life cycle of the sister lineage to diatoms, the Bolidophytes, has also remained an enigma. Here, to address yet unresolved questions about the origin of the diatom life cycle, transcriptomes for *Bolidomonas pacifica* (Bolidophyceae) and two diploid diatoms, *Leptocylindrus danicus* and *Hemiaulus sinensis* (Bacillariophyceae) were sequenced and analyzed. A novel approach of quantifying the degree of heterozygosity across the genome was used to infer the ploidy in *Bolidomonas*, whose ploidy level is unknown. Lack of heterozygous alleles in Bolidomonas strongly suggests that it is a vegetatively haploid organism, and the presence of several meiosis-specific genes indicate that it is likely capable of sexual reproduction as well. These results suggest that *Bolidomonas* represents the haploid phase of a haplodiplonitic life cycle, in which silicified Parmaleans are the diploid phase.

TABLE OF CONTENTS

I. INTRODUCTION 1
a. Contributions 11

II. TRANSCRIPTOMIC INSIGHTS INTO THE LIFE HISTORY OF BOLIDOPHYTES, THE SISTER LINEAGE TO DIATOMS 12
a. Abstract 13
b. Introduction 14
c. Materials & Methods 15
d. Results 17
e. Discussion 20
f. Acknowledgements 23
g. References 24
h. Figures 28
i. Appendix 32

III. CONCLUDING REMARKS 33

INTRODUCTION

Diatoms account for as much as 20 percent of global primary production and form the base of many ecologically and economically important ecosystems (Falkowski et al. 1998, Field et al. 1998). Although the widespread existence of diatoms has long been documented, with the earliest description of diatoms occurring in 1703 (Round et al. 1990), genome sequences of two diatom species, *Thalassiosira pseudonana* (Armbrust et al. 2004) and *Phaeodactylum tricornutum* (Bowler et al. 2008), has greatly altered our understanding of this lineage.

These studies have focused greater attention onto diatom metabolism, physiology, and evolution. The genomes have revealed several factors that are thought to have contributed to the evolutionary success of diatoms, including the presence of complete urea cycle (Armbrust et al. 2004) with associated nitrogen recycling pathways (Allen et al. 2006, Allen et al. 2011) and the presence of hundreds of bacterially derived genes within diatom genomes (Bowler et al. 2008). Their silicified cell walls (frustules) are the most commonly cited feature of diatoms. These unique bipartite frustules are cited as important in the evolution of diatoms for a variety of reasons, but perhaps most intriguing is the hypothesis that it has co-evolved alongside the life-history strategy utilized by diatoms and acts as an intrinsic trigger for sexual reproduction (Lewis 1984).

An organism's life cycle is a direct result of its evolutionary past and it reflects fundamental aspects of its ecology. Yet linking life-history stages through direct observations is challenging, if not outright impossible. As a result, life histories may often be poorly or incompletely described (von Dassow and Montresor 2011). Unless one fortuitously happens upon an organism while it is undergoing a life phase shift in nature, or the exact triggers to induce a phase transition within the laboratory are known, it is unlikely an organism's life cycle will be observed directly. Although a variety of hypotheses have been proposed about the life cycle of the ancestral diatom, these difficulties in observation, combined with the variety of strategies found throughout the photosynthetic stramenopiles (Figure 1), have not allowed for the construction of predictive evolutionary frameworks from which to make inferences about ancestral or derived character states (Richerd et al. 1993, Sims et al. 2006).

Stramenopiles display a haphazard assemblage of haplontic life cycles in which mitosis occurs only in the haploid phase (Figure 2), haplodiplontic life cycles where both haploid and diploid phases are

mitotic (Figure 3), and diplontic life cycles whereby only the diploid phase undergoes mitosis (Figure 4). Most of these algae have either haplontic or haplodiplontic life cycles (Van den Hoek et al. 1995, von Dassow and Montresor 2011), which makes the ubiquitous presence of a diplontic life cycle in diatoms a point of interest.

Although estimates place the number of extant diatoms to be in excess of 200,000 species (Mann and Droop 1996) and life-cycles have only been examined for a few hundred of them (Edlund and Stoermer 1997), the uniformity of the diplontic life cycle and sex in diatoms (Mann 1999) is well supported. It is important to note that, in diatoms, sex is primarily linked to the restoration of cell size after diminution resulting from serial asexual (mitotic) divisions (Figure 5) (Macdonald 1869, Pfitzer 1871), and typically represents a single brief stage of the overall life cycle (Figure 6). Meiosis and subsequent reproduction are induced after a cell reaches a critical size threshold—typically thought to be approximately 30% of the enlarged cell size—following numerous rounds of mitotic cell division. If induced, the haploid gametes join to form a zygotic cell (the "auxospore") (Crawford 1974, Hoops and Floyd 1979). This auxospore subsequently develops into a new and enlarged diatom cell (Round et al. 1990). Failure to induce meiosis and restore cell size may result in a cell shrinking to a level at which reproduction and auxosporulation is no longer possible, resulting in extirpation of that cell line. Once the enlarged cell emerges from the auxospore the vegetative diploid mitotic stage can last for years or even decades (Mann 1988), while the whole process of gametogenesis, fertilization, and auxosporulation can occur in as little as a few hours (Chepurnov et al. 2004).

Further adding to the difficulty in resolving the origin of the diplontic life cycle in diatoms is the lack of life-cycle information for the closest known relatives of diatoms, which was only discovered in the late 1990s (Guillou et al. 1999). This lineage—the bolidophyceae—consists of fewer than 20 described species (Guiry 2013) and therefore stands in stark contrast to the diversity found with the diatom lineage (Mann and Droop 1996, Vardi et al. 2008). The close relationship of diatoms and bolidophytes, coupled with this disparity in species richness provides a framework for investigating the evolutionary success of diatoms.

As previously mentioned, making direct links between alternating life history stages can be challenging, and unfortunately, very little is known about the life cycle stages within Bolidophyceae. In the

absence of direct observations of life-history phases, it is important to find and leverage alternate methods of describing and characterizing these different stages. Although genomic studies have contributed greatly to our understanding of diatoms, these types of data are rarely used to better characterize and understand life cycles and their evolution. Furthermore, the amount of publically available genomic data for diatoms is still low given the estimated number of species. Insights from genomic data have also been crippled by a lack of genome-scale data from the bolidophytes. In order to characterize and resolve the life cycles of bolidophytes, which will provide a framework understanding the origin of the diatom life cycle, I gathered transcriptomic data from *Bolidomonas pacifica* (Bolidophyceae), whose ploidy level is unknown, and two diploid diatoms, *Leptocylindrus danicus* and *Hemiaulus sinensis*, and used a novel genomic approach to characterize their ploidy levels, using the degree of heterozygosity as a surrogate measure of genomic ploidy. For a diploid organism, the null expectation is that detectable levels of heterozygosity should exist. In contrast, a haploid organism, by definition, should not exhibit allelic heterozygosity since only a single allele will be present per locus. My null hypothesis was that *Bolidomonas* is the haploid stage of a haplodiplontic life cycle in which the silicified Parmalean algae are the diploid phase, and the data presented here were consistent with that hypothesis.

References

Allen A.E., Vardi A., Bowler C. (2006). An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr. Opin. Plant Biol.* 9: 264-273.

Allen A.E., Dupont C.L., Obornik M., Horak A., Nunes-Nesi A., McCrow J.P., Zheng H. et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473: 203-207.

Armbrust E.V., 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.

Bowler C., Allen A.E., Badger J.H., Grimwood J., Jabbari K., Kuo A., Maheswari U. et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456: 239-244.

Chepurnov V.A., Mann D.G., Sabbe K., Vyverman W. (2004). Experimental studies on sexual reproduction in diatoms. *Int. Rev. Cytol.* 237: 91-154.

Crawford R.M. (1974). The auxospore wall of the marine diatom Melosira nummuloides (Dillw.) C. Ag. and related species. *Br. Phycol. J.* 9: 9-20.

Edlund M.B., Stoermer E.F. (1997). Ecological, evolutionary, and systematic significance of diatom life histories. *J. Phycol.* 33: 897-918.

Falkowski P.G., Barber R.T., Smetacek V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science* 281: 200-206.

Field C.B., Behrenfeld M.J., Randerson J.T., Falkowski P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281: 237-240.

Guillou L., Chretiennot-Dinet M.-J., Medlin L.K., Claustre H., Loiseaux-de G.S., Vaulot D. (1999). *Bolidomonas*: A new genus with two species belonging to a new algal class, the Bolidophyceae (Heterokonta). *J. Phycol.* 35: 368-381.

Guiry M.D., Guiry, G.M. (2013). AlgaeBase. World-wide electronic publication. National University of Ireland, Galway.

Hoops H.J., Floyd G.L. (1979). Ultrastructure of the centric diatom, Cyclotella meneghiniana: vegetative cell and auxospore development*. *Phycologia* 18: 424-435.

Lewis W.M. (1984). The Diatom Sex Clock and Its Evolutionary Significance. *Am. Nat.* 123: 73-80.

Macdonald J.D. (1869). I.—On the structure of the Diatomaceous frustule, and its genetic cycle. *Ann. Mag. Nat. Hist* 3: 1-8.

Mann D.G. (1988). Why didn't Lund see sex in Asterionella? A discussion of the diatom life cycle in nature. *Algae and the aquatic environment* 29: 385-412.

Mann D.G., Droop S.J.M. (1996). Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia* 336: 19-32.

Mann D.G. (1999). The species concept in diatoms. *Phycologia* 38: 437-495.

Pfitzer E.H.H. (1871). *Untersuchungen über Bau und Entwicklung der Bacillariaceen (Diatomaceen)*. Marcus.

Richerd S., Couvet D., Valéro M. (1993). Evolution of the alternation of haploid and diploid phases in life cycles. II. Maintenance of the haplo-diplontic cycle. *J. Evol. Biol.* 6: 263-280.

Round F.E., Crawford R.M., Mann D.G. (1990). *The Diatoms: Biology & Morphology of the Genera*. Cambridge University Press: Cambridge.

Sims P.A., Mann D.G., Medlin L.K. (2006). Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* 45: 361-402.

Van den Hoek C., Mann D., Jahns H.M. (1995). An Introduction to Phycology. Cambridge University Press, Cambridge.

Vardi A., Thamatrakoln K., Bidle K.D., Falkowski P.G. (2008). Diatom genomes come of age. *Genome Biol.* 9: 5.

von Dassow P., Montresor M. (2011). Unveiling the mysteries of phytoplankton life cycles: patterns and opportunities behind complexity. *J. Plankton Res.* 33: 3-12.

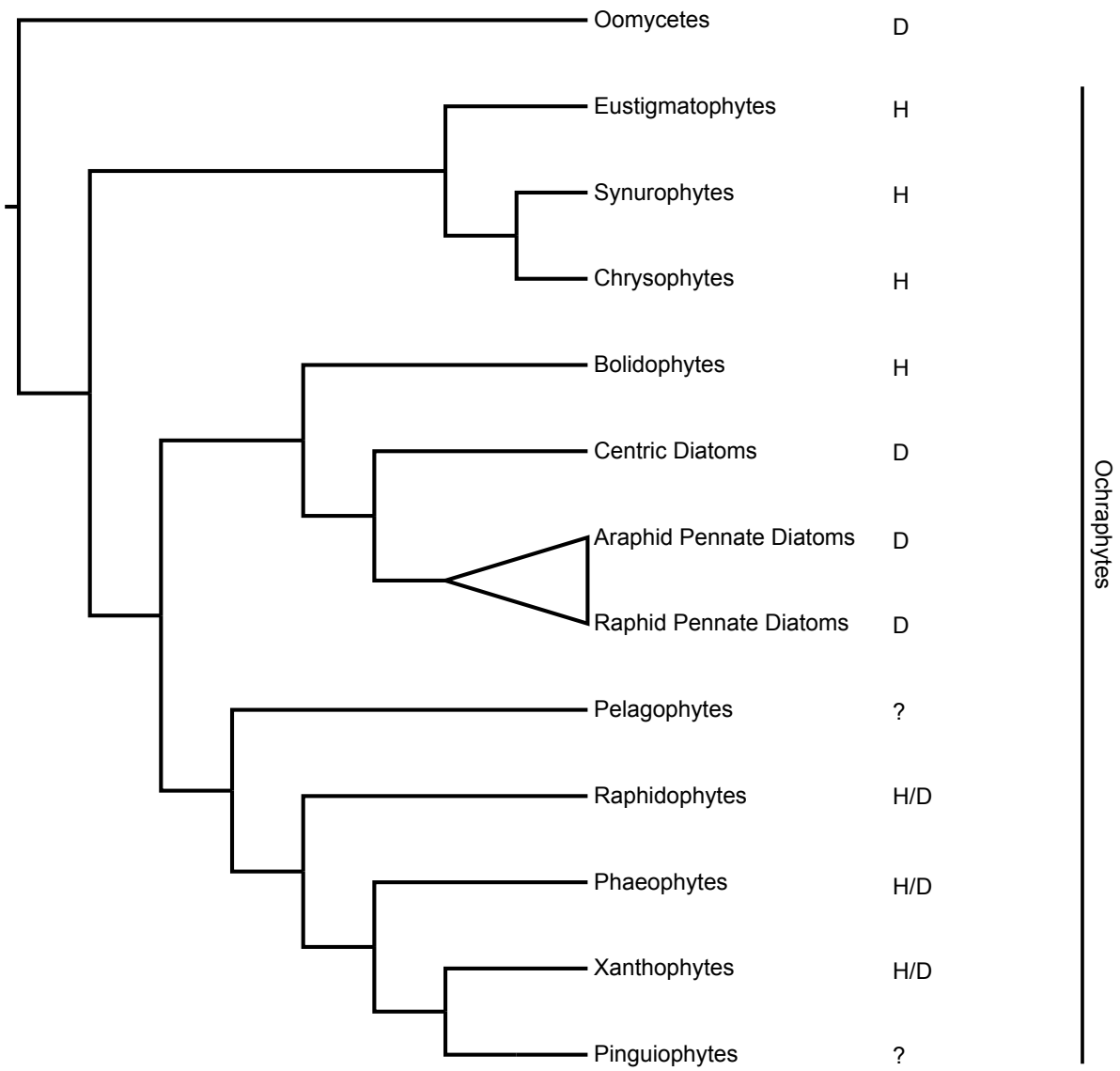


Figure 1. The known distributions of diplontic (D) and haplontic or haplodiplontic (H) life histories in photosynthetic stramenopiles and oomycetes.

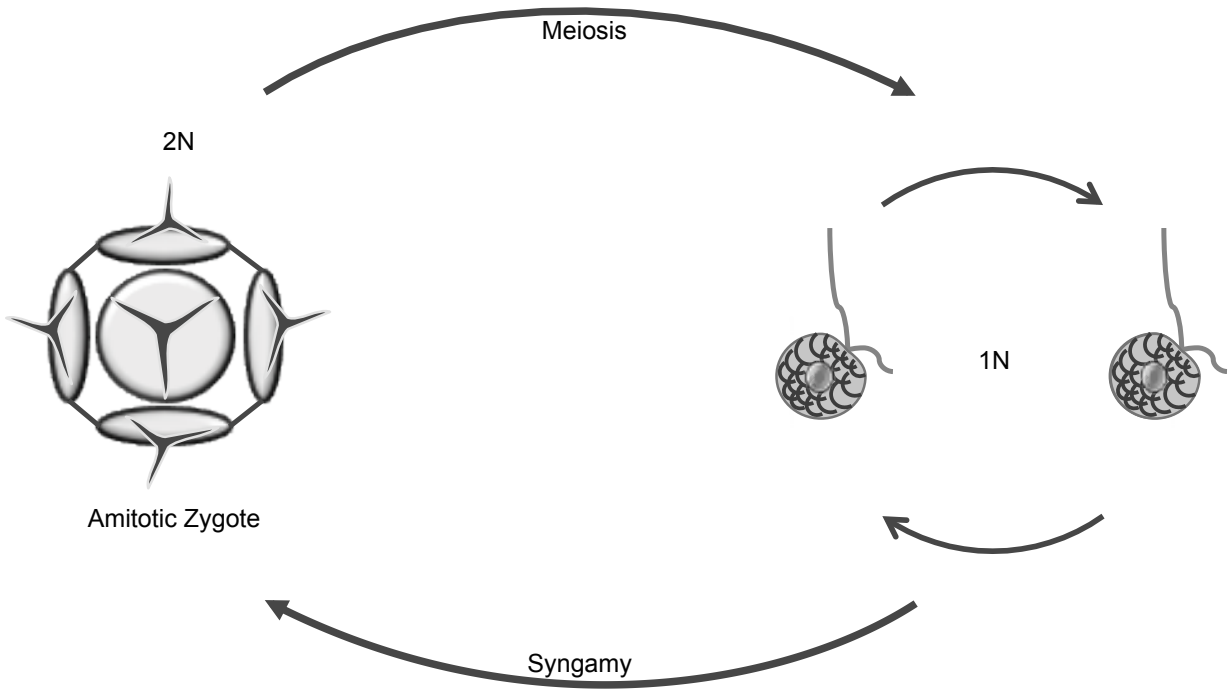


Figure 2. A haplontic life cycle, in which the haploid phase is the vegetative mitotic phase and the diploid phase is an amitotic zygotic form that undergoes meiosis to form additional haploid mitotic organisms.

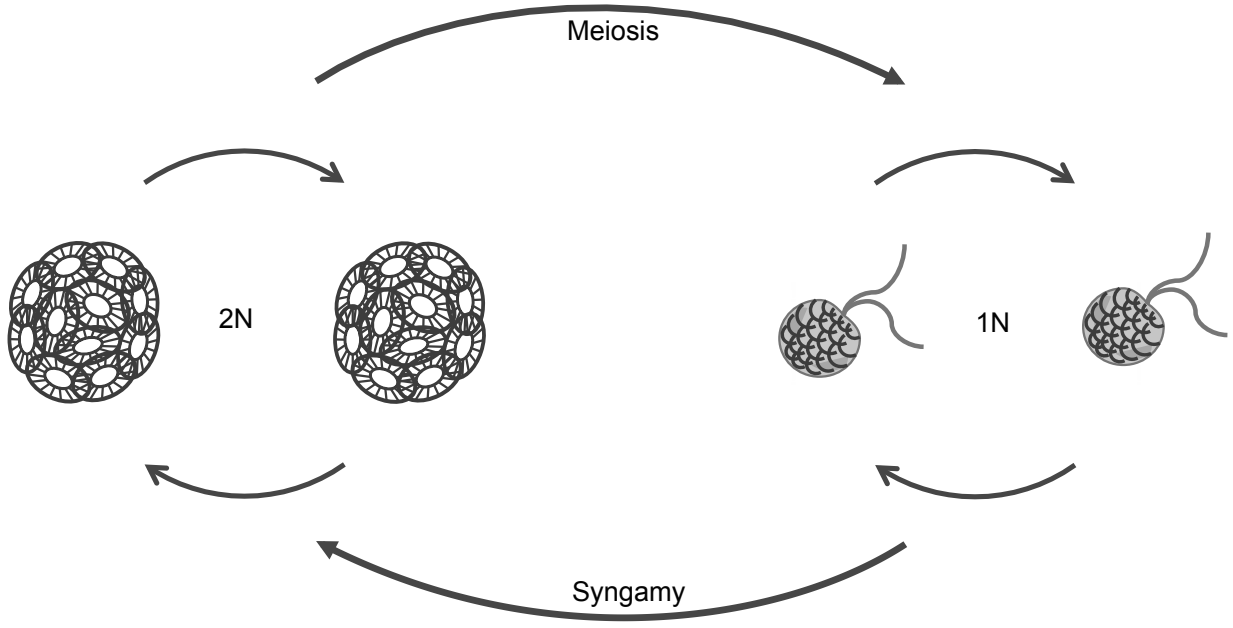


Figure 3. A haplodiplontic life cycle in which both the haploid and diploid phases are capable of undergoing vegetative mitosis.

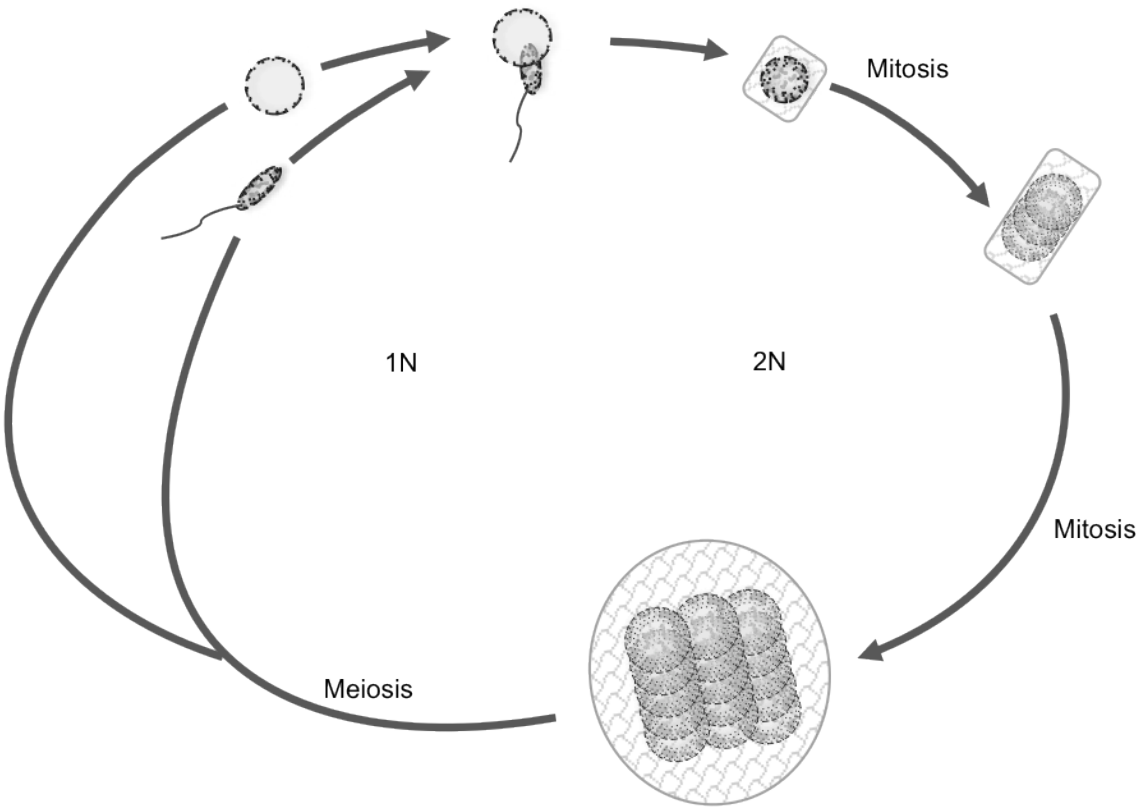


Figure 4. A diplontic life cycle where only the diploid phase is capable of mitosis and intermittent meiosis produces a haploid phase that is typically brief and solely for reproduction.

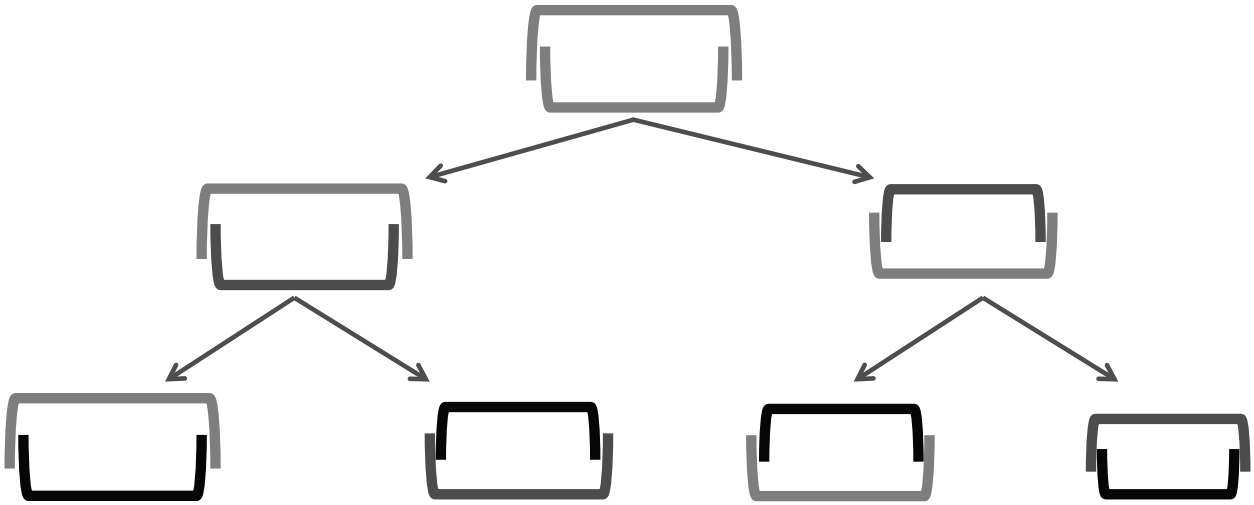


Figure 5. The diminution effect during the mitotic phase of a diatom’s life cycle. This occurs as a result of the semi-conservative nature of cell division, in which each half ("valve") of the parent cell becomes the larger "epivalve" in the daughter cells.

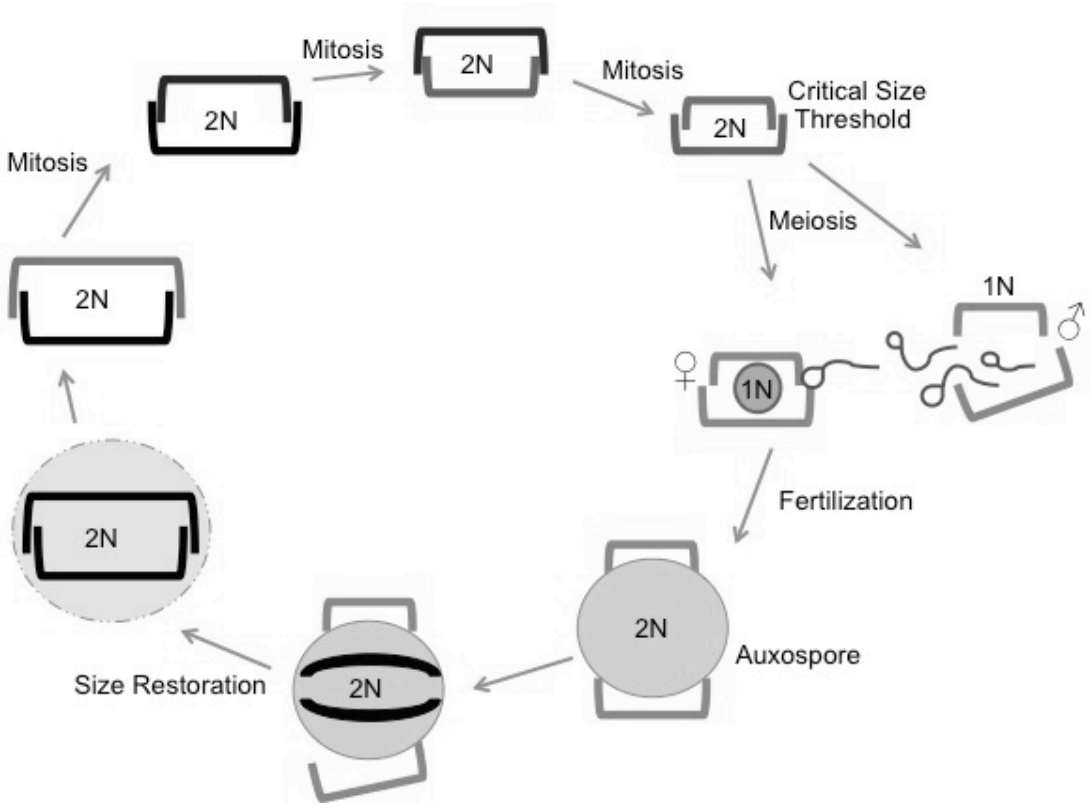


Figure 6. A typical diatom life cycle exhibiting mitotic division until reaching a critical size threshold that induces meiosis, which is followed by fertilization. The zygotic “auxospore” phase then swells and allows for the generation of a new enlarged diploid cell.

CONTRIBUTIONS

The subsequent body of work was done through the collaborative efforts of a number of individuals from numerous academic institutions. The collaborators and contributions are as listed below:

Collaborator contributions:

Andrew J. Alverson: Experiment conception, diatom insights, literature review, writing.

Norman J. Wickett: Custom scripts, translations, transcriptomic insights.

Andrew M. Schurko: Identification of meiosis-specific homologs, insight into meiosis.

Elizabeth C. Ruck: Cell culturing, RNA extraction, insight into diatoms, lab authority.

My contributions to this body of work:

Transcriptome assembly, annotation, translation, analysis of heterozygosity, identification of meiosis-specific homologs, literature review, and writing. For verification, a statement of endorsement acknowledging my status as primary author and contributor has been included (see Ch. 2 Appendix).

TRANSCRIPTOMIC INSIGHTS INTO THE LIFE HISTORY OF BOLIDOPHYTES, THE SISTER LINEAGE TO DIATOMS

Colton R. Kessenich, Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701, USA

Elizabeth C. Ruck, Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701, USA

Andrew M. Schurko, Department of Biology, Hendrix College, Conway, AR 72032, USA

Norman J. Wickett, Chicago Botanic Garden, Glencoe, IL 60022, USA; Program in Biological Sciences, Northwestern University, Evanston, IL 60208

Andrew J. Alverson, Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701, USA

Under review in Journal of Phycology; submitted 03/09/2014

Abstract

Diatoms are perhaps the most diverse lineage of eukaryotic algae. Their unique siliceous cell wall and diplontic life history are thought to have played important roles in their evolutionary success. The characteristic diminution of the diatom cell wall over the course of vegetative growth provides a reliable, intrinsic trigger for sexual reproduction, establishing a direct link between the evolution of cell-wall and life-history features. It is unclear, however, whether the diplontic life cycle of diatoms represents an ancestral or derived trait. This uncertainty is based in part on our lack of understanding of the life cycle of the sister lineage to diatoms, which includes a mix of two free-living forms: naked biflagellate unicells in the genus *Bolidomonas* and silicified forms in the order Parmales. These two forms might represent different life-history stages, though directly establishing such links often proves difficult. We sequenced transcriptomes for *Bolidomonas* and two diatoms and found that approximately 0.1% of the coding regions in the two diploid diatoms are heterozygous, whereas *Bolidomonas* is virtually devoid of heterozygous alleles, consistent with expectations for a haploid genome. We also identified homologs of meiotic genes in the *Bolidomonas* transcriptome, indicating a capacity for sexual reproduction despite the evidently haploid genome. These results suggest that *Bolidomonas* is haploid and predict that parmaleans represent the diploid phase of a haplodiplontic life cycle. These data fill an important gap in our understanding of the origin of the diplontic life history of diatoms, which may represent an evolutionarily derived, adaptive feature.

Introduction

Since their origin some 250 million years ago (Sorhannus 2007), diatoms have become one of the most diverse lineages of eukaryotic algae, with conservative diversity estimates numbering in the hundreds of thousands of species (Mann and Droop 1996, Mann and Vanormelingen 2013). This has motivated a great deal of research focused on identifying the key traits underlying their origin and subsequent diversification. Common hypotheses focus on the many possible benefits of their unique, bipartite siliceous cell walls. Possible advantages include reduced energetic costs compared to carbon-based cell walls (Raven 1983), mechanical protection against predation (Hamm et al. 2003), and an enhanced photosynthetic apparatus (Fuhrmann et al. 2004). Efforts to understand the evolutionary

success of diatoms have also focused on their life history, in part because it is so inextricably linked to the origin and evolution of their cell walls (Lewis 1984, Mann and Marchant 1989).

Diatoms are uniformly diplontic, a life cycle in which all stages are diploid save the haploid gametes produced during brief and intermittent periods of sexual reproduction (Chepurnov et al. 2004, Drebes 1977). Although most other algae are either haplontic or haplodiplontic (Van den Hoek et al. 1995, von Dassow and Montresor 2011), the heterokont lineage (which includes diatoms) contains a smattering of other diplonts, including oomycetes and subsets of xanthophytes, phaeophytes and raphidophytes (Bell 1997, Cronberg 2005, Van den Hoek et al. 1995, Yamaguchi and Imai 1994). However few, the broad phylogenetic distribution of diplontic heterokonts coupled with uncertainties about the life histories of several heterokont classes, make it unclear whether the diplontic life cycle of diatoms represents an ancestral or derived condition (Kooistra et al. 2007, Mann and Marchant 1989, Medlin et al. 1997, Medlin et al. 1993). Critically, reconstructing the origin of the diatom life history requires characterization of the life history of their closest known relatives, the Bolidophyceae.

Bolidophytes were only discovered in the late 1990s, and given that the heterokont lineage includes several other non-diatoms with silicified cells (e.g., synurids and chrysophytes; Van den Hoek et al. 1995), it was surprising that the sister taxon to diatoms was a group of relatively nondescript, naked, flagellated unicells (Guillou et al. 1999). Nevertheless, single and multi-gene phylogenetic analyses have consistently, and strongly, supported this relationship (Daugbjerg and Guillou 2001, Goertzen and Theriot 2003, Guillou et al. 1999, Yang et al. 2012). Earlier hypotheses suggested a possible relationship between diatoms and another small, relatively poorly known, and then phylogenetically unresolved lineage, the Parmales (Mann and Marchant 1989). Originally described as chrysophytes, parmaleans are encased by 3–8 siliceous plates, which resemble chrysophyte scales and, compellingly, the auxospore scales of some diatoms (Booth and Marchant 1987). A number of other similarities in the development, morphology, and arrangement of parmalean cell plates suggested a close affinity to the inferred common ancestor of diatoms as well (Mann and Marchant 1989, see also Round and Crawford 1981). Some 25 years after they were described (Booth and Marchant 1987), the first parmaleans were isolated and cultured, and phylogenetic analysis of nuclear-encoded *SSU* rDNA and plastid *rbcL* sequences placed them as sister to diatoms and fully interspersed among bolidophytes (Ichinomiya et al. 2011), suggesting

that the two groups might represent different life-history stages of the same organism (Guillou 2011). A determination of the ploidy levels of bolidophytes and parmaleans is central to testing this hypothesis and, by extension, resolving the origin of the diplontic life cycle in diatoms (Chepurnov et al. 2004, Guillou et al. 1999, Guillou 2011, Ichinomiya et al. 2011, Mann and Marchant 1989, Sims et al. 2006).

Establishing direct links between alternating life history phases through direct observations of cell cultures or natural populations can be challenging, if not impossible (Chepurnov et al. 2004, Mann and Chepurnov 2004, von Dassow and Montresor 2011). Illustrative of this, life cycles have been characterized for just a few hundred diatom species (Chepurnov et al. 2004, Edlund and Stoermer 1997). These difficulties have prompted the development of creative genetic and genomic approaches for characterizing complex life histories. We sequenced high-depth transcriptomes for *Bolidomonas pacifica* Guillou et Chrétiennot-Dinet and two distantly related diatoms, *Leptocylindrus danicus* Cleve and *Hemiaulus sinensis* Greville—both known diploids—and used these data to characterize each organism's capacity for sexual reproduction and degree of heterozygosity in the genome. The latter was used as a surrogate measure of each organism's ploidy level. The results shed new light on the relationship between bolidophytes and parmaleans and, in addition, the origin of the diplontic life cycle of diatoms.

Materials and Methods

Cell Culturing, RNA Extraction, and Sequencing

Clonal cultures of two diatoms, *Leptocylindrus danicus* (strain ECT3929) and *Hemiaulus sinensis* (strain 24i10-1A), were grown in f/2 medium (Guillard and Ryther 1962, Guillard 1975), and *Bolidomonas pacifica* (strain CCMP1866) was grown in PRO99 medium (Moore et al. 2002). All cultures were grown at 22°C on a 12:12 light:dark cycle. Daily growth rates of batch cultures were estimated based on chlorophyll *a* fluorescence with a Trilogy Laboratory Fluorometer (Turner Designs, CA, USA). Cells were harvested during exponential growth and concentrated by centrifugation (1,272 g for diatoms, 2,348 g for *Bolidomonas*) for 10 minutes at 2°C. RNA was isolated from undisrupted cells with a Qiagen RNeasy kit (Qiagen, Venlo, Netherlands). DNase-treated RNA was quantified with a Qubit 2.0 (Life Technologies, NY, USA), and RNA quality was assessed with a Bioanalyzer (Agilent Technologies, CA, USA). At least 10 µg of RNA was sent to the Beijing Genomics Institute in Shenzhen, China, for DNA sequencing on the

Illumina HiSeq2000 platform. A total of 33–36 Gb of 90-bp paired-end reads was sequenced for each species.

Transcriptome Assembly and Annotation

FastQC (Andrews 2009) was used to assess per-base quality scores on all sequencing reads prior to trimming the ends to an average per-base Phred score of at least 30 for each position in the read. Trimmed reads were assembled with Trinity using default parameters (Grabherr et al. 2011).

To translate assembled contigs into amino acid sequences, the transcriptomes were first searched against the RefSeq protein database maintained by the National Center for Biotechnology Information (NCBI) with NCBI BLASTX (ver. 2.2.28+). The BLASTX output was used as a reference to build the translations with Genewise (ver. 2.4.1) (Birney et al. 2004). Custom Perl scripts were used alongside the Genewise output to reconcile internal stop codons, introns, and frameshifts resulting from assembly artifacts. After this step, all contigs ≤200 bp in length were also discarded from further analyses. The remaining contigs were then searched against the NCBI RefSeq protein database with NCBI BLASTX and these results were used to assign annotations with b2g4pipe, which is part of the Blast2GO (ver. 2.7.0) software package (Conesa et al. 2005). Annotations were based on an e-value cutoff of $1.0e^{-5}$ and an annotation cutoff value of 35. All raw sequence reads and transcriptome assemblies are available on NCBI/GenBank under project number PRJNA237965.

Heterozygosity Analyses

For each transcriptome, trimmed reads were mapped back to the assembled contigs with Bowtie (ver. 0.12.8) (Langmead et al. 2009), and mismatches [i.e., single nucleotide polymorphisms (SNPSs)] between individual reads and contigs were identified with Samtools (ver. 0.1.19) and bcf tools (Li et al. 2009). Only SNP calls with a Phred-like mismatch score of ≥20 were included in subsequent analyses. The effect of sequencing read depth on SNP calling was evaluated by applying depth thresholds ranging from 4–200 reads. Inferences were based on two final SNP counts: one that required a flat depth of ≥20 reads at the SNP position, and one that ignored the lowest coverage quartile of data in each transcriptome. Final SNP densities were based only on regions of the transcriptome that met our

minimum depth requirements. The small number of SNPs in the *Bolidomonas* transcriptome were examined manually in Tablet (ver. 1.13.08.05) (Milne et al. 2013) to further establish their read locations, contig locations, and sequencing depth-of-coverage.

Identification of Meiosis Genes

A set of 12 genes encoding proteins known to function in meiotic recombination across a diverse set of eukaryotes was compiled as a reference database of meiosis genes (Hanson et al. 2013, Malik et al. 2008). Reciprocal BLASTP and TBLASTN searches against nucleotide and protein databases at NCBI, the Joint Genome Institute (JGI), and the Online Resource for Community Annotation of Eukaryotes (ORCAE) were used to identify contigs in our assemblies with homologs to meiosis-specific genes in the model angiosperm *Arabidopsis thaliana,* the brown alga *Ectocarpus siliculosus*, and the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. Orthologs from these four taxa were used as queries for TBLASTN searches against the three newly sequenced transcriptomes. Significant hits (e-value cutoff of $1.0e^{-5}$) were used in reciprocal BLASTX searches against the NCBI non-redundant protein (nr) database to further validate their identities.

Results

Heterozygosity

For a clonal culture of a diploid organism, high-quality nucleotide mismatches (SNPs) between a sequencing read and an assembled contig represent heterozygous alleles. Non-inbred diploid organisms should harbor detectable levels of heterozygosity, whereas haploid organisms should not. Based on these null expectations, we assessed the ploidy level in two diatoms, which are known to be vegetatively diploid, and *Bolidomonas*, whose ploidy level is unknown. The two diatom transcriptomes contain large numbers of SNPs spread across large proportions of the contigs (Figure 1A). A total of 59% of the 16,858 contigs in the *Hemiaulus* transcriptome contained at least one SNP. Likewise, the *Leptocylindrus* transcriptome assembled into 15,701 contigs, 45% of which contained at least one SNP. In sharp contrast, the overwhelming majority of the 13,373 contigs in the *Bolidomonas* assembly were fully homozygous, with just 556 of them (4%) containing one or more SNPs.

The statistical significance of a SNP call is based, in part, on the overall sequencing depth at the position as well as a user-defined depth requirement for a SNP to be called. If reads are mapped accurately and more-or-less evenly across the assembly, then increases in the depth cutoff should result in a corresponding decrease in the number of SNP calls. To rule out possible adverse effects of our predefined read-depth cutoffs on the inferred SNP densities, we calculated the total number of SNPs at depths ranging from 4 to 200 reads (Supplemental Figure 1). These results revealed no apparent biases related to our depth cutoff, so we report SNP densities using two read-depth criteria, one requiring a total of ≥20x coverage at the SNP position and another that ignored SNPs in sites falling within the lowest quartile of coverage for that particular assembly (Figure 1B). Results were similar across the two depths, with *Hemiaulus* containing 1.3–1.4 SNPs per kb and *Leptocylindrus* containing 0.8–0.9 SNPs per kb (Figure 1B). *Bolidomonas* had a substantially lower SNP density than both diatoms, with just 0.03–0.06 SNPs per kb, which translates to roughly 2–7% of the overall variation observed in diatoms (Figure 1B).

To further characterize the small number of SNPs in *Bolidomonas*, we performed a BLASTN search of the 556 SNP-containing contigs against themselves to identify SNPs discriminating recent gene duplicates rather than heterozygous alleles. A total of 184 of the 556 SNP-containing contigs grouped into 68 single-linkage clusters representing putative families of two or more contigs. Fully 449 of the 1144 SNPs (39%) fall within these families, indicating that these likely represent substitutions between paralogous genes rather than allelic polymorphisms (Malhis and Jones 2010, Pavy et al. 2006). Manual examination of the remaining 372 SNP-containing contigs revealed that polymorphisms occurred largely in three regions of the assembly: within 20 bp of the ends of contigs, in regions of repetitive or low-complexity sequence, and/or in areas of reduced coverage compared to the overall mean coverage in the assembly—a distribution that strongly implicates known sources of error related to the Illumina sequencing platform, sequence assembly, and SNP calling (Dohm et al. 2008, Nakamura et al. 2011). In addition, many *Bolidomonas* SNPs were called in just one read direction. Taken together, these data strongly suggest that most, or probably all, of the very few SNPs called for *Bolidomonas* are false-positives.

Meiotic Genes in Bolidomonas

The 12 genes we surveyed are known to have meiosis-specific functions in model eukaryotes (Carr et al. 2010, Hanson et al. 2013, Malik et al. 2008). These 12 genes were divided into four categories: chromosome structure, double-strand break formation, crossover formation and resolution, and homologous-recombination-mediated DNA repair (Figure 2). Homologs for *REC8*, which is involved in chromosome cohesion and condensation, were absent from all surveyed heterokonts (Figure 2), consistent with previously described absences from many other protists that nevertheless undergo sexual reproduction (Malik et al. 2008, Ramesh et al. 2005). Homologous recombination is often necessary for accurate chromosome segregation during meiosis, and the initiation of this process usually involves formation of double-strand breaks. Homologs of meiosis-specific *SPO11-2* (involved in double-strand break formation in plants and some protists; Malik et al. 2007) were identified in all species, including *Bolidomonas* (Figure 2). Following double-strand break formation, synapsis is usually mediated by the synaptonemal complex, which connects homologs along their lengths. However, meiosis-specific homologs of *HOP1* and *ZYP1*, which are involved in this process, are missing from heterokonts and several other meiotic eukaryotes, suggesting that they may not be absolutely required for meiosis (Malik et al. 2008). Several other genes encoding proteins involved in strand exchange, crossover formation/resolution and homologous-recombination-mediated DNA repair were also present in *Bolidomonas* and diatoms (Figure 2). On the whole, *Bolidomonas* has all but one (*MSH4*) of the meiosis specific genes found in *Ectocarpus* and the four diatoms, all of which are known to undergo sexual reproduction (Figure 2). Importantly, absence of a gene from a transcriptome does not necessarily indicate absence from the genome, but instead that there were too few transcripts to be detected. Indeed, transcript levels of meiosis-related genes were universally low—in most cases less than 5% the level of a typical housekeeping gene—in *Bolidomonas* and the two diatoms, none of which showed signs of sexual reproduction at the time their RNA was harvested (Supplemental Figure 2). The presence of transcripts for meiotic genes in vegetatively dividing cells either reflects spurious, background-level transcription (possibly without translation) or suggests that some of the genes might be involved in a distinct non-meiotic pathway, such as DNA repair. Notably, expression of meiosis-specific genes in non-meiotic tissues has been reported in animals (Guikema et al. 2008, Zierhut et al. 2004), underscoring that transcription is not direct evidence for corresponding protein production and gene function.

Discussion

Their extraordinary species diversity and exclusively diplontic life history together set diatoms apart from all other algae, prompting decades of interest and speculation about the potential importance of their life cycle to their origin and diversification (Edlund and Stoermer 1997, Guillou 2011, Kooistra et al. 2007, Lewis 1984, Mann and Marchant 1989, Medlin et al. 1997, Sims et al. 2006). Despite exhibiting a broad range of variation in reproductive biology, diatoms all share the same fundamental diplontic life history, in which the short-lived, meiotically derived gametes (which do not undergo mitosis) represent the only haploid stage (Chepurnov et al. 2004, Drebes 1977). It is clear, therefore, that the common ancestor of diatoms was diplontic, but it is far less clear whether this is an ancestral or derived feature within heterokonts (Kooistra et al. 2007, Sims et al. 2006)—a question that requires knowledge of the life cycle of bolidophytes, the sister lineage to diatoms (Mann and Marchant 1989, Sims et al. 2006).

Mann and Marchant (1989) presented a detailed set of hypotheses about the evolutionary transition from an early haplontic "pre-diatom" to the diplontic common ancestor of diatoms. Their scenario included the prediction that Parmales, then classified as chrysophytes, is the closest living relative of diatoms. This part of their hypothesis was validated by molecular phylogenetic analyses, which resolved Parmales and bolidophytes, together, as sister to diatoms (Guillou et al. 1999, Ichinomiya et al. 2011). Although bolidophytes had not been discovered at the time Mann and Marchant (1989) formulated their hypotheses, they predicted that the transition from a haplontic ancestor to the diplontic life cycle of diatoms involved a haplodiplontic intermediate with two free-living, mitotically capable stages, one of which was a flagellated haploid cell and the other an unflagellated diploid cell.

Despite longstanding recognition that knowledge of the ploidy level of *Bolidomonas* is critical to unraveling the origin of the diatom life cycle (Guillou et al. 1999, Guillou 2011, Sims et al. 2006), the life history and ploidy of *Bolidomonas* are as poorly known today as when the first cultures were established some two decades ago (Guillou et al. 1999). With seemingly little hope of directly observing life-cycle transitions, the goal of this study was to provide the first estimate the ploidy level of *Bolidomonas*. To do so, we sequenced high-depth transcriptomes for *Bolidomonas* and two diploid diatoms, *Leptocylindrus* and *Hemiaulus*. We measured the heterozygosity in each transcriptome, with the hypothesis that diploids

will exhibit measurable levels of heterozygosity and haploids, by definition, will not. Indeed, roughly 0.1% of the nucleotides in the coding regions of the two diploid diatoms were polymorphic. The *Bolidomonas* transcriptome was, by contrast, devoid of allelic polymorphisms, consistent with expectations for a haploid genome. In addition, the presence or absence of a set of meiotic genes (rather than an individual gene) can serve as a reliable marker for the presence of meiosis and sexual reproduction in an organism (Schurko and Logsdon 2008). The *Bolidomonas* transcriptome contained a set of meiosis-related genes that point strongly to a capacity for sexual reproduction (Malik et al. 2008, Ramesh et al. 2005), despite no observations of sex either in either nature or in culture. Although certain meiotic genes were absent in *Bolidomonas,* these genes are also absent in other sexual taxa with known meiosis.

The absence of allelic polymorphisms in *Bolidomonas* could also be indicative of fully homozygous diploid genome. Indeed, there is evidence pointing to a loss of heterozygosity in some diatoms after just six months in culture (Murphy 1978). For diatoms that undergo sexual reproduction in culture, homozygotes could eventually become fixed either through stochastic processes or by selection for the particular culture conditions (Murphy 1978). Although the *Bolidomonas* culture sequenced here (CCMP1866) was first established 20 years prior to this study, we consider it unlikely to be a homozygous diploid for a several reasons. First, there are no reports of sexual reproduction in *Bolidomonas* cultures, which is the requisite first step in the formation of new homozygous genotypes and subsequent loss of parental heterozygotes from a clonal culture (Chepurnov et al. 2004, Murphy 1978). Second, 0.75% of sites in the genome of *T. pseudonana* remain polymorphic (Armbrust et al. 2004) despite >50 years in culture (Guillard and Ryther 1962). *Thalassiosira pseudonana* is, however, unusual in many respects (Alverson et al. 2011), including that it does not undergo size reduction during vegetative growth and so, consequently, may have gone without sexual reproduction during all its years in culture (Chepurnov et al. 2008). Assuming for a moment that *Bolidomonas* is diploid, the different genomic consequences of long-term culture between it and *T. pseudonana* could not be more striking. Although both species are likely capable of meiosis and sexual reproduction (Figure 2), neither appears to have undergone sex in culture. All the while *T. pseudonana* has remained highly heterozygous, and *Bolidomonas* has suffered a *complete* loss of heterozygosity. The simplest explanation for the difference in heterozygosity is their

underlying base difference in ploidy level, not the result of drastically different genomic trajectories while in culture.

Phylogenetic analyses and DNA sequence comparisons leave little doubt that bolidophytes and parmaleans are one and the same (Ichinomiya et al. 2011), with some culture strains from the two groups showing 100% identity in their SSU rDNA genes (e.g., GenBank GIs 321116571 and 554595637). The results presented here shed new important light on the nature of their relationship, suggesting that the two forms are different phases of a single heteromorphic, haplodiplontic life cycle, in which bolidophytes are the vegetatively haploid phase and parmaleans are the vegetatively diploid phase. The broad outlines of this hypothesis were first laid out clearly by Mann and Marchant (1989), and with knowledge of bolidophytes and their close relationship to Parmales, the hypothesis was fully articulated by Guillou (2011). If correct, a sequenced Parmales genome will be found to harbor heterozygous alleles at positions masquerading here as homozygous in *Bolidomonas*.

Despite this step forward, other critical observations are still missing and may go undiscovered unless phase transitions in the bolidphyte/parmalean life cycle are fortuitously happened upon in the wild or whether life-cycle transitions can be manipulated in culture. It remains unclear, for example, whether the life cycle includes a resting stage, what a resting stage might look like, and whether it is mitotically or meiotically regenerated. Less clear still is whether the origin of the diplontic life cycle of diatoms included all of the transitional stages outlined by Mann and Marchant (1989). Stated differently, is Parmales *the* sister lineage—either extant or extinct—to diatoms? The answer to this question obviously has profound consequences for understanding the origin of diatoms. Marine metagenomic studies have yet to reveal a closer living relative, and no obvious candidates have emerged from stratigraphic studies. Given how long it took to resolve the relationship between diatoms and parmaleans (Guillou et al. 1999, Ichinomiya et al. 2011), it seems reasonable to expect that important gaps in our knowledge of the origin of diatoms remain unfilled.

Here, we present one in a growing number of studies that have leveraged genomic data to understand basic aspects of the biology of organisms that are not easily cultured or characterized using traditional methods. For example, the complete genome of an uncultured Archaean culled from a large marine shotgun metagenomic dataset gave insights into the physiology, metabolism, and evolution of this

poorly understood, but functionally important, group of marine microbes (Iverson et al. 2012). More directly related to the results presented here, a transcriptome-based study of another haplodiplontic phytoplankton species, *Emiliana huxleyi*, revealed transcriptional differences across thousands of genes between the haploid and diploid stages (von Dassow et al. 2009). The divergent, phase-specific transcription profiles found in *E. huxleyi* likely extend to the *Bolidomonas*/Parmales life cycle as well. For example, transcripts of sexually induced genes (*SIGs*), which encode flagellar mastigoneme proteins in heterokonts (Honda et al. 2007), were abundant in the *Bolidomonas* transcriptome but present only at very low levels in the two diatoms (not shown), which only express mastigonemes in male gametes. In the other direction, the *Bolidomonas* transcriptome contained no transcripts for silicon transporter (*SIT*) genes, which were present in high abundance in both diatoms (not shown). In light of this, the *Bolidomonas* transcriptome reported here might represent a fraction of the overall genes present in the *Bolidomonas*/Parmales genome.

Resolving the ploidy level of *Bolidomonas* represents a considerable step forward towards understanding the origin of the diplontic life cycle of diatoms. However, a complete understanding of the adaptive values, if any, of the cell wall and diplontic life history of diatoms requires a better understanding of the co-evolution of these two traits (Lewis 1984), which are intimately linked by the size diminution that governs the timing of sexual reproduction in most species. Large-scale genomic and metagenomic datasets show great promise in motivating the development of new tools and novel approaches for addressing these and other difficult, longstanding questions that have evaded answer by traditional means.

Acknowledgments

We thank Matt Ashworth and Ed Theriot for sharing cultures of *Hemiaulus* and *Leptocylindrus*. We thank Bob Andersen for sharing insights into heterokont life histories, and we thank Lizzie Goodwin-Horn for helping with analyses of the meiosis genes. Computational analyses were conducted using resources available through the Arkansas High Performance Computing Center (AHPCC) at the University of Arkansas. Resources managed by AHPCC are supported in part by NSF grants MRI 0722625, MRI-R2

0959124, and a grant from the Arkansas Science and Technology Authority. This research was funded with start-up funds from the University of Arkansas to Andrew Alverson.

References

Alverson A.J., Beszteri B., Julius M.L., Theriot E.C. (2011). The model marine diatom *Thalassiosira pseudonana* likely descended from a freshwater ancestor in the genus *Cyclotella*. *BMC Evol. Biol.* 11: 125.

Andrews S. (2009). FastQC: A quality control tool for high throughput sequence data.

Armbrust E.V., 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.

Bell G. (1997). The evolution of the life cycle of brown seaweeds. *Biol. J. Linn. Soc.* 60: 21-38.

Birney E., Clamp M., Durbin R. (2004). GeneWise and Genomewise. *Genome Res.* 14: 988-995.

Booth B.C., Marchant H.J. (1987). Parmales, a new order of marine Chrysophytes, with descriptions of three new genera and seven new species1, 2. *J. Phycol.* 23: 245-260.

Carr M., Leadbeater B.S., Baldauf S.L. (2010). Conserved meiotic genes point to sex in the choanoflagellates. *J. Eukaryot. Microbiol.* 57: 56-62.

Chepurnov V.A., Mann D.G., Sabbe K., Vyverman W. (2004). Experimental studies on sexual reproduction in diatoms. *Int. Rev. Cytol.* 237: 91-154.

Chepurnov V.A., Mann D.G., von Dassow P., Vanormelingen P., Gillard J., Inze D., Sabbe K. et al. (2008). In search of new tractable diatoms for experimental biology. *BioEssays* 30: 692-702.

Conesa A., Gotz S., Garcia-Gomez J.M., Terol J., Talon M., Robles M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.

Cronberg G. (2005). The life cycle of *Gonyostomum semen* (Raphidophyceae). *Phycologia* 44: 285-293.

Daugbjerg N., Guillou L. (2001). Phylogenetic analyses of Bolidophyceae (Heterokontophyta) using *rbc*L gene sequences support their sister group relationship to diatoms. *Phycologia* 40: 153-161.

Dohm J.C., Lottaz C., Borodina T., Himmelbauer H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* 36: e105.

Drebes G. (1977). Sexuality. Werner, D.(ed.) The Biology of Diatoms: 250-283, University of California Press.

Edlund M.B., Stoermer E.F. (1997). Ecological, evolutionary, and systematic significance of diatom life histories. *J. Phycol.* 33: 897-918.

Fuhrmann T., Landwehr S., El Rharbi-Kucki M., Sumper M. (2004). Diatoms as living photonic crystals. *Appl. Phys. B* 78: 257-260.

Goertzen L.R., Theriot E.C. (2003). Effect of taxon sampling, character weighting, and combined data on the interpretation of relationships among the heterokont algae. *J. Phycol.* 39: 423-U424.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X. et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29: 644-U130.

Guikema J.E.J., Schrader C.E., Leus N.G.J., Ucher A., Linehan E.K., Werling U., Edelmann W. et al. (2008). Reassessment of the Role of Mut S Homolog 5 in Ig Class Switch Recombination Shows Lack of Involvement in cis- and trans-Switching. *J. Immunol.* 181: 8450-8459.

Guillard R.R.L., Ryther J.H. (1962). Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt and *Detonula confervacea* (Cleve) Gran. *Can. J. Microbiol.* 8: 229-239.

Guillard R.R.L. (1975). Culture of phytoplankton for feeding marine invertebrates. In: Smith WL, Chanley MH (eds). *Culture of Marine Invertebrate Animals*. Plenum Press: New York. pp 26-60.

Guillou L., Chretiennot-Dinet M.-J., Medlin L.K., Claustre H., Loiseaux-de G.S., Vaulot D. (1999). *Bolidomonas*: A new genus with two species belonging to a new algal class, the Bolidophyceae (Heterokonta). *J. Phycol.* 35: 368-381.

Guillou L. (2011). Characterization of the Parmales: Much more than the resolution of a taxonomic enigma. *J. Phycol.* 47: 2-4.

Hamm C.E., Merkel R., Springer O., Jurkojc P., Maier C., Prechtel K., Smetacek V. (2003). Architecture and material properties of diatom shells provide effective mechanical protection. *Nature* 421: 841-843.

Hanson S.J., Schurko A.M., Hecox-Lea B., Welch D.B.M., Stelzer C.-P., Logsdon J.M. (2013). Inventory and phylogenetic analysis of meiotic genes in monogonont rotifers. *J. Hered.* 104: 357-370.

Honda D., Shono T., Kimura K., Fujita S., Iseki M., Makino Y., Murakami A. (2007). Homologs of the Sexually Induced Gene 1 (*sig1*) Product constitute the Stramenopile Mastigonemes. *Protist* 158: 77-88.

Ichinomiya M., Yoshikawa S., Kamiya M., Ohki K., Takaichi S., Kuwata A. (2011). Isolation and characterization of Parmales (Heterokonta/Heterokontophyta/Stramenopiles) from the Oyashio region, Western North Pacific. *J. Phycol.* 47: 144-151.

Iverson V., Morris R.M., Frazar C.D., Berthiaume C.T., Morales R.L., Armbrust E.V. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* 335: 587-590.

Kooistra W., Gersonde R., Medlin L.K., Mann D.G. (2007). The origin and evolution of the diatoms: their adaptation to a planktonic existence. *Evolution of primary producers in the sea*: 207-249.

Langmead B., Trapnell C., Pop M., Salzberg S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10: R25.

Lewis W.M. (1984). The Diatom Sex Clock and Its Evolutionary Significance. *Am. Nat.* 123: 73-80.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G. et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078-2079.

Malhis N., Jones S.J. (2010). High quality SNP calling using Illumina data at shallow coverage. *Bioinformatics* 26: 1029-1035.

Malik S.-B., Ramesh M.A., Hulstrand A.M., Logsdon J.M. (2007). Protist homologs of the meiotic Spo11 gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. *Mol. Biol. Evol.* 24: 2827-2841.

Malik S.-B., Pightling A.W., Stefaniak L.M., Schurko A.M., Logsdon Jr J.M. (2008). An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis*. *PLoS One* 3: e2879.

Mann D., Marchant H. (1989). The origins of the diatom and its life cycle. *The chromophyte algae: problems and perspectives. Clarendon Press, Oxford*: 307-323.

Mann D.G., Droop S.J.M. (1996). Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia* 336: 19-32.

Mann D.G., Chepurnov V.A. (2004). What have the Romans ever done for us? The past and future contribution of culture studies to diatom systematics. *Nova Hedwigia* 79: 1-2.

Mann D.G., Vanormelingen P. (2013). An inordinate fondness? The number, distributions, and origins of diatom species. *J. Eukaryot. Microbiol.* 60: 414-420.

Medlin L., Kooistra W., Gersonde R., Sims P., Wellbrock U. (1997). Is the origin of diatoms related to the end-Permian mass extinction?

Medlin L.K., Williams D.M., Sims P.A. (1993). The evolution of the diatoms (Bacillariophyta). I. Origin of the group and assessment of the monophyly of its major divisions. *Eur. J. Phycol.* 28: 261-275.

Milne I., Stephen G., Bayer M., Cock P.J., Pritchard L., Cardle L., Shaw P.D. et al. (2013). Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* 14: 193-202.

Moore L.R., Post A.F., Rocap G., Chisholm S.W. (2002). Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* 47: 989-996.

Murphy L.S. (1978). Biochemical taxonomy of marine phytoplankton by electrophoresis of enzymes. II. Loss of heterozygosity in clonal cultures of centric diatoms *Skeletonema costatum* and *Thalassiosira pseudonana*. *J. Phycol.* 14: 247-250.

Nakamura K., Oshima T., Morimoto T., Ikeda S., Yoshikawa H., Shiwa Y., Ishikawa S. et al. (2011). Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* 39: e90.

Pavy N., Parsons L.S., Paule C., MacKay J., Bousquet J. (2006). Automated SNP detection from a large collection of white spruce expressed sequences: contributing factors and approaches for the categorization of SNPs. *BMC Genomics* 7: 174.

Ramesh M.A., Malik S.-B., Logsdon Jr J.M. (2005). A phylogenomic inventory of meiotic genes: evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr. Biol.* 15: 185-191.

Raven J.A. (1983). The Transport and Function of Silicon in Plants. *Biol. Rev. Camb. Philos. Soc.* 58: 179-207.

Round F.E., Crawford R.M. (1981). The Lines of Evolution of the Bacillariophyta .1. Origin. *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* 211: 237-+.

Schurko A.M., Logsdon J.M. (2008). Using a meiosis detection toolkit to investigate ancient asexual “scandals” and the evolution of sex. *BioEssays* 30: 579-589.

Sims P.A., Mann D.G., Medlin L.K. (2006). Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* 45: 361-402.

Sorhannus U. (2007). A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. *Mar. Micropaleontol.* 65: 1-12.

Van den Hoek C., Mann D., Jahns H.M. (1995). An Introduction to Phycology. Cambridge University Press, Cambridge.

von Dassow P., Ogata H., Probert I., Wincker P., Da Silva C., Audic S., Claverie J.-M. et al. (2009). Transcriptome analysis of functional differentiation between haploid and diploid cells of *Emiliania huxleyi*, a globally significant photosynthetic calcifying cell. *Genome Biol.* 10: R114.

von Dassow P., Montresor M. (2011). Unveiling the mysteries of phytoplankton life cycles: patterns and opportunities behind complexity. *J. Plankton Res.* 33: 3-12.

Yamaguchi M., Imai I. (1994). A microfluorometric analysis of nuclear DNA at different stages in the life history of *Chattonella antiqua* and *Chattonella marina* (Raphidophyceae). *Phycologia* 33: 163-170.

Yang E.C., Boo G.H., Kim H.J., Cho S.M., Boo S.M., Andersen R.A., Yoon H.S. (2012). Supermatrix data highlight the phylogenetic relationships of photosynthetic stramenopiles. *Protist* 163: 217-231.

Zierhut C., Berlinger M., Rupp C., Shinohara A., Klein F. (2004). Mnd1 is required for meiotic interhomolog repair. *Curr. Biol.* 14: 752-762.

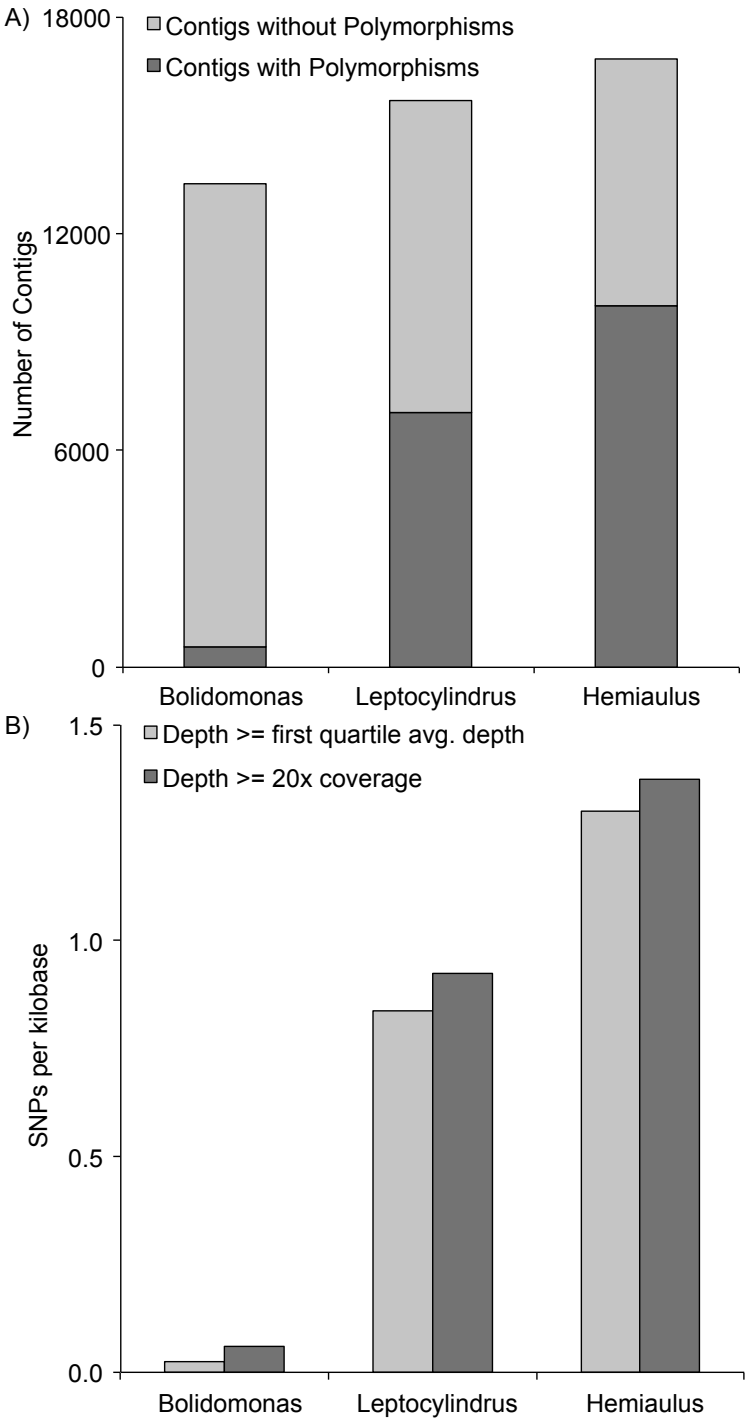


Figure 1. Allelic single nucleotide polymorphisms (SNPs) in the transcriptomes of *Bolidomonas pacifica* and two diatoms, *Leptocylindrus danicus* and *Hemiaulus sinensis*. A) The number of contigs (i.e., genes) with at least one SNP as a subset of the total contigs in each assembly. B) The SNP density in each transcriptome.

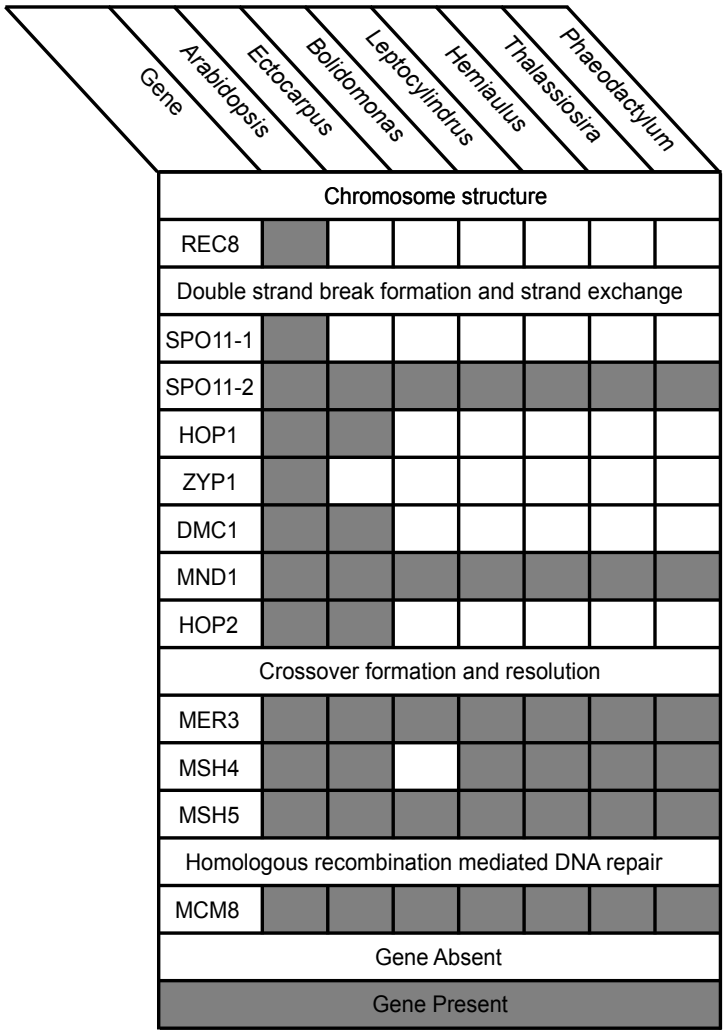
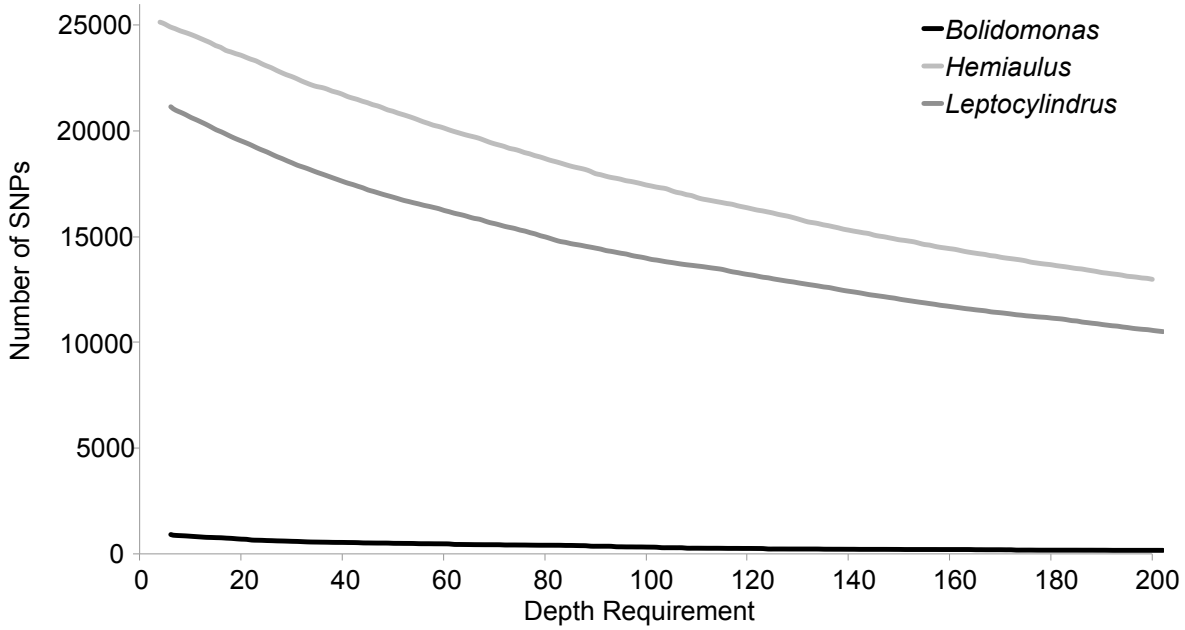
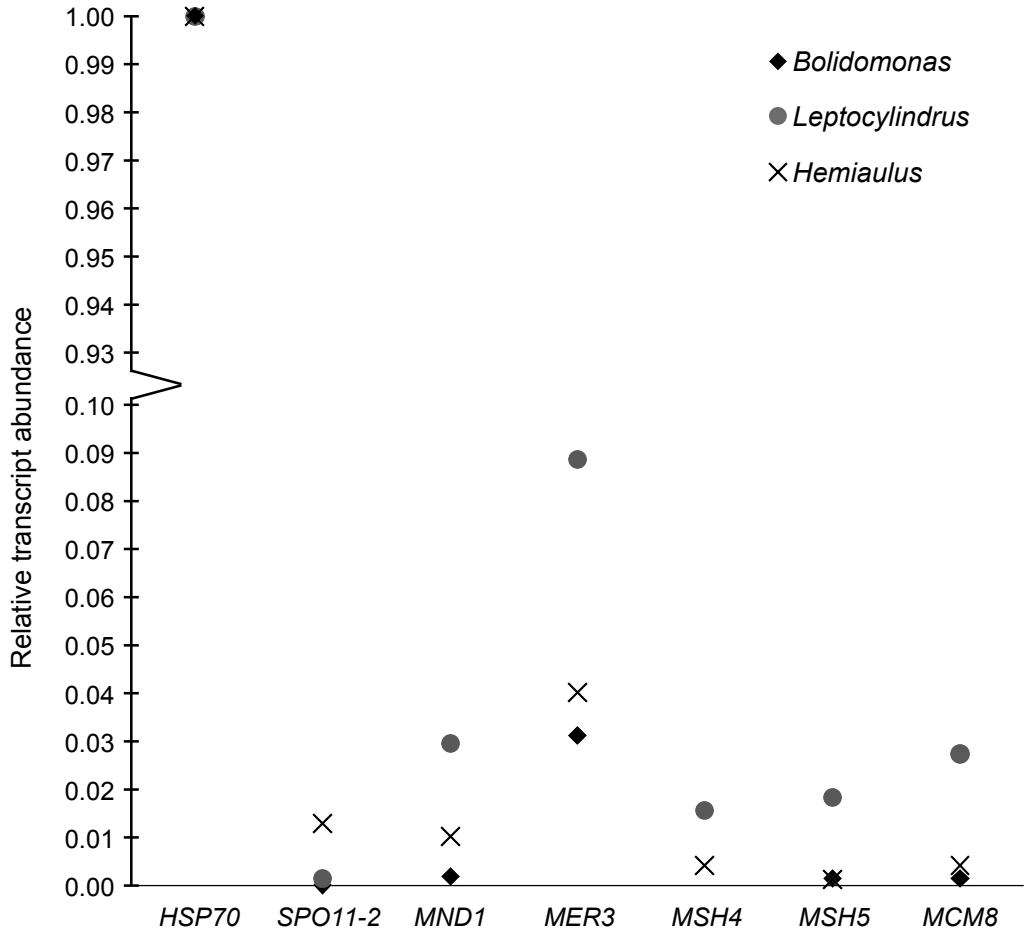


Figure 2. An inventory of meiosis genes, grouped according to function, in the model angiosperm *Arabidopsis thaliana*, the brown alga *Ectocarpus siliculosus*, *Bolidomonas pacifica,* and four diatoms: *Leptocylindrus danicus*, *Hemiaulus sinensis, Thalassiosira pseudonana,* and *Phaeodactylum tricornutum*.



Supplementary Figure 1. The total number of allelic polymorphisms at different read-depth requirements, ranging from a minimum 4 to 200 reads necessary to infer a SNP at a given position.



Supplementary Figure 2. Transcript levels for meiosis-specific genes relative to the general housekeeping gene, HSP70, which was set to a value of 1.

Appendix

UNIVERSITY OF ARKANSAS

J. William Fulbright College of Arts and Sciences
Department of Biological Sciences

11 March 2014

To Whom It May Concern:

Colton Kessenich submitted an article entitled "*Transcriptomic insights into the life history of bolidophytes, the sister lineage to diatoms*" to the Journal of Phycology on 09 March 2014. This manuscript represent the bulk of the work for his Master's thesis.

Although this article was authored through the collaborative efforts of a number of individuals, Colton Kessenich is the primary author of the publication and contributed well in excess of 51% of the work, including both data analysis and writing of the manuscript.

Thank you,

Andrew Alverson, Ph.D.
Thesis Director

Science Engineering, Room 601 • Fayetteville, AR 72701-1201 • 479-575-3251 • Fax: 575-4010 • www.uark.edu

CONCLUDING REMARKS

The linking of life-history stages ideally integrates morphological, genomic, ecological, and environmental data. Unfortunately it is commonplace to have only a limited subset of these types of data, rendering a comprehensive analysis difficult or impossible (Richerd et al. 1993, von Dassow and Montresor 2011). In this study, I provided new insights into the life cycle of bolidophytes, the closest known relatives to diatoms—namely that bolidomonas appears to be haploid, which when taken with our current knowledge of the other species in the genus, supports the idea of a haplodiplontic life cycle. As the number of sequenced transcriptomes and genomes of diatoms and their relatives increases, new insights into the biology of both diatoms and bolidophytes will be possible.

Further characterization of *Bolidomonas*—or the Parmales—from either observational or genomic data, will continue to help resolve the events that led to the origin and evolutionary diversification of diatoms, which remains a central focus research on diatoms (e.g. Allen et al. 2006, Allen et al. 2011, Alverson et al. 2007, Alverson et al. 2011, Armbrust et al. 2004, Bowler et al. 2008). The diverse life-history strategies found within heterokonts make it difficult to separate ancestral vs. derived conditions. Resolution of the haplodiplontic of bolidophytes adds some clarity, but more work is needed to determine whether the diplontic life cycle of diatoms is a an ancestral or derived trait.

References

Allen A.E., Vardi A., Bowler C. (2006). An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr. Opin. Plant Biol.* 9: 264-273.

Allen A.E., Dupont C.L., Obornik M., Horak A., Nunes-Nesi A., McCrow J.P., Zheng H. et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473: 203-207.

Alverson A.J., Jansen R.K., Theriot E.C. (2007). Bridging the Rubicon: Phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Mol. Phylogenet. Evol.* 45: 193-210.

Alverson A.J., Beszteri B., Julius M.L., Theriot E.C. (2011). The model marine diatom *Thalassiosira pseudonana* likely descended from a freshwater ancestor in the genus *Cyclotella*. *BMC Evol. Biol.* 11: 125.

Armbrust E.V., 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.

Bowler C., Allen A.E., Badger J.H., Grimwood J., Jabbari K., Kuo A., Maheswari U. et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456: 239-244.

Richerd S., Couvet D., Valéro M. (1993). Evolution of the alternation of haploid and diploid phases in life cycles. II. Maintenance of the haplo-diplontic cycle. *J. Evol. Biol.* 6: 263-280.

von Dassow P., Montresor M. (2011). Unveiling the mysteries of phytoplankton life cycles: patterns and opportunities behind complexity. *J. Plankton Res.* 33: 3-12.