5-2022

Impact of Genetic Variation and Timescale on Diatom Salinity Stress Response

Kala M. Downey
University of Arkansas, Fayetteville

Impact of Genetic Variation and Timescale on Diatom Salinity Stress Response

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Biology

by

Kala M. Downey
Converse College
Bachelor of Science in Biology and Chemistry, 2012
Austin Peay State University
Master of Science in Biology, 2016

May 2022
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council

Andrew Alverson, Ph.D.
Dissertation Director

Jeffery Lewis, Ph.D.
Committee Member

Adam Siepielski, Ph.D.
Committee Member

Burt Bluhm, Ph.D.
Committee Member

Jeremy Beaulieu, Ph.D.
Committee Member

Abstract

Natural environments are dynamic, and organisms must sense and respond to changing conditions. One common way organisms deal with stressful environments is through gene expression changes, allowing for stress acclimation and resistance which occurs over varying time spans in different species. The recent evolutionary history of populations could greatly influence their ability to respond successfully. An evolutionary history in disturbed or fluctuating conditions could promote increased resistance or a more rapid response to these environmental stressors. To understand the impact of genotypic variation and timescales on response and acclimation to salinity changes, we have been exploiting the abilities of euryhaline diatoms in the order Thalassiosirales.

This dissertation explores the mechanisms two species of Thalassiosirales use to mitigate short- and long-term effects of salinity stress. We first clarified the phylogenetic relationships of *Cyclotella*, one of the largest clades in the order containing numerous marine—freshwater transitions, reclassifying the relatively new genus *Spicaticribra* as a member of *Cyclotella* based on phylogenetic analyses of the genes *rbcL, psbC, SSU*, and *LSU*. We then determined that variation derived from genotypic differences between strains of *Cyclotella cryptica* had a greater impact then that imposed by gene expression changes following acclimation to different salinity conditions. When pooled together, the primary transcriptome modifications were related to the regulation of compatible solutes and ion transporters in an effort to maintain the osmotic gradient in suboptimal salinities. Subsequently, we acclimated multiple strains of *Skeletonema marinoi*, another euryhaline member of Thalassiosirales, to a range of salinities. With sufficient technical replication of these strains, we were able to determine that the variation between strains was due largely to degree of differential expression of genes, rather than strains regulating different genes entirely. Including genetic variation allowed us to identify more genes impacted by salinity changes than any one strain would have revealed and, additionally, we found a small set of 27 shared genes that were differentially

expressed in all strains. We argue that this core set may have played an important role in ancestral marine—freshwater transitions for *S. marinoi*. Similar to *C. cryptica*, averaged response across all strains relied heavily on long-term maintenance of the cellular osmotic gradient, but there was little other indication of severe stress in the salinity treatments we used.

We found considerable disparity in comparing the results of our post-acclimated *C. cryptica* behavior with previously reported short-term stress behavior in the species. Thus, we conducted a short-term time series experiment looking at *C. cryptica'*s gene expression response to a rapid exposure to freshwater. Differential gene expression analysis concluded *C. cryptica* responds to freshwater shock by temporarily halting growth and downregulating genes associated with maintenance of cellular division, such as ribosome biogenesis, transcription, and translation. Genes involved in reactive oxygen species scavenging and regulation of the osmotic gradient (i.e., osmolytes and ion transporters) are upregulated. Although limited, our comparison of results from different timescales found little overlap between the regulated genes during the peak period of stress, initial acclimation, and post-acclimation. This suggests that there is wide variation in the genes responsible for various stages of the stress response.

These experiments highlight the power of including genetic variation to uncover mechanisms utilized in environmental stress response, as well as the importance of considering how that response changes from exposure to acclimation in order to develop a complete picture of how organisms cope with stress.

Table of Contents

List of Tables

List of Figures

red gradient arrow), distinguishing genes that differed significantly in either the magnitude or direction of their response to low salinities. The first category includes genes that were DE in one strain but not the others, or that were DE in multiple strains but with significant differences in logFC values in the same direction. Genes of the second category were significantly upregulated in some strains, whereas they were significantly downregulated in other strains.

Fig. 4.2. Growth response of Baltic *S. marinoi* in low salinities. Growth rates of the eight *S. marinoi* strains examined in this study at three different salinities. The letters in the individual panels correspond with the sampling locations in Table 1 ('Collection ID') and Fig. 1A. Each point represents a single estimate of the slope of the natural logarithm of in vivo relative fluorescence against time for each sequential transfer, using a horizontal jitter of points to avoid overplotting. 74

Fig. 4.3. Transcriptome response of Baltic *S. marinoi* to low salinities. a Number of DE genes at a 5 % FDR-level in the average response and the individual strains. The number of DE genes is indicated separately for each contrast, distinguishing between genes that are up- or downregulated. **b** Direction of DE in the top-100 genes of the average response and individual strains as selected by P-value or logFC. For each contrast in the average and individual strains (vertical black bar), the direction of DE is indicated for the top-100 genes selected by stageR's FDR-adjusted P-value of the global null hypothesis (Padjscreen). Thus, although a gene can have a high P-value on a dataset-wide level, it is not necessarily DE in each individual contrast. In addition, we show the top-100 genes selected by logFC (topconfects [100]) and the contrast-specific 5 % FDR-controlled P-value (Padj) for the 8-24 contrast of the average effects, as this contrast showed the greatest number of DE genes in **a**. **c** Number of enriched GO terms for the ORA and GSEA analyses. The number of up- and downregulated GSEA GO terms represents the output classification by CAMERA. The number of enriched GO terms includes Biological Process, Molecular Function and Cellular Component GO terms, prior to removal of redundant GO terms by REVIGO. 75

Fig 4.4. GO enrichment on the average response of *S. marinoi* to low salinities: Biological Process. The results of two types of GO enrichment analyses are shown: ORA (in topGO, Fisher's exact test, *elim* algorithm) and GSEA (in CAMERA), after removal of redundant terms by REVIGO. For ORA, we classified the total set of DE genes in the average response into two categories, distinguishing between genes that are up- or downregulated in low salinities, regardless of salinity contrast (see Supplementary Methods for details). For CAMERA, we performed GSEA analyses on each individual contrast separately, showing only the 8-24 contrast in this figure. Barplot height indicates the proportion of genes that are DE with a given GO term to the total number of genes with this GO term in the genome of *S. marinoi*. The barplots are colored according to P-value. Within the set of up- and downregulated genes, the GO terms are ranked from lowest to highest P-value, using the lowest of two P-values from ORA or GSEA. Symbols indicate major categories of cellular processes to which a GO term belongs. Only Biological Process GO terms are shown. 78

Fig. 4.5. Set of genes that are DE in at least one contrast of each strain: the core response. The heatmap shows logFC values for the individual strains and average response of the 27 core response genes. The three salinity combinations are indicated on top of the figure. Contrasts that were significant are outlined in black. Row names 79

List of Published papers

Chapter 2:

Downey KM, Julius ML, Theriot EC, Alverson AJ. 2021. Phylogenetic analysis places *Spicaticribra* within *Cyclotella.* Diatom Research 36(2):93-99.

Chapter 3:

Nakov T, Judy KJ, Downey KM, Ruck EC, Alverson AJ. 2020. Transcriptional response of osmolyte synthetic pathways and membrane transporters in a euryhaline diatom during long-term acclimation to a salinity. Journal of Phycology 56:1712-1728.

Chapter 4:

Pinseel E, Nakov T, Van den Berge K, Downey KM, Judy KJ, Kourtchenk O, Kremp A, Ruck EC, Sjöqvist C, Töpel M, Godhe A, Alverson AJ. Strain-specific transcriptional responses overshadow salinity effects in a marine diatom sampled along the Baltic Sea salinity cline. The ISME Journal. https://doi.org/10.1038/s41396-022-01230-x

Chapter 1 Introduction

1.1 Diatoms in stress research.

Diatoms are a group of ubiquitously occurring unicellular brown algae with a distinct silica-based cell wall. Diatoms alone account for ≥ 40% of marine primary production, which translates to roughly one-quarter of global net primary production (Werner 1977, Nelson et al. 1995). With natural environments being dynamic, organisms must be able to recognize and respond to various environmental conditions throughout their life. While multicellular organisms have specialized organs and tissues to assist with this, microorganisms, such as diatoms, have evolved mechanisms for adapting to diverse environmental conditions. During stress, cells must be able to switch between cell growth, survival, and death. One way to survive stress and adjust to long-term change is to alter gene expression.

The salinity gradient between marine and freshwaters is of particular interest in the microbial field. Changes in osmotic pressure between these environments represents one of the major ecological divides structuring microbial diversity (Lozupone and Knight 2007), with rare transitions taking place over considerable evolutionary timescales and often succeeded by diversification events (Logares et al. 2009, Cavalier-Smith 2009, Nakov et al. 2019). As such, establishing how marine-freshwater transitions are accomplished is vital to our comprehension of lineage diversification on evolutionary scales. In shorter timescales, increases in precipitation and ice melt are causing salinity declines on unprecedented rapid time scales (Rabe et al. 2011, Durack et al. 2012, McCrystall et al. 2021).

Numerous stress studies at varying timescales have focused primarily on how model organisms (e.g.*, Saccharomyces cerevisiae, Escheria coli,* and *Phaeodactylum tricornutum*) respond to different stress conditions, including salinity (Gasch et al. 2000, Causton et al. 2001, Cánovas et al. 2007, Arense et al. 2010, Jozefczuk et al. 2010, De Martino et al. 2011, Levitan et al. 2015), but increasing efforts are being made to include non-model organisms in these

studies. One of the shortcomings of model organisms is that, because they are so well-suited to laboratory settings, findings based on these organisms are not always generalizable even to members of their own genus. For instance, *P. tricornutum* is treated as a model for diatom research. However, it differs from other diatoms in several ways, most notably the very low quantity of silica in the cell wall which is of great importance in diatom salinity responses (Costaouëc et al. 2017), making it difficult to generalize findings across multiple diatom species. To develop a broader understanding of how different diatoms respond to salinity changes in the environment, we focused our work across a broader range of related species to observe gene expression behavior. The diatom order Thalassiosirales has an extensive evolutionary history of transitions between marine and freshwater (Alverson et al. 2007, Alverson et al. 2011), which can greatly impact the ability of a species to respond to rapid and prolonged salinity change (James et al. 2003, Logares et al. 2009).

1.1.1 Importance of accurate taxonomic classifications in stress studies.

The recent evolutionary history of populations could greatly influence their ability to survive rapid salinity change (James et al. 2003, Logares et al. 2009). An evolutionary history in disturbed or fluctuating conditions could lead to the evolution of life history traits that promote survival under environmental change (Lee and Gelembiuk 2008). Thus, it is vital that taxonomic nomenclature provides an accurate picture of species relationships. Thalassiosirales is a large order of hundreds of diatom species with a complex taxonomic history with multiple instances of monotypic genera and polyphyletic clades requiring reclassification to meld morphological and phylogenetic data. Polyphyletic clades are problematic when considering related traits on an evolutionary timescale as taxonomic classifications that do not accurately portray ancestral relationships can lead to misinterpretations. As such, it is vital that existing taxonomic classifications accurately reflect relationships among organisms.

1.1.2 Salinity stress in diatoms.

Cell exposed to suboptimal salinity environments experience both osmotic and ionic stress. The gene expression changes induced by osmotic and salt stress are very similar (Rep et al. 2000, Causton et al. 2001). Increased concentrations in salt (hyperosmotic) outside the cells will immediately cause cells to lose intracellular water causing cell shrinkage. Decreased salt concentrations (hypoosmotic) outside the cell lead to water uptake causing cells to swell with the potential to burst, thus cells have evolved ways to combat changes in osmolarity to maintain cellular homeostasis (Wood 1999). Osmotic stress has multiple effects on cells: morphological, transport, and metabolic adjustments (Tamás and Hohmann 2003).

In algae, reaction to osmotic begins with a very rapid change (seconds) in turgor pressure caused by water fluxes in or out of the cells following the osmotic gradient (Scholz and Liebezeit 2012). The expansion or shrinkage of the plasma membrane triggers the activation of ionic transporters, especially Na+/K+ pumps. Growth in elevated NaCl concentrations changes the intracellular Na+/K+ ratio and high ratios are toxic. If ionic levels are too low this causes damage to the cell and inhibits growth (Wadskog and Adler 2003, Pozdnyakov et al. 2020). Thus, in hypoosmotic conditions, the cell must carefully modulate the export of these ions when balancing the osmotic gradient. In general, changes in the concentrations of these inorganic ions precede the accumulation of compatible solutes, either through biosynthesis or absorption from the surrounding environment, which tends to occur over a span of minutes or hours (Van Bergeijk et al. 2003). These organic solutes are low-molecular-weight compounds that are osmotically active, are able to stabilize macromolecules such as enzymes, and are compatible with cell metabolism (Burg 1995, Yancey 2005). In diatoms, some of the most common osmolytes utilized in response to osmotic stress include dimethylsulfoniopropionate (DMSP), glycine-betaine, and proline (Liu and Hellebust 1974, Jackson et al. 1992, Nakov et al. 2020). The increasing number of studies investigating how diatoms respond to rapid and gradual salt fluctuations in the environment on both short- and long-term time frames are identifying

variations that will provide valuable insight into historical transitions between marine and freshwaters (Rijstenbil et al. 1989, Lyon et al. 2016, Nakov et al. 2020). Transcriptome-based research provides the opportunity to characterize and compare the changes in gene expression that underlie these processes.

1.2 Dissertation outline

The dissertation presents unique research focusing on using gene expression in two species of diatoms to compare initial shock vs post-acclimation transcriptome data in response to salinity change and highlight the impact of genotypic variation on the salinity response. Chapter 2 reclassifies *Spicaticribra kingstonii* as a *Cyclotella*, along with nine other members of the *Spicaticribra* genus. This reestablishes the monophyly of *Cyclotella* and aligns known phylogenetic relationships with taxonomic nomenclature. Chapter 3 identifies the confounding role of genotypic variation in gene expression studies and details the post-acclimation response of the euryhaline *Cyclotella cryptica* to high and low salinity after 120 days. Chapter 4 further investigates the magnitude of the impact of genetic variation in diatom salinity responses following acclimation to new environments using another euryhaline species, *Skeletonema marinoi*. In doing so, we were able to identify a small set of shared differentially expressed genes that may have been an essential part of the ancestral mitigation of salinity stress. Chapter 5 focuses on the transcriptomic alterations *C. cryptica* experiences on a shorter timescale of only 10 hours in full freshwater. We find that there is considerable disparity between the genes regulated during the peak period of stress response and those that enable the initial stages of acclimation and post-acclimation persistence.

1.3 References

Alverson, Andrew J., Bánk Beszteri, Matthew L. Julius, and Edward C. Theriot. 2011. “The Model Marine Diatom Thalassiosira Pseudonana Likely Descended from a Freshwater Ancestor in the Genus Cyclotella.” BMC Evolutionary Biology 11: 125.

Alverson, Andrew J., Robert K. Jansen, and Edward C. Theriot. 2007. “Bridging the Rubicon: Phylogenetic Analysis Reveals Repeated Colonizations of Marine and Fresh Waters by Thalassiosiroid Diatoms.” Molecular Phylogenetics and Evolution 45 (1): 193–210.

Arense, Paula, Vicente Bernal, José L. Iborra, and Manuel Cánovas. 2010. “Metabolic Adaptation of Escherichia Coli to Long-Term Exposure to Salt Stress.” Process Biochemistry 45 (9): 1459–67.

Burg, M. B. 1995. “Molecular Basis of Osmotic Regulation.” The American Journal of Physiology 268 (6 Pt 2): F983–96.

Cánovas, M., V. Bernal, A. Sevilla, T. Torroglosa, and J. L. Iborra. 2007. “Salt Stress Effects on the Central and Carnitine Metabolisms of Escherichia Coli.” Biotechnology and Bioengineering 96 (4): 722–37.

Causton, H. C., B. Ren, S. S. Koh, C. T. Harbison, E. Kanin, E. G. Jennings, T. I. Lee, H. L. True, E. S. Lander, and R. A. Young. 2001. “Remodeling of Yeast Genome Expression in Response to Environmental Changes.” Molecular Biology of the Cell 12 (2): 323–37.

Cavalier-Smith, Thomas. 2009. “Megaphylogeny, Cell Body Plans, Adaptive Zones: Causes and Timing of Eukaryote Basal Radiations.” The Journal of Eukaryotic Microbiology 56 (1): 26–33.

Costaouëc, Tinaïg Le, Tinaïg Le Costaouëc, Carlos Unamunzaga, Lalia Mantecon, and William Helbert. 2017. “New Structural Insights into the Cell-Wall Polysaccharide of the Diatom Phaeodactylum Tricornutum.” Algal Research. https://doi.org/10.1016/j.algal.2017.07.021.

De Martino, Alessandra, Ana Bartual, Anusuya Willis, Agnes Meichenin, Beatriz Villazán, Uma Maheswari, and Chris Bowler. 2011. “Physiological and Molecular Evidence That Environmental Changes Elicit Morphological Interconversion in the Model Diatom Phaeodactylum Tricornutum.” Protist 162 (3): 462–81.

Durack, Paul J., Susan E. Wijffels, and Richard J. Matear. 2012. “Ocean Salinities Reveal Strong Global Water Cycle Intensification during 1950 to 2000.” Science 336 (6080): 455–58.

Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown. 2000. “Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes.” Molecular Biology of the Cell 11 (12): 4241–57.

Jackson, Anne E., Stephen W. Ayer, and Maurice V. Laycock. 1992. “The Effect of Salinity on Growth and Amino Acid Composition in the Marine Diatom Nitzschia Pungens.” Canadian Journal of Botany. Journal Canadien de Botanique 70 (11): 2198–2201.

James, Kimberley R., Belinda Cant, and Tom Ryan. 2003. “Responses of Freshwater Biota to Rising Salinity Levels and Implications for Saline Water Management: A Review.” Australian Journal of Botany 51 (6): 703–13.

Jozefczuk, Szymon, Sebastian Klie, Gareth Catchpole, Jedrzej Szymanski, Alvaro Cuadros-Inostroza, Dirk Steinhauser, Joachim Selbig, and Lothar Willmitzer. 2010. “Metabolomic and Transcriptomic Stress Response of Escherichia Coli.” Molecular Systems Biology 6 (May): 364.

Lee, Carol Eunmi, and Gregory William Gelembiuk. 2008. “Evolutionary Origins of Invasive Populations.” Evolutionary Applications 1 (3): 427–48.

Levitan, Orly, Jorge Dinamarca, Ehud Zelzion, Desmond S. Lun, L. Tiago Guerra, Min Kyung Kim, Joomi Kim, Benjamin A. S. Van Mooy, Debashish Bhattacharya, and Paul G. Falkowski. 2015. “Remodeling of Intermediate Metabolism in the Diatom Phaeodactylum Tricornutum under Nitrogen Stress.” Proceedings of the National Academy of Sciences of the United States of America 112 (2): 412–17.

Liu, M. S., and J. A. Hellebust. 1974. “Uptake of Amino Acids by the Marine Centric Diatom Cyclotella Cryptica.” Canadian Journal of Microbiology 20 (8): 1109–18.

Logares, Ramiro, Jon Bråte, Stefan Bertilsson, Jessica L. Clasen, Kamran Shalchian-Tabrizi, and Karin Rengefors. 2009. “Infrequent Marine–freshwater Transitions in the Microbial World.” Trends in Microbiology 17 (9): 414–22.

Lozupone, Catherine A., and Rob Knight. 2007. “Global Patterns in Bacterial Diversity.” Proceedings of the National Academy of Sciences of the United States of America 104 (27): 11436–40.

Lyon, Barbara R., Jennifer M. Bennett-Mintz, Peter A. Lee, Michael G. Janech, and Giacomo R. DiTullio. 2016. “Role of Dimethylsulfoniopropionate as an Osmoprotectant Following Gradual Salinity Shifts in the Sea-Ice Diatom Fragilariopsis Cylindrus.” Environmental Chemistry 13 (2): 181–94.

McCrystall, Michelle R., Julienne Stroeve, Mark Serreze, Bruce C. Forbes, and James A. Screen. 2021. “New Climate Models Reveal Faster and Larger Increases in Arctic Precipitation than Previously Projected.” Nature Communications. https://doi.org/10.1038/s41467-021-27031-y.

Nakov, Teofil, Jeremy M. Beaulieu, and Andrew J. Alverson. 2019. “Diatoms Diversify and Turn over Faster in Freshwater than Marine Environments.” Evolution; International Journal of Organic Evolution 73 (12): 2497–2511.

Nakov, Teofil, Kathryn J. Judy, Kala M. Downey, Elizabeth C. Ruck, and Andrew J. Alverson. 2020. “Transcriptional Response of Osmolyte Synthetic Pathways and Membrane Transporters in a Euryhaline Diatom during Long-Term Acclimation to a Salinity Gradient.” Journal of Phycology. https://onlinelibrary.wiley.com/doi/abs/10.1111/jpy.13061?casa_token=WxPO1IBOAuoAAAAA:zcF3_ZZSO99Xcn-uy8dT5HVtuy9WnS1evYKF-flWvqHBudKtOHEHuAHjpXe_81xkpHowxkbBIu1krg.

Nelson, David M., Paul Tréguer, Mark A. Brzezinski, Aude Leynaert, and Bernard Quéguiner. 1995. “Production and Dissolution of Biogenic Silica in the Ocean: Revised Global Estimates, Comparison with Regional Data and Relationship to Biogenic Sedimentation.” Global Biogeochemical Cycles. https://doi.org/10.1029/95gb01070.

Pozdnyakov, Ilya, Pavel Safonov, and Sergei Skarlato. 2020. “Diversity of Voltage-Gated Potassium Channels and Cyclic Nucleotide-Binding Domain-Containing Channels in Eukaryotes.” Scientific Reports. https://doi.org/10.1038/s41598-020-74971-4.

Rabe, Benjamin, Michael Karcher, Ursula Schauer, John M. Toole, Richard A. Krishfield, Sergey Pisarev, Frank Kauker, Rüdiger Gerdes, and Takashi Kikuchi. 2011. “An Assessment of Arctic Ocean Freshwater Content Changes from the 1990s to the 2006–2008 Period.” Deep Sea Research Part I: Oceanographic Research Papers. https://doi.org/10.1016/j.dsr.2010.12.002.

Rep, Martijn, Marcus Krantz, Johan M. Thevelein, and Stefan Hohmann. 2000. “The Transcriptional Response of Saccharomyces Cerevisiae to Osmotic Shock: Hot1p and msn2p/msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes* 210.” The Journal of Biological Chemistry 275 (12): 8290–8300.

Rijstenbil, J. W., J. A. Wijnholds, and J. J. Sinke. 1989. “Implications of Salinity Fluctuation for Growth and Nitrogen Metabolism of the Marine Diatom Ditylum Brightwellii in Comparison with Skeletonema Costatum.” Marine Biology 101 (1): 131–41.

Scholz, Bettina, and Gerd Liebezeit. 2012. “Compatible Solutes in Three Marine Intertidal Microphytobenthic Wadden Sea Diatoms Exposed to Different Salinities.” European Journal of Phycology 47 (4): 393–407.

Tamás, Markus J., and Stefan Hohmann. 2003. “The Osmotic Stress Response of Saccharomyces Cerevisiae.” In Yeast Stress Responses, edited by Stefan Hohmann and Willem H. Mager, 121–200. Berlin, Heidelberg: Springer Berlin Heidelberg.

Van Bergeijk, Stef A., Claar Van der Zee, and Lucas J. Stal. 2003. “Uptake and Excretion of Dimethylsulphoniopropionate Is Driven by Salinity Changes in the Marine Benthic Diatom Cylindrotheca Closterium.” European Journal of Phycology 38 (4): 341–49.

Wadskog, Ingrid, and Lennart Adler. 2003. “Ion Homeostasis in Saccharomyces Cerevisiae under NaCl Stress.” In Yeast Stress Responses, edited by Stefan Hohmann and Willem H. Mager, 201–39. Berlin, Heidelberg: Springer Berlin Heidelberg.

Werner, Dietrich. 1977. “Introduction with a Note on Taxonomy.” Botanical Monographs. https://agris.fao.org/agris-search/search.do?recordID=US201302980574.

Wood, J. M. 1999. “Osmosensing by Bacteria: Signals and Membrane-Based Sensors.” Microbiology and Molecular Biology Reviews: MMBR 63 (1): 230–62.

Yancey, Paul H. 2005. “Organic Osmolytes as Compatible, Metabolic and Counteracting Cytoprotectants in High Osmolarity and Other Stresses.” The Journal of Experimental Biology 208 (Pt 15): 2819–30.

Chapter 2 Phylogenetic analysis places Spicaticribra within Cyclotella

Kala M. Downey[1], Matthew L. Julius[2], Edward C. Theriot[3], and Andrew J. Alverson[1]

[1]University of Arkansas, Department of Biological Sciences, 1 University of Arkansas, Fayetteville, AR, USA

[2]St. Cloud State University, Department of Biological Sciences, St. Cloud, MN, USA

[3]The University of Texas at Austin, Section of Integrative Biology, Austin, TX, USA

Author Contributions: I collected *Cyclotella kingstonii*, sequenced the LSU, SSU, rbcL, and psbC genes, and performed the phylogenetic analyses. Andrew J. Alverson justified the reclassification. Matthew L. Julius and Edward C. Theriot consulted on the reclassification of the genus.

This chapter has been published:

Downey KM, Julius ML, Theriot EC, Alverson AJ. 2021. Phylogenetic analysis places *Spicaticribra* within *Cyclotella.* Diatom Research 36(2):93-99

2.1 Abstract

A strong consensus has emerged that taxonomic classifications should be based on an underlying phylogenetic hypothesis. According to this view, named groups should be monophyletic, ensuring that a name uniquely matches the evolutionary history and biological attributes of a group of taxa. As originally conceived, the diatom genus *Cyclotella* is a large and morphologically diverse assemblage of taxa that we now know consists of several distantly related lineages. Considerable progress has been made in assigning these lineages into different monophyletic genera. The genus *Spicaticribra* was originally described as monotypic and has features that suggest a close relationship to *Cyclotella* and, in addition, has several ancestral features that appear to differentiate it from *Cyclotella*. We sequenced two nuclear and two plastid genes to resolve the phylogenetic position of *Spicaticribra* and show that it is embedded within a clade that includes the type species of *Cyclotella* and, further, that maintaining the name *Spicaticribra* renders *Cyclotella* nonmonophyletic. We transfer *Spicaticribra* species into *Cyclotella*, resolve other nomenclatural issues, and caution against using ancestral characters and character states for taxonomic classification.

2.2 Introduction

Among the 820 accepted species names in Thalassiosirales, nearly half of them have been classified as either *Thalassiosira* (193) or *Cyclotella* (170) (AlgaeBase, accessed 12 February 2021, Guiry & Guiry 2010). These numbers reflect similar and longstanding taxonomic challenges presented by these two genera. Both of them encompass a broad range of morphological diversity, and their "defining" morphological features include characters that, in some cases, appear to have been present since the origin of Thalassiosirales. The use of ancestral characters for classification can lead to the bloating of taxonomic groups over time due to the inclusion of distantly related species. This situation applies to *Thalassiosira*, a large genus that phylogenetic analyses have shown to be polyphyletic (Alverson et al. 2007). Efforts to diagnose and name clades within the broadly and loosely defined *Thalassiosira* (Alverson et al. 2006, Stachura-Suchoples & Williams 2009) have left the remaining species polyphyletic, which we view as a transitional state of progress toward establishing a natural classification of this genus.

The challenge of *Cyclotella* is somewhat simpler because although it, too, has accumulated a large set of morphologically diverse species, decades of research have revealed an information-rich set of characters that have made it easier to first delineate phenetic morphological groups (e.g., Lowe 1975) and later to distinguish between ancestral and derived character states. As a result, many extant species can now be classified into one of several monophyletic genera that formerly fell under the *Cyclotella* umbrella. Examples include *Discostella* (Houk & Klee 2004), *Lindavia* (Nakov et al. 2015), and *Cyclotella sensu stricto* (Alverson et al. 2011). Many challenges remain, however. For example, it can be hard to classify fossil taxa that have a mix of ancestral and derived character states (Stone et al. 2020), and the failure to make this distinction can lead to descriptions of new genera that make existing ones nonmonophyletic.

Spicaticribra is one such genus. It was first described as monotypic and was defined by the following diagnostic characters: continuous “spicate” cribra on the interior valve face, absence of central strutted processes, and absence of external extensions of the marginal strutted processes (Johansen et al. 2008). Johansen et al. (2008) suggested that the lack of distinct regions on the valve face appeared to place it outside of *Cyclotella*, whereas the presence of continuous cribra indicated a closer relationship to another polyphyletic genus, *Thalassiosira*. Without a home for the species, it was placed into its own genus. Other species were later placed into *Spicaticribra*, and the online nomenclatural database, AlgaeBase, lists a total of 10 *Spicaticribra* species (accessed 12 February 2021, Guiry & Guiry 2010).

We collected and cultured the type species of *Spicaticribra*, *S. kingstonii*, from its type locality. Phylogenetic analyses of four DNA markers place *S. kingstonii* within *Cyclotella sensu stricto*. In an effort to preserve the phylogenetic integrity of *Cyclotella*, we transfer *S. kingstonii* and other *Spicaticribra* into *Cyclotella*.

2.3 Methods

We collected near-surface phytoplankton with a 10 µM mesh plankton net from the Tuckasegee River, North Carolina, USA, (35.439933, -83.55145) on 24 April 2017. We isolated individual cells with a micropipette and grew them at 21℃ in WC medium (Guillard & Lorenzen 1972). We cleaned clonally cultured cells with nitric acid and rinsed them with deionized water until the solution reached a neutral pH. Cleaned cells were transferred onto coverslips and allowed to evaporate overnight before permanently affixing them onto microscope slides with Naphrax®. We identified and photographed cells at 600x magnification with a Zeiss compound microscope. For scanning electron microscopy, cleaned cells were dried onto 12 mm diameter coverslips and coated with 15 nm of iridium with a Cressington 208 Bench Top Sputter Coater (Cressington Scientific Instruments, Watford, UK). Scanning electron micrographs were taken

with a Zeiss SUPRA 40 VP scanning electron microscope (Carl Zeiss Microscopy, Thornwood, NY, USA).

We collected live cells from a single clonal culture by centrifugation, vortexed them with 1.0 mm glass beads, and extracted DNA with a Qiagen DNeasy Plant Kit. We sequenced two nuclear (*SSU* and partial *LSU* rDNA) and two plastid (*rbcL* and *psbC*) genes. Primers, PCR conditions, and Sanger sequencing followed Alverson et al. (2007). We added *Spicaticribra* sequences to gene alignments from Alverson et al. (2011), using SSU-ALIGN version 0.1.1 (Nawrocki et al. 2009) to align rDNA sequences with the covariance models included in the program for the *SSU* alignment and a heterokont-based covariance model for the *LSU* alignment (Nakov et al. 2014). We removed poorly aligned sections of the alignment with SSU-MASK using the default settings, which retains columns with a Bayesian Posterior Probability of 0.95 of being correctly aligned. Multiple sequence alignment of the *psbC* and *rbcL* plastid genes was performed manually in AliView version 1.25 (Larson 2014). We used trimAl version 1.4 (Capella-Gutiérrez et al. 2009) to remove alignment columns with gaps in more than 20% of the sequences. We concatenated sequences for all four genes into a single alignment with AMAS (Borowiec 2016) and used IQ-TREE version 1.6.4 (Nguyen et al. 2015) to reconstruct phylogenetic relationships. The concatenated sequences were partitioned by gene, and the best-fit substitution model for each partition was inferred using the ModelFinder algorithm implemented in IQ-TREE. We inferred the maximum-likelihood tree using the edge-linked partition model in IQ-TREE and applied the TIM2+F+R3 model to the *LSU* partition, TN+F+R3 to *SSU*, GTR+F+I+G4 to *psbC*, and the GTR+F+R4 model to the *rbcL* partition. Branch support was assessed with 100,000 bootstrap replicates using Ultrafast Bootstrap Approximation (Hoang et al. 2018) and the setting '-bnni' to guard against overestimation of branch support.

Newly generated DNA sequences are available from the National Center for Biotechnology Information's GenBank database under accession numbers MW327042,

MW327043, MW326755, and MW326756. Multiple sequence alignments and tree files are available from a Zenodo online repository (10.5281/zenodo.4313346).

2.4 Results

We collected and cultured *S. kingstonii* from the Tuckasegee River, USA, which is part of the same riverine/reservoir system as Fontana Lake, the type locality for this species and for the genus *Spicaticribra* (Johansen et al. 2008). Light microscope images confirmed the identity as *S. kingstonii* (Fig. 1). Phylogenetic analyses of two nuclear and two plastid genes placed *S. kingstonii* as sister to a clade that includes *Cyclotella distinguenda* Hustedt—the type species of the genus *Cyclotella*—and other *Cyclotella* species (Fig. 2). *Cyclotella nana* Hustedt was sister to the *Spicaticribra*+*Cyclotella* clade (Fig. 2). By separating *C. nana* from the remaining *Cyclotella*, *Spicaticribra* renders *Cyclotella* nonmonophyletic.

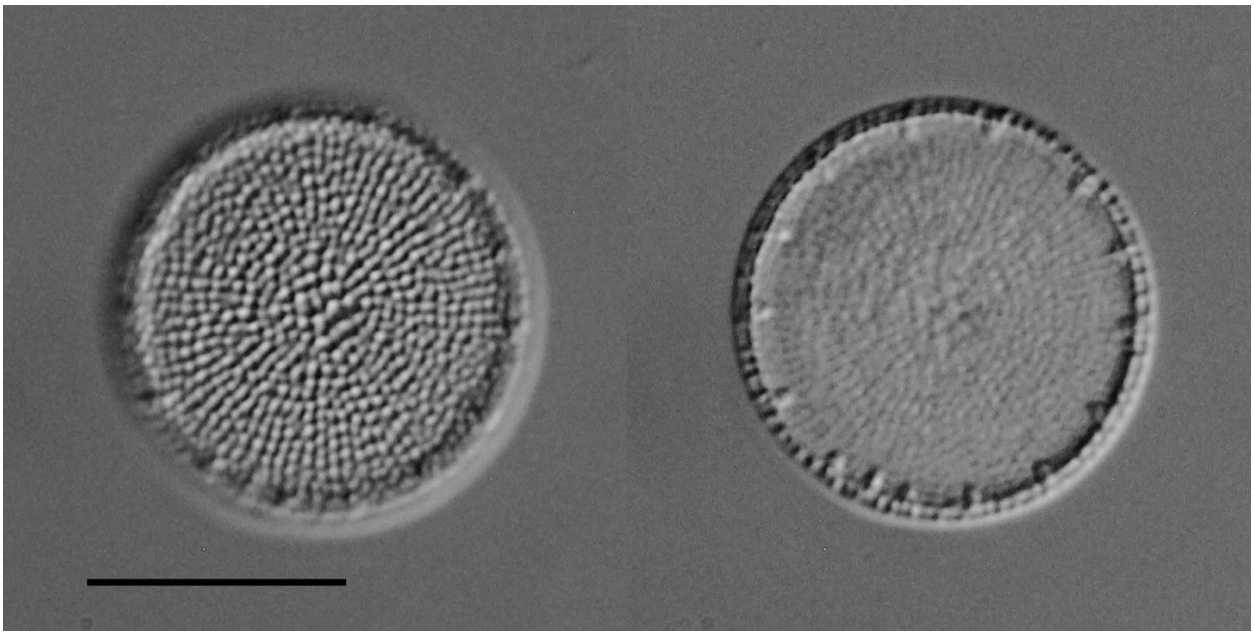


Fig 2.1. Light micrographs of *Spicaticribra kingstonii* culture strain AJA246-42, collected from Tuckasegee River, USA. The photographs show the same specimen at different focal planes to illustrate the areolar pattern (left) and the marginal ring of strutted processes (right). Scale bar = 10 μM.

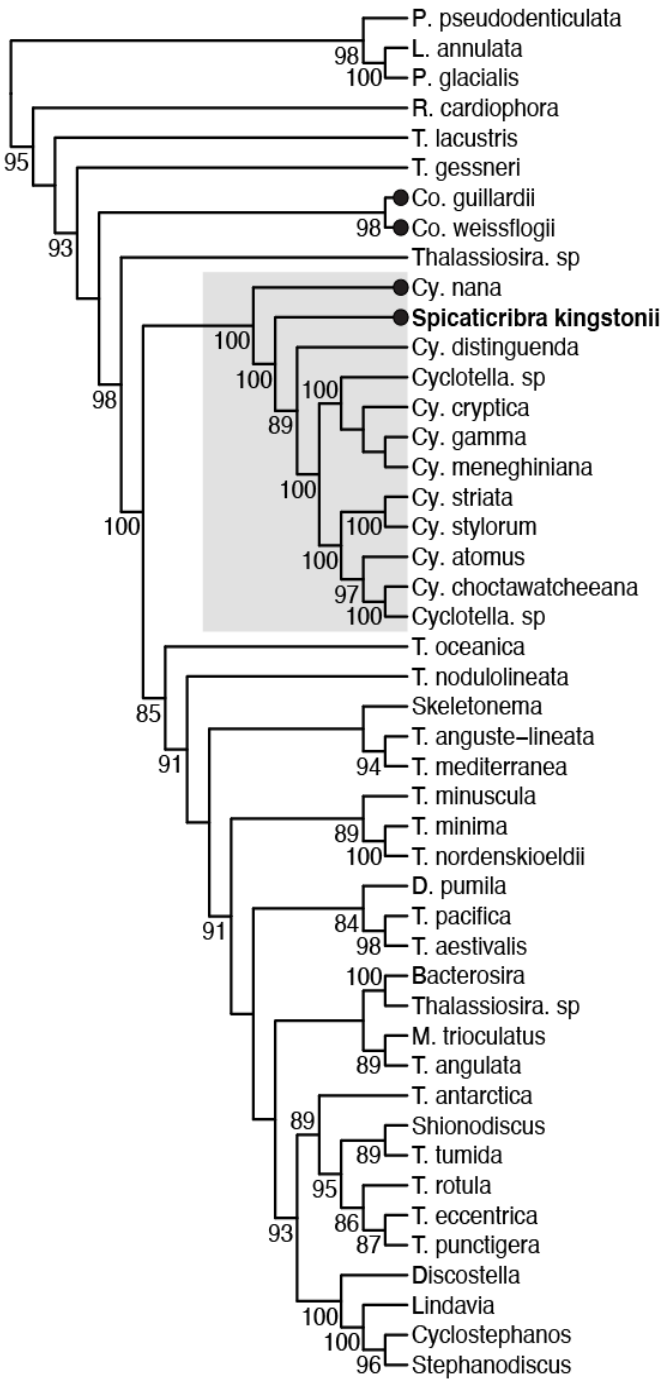


Fig 2.2. Phylogenetic position of *Spicaticribra kingstonii* based on maximum likelihood analysis of two nuclear and two plastid genes. Ultrafast bootstrap support values >80% are shown. The gray box delimits the *Cyclotella* clade, including *S. kingstonii*. Black dots show a polyphyletic set of taxa that have been transferred into *Spicaticribra* since its inception. For simplicity, outgroups are not shown, and some monophyletic genera were collapsed into a single branch. Genus names are abbreviated as follows: *Porosira* (P), *Lauderia* (L), *Conticribra* (Co), *Thalassiosira* (T), *Cyclotella* (Cy), *Detonula* (D), *Minidiscus* (M).

2.5 Discussion

A strong consensus has emerged that taxonomic classifications should be natural, meaning that named groups are monophyletic, which ensures that a name corresponds to a shared, unique evolutionary history among the taxa bearing that name (Kociolek et al. 1989, Williams & Kociolek 2011, Kociolek & Williams 2015). A great deal of progress has been made in using phylogenetic trees and phylogenetic character interpretations to subdivide Thalassiosirales into monophyletic genera. Most of this work, and much of the work that remains, concerns the two largest genera, *Cyclotella* and *Thalassiosira*. Molecular phylogenies have facilitated these efforts, but a phylogenetic tree is a hypothesis that can change—along with the classification that accompanies it—as new taxa are discovered (Kociolek & Williams 2015). A natural genus-level classification can also be disrupted in unintended ways by the discovery of new taxa without a clear generic affiliation—this is especially problematic when character interpretations of the new taxa are not made with respect to phylogenetic relationships (Kociolek & Williams 2015). New and challenging taxa are often placed in monotypic genera, either because of ambiguity about their relationship to other genera (Johansen et al. 2008), because they exceed some arbitrary threshold of difference to known genera (Williams 2009, 2013), or to intentionally communicate uncertainty about their relationship to other taxa (Williams 2013, Stone et al. 2020). Without "evaluation of all available evidence in terms of monophyly and synapomorphy" (Williams & Kociolek 2011 p. 51), a newly described monotypic genus can make existing ones nonmonophyletic, thereby undermining the *primum non nocere* principle that seems to underlie many of the guiding principles laid out by Kociolek & Williams (2015), an approach we see as having greatly improved our approach to diatom systematics and classification.

Spicaticribra was defined originally by the presence of a "spicate" pattern of continuous cribra internally, absence of central strutted processes, and absence of external extensions of the marginal strutted processes (Johansen et al. 2008). Unless phylogenetic analyses

demonstrate derived secondary loss of a character, we agree that character absences generally do not provide compelling evidence in support of a genus (Kociolek & Williams 2015), so we focus here on the internal cribra. Prasad et al. (2011) noted similarities between the cribral pattern of *Spicaticribra* and many *Conticribra* and *Thalassiosira* species. Based on this similarity, Khursevich & Kociolek (2012) later suggested that many of these species might eventually warrant transfer into *Spicaticribra*. A similar cribral pattern is present across a diversity of Thalassiosirales, including *Cyclotella nana* (formerly *Thalassiosira pseudonana*; Alverson et al. 2011) and many *Conticribra* and *Thalassiosira* species (Prasad et al. 2011, Khursevich & Kociolek 2012). One possible distinction is whether a stria consists of multiple parallel rows of areolar pores (e.g., *T. gessneri* and *T. livingstoniorum*) or single pores most evident in *S. kingstonii*, some *Conticribra guillardii*, and some *Cyclotella nana*. Another possible distinguishing feature of *Spicaticribra* is that the internal cribra do not align with external openings of the areolae (Johansen et al. 2008).

The diagnosis of *Spicaticribra* was later broadened by Khursevich & Kociolek (2012) to include the following characters: loculate areolae with continuous or semicontinuous cribra and external foramina; plicated valve face—or not; one or more labiate processes that extend outwardly from the frustule—or not; strutted processes with 2–4 satellite pores that extend outwardly from the frustule—or not; absence of strutted processes on the valve face—or rarely not. By relaxing the required absence of central strutted processes and external extensions of the strutted processes and allowing for presence or absence of other characters, the amended definition captures genera and species from across almost the whole of Thalassiosirales. Several species were identified as candidates for transfer to *Spicaticribra* under this new definition (Khursevich & Kociolek 2012), including *T. lacustris*, *T. gessneri*, *Conticribra guillardii*, *Conticribra weissflogii*, and *Cyclotella nana*—a group of taxa whose common ancestor traces back nearly to the root node of Thalassiosirales (Fig. 2). Several of these (*Conticribra guillardii, Conticribra weissflogii*, and *Cyclotella nana*) were later transferred (Khursevich & Svirid 2013)

despite known phylogenetic positions which showed them to be polyphyletic (Alverson et al. 2007). These transfers rendered *Conticribra*, *Cyclotella*, and *Spicaticribra* nonmonophyletic (Fig. 2).

A large body of research, dating back to early applications of the scanning electron microscope for describing frustule morphology, has highlighted the importance of strutted process ultrastructure in understanding the phylogeny (Theriot & Serieyssol 1994, Shiono 2000) and classification of Thalassiosirales (Fryxell & Hasle 1979, 1980). *Spicaticribra* has a strutted process ultrastructure not found outside of the cyclotelloid and cyclostephanoid lineages of Thalassiosirales (Alverson et al. 2007). The large strutted processes of *Spicaticribra*, which feature robust cowlings and broad satellite pore covers (Fig. 3), resemble those of many *Cyclotella* species (e.g., *C. gamma*, *C. choctawhatcheeana*, and *C. distinguenda*) and clearly indicate a closer relationship with these species than any *Thalassiosira* (Fig. 3). Johansen et al. (2008) noted similarities between *S. kingstonii* and *T. pseudonana*. Phylogenetic analyses placed *T. pseudonana* as sister to a clade of *Cyclotella* species, compelling resurrection of its original name, *Cyclotella nana* (Alverson et al. 2011). These results also helped circumscribe a monophyletic *Cyclotella sensu stricto*. Although ultimately discounted, comparisons of *Spicaticribra* to *T. pseudonana*/*C. nana* were prescient, as phylogenetic analyses of both morphological (Alverson et al. 2011) and molecular (Fig. 2; Tuji et al. 2012) datasets have shown that *S. kingstonii* falls in the middle of grade between *C. nana*, *C. distinguenda* (the generitype), and the rest of *Cyclotella*—a result that we again recovered with our analyses of *S. kingstonii* from its type locality (Fig. 2). In short, *Spicaticribra* has rendered *Cyclotella* nonmonophyletic, a result that closely matches one of the hypothetical consequences of describing monotypic genera without reference to a phylogenetic hypothesis (Kociolek & Williams 2015).

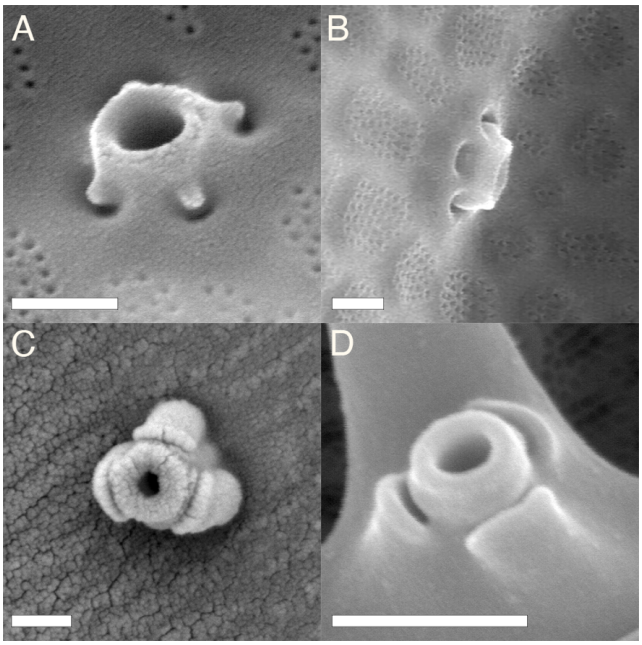


Fig 2.3. Scanning electron micrographs showing the contrasting strutted process ultrastructure between two representative *Thalassiosira* species, *T. pacifica* (A) and *T. nordenskioeldii* (B), and two species in the *Cyclotella* clade, *Spicaticribra kingstonii* (C) and *C. distinguenda* (D). Scale bar = 250 nm.

Two options are available for preserving a monophyletic *Cyclotella*: retain *Spicaticribra* and place *C. nana* into a monotypic genus, or include *Spicaticribra* in *Cyclotella* (Fig. 2). The nomenclatural history of *C. nana* has been one of instability, which has led to confusion about the biology and natural history of this important model species (Alverson et al. 2011). The placement of *C. nana* in *Cyclotella* dates back to its original description by Hustedt (1957), who correctly diagnosed its placement in *Cyclotella* despite the limited information available to him at the time. A solution to the conflict resulting from the creation and later expansion of *Spicaticribra* should prioritize historical precedent, taxonomic continuity, and maintain recent progress in stabilizing the taxonomy of *Cyclotella.* To this end, we recommend that *Spicaticribra* should not be recognized as a separate genus, and we therefore transfer the following species to *Cyclotella*:

Cyclotella inlandica M L Julius, K M Downey, E C Theriot, *et* A J Alverson *comb. nov.*

Basionym: *Thalassiosira inlandica* Hayashi *in* Hayashi & al. 2007, A fossil freshwater *Thalassiosira*, *T. inlandica* sp. nov. (Bacillariophyta), with semicontinuous cribra and elongated marginal fultoportulae. *Phycologia,* 46: 354, figs 2-40.

Cyclotella kingstonii M L Julius, K M Downey, E C Theriot, *et* A J Alverson *comb. Nov.*

Basionym: *Spicaticribra kingstonii* J.R.Johansen, Kociolek & R.L.Lowe 2008, *Spicaticribra kingstonii*, gen. nov. et sp. nov. (Thalassiosirales, Bacillariophyta) from Great Smoky Mountains National Park, U.S.A. *Diatom Research,* 23:369, figs 1–24.

Cyclotella kodaikanaliana M L Julius, K M Downey, E C Theriot, *et* A J Alverson *comb. nov.*

Basionym: *Spicaticribra kodaikanaliana* B.Karthick *et* Kociolek 2011, Four new centric diatoms (Bacillariophyceae) from the Western Ghats, South India, *Phytotaxa* 22:29, figs 3 and 4.

Cyclotella tanaka M L Julius, K M Downey, E C Theriot, *et* A J Alverson *nom. nov.*

Replaced synonym: *Stephanodiscus minuta* Pantcsek 1889*, Beiträge zur Kenntniss der fossilen Bacillarien Ungarns. II Thiel: Brackwasser Bacillarien*. Nagy-Tapolcsány. Buchdruckerei von Julius Platzko, p.114, Tafel XII, fig. 213.

Note 1: The replacement name *Cyclotella tanaka* is proposed for *Stephanodiscus minutus/ Spicaticribra minuta* because its transfer to *Cyclotella* would create a later homonym to *Cyclotella minuta* (Skvortsov) Antipova 1956 (ICN Art. 53.1, also ICN Art. 11.8 *Note 5* – “in accordance with Art. 53, later homonyms are illegitimate whether the type is fossil or non-fossil”).

Note 2: The species is renamed after Dr. Hiroyuki Tanka in honor of his many research contributions involving the Thalassiosirales.

Cyclotella rudis M L Julius, K M Downey, E C Theriot, *et* A J Alverson *comb. nov.*

Basionym: *Thalassiosira rudis* P.I.Tremarin, T.A.V.Ludwig, V.Becker *et* L.C.Torgan in Ludwig *et al,* 2008, *Thalassiosira rudis* sp. nov. (Coscinodiscophyceae): A new freshwater species. *Diatom Research* 23:391, figs 1–57.

Additional taxa have been transferred to *Spicaticribra* that either do not possess the diagnostic characters of *Cyclotella* or lack sufficient illustration in the literature to determine if the transfer was warranted. These include:

Spicatacribra guillardi (Hasle) Khursevich *et* Kociolek

Note: Should be maintained as ***Conticribra guillardii* (Hasle) K.Stachura-Suchoples & D.M.Williams** as its marginal strutted processes do not conform with those expressed in *Cyclotella.*

Spicatacribra weissflogii (Grunow) Khursevich *et* Kociolek

Note: Should be maintained as ***Conticribra weissflogii* (Grunow) Stachura-Suchoples & D.M.Williams** as its marginal strutted processes do not conform with those expressed in *Cyclotella.*

Spicaticribra kamszatica (Lupikina) Khursevich *et* Kociolek

Note: Should be maintained as ***Thalassiosira kamszatica* (Lupikina) Lupikina *et* Khursevich** as sufficient illustrations do not exist in literature to accurately diagnose this taxon.

Spicaticribra kilarskii (Kaczmarska) Kociolek *et* Khursevich

Note: Should be maintained as ***Thalassiosira kilarskii* I.Kaczmarska** as sufficient illustrations do not exist in literature to accurately diagnose this taxon.

Spicaticribra nevadica (Khursevich & VanLandingham) Khursevich & Kociolek

Note: Should be maintained as ***Thalassiosira nevadica* Khursevich *et* VanLandingham** as its marginal strutted processes do not conform with those expressed in *Cyclotella.*

Spicaticribra patagonica (N.Maidana) Kociolek *et* Khursevich

Note: Should be maintained as ***Thalassiosira patagonica*** **N.I.Maidana** as its marginal strutted processes do not conform with those expressed in *Cyclotella.*

Spicaticribra tricircularis (Stachura-Suchoples & D.M.Williams) Kociolek *et* Khursevich[1]

Note: Should be maintained as ***Conticribra tricircularis*** **Stachura-Suchoples *et* D.M.Williams** as its marginal strutted processes do not conform with those expressed in *Cyclotella.*

2.6 Acknowledgements

This work was supported by the National Science Foundation under grant number DEB-1651087. We thank Elizabeth Ruck for providing light micrographs and Matt Ashworth for providing SEM images. We thank two anonymous reviewers for constructive comments on an earlier version of the manuscript.

2.7 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 Evolutionary Biology 11: 125.

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. Molecular Phylogenetics and Evolution 45: 193–210.

Alverson A. J., Kang S.-H. & Theriot E. C. 2006. Cell wall morphology and systematic importance of Thalassiosira ritscheri (Hustedt) Hasle, with a description of Shionodiscus gen. nov. Diatom Research 21: 251–262.

Capella-Gutiérrez S., Silla-Martínez J. M. & Gabaldón T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses.

Fryxell G. A. & Hasle G. R. 1979. The genus Thalassiosira: Species with internal extensions of the strutted processes. Phycologia 18: 378–393.

Fryxell G. A. & Hasle G. R. 1980. The marine diatom Thalassiosira oestrupii: Structure, taxonomy and distribution. American Journal of Botany 67: 804–814.

Guillard R. R. L. & Lorenzen C. J. 1972. Yellow-green algae with Chlorophyllide c. Journal of Phycology 8: 10–14.

Guiry M. D. & Guiry G. M. 2010. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway.

Hoang D. T., Chernomor O., von Haeseler A., Minh B. Q. & Vinh L. S. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Molecular Biology and Evolution 35: 518–522.

Houk V. & Klee R. 2004. The stelligeroid taxa of the genus Cyclotella (Kützing) Brébisson (Bacillariophyceae) and their transfer into the new genus Discostella gen. nov. Diatom Research 19: 203–228.

Hustedt F. 1957. Die Diatomeenflora des Flußsystems der Weser im Gebiet der Hansestadt Bremen [Diatom flora of the tributaries of the Weser near the city of Bremen]. Abhandlungen. Naturwissenschaftlicher Verein zu Bremen 34: 181–440.

Johansen J., Kociolek P. & Lowe R. 2008. Spicaticribra kingstonii, gen. nov. et sp. nov. (Thalassiosirales, Bacillariophyta) from Great Smoky Mountains National Park, U.S.A. Diatom Research 23: 367–375.

Khursevich G. & Kociolek J. P. 2012. A preliminary, worldwide inventory of the extinct, freshwater fossil diatoms from the orders Thalassiosirales, Stephanodiscales, Paraliales, Aulacoseirales, Melosirales, Coscinodiscales, and Biddulphiales. Nova Hedwigia 141: 315–364.

Khursevich G. K. & Svirid A. A. 2013. The audit of diatom algae in the families Thalassiosiraceae and Stephanodiscaceae from waters in Belarus. Vestsi BDPU. Series 3. Physics, Mathematics, Informatics, Biology, Geography 3: 3–9.

Kociolek J. P., Theriot E. C. & Williams D. M. 1989. Inferring diatom phylogeny: A cladistic perspective. Diatom Research 4: 289–300.

Kociolek J. P. & Williams D. M. 2015. How to define a diatom genus? Notes on the creation and recognition of taxa, and a call for revisionary studies of diatoms. Acta Botanica Croatica 74: 195–210.

Lowe R. L. 1975. Comparative ultrastructure of the valves of some Cyclotella species (Bacillariophyceae). Journal of Phycology 11: 415–424.

Nakov T., Guillory W., Julius M., Theriot E. & Alverson A. 2015. Towards a phylogenetic classification of species belonging to the diatom genus Cyclotella (Bacillariophyceae): Transfer of species formerly placed in Puncticulata, Handmannia, Pliocaenicus and Cyclotella to the genus Lindavia. Phytotaxa 217: 249–264.

Nguyen L.-T., Schmidt H. A., von Haeseler A. & Minh B. Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32: 268–274.

Prasad A. K. S., Nienow J. A. & Hargraves P. 2011. Plicate species of the diatom genus Thalassiosira (Bacillariophyta) from the Atlantic and Gulf Coasts of Southeastern United States, with the description of T. livingstoniorum sp. nov. Proceedings of the Academy of Natural Sciences of Philadelphia 161: 1–34.

Shiono M. 2000. Three new species in the Thalassiosira trifulta group in Late Neogene sediments from the northwest Pacific Ocean. Diatom Research 15: 131–148. Taylor & Francis.

Stachura-Suchoples K. & Williams D. M. 2009. Description of Conticribra tricircularis , a new genus and species of Thalassiosirales, with a discussion on its relationship to other continuous cribra species of Thalassiosira Cleve (Bacillariophyta) and its freshwater origin. European journal of phycology 44: 477–486. Taylor & Francis.

Stone J. R., Edlund M. B. & Alverson A. J. 2020. Fascinorbis gen. nov., a new genus of Stephanodiscaceae (Bacillariophyta) from a Late Miocene lacustrine diatomite. Diatom Research: 1–13.

Theriot E. & Serieyssol K. 1994. Phylogenetic systematics as a guide to understanding features and potential morphological characters of the centric diatom family Thalassiosiraceae. Diatom Research 9: 429–450.

Tuji A., Leelahakriengkrai P. & Peerapornpisal Y. 2012. Distribution and phylogeny of Spicaticribra kingstonii-rudis species complex. Memoirs of National Museum Natural Science of Tokyo 48: 139–148.

Williams D. M. 2009. ‘Araphid’ diatom classification and the ‘absolute standard’. Acta Botanica Croatica 68: 455–463.

Williams D. M. 2013. Why is Synedra berolinensis so hard to classify? More on monotypic taxa. Phytotaxa 127: 113–127.

Williams D. M. & Kociolek J. P. 2011. An overview of diatom classification with some prospects for the future. In: The Diatom World (Ed. by J. Seckbach & P. Kociolek), pp. 47–91. Springer Netherlands, Dordrecht.

Chapter 3 Transcriptional response of osmolyte synthetic pathways and membrane transporters in a euryhaline diatom during long-term acclimation to a salinity gradient

Teofil Nakov, Kathryn J. Judy, Kala M. Downey , Elizabeth C. Ruck , and Andrew J. Alverson

[1]Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas, USA

Author Contributions: Teofil Nakov, Elizabeth Ruck, and Andrew Alverson design the study. Teofil and Elizabeth performed experiments and collected data. Teofil, Kala Downey, and Kathryn Judy analysed the data. Teofil, Kathryn, Kala, and Andrew wrote the paper.

This chapter has been published:

Nakov T, Judy KJ, Downey KM, Ruck EC, Alverson AJ. 2020. Transcriptional response of osmolyte synthetic pathways and membrane transporters in a euryhaline diatom during long-term acclimation to a salinity. Journal of Phycology 56:1712-1728

3.1 Abstract

How diatoms respond to fluctuations in osmotic pressure is important from both ecological and applied perspectives. It is well known that osmotic stress affects photosynthesis and can result in the accumulation of compounds desirable in pharmaceutical and alternative fuel industries. Gene expression responses to osmotic stress have been studied in short-term trials, but it is unclear whether the same mechanisms are recruited during long-term acclimation. We used RNA-seq to study the genome-wide transcription patterns in the euryhaline diatom, *Cyclotella cryptica*, following long-term acclimation to salinity that spanned the natural range of fresh to oceanic water. Long-term acclimated *C. cryptica* exhibited induced synthesis or repressed degradation of the osmolytes glycine betaine, taurine and dimethylsulfoniopropionate (DMSP). Although changes in proline concentration is one of the main responses in short-term osmotic stress, we did not detect a transcriptional change in proline biosynthetic pathways in our long-term experiment. Expression of membrane transporters showed a general tendency for increased import of potassium and export of sodium, consistent with the electrochemical gradients and dependence on co-transported molecules. Our results show substantial between-genotype differences in growth and gene expression reaction norms and suggest that the regulation of proline synthesis important in short-term osmotic stress might not be maintained in long-term acclimation. Further examination using time-course gene expression experiments, metabolomics and genetic validation of gene functions would reinforce patterns inferred from RNA-seq data.

3.2 Introduction

The physiological mechanisms involved in the response by diatoms to changes in osmotic pressure due to salinity, drought, or freezing, are important for understanding diatom ecology and evolution, population and community response to environ- mental change, and the use of diatoms in industrial applications. The contrast between marine and freshwaters is important for understanding diatom community composition (Potapova 2011), functional diversity (Edwards et al. 2015), cell size distribution (Litchman et al. 2009, Nakov et al. 2014), colonization patterns (Mann 1999, Alverson et al. 2007, Ruck et al. 2016) and lineage diversification (Nakov et al. 2019). In a global context, understanding how diatoms will respond to the freshening of marine environments (Wadhams and Munk 2004, Swart et al. 2018), whether by changes in abundance, phenology, geographic distribution or even local extirpation, is of paramount importance for primary production and carbon burial and could have cascading upward effects on marine food webs (Li et al. 2009, Coupel et al. 2015). From an applied perspective, osmotic stress can induce accumulation of secondary compounds with pharmacological and nutritional value (Cheng et al. 2014, Chen et al. 2017, Sayanova et al. 2017) as well as energy-rich oil compounds relevant to the alternative fuel industry (d'Ippolito et al. 2015, Tanaka et al. 2015, Traller et al. 2016). Advancing our understanding of physiological responses by diatoms to short- and long-term osmotic stress would clearly shed light on several important questions in diatom biology.

Coping with osmotic stress is commonly achieved through regulation of the intracellular concentration of small organic molecules known as compatible solutes (osmolytes or osmoprotectants). Adjusting the concentration of osmolytes in response to external osmotic pressure, either through synthesis or membrane transport, balances the osmotic difference between the cell and its environment, and prevents cytosol dehydration in hyperosmotic or dilution in hypoosmotic conditions. Biochemical studies of osmotic stress in diatoms date back to at least the 1970s and have generally focused on species of ecological (e.g., sea-ice

diatoms) or applied interest (e.g., oleaginous species; Liu and Hellebust 1976a, Hellebust 1985). Despite the relatively narrow phylogenetic and ecological breadth of species studied so far, this work has generated a wealth of information regarding the diversity of compatible solutes used by diatoms and the kinetics of osmoprotectant uptake under various conditions.

Diatom osmoprotectants, like those in bacteria and plants, are generally zwitterionic compounds of small molecular weight. They can be accumulated safely to high concentrations to counteract osmotic differences (Yancey 2005, Burg and Ferraris 2008), and some of them might also contribute toward scavenging of reactive oxygen species (Sunda et al. 2002). The list of experimentally determined com- pounds with osmolyte properties in diatoms includes free proline and other amino acids (Liu and Hellebust 1976a,b, Dickson and Kirst 1987, Krell et al. 2008), quaternary ammonium com- pounds (e.g., glycine betaine; Dickson and Kirst 1987, Kageyama et al. 2018b, Torstensson et al. 2019), sulphonic acid compounds (e.g., taurine; Jackson et al. 1992) and organosulfur compounds (e.g., dimethylsulfoniopropionate, DMSP; Lyon et al. 2011, Kettles et al. 2014, Lavoie et al. 2018, Kageyama et al. 2018a).

Diatoms can import a variety of exogenous compatible solutes, including DMSP, glycine betaine and various amino acids (Liu and Hellebust 1974, Lavoie et al. 2018, Petrou and Nielsen 2018, Torstensson et al. 2019), and there appear to be interactions between different osmolytes and import versus synthesis strategies (Kageyama et al. 2018a). For example, osmolyte addition experiments indicate that in the presence of exogenous glycine betaine, cells under hyperosmotic stress import or synthesize less DMSP than they would without exogenous glycine betaine under the same stress conditions. The opposite, addition of DMSP during hyperosmotic stress, does not affect the intracellular concentration of glycine betaine (Kageyama et al. 2018a), suggesting that when the cell has a choice of exogenous osmolytes, glycine betaine is preferred over DMSP. In general, short-term acclimation to osmotic pressure by diatoms involves multiple biosynthetic pathways and membrane transport systems, and it is

also possible that recruitment of different osmotic stress response mechanisms depends on other factors, including temperature, nutrient availability or community composition.

Although we have a relatively detailed picture of the array of osmoprotectants used by diatoms, less is known about how the intracellular concentrations of compatible solutes are regulated, either by synthesis or transport. Transcriptomic and proteomic studies in the sea-ice diatom Fragilariopsis cylindrus have revealed that short-term exposure to hyperosmotic conditions, that is, in vitro transfer from normal seawater salinity (~30) to hypersaline water (70), induces proline and DMSP biosynthesis (Krell et al. 2007, 2008, Lyon et al. 2011, 2016). *Fragilariopsis cylindrus* can also import exogenous glycine betaine, suggesting the presence of at least three mechanisms for coping with changes in ambient salinity (Torstensson et al. 2019). Studies of temperate diatom species have focused on general transcriptional patterns rather than the specific osmoprotectant pathways and have shown that the response to salinity includes changes in the expression of genes for nitrogen, carbon and fatty acid metabolism as well as transcription factors (Cheng et al. 2014, Bussard et al. 2017). Targeted molecular studies of synthetic pathways in *Cyclotella nana* have shown that synthesis of both glycine betaine and DMSP is induced during osmotic stress and that this induction is regulated both at the transcript and protein level (Kageyama et al. 2018a,b). However, the majority of work to date has focused on short-term stress responses across a few salinity treatments. The regulation of osmolyte concentrations in long-term acclimation trials, especially over a broad range of salinity treatments, is poorly understood.

We use RNA-seq to study patterns of gene expression across a gradient spanning the natural range of salinity found in most inland and oceanic habitats. We used the diatom species *Cyclotella cryptica* (Reimann et al. 1963) because it is capable of sustaining relatively fast growth both in freshwater (0) and full ocean salinity (30-36). Euryhaline species like *C. cryptica* are ideally suited to study the cellular response to changes in ambient salinity because: (i) from an evolutionary perspective, they are most likely to be distributed across habitats with different

salinities and therefore most likely to transition between marine and freshwaters, (ii) from an ecological perspective, their populations are less likely to experience severe stress from freshening of marine habitats and (iii) from an applied perspective, they are common models for industrial applications. We expanded upon previous studies by focusing on changes in gene expression quantifiable after long-term (120 d) acclimation to a gradient of five salinity treatments, rather than the short-term (h to d) response to osmotic stress.

3.3 Materials and Methods

3.3.1Salinity experiment and RNA-seq.

We grew four strains of *Cyclotella cryptica* (Fig. 1) in triplicate at five salinity (S) treatments (0, 2, 12, 24, 36) for a total sample size of 60 (4 strains x 3 replicates x 5 treatments). We used an artificial seawater recipe that kept nutrient levels constant across salinity treatments while varying the amounts of salts depending on the desired salinity (Table S1 in the Supporting Information). We used a semi-continuous batch culture design and transferred the cultures into fresh media every 5 or 6 d, ensuring low cell density and maintaining exponential growth from one batch to the next. We carried out the experiment on a single shelf of a Percival incubator at 15°C with a 12:12 h light:dark regime and 17.6 µmol photons m^2 s^1 light intensity. To avoid potential batch effects, we randomized the culture tubes daily over a 6 x 3 grid of racks. We used a Trilogy fluorometer (Turner designs) to monitor growth by daily measurement of in vivo relative chlorophyll a fluorescence as a proxy for biomass. This setup provided 5 or 6 relative fluorescence measurements per transfer, all taken during the exponential growth phase. Exponential growth on a semi-log plot is linear, so to estimate growth rates we calculated the slope of a linear regression of the logarithm of relative fluorescence by day (Wood et al. 2005). We ran the experiment for ~120 d, so we were able to obtain repeated estimates of growth rates at each salinity to more accurately estimate the optimal salinity and shape of the reaction norm for each strain. The growth data can be

visualized through an interactive application available at https://diatom.shinyapps.io/Diatoms_and_Salinity/.

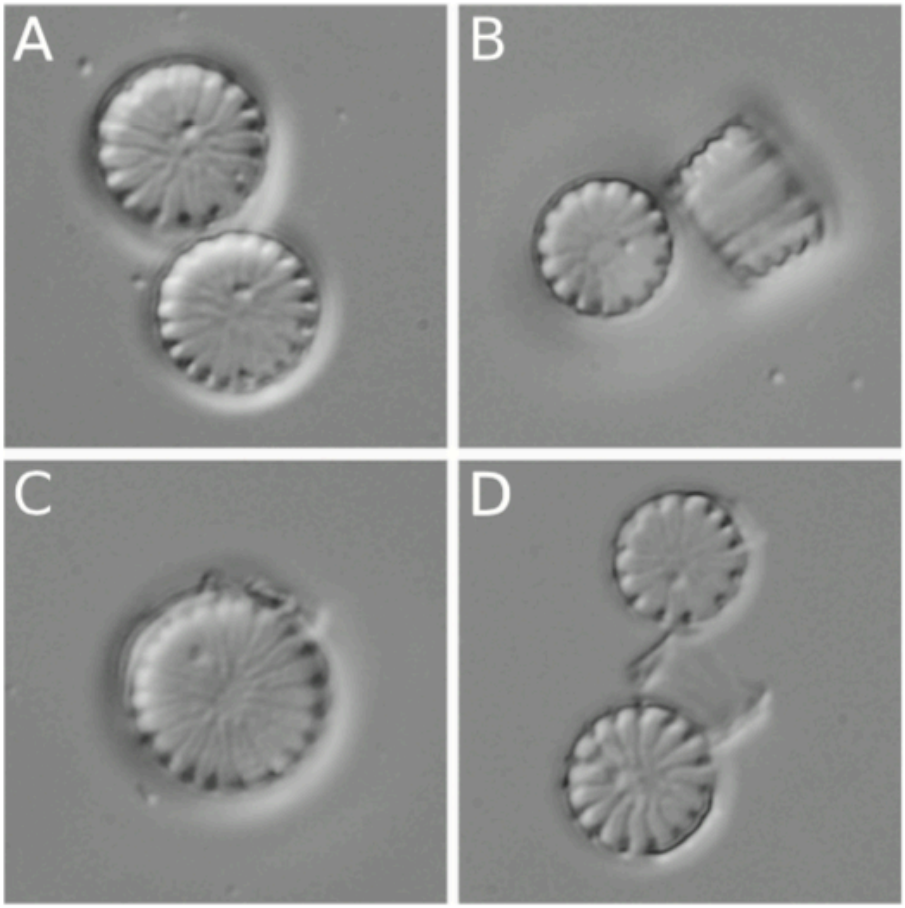


Fig. 3.1. Light microscope images of *Cyclotella cryptica* strains: (A) CCMP331, (B) CCMP332, (C) CCMP333 and (D) CCAC1263b. Note that although these exemplar images show differences in size, the size ranges of the cultures overlapped throughout the experiment. Scale bar equals 10 μm (bottom right).

At the end of the experiment, we harvested cells from one of the three replicates by filtration, immediately extracted total RNA (ZR Fungal/Bacterial RNA MicroPrep, Zymo Research) and froze the RNA at 80°C. We prepared indexed RNA-seq libraries with the Illumina TruSeq kit and sequenced all libraries together on a single lane of Illumina HiSeq2000 (2 9 100 paired-end reads). The RNA-seq data are available at the Short Read Archive (National Center for Biotechnology Information) under bioproject PRJNA589195.

3.3.2 Transcriptome assembly, annotation and read mapping.

We assembled the transcriptomes closely following the procedure described in Parks et al. (2017). Briefly, we corrected sequencing errors with Rcorrector (Song and Florea 2015) before trimming adapters, low-quality bases and low-quality reads with Trimmomatic (Bolger et al. 2014). We pooled the reads originating from libraries of the same strain and assembled a transcriptome for each strain de novo using Trinity (Grabherr et al. 2011). We extracted predicted transcripts and proteins using Transdecoder and annotated the assemblies with the Trinotate pipeline (Bryant et al. 2017) and custom sequence similarity searches with BLAST (Altschul et al. 1990) or Diamond BLAST (Buchfink et al. 2015) against the SwissProt, Uniref90 and Trembl databases. We also per- formed Hidden Markov Model searches using HMMER (Finn et al. 2011) and predicted signaling peptides and transmembrane domains with SignalP (Petersen et al. 2011) and TMHMM (Krogh et al. 2001) respectively. Finally, for the reference transcriptome (CCMP332), we also carried out a KEGG pathway annotation using the BlastKoala web server (Kanehisa et al. 2016b) and the standalone KofamKoala program (Aramaki et al. 2019). We used AFAFind (Gruber et al. 2015) to identify proteins that are targeted to the chloroplast.

We mapped reads to the reference transcriptome using Salmon (Patro et al. 2017) with default settings and assessed the variance around abundance estimates with Salmon's Gibbs sampling option with 1,000 samples. We elected to map against the transcriptome because it is less computationally demanding compared to alignment and quantification against the highly fragmented genome assembly of Cyclotella cryptica (Traller et al. 2016). We converted transcript-level abundance estimates into gene-level counts using the R/Bio-conductor package tximport (Love et al. 2019).

3.3.3 Differential gene expression.

We analysed gene-level expression data using the R/Bioconductor package DESeq2 (Love et al. 2014). We started with an analysis where salinity was a factor with five levels

corresponding to our five salinity treatments. Each treatment had four biological replicates corresponding to the four strains of C. cryptica. Preliminary results with this design showed that differences in gene expression originating from differences between strains/genotypes overwhelmed the signal of the salinity effect, thereby reducing the statistical power necessary to detect differential gene expression due to salinity. Further evidence of divergence among the four strains came from: (i) the mapping rates to the reference transcriptome, (ii) sequence divergence in the full-length large ribosomal subunit rRNA and (iii) shapes of the growth reaction norms.

To improve our statistical power, we pooled the salinity treatments into low salinity (S = 0 and S = 2) with a total of eight “replicates” and high salinity (S = 12, S = 24 and S = 36) with a total of 12 “replicates” (Table 1). Although no grouping of treatments is ideal, this strategy seemed most natural for a euryhaline species most commonly recorded in waters of higher salinity because it measures the difference in gene expression between essentially freshwater (S = 0 and S = 2) and brackish + marine water (salinity S = 12, S = 24 and S = 36). Using this design, we ran two analyses, one naive to potential genotype effects (design: abundance salinity) and another that included genotype as a fixed-effect term (design: abundance salinity + genotype). In the latter case, we had two replicates per treatment per genotype for the low salinity treatment and three for the high salinity treatment.

A downside of converting the five salinity treatments into two conditions, low versus high salinity, was that the variation within the pooled treatments could become inflated, possibly reducing power to detect differentially expressed genes. To address this, we treated genotype as a nuisance effect and tried to remove its influence using functions available in the R/Bioconductor package RUVSeq (Risso et al. 2014). We applied the RUVg routine that uses a set of negative control genes (known not to be differentially expressed between the relevant treatments) to estimate the amount of unwanted variation, that is, variation originating from

nuisance factors instead of the treatment effects. We constructed an empirical set of negative control genes by sorting all expressed genes by their adjusted P-values from a genotype-naive

Table 3.1. Number of differentially expressed genes or orthogroups in different analyses with four and three genotypes. In the three genotype analysis, we removed the genotype CCAC1263b because its genetic divergence, resulting in a lower read mapping rate and different reaction norm, which together might have disproportionately influenced the results. We used an absolute LFC ≥ 1 and P-value ≤ 0.05 after adjusting for multiple tests. S = salinity expressed as grams salt per litre of water.

Analysis	Contrast	Genes: four genotypes	Genes: three genotypes	Orthogroups: four genotypes
(S=12+S=24+S=36) vs (S=0+S=2)	High: low	263	420	155
(S=12+S=24+S=36) vs (S=0+S=2) + genotype	High: low	534	1346	344
Five treatments	0:12	105	220	69
Genotype-naïve analysis	2:12	37	160	23
	24:12	3	5	5
	36:12	24	47	21
Five treatments + RUVg with two (four strains analysis) and one (three strains analysis) factors of unwanted variation	0:12	105	174	226
	2:12	157	477	87
	24:12	456	738	12
	36:12	482	662	105

analysis (design: abundance salinity) and taking a set of 6,757 genes that were least likely to be differentially expressed. Then, we estimated three factors of unwanted variation meant to be included as covariates in downstream analyses. We constructed diagnostic PCA plots for the first, first and second together and all three factors of unwanted variation combined to assess whether salinity is the main source of variation in gene expression after accounting for nuisance effects (i.e., samples cluster by treatment rather than strain in a PCA biplot). An analysis with two factors of unwanted variation was sufficient to capture the salinity effect, so downstream analyses were performed with these settings (design: abundance – salinity + first factor of unwanted variation + second factor of unwanted variation).

We calculated the total number of differentially expressed genes in various comparisons by filtering for absolute log2 fold change (LFC) ≥1 and P-value ≤0.05 (adjusted for multiple tests). To assess the importance of genes to the given treatment combinations, we ranked the

genes using three methods: (i) directly by using raw LFC, (ii) the effect-size shrinking procedure (Stephens 2017) and (iii) the Topconfects method (Harrison et al. 2019). In general, ranking by raw LFC is less conservative than the two methods that produce moderated effect sizes. Thereafter we looked more closely at the functional annotation of the top 50 differentially expressed genes for their potential roles in the response to salinity change. We also performed functional comparisons by counting the frequencies of GO categories among the top 500 differentially expressed genes in various comparisons Orthogroup-level analysis. As an alternative to the reference genotype approach, where we mapped all libraries to the assembly of CCMP332, we performed an analysis where we mapped reads and quantified abundance from each genotype to the de novo assembly of the transcriptome of the same genotype. This approach offset any potential biases introduced by mapping reads from four different genotypes to the transcriptome assembly from a single reference genotype. However, this approach required an alternative way to combine the genes and read counts. To do this, we ran OrthoFinder (Emms and Kelly 2015, 2019) with default settings to infer orthogroups for the four transcriptomes, changing the unit of comparison from a “gene” to “orthogroup.” In the vast majority of orthogroups, multiple Trinity genes of a genotype were assigned to the same orthogroup. To take all these data into account, we summed the expression of reads across members of an orthogroup and assigned this sum as the expression level of the orthogroup for a particular genotype. Downstream analyses were identical to those described for the gene-level expression analyses.

3.3.4 Biosynthetic pathways and transporter genes.

We identified pathways previously proposed to be involved in the response to salinity stress by diatoms and assessed whether the expression of key genes across salinity treatments was consistent with expectations based on previous studies. We focused on the biosynthetic pathways of four osmoprotectant molecules: proline (Liu and Hellebust 1976a,b), glycine betaine (Dick- son and Kirst 1987), taurine (Jackson et al. 1992) and DMSP (Kirst 1996, Lyon et

al. 2011). We predicted that genes involved in the synthesis of these compounds would be induced at high salinity, whereas genes involved in the catabolism of these compounds would either be repressed or not differentially expressed. For proline, glycine betaine and taurine pathways, we identified relevant genes based on KOFAM mapping of Trinity genes to members of the respective pathway from the reference KEGG database (Kanehisa et al. 2016a). For genes associated with DMSP synthesis, we downloaded protein sequences for the candidate genes from *Fragilariopsis cylindrus* and *C. nana* as circumscribed in Lyon et al. (2011) and Kageyama et al. (2018a), respectively, and used BLAST searches against the predicted proteins from the de novo assembly of *C. cryptica* strain CCMP332 to identify putative homologs.

We also investigated the expression levels of genes encoding proteins involved in membrane transport of osmolytes (amino acids, glycine betaine, DMSP), sodium (Na^+) and potassium (K^+) cations, and water and small non-polar molecules (aquaporins). We identified genes that match transporters through protein domain searches against the PFAM and KEGG databases. In all cases, we used e-value cut-off of $1e^{-10}$.

3.4 Results

3.4.1 Salinity reaction norms.

We obtained a total of 1,560 growth rate estimates for four strains of C. cryptica, each grown in triplicate at five salinity treatments ranging from S = 0 to S = 36 salinity (78 estimates per strain + treatment combination). In culture collections, three of the strains (CCMP 331, 332, and 333, National Center for Marine Aquaculture) have been maintained at S = 30, whereas the fourth (CCAC1263b, Culture Collection of Algae at the University of Cologne) has been maintained at S = 16 salinity. To determine the optimal salinity level for growth and the salinity reaction norm for each strain, we summarized the variation in growth rates across replicates and transfers (sequential replicates; Fig. 2). All four strains of *C. cryptica* grew across the entire

range of tested salinities, and the optimal salinity for all strains was lower than the conditions of their long-term maintenance. Strains CCMP331–333, isolated from saline inland pools on Martha's Vineyard, USA, had hump-shaped reaction norms, with the fastest growth rates at intermediate salinity (S = 12) and slower growth rates at lower and higher salinities. Strain CCAC1263b, isolated from the Serchio River estuary in Tuscany, Italy, grew fastest at S = 0 and S = 2 with slower growth rates at higher salinities. Among the NCMA strains, the mapping reference strain, CCMP332, had slightly lower growth rates at all salinity treatments. The summarized growth rates indicated that, at the preferred salinity, the cultures doubled on average about four times in 5 d (Fig. 2).

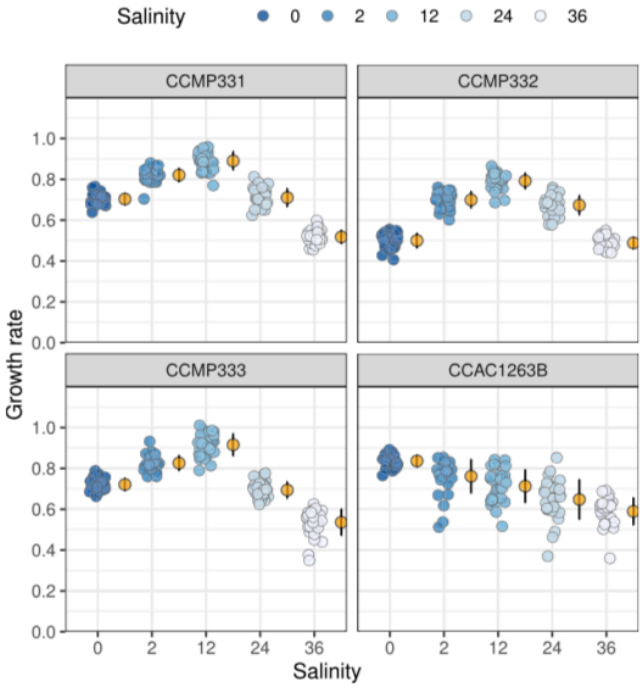


Fig 3.2. Growth rates for four strains of Cyclotella cryptica at different salinities. Each point represents a single estimate of the slope of the logarithm of in vivo relative fluorescence against time (d). The points are plotted with a horizontal jitter to avoid overplotting. Higher salinities are represented by lighter shades. The circles to the right of a cluster show the mean growth rate across replicates and sequential transfers. The error bars show the mean SD [Colour figure can be viewed at wileyonlinelibrary.com]

3.4.2 Transcriptome assembly, sequence divergence and read mapping rates.

We used a de novo assembly of the transcriptome of strain CCMP332 as a reference for read mapping. We combined the libraries prepared for each treatment into a single read pool of 99 million read pairs and assembled these using Trinity with default settings. The final assembly consisted of 30,497 nuclear contigs (Trinity transcripts) belonging to 13,514 Trinity genes.

Although all of our strains were identified as *C. cryptica*, shapes of the salinity reaction norms indicated that one of the strains, CCAC1263b, in our experiment responds to salinity differently than the other three strains (Fig. 2) and so might be quite different ecologically. To assess the degree of divergence among our strains, we aligned the longest large ribosomal subunit rRNA transcripts per assembly, trimmed the alignment to a final length of 3,438 sites to avoid counting indels due to possible incomplete assembly at the ends of contigs, and calculated the pairwise number of substitutions. Compared to the reference strain (CCMP332), CCMP331 had six substitutions (three transversions), CCMP333 had seven substitutions (four transversions), and CCAC1263b had 13 substitutions (3 transversions). The same analysis for the small ribosomal subunit rRNA revealed six to eight substitutions across pairwise comparisons to the reference. The strains CCMP331 and CCMP333 had a single substitution in the LSU rRNA contigs and identical SSU rRNA contigs.

Consistent with these results, mapping rates of libraries against the reference transcriptome also varied. Reads of the different libraries of the reference strain CCMP332 mapped to the reference assembly at a rate of 95%. Reads from CCMP331 and CCMP333 at a rate of 88%–89%, and reads from CCAC1263 mapped at a rate of 85%–86%. In total, after summarizing transcript-level abundances to Trinity “genes,” we obtained counts for 13,512 genes.

3.4.3 Differential gene expression.

A common experimental design for clonally dividing unicellular organisms is to replicate the experiment using the same genotype. This design tests whether multiple aliquots of the

same culture strain (genotype) respond to some treatment in the same way, either simultaneously or asynchronously. Our experiment differed from this design in that our unit of replication was not sub- samples of the same culture, but different genotypes of the same species. This approach had two main consequences. First, we expected to see more variation among the replicates compared to experiments where the unit of replication is a single culture strain (genotype). Second, the patterns we observed should be more robust, because they would be strong enough to be detected despite the differences inherent due to different genomes, culture origin, ecology and long-term culture maintenance. In other words by using multiple genotypes the experiment arguably better approximates natural situations. Since we have gene expression data for four genotypes, our replication design allowed us to assess the overall effect of salinity on relevant genes and pathways. However, because we do not have within-genotype replication for gene expression, we could not address possible interactions between genotypes and salinity treatments.

In our first analysis, we treated the four strains as biological replicates and salinity as a categorical predictor with five levels corresponding to the five treatments. As the majority of strains had fastest growth rates at S = 12, we considered this the salinity optimum and used it as the reference condition. The reported number of differentially expressed genes are therefore relative to S = 12 unless otherwise noted (Table 1). A principal components analysis of the per-gene abundance data showed that libraries clustered by strain rather than treatment, with CCMP332 separated from CCMP331 and CCMP333 along PC1 and CCAC1263b separated from the remaining three strains along PC2 (Fig. 3A). Given that most variation in per-gene counts was related to genotype effects, we expected a low number of differentially expressed genes between salinity treatments. Indeed, we found 105 differentially expressed genes in the comparison of S = 12 to S = 0 and 24 differentially expressed genes in the S = 12 to S = 36 contrast [absolute LFC ≥ 1 and adjusted P-value (Padj) ≤ 0.05]. The salinity stress response commonly involves a much greater portion of the transcriptome (Cheng et al. 2014, Bussard et

al. 2017), so we performed two additional analyses, either to directly account for potential genotype effect on gene expression or to control genotype effects as an unwanted, nuisance factor.

In a genotype-naive analysis with this setup (abundance – salinity) we found 111 upregulated and 152 downregulated genes in low relative to high salinity (absolute LFC ≥ 1; Padj ≤ 0.05). We then accounted for potential genotype effects with the model abundance – salinity + genotype with two replicates per genotype for the low salinity and three for the high salinity treatment. With this model, we found 256 up- and 278 downregulated genes in the low versus high salinity treatment. In contrast, the genotype effect for the comparisons between the reference strain and any of the three additional strains resulted in a minimum of 3,140 differentially expressed genes (CCMP333 to CCMP332 contrast). These results suggest that accounting for differences between genotypes substantially increased our power to detect differentially expressed genes between salinity treatments and that many more genes were differentially expressed between genotypes than between salinity treatments.

The above analysis has the caveat that we could not test for differential expression between salinity treatments within the pooled treatments (e.g., S = 12 vs. S = 36). The second alternative for dealing with the genotype effect was to treat it as an unwanted factor and control for its effect while simultaneously estimating the effect of salinity using the RUV method (Risso et al. 2014). We performed exploratory analyses accounting for up to three factors of unwanted

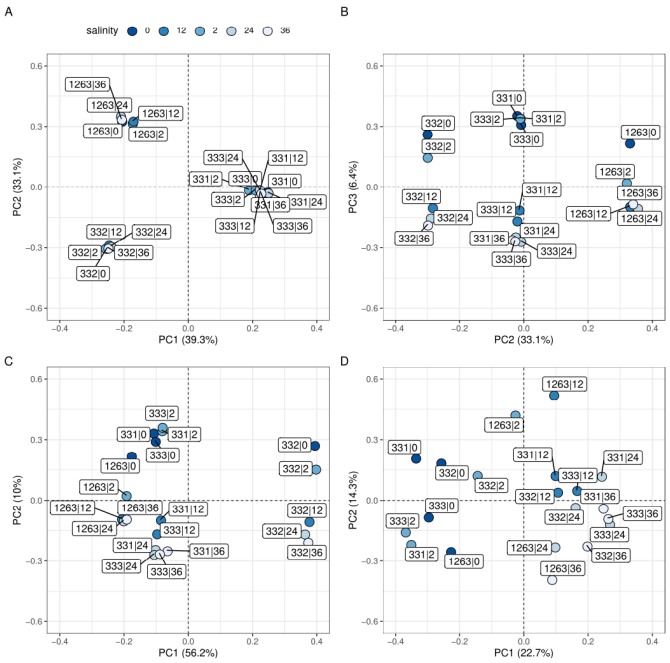


Fig. 3.3. Principal Component Analysis biplots of gene expression counts. Shown are principal component 1 (PC1) and 2 (PC2) for normalized counts (A), PC2 and PC3 for normalized counts (B), PC1 and PC2 for RUVg (Removal of Unwanted Variation using control genes) analysis with one factor of unwanted variation (C), and PC1 and PC2 for RUVg analysis with two factors of unwanted variation (D). Numbers in parentheses next to axis titles are the percent variance explained by that principal component. The samples are shaded by salinity, and the sample labels indicate the strain and salinity treatment separated by a vertical bar. Differences in expression are dominated by genotype effects in plots A and C. In plot B the target contrast between low (S = 0 and S = 2) versus high salinity (S = 12, S = 24, S = 36) is loaded on PC3, while in plot D, the target contrast is loaded on PC1 (i.e., all low salinity samples have negative PC1 scores and all high salinity samples have positive PC1 scores).

variation (Fig. 3, B and C). With one factor of unwanted variation, we found that samples still clustered primarily by strain with 56.15% of the variance (PC1) explained by differences between the reference strain and all other strains, but PC2 (9.99% of variance) separated the samples by high and low salinity. Adding a second factor of unwanted variation to the model further improved the clustering in that PC1 (22.71% of variance) now separated the samples based on salinity, with S=0 and S=2 in one group (negative PC1 scores) and S=12, S=24 and S=36 in another (positive PC1 scores). DESeq2 analyses with a single factor of unwanted variation showed that the number of differentially expressed genes ranged from 23 between S=0 and S=12 salinity to 255 in the S = 24 to S = 12 comparison (Table 1). With two factors of

unwanted variation these numbers increased to 105 for the S = 0 to S = 12 comparison and 482 for the S = 36 to S = 12 comparison, again showing increased power to detect differential expression associated with salinity after accounting for or removing the effect of genotype.

We also analysed the data after removing strain CCAC1263b, which had the lowest mapping rate to the reference transcriptome and a differently shaped salinity reaction norm. With this three-strain dataset one factor of unwanted variation was sufficient to capture the salinity treatments on PC1 (28.47% variance). The results otherwise agreed that accounting for genotype, either as a predictor or nuisance factor, increased our power to detect differential expression (Table 1).

3.4.4 Ranking of differentially expressed genes.

We ranked genes by (i) raw LFC, (ii) using the adaptive shrink- age method of Stephens (2017) available in DESeq2 and (iii) the Topconfects method (Harrison et al. 2019). Moderated effect-size estimation by the adaptive shrinkage and Topconfects methods can pro- vide very different gene rankings and, consequently, emphasize different biological functions compared to raw LFC values (McCarthy and Smyth 2009, Love et al. 2014, Harrison et al. 2019). Using these three approaches, we found markedly different gene sets and biological processes represented among the top 50 differentially expressed genes. One clear example comes from the analysis with salinity recoded into low versus high with genotype as the predictor (design: abundance – salinity + genotype). The discrepancy between the ranked gene sets was evident from the median expression of the top 50 ranked genes obtained by the three methods. For raw LFC ranking, the median expression was just 22.7 reads, for shrunken LFC (LFCshrink) ranking 150 reads, and for Topconfects ranking 534 reads, revealing that lowly expressed genes were enriched in raw LFC rankings. The overlap between the three top-50 sets showed that just 27–29 genes were shared among the raw LFC ranking and the other two methods, whereas the LFCshrink and Topconfects ranked lists shared 38 of the 50 genes, suggesting better agreement when using methods that moderate effect sizes. With respect to functional

annotations, only 15 out of the top 50 genes ranked by raw LFC had hits to the PFAM and KEGG databases (e- value < $1e^{-10}$), compared to 23 and 27 for LFCshrink and Topconfects ranking respectively. Similarly, the overlap in functional annotations represented among the top 50 ranked genes was higher between the LFCshrink and Topconfects ranking methods (17 PFAM and 18 KEGG accessions shared) relative to the overlap between either of these and raw LFC ranking (13 PFAM and 8 KEGG accessions shared).

3.4.5 Biological functions represented among the top ranked genes.

Here, we focus on results from the two analyses that accounted for genotype effects, either as a predictor or nuisance factor, and on functional annotation of the top 500 differentially expressed genes as ranked by the Topconfects method. For the analysis with salinity recoded into low versus high with genotype as predictor, gene ontology classification based on crossreferencing PFAM domains against GO categories showed enrichment for integral membrane components, transmembrane transport, oxidation-reduction processes and ATP binding/protein phosphorylation activity (Fig. 4A). On a per-gene basis, the top 50 ranked genes with KEGG hits included two enzymes from the glycine/betaine biosynthesis pathway, a proline/betaine transporter gene, four sodium or proton-de- pendent solute transporters (for amino acids, phosphate or dicarboxylate), two cysteine-rich secretory proteins and genes likely involved in cell signalling pathways and transcriptional/translational regulation. We found similar results for the RUVg analysis of the five-treatment dataset, specifically the contrast between S = 12 (reference condition) and S = 36 salinity, where genotype was a nuisance effect. In addition to the GO terms found in the high versus low salinity analysis, the S = 12 versus S = 36 contrast also had a higher number of regulated genes with functions in sequence-specific DNA binding and transcriptional regulation (Fig. 4B). On a per-gene basis, the top KEGG hits for the two analyses were nearly identical, with the addition of a cation-transporting ATPase, two protein kinases and a STEAP4 metalloreductase in the RUVg analysis. Reanalysis with strain

CCAC1263 removed gave results that were qualitatively similar in terms of gene ranking and GO category representation among the top 500 ranked genes.

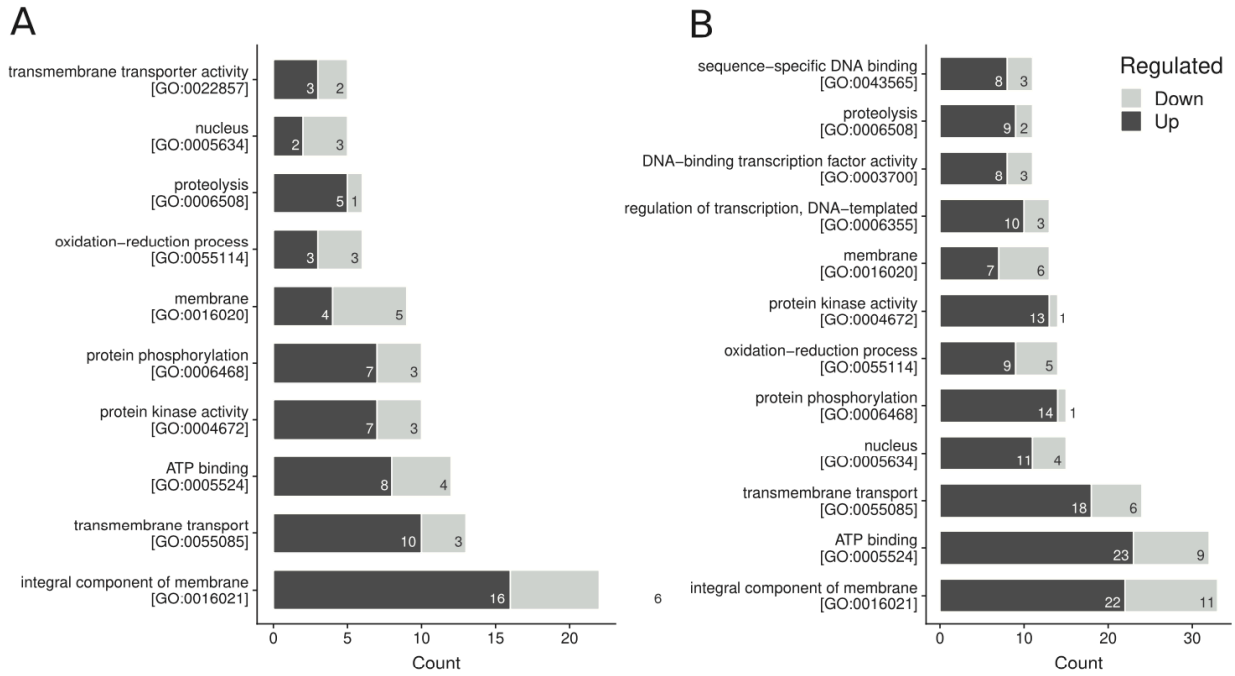


Fig. 3.4. The 10 most common gene ontology categories represented among the top 500 differentially expressed genes across genotypes and salinity treatments. (A) Results from the analysis with salinity recoded into low versus high with genotype as predictor. (B) Results from the analysis with five treatments and genotype as a factor of unwanted variation.

3.4.6 Orthogroup-level differential expression.

Finally, motivated by the differences in sequence (for ribosomal rRNA genes) and lower mapping rates to the CCMP332 reference transcriptome, we also performed an orthogroup-level analysis where we mapped each library to the transcriptome assembly of its corresponding genotype (i.e., we mapped CCMP331 grown at S = 12 to the assembly of all CCMP331 libraries, not to the reference CCMP332 assembly). To construct a common set of units for which we measured expression, we inferred orthogroups for the four transcriptomes and compared the expression of a particular orthogroup across treatments to assess differential expression. We annotated the orthogroups based on the gene- level annotation described earlier. As the genes are nested in orthogroups, the annotations of the genes belonging to an

orthogroup can be assigned to the orthogroup itself. Thereafter, we performed the same procedures described for the gene-level analysis.

In total, we had 10,018 orthogroups represented across all libraries. There were generally fewer differentially expressed orthogroups than genes for a particular analysis (Table 1). This was expected because we had fewer orthogroups than genes (in the OrthoFinder procedure genes within a genotype can be clustered within the same orthogroup). However, despite the lower count of differentially expressed orthogroups, there was close agreement between the orthogroup- and gene-level analyses across the relevant osmolyte biosynthetic pathways and transporter genes (Figs. S1–S13 in the Supporting Information). For brevity, we focus our remaining discussion on the gene-level results.

3.5 Discussion

3.5.1 Osmolyte biosynthetic pathways.

In *C. cryptica*, experimental data have shown that increasing the concentration of free intracellular proline is a primary mechanism for mitigating hyperosmotic stress (Liu and Hellebust 1976a, Hellebust 1985). The use of glycine betaine and DMSP in osmoregulation has not been shown in *C. cryptica*, though they are present in the model diatom species, *C. nana* (formerly *Thalassiosira pseudonana*; Kageyama et al. 2018). Taurine is an osmoprotectant in some species of *Pseudonitzschia* (Jackson et al. 1992), a distantly related raphid pennate diatom. Importantly, all of these findings pertain to short-term (h to d) osmotic shock whereas our experiment assessed the longer-term acclimation (4 months or about 96 generations) to altered osmotic conditions.

No evidence for transcriptional regulation of proline biosynthesis genes in long-term acclimation trial. The KEGG reference pathway for arginine and proline synthesis describes four potential pathways for synthesis of free proline. Two of these involve single reactions: (i) cleavage of N-terminal proline from peptides by proline iminopeptidase and (ii) direct conversion

of ornithine to proline by ornithine cyclodeaminase. Proline synthesis can also occur in a stepwise fashion through the intermediate 1-pyrroline 5-carboxylate from either (iii) glutamate, catalysed by 1-pyrroline-5-carboxylate dehydrogenase or (iv) ornithine, via ornithine aminotransferase. It is unknown if all of these pathways are present in diatoms as only the latter two have been described in previous studies. KOFAM searches using the predicted proteins from our reference strain CCMP332 indicated that enzymes from each of these pathways were present in C. cryptica (e-values < $1e^{-28}$). We examined expression patterns of these genes as well as ones involved in closely related reactions (e.g., arginase and glutamate 5-kinase). Since the function of these enzymes is synthesis of free proline, we predicted that their expression should be induced at higher salinity. Although several genes had expression patterns consistent with this prediction, none of the differences were statistically significant (Figs. S1 and S2).

Previous studies on *C. cryptica* (Liu and Hellebust 1974, 1976a,b) and *F. cylindrus* (Krell et al. 2007, 2008, Lyon et al. 2011) have shown that proline is an important osmoprotectant. Gene expression studies have further shown increased expression of the most relevant enzyme, pyrroline-5-carboxylate reductase (enzyme commission [EC]:1.5.1.2), which catalyses the last reaction in stepwise proline synthesis pathways 3 and 4 described above (Krell et al. 2007, 2008, Lyon et al. 2011). Our result is not necessarily incompatible with these studies, however, as they focused more on the short-term response to osmotic stress (at the scale of h to d), whereas our 4-month experiment measured the acclimated response to osmotic stress. It is possible that the build-up of intracellular proline is part of an immediate coping strategy, one missed by our experiment, and is later replaced or augmented by alternative mechanisms that were captured by our longer-term experiment. It is also possible that transcript and protein abundances are disconnected for proline synthesis genes, as has been found for glycine betaine synthesis genes in response to hyperosmotic stress in *C. nana* (Kageyama et al. 2018b).

3.5.2 Methyltransferases play key roles in modulating glycine betaine and DMSP concentrations.

Glycine betaine: The biosynthetic pathway for glycine betaine (trimethylglycine) has been studied in the model diatom *C. nana*, which is also capable of growing across a broad salinity range (Brand 1984). The genome of *C. nana* contains enzymes for two glycine betaine synthesis pathways, either via oxidation of choline (choline dehydrogenase and aldehyde dehydrogenase) or through stepwise methylation of glycine (glycine/sarcosine/dimethylglycine methyltransferase). The key enzyme in the glycine methylation pathway in *C. nana* (annotated as TpGSDMT) contains tandemly repeated methyltransferase domains with different substrate specificity for the precursors glycine, sarcosine and dimethylglycine (Kageyama et al. 2018b). As a result, this enzyme is potentially sufficient to catalyse the entire set of reactions and is induced under hyperosmotic conditions in *C. nana* (Kageyama et al. 2018b). Methyltransferases are also induced in response to osmotic stress in the psychrophilic diatom *F. cylindrus* (Lyon et al. 2011).

Similar to *C. nana*, the transcriptome of *C. cryptica* contains genes for both betaine synthesis pathways. The choline dehydrogenase gene was not significantly differentially expressed. This result is also consistent with experiments for choline and glycine betaine uptake in sea-ice diatoms, which do not actively uptake exogenous choline even though it could be converted to glycine betaine. In contrast, a gene from the alternative, glycine betaine pathway, best annotated as a sarcosine/dimethylglycine methyltransferase, was upregulated 7.6-fold (LFCshrink; Padj = $1.00e^{-85}$) in high versus low salinity (Fig. 5, Fig. S3). As in C. nana, the homolog of this gene in *C. cryptica* contains tandemly repeated methyltransferase domains, suggesting that it, too, can catalyse the sequential methylation of glycine, sarcosine and dimethylglycine to produce betaine.

Interestingly, we also detected a sixfold induction of a sarcosine oxidase/pipecolate oxidase gene that in bacteria and mammals catalyses a reverse sarcosine-glycine reaction

(LFCshrink; Padj = $5.47e^{-51}$; Fig. 5, Fig. S3). This seems counterintuitive, as the methyltransferase and oxidase would be competing for sarcosine as a substrate and then converting it in oppo- site directions, one to dimethylglycine and thereafter betaine, the other back to the unmethylated glycine. However, the gene identified as sarcosine oxidase had a 20-amino-acid signal peptide and a heterokont-specific chloroplast target peptide, which suggests that it is transported through the plastid membranes and functions in the stroma rather than the cytosol. In the chloroplasts of *Arabidopsis*, this enzyme targets sarcosine only opportunistically, and its primary substrate is pipecolic acid (Goyer et al. 2004), which is a component of the plant immune response (Hartmann et al. 2018). If this sarcosine/pipecolate oxidase has a similar function in the chloroplast of diatoms, then it is unlikely to compete for substrate with the cytoplasmic glycine/sarcosine/dimethylglycine methyltransferase, both due to its physical separation by the four mem- branes of the diatom plastid and the different affinities of the two enzymes for pipecolate versus sarcosine.

DMSP: Studies in both *C. nana* and *F. cylindrus* have shown that the intracellular concentration of DMSP increases when cells are exposed to higher osmotic pressure (Lyon et al. 2011, Kettles et al. 2014, Kageyama et al. 2018a). The algal DMSP synthesis pathway (Gage et al. 1997), which is present in diatoms (Lyon et al. 2011, Kageyama et al. 2018b), begins with transamination of methionine and is further facilitated by reductase, methyltransferase and decarboxylase enzymes, all of which have candidate genes in *C. nana* and *F. cylindrus*. In *C. nana*, the methyltransferase shows elevated expression at both the RNA and protein levels (Kageyama et al. 2018a), while in *F. cylindrus*, proteomic data indicate that the enzymes catalysing each of these four steps are upregulated during hyperosmotic stress (Lyon et al. 2011).

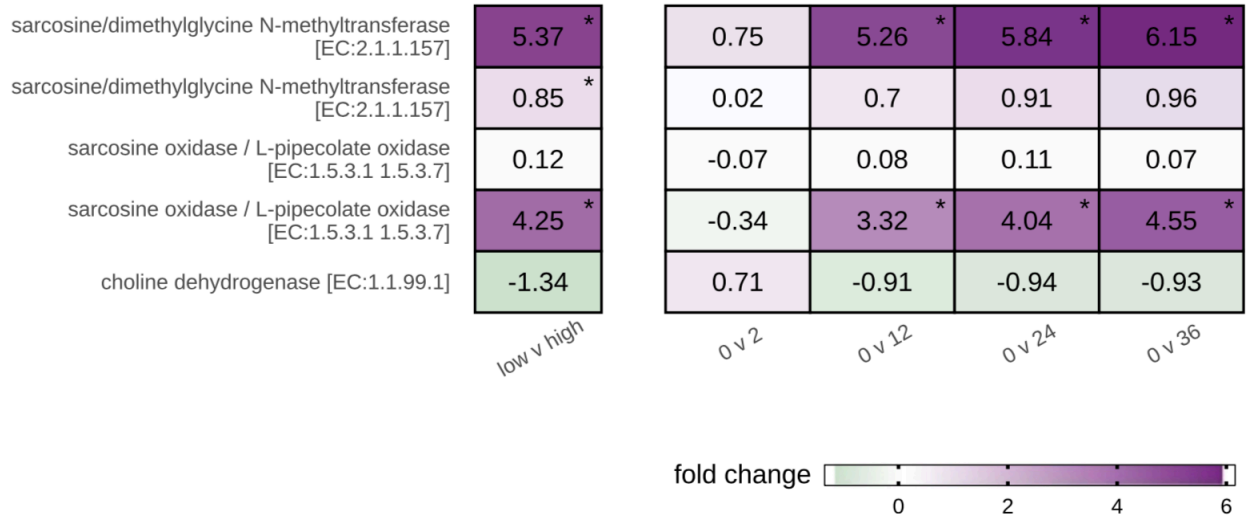


Fig. 3.5. Expression of genes from the glycine betaine and DMSP biosynthetic pathways. Shown is the ratio of transcripts per million (TPM) of a treatment against the TPM at salinity 0. The first methyltransferase is homologous to the methyltransferase involved in DMSP biosynthesis in *Cyclotella nana* (Kageyama et al. 2018a). The second methyltransferase is homologous to the glycine betaine synthesis gene in C. nana with tandem methyltransferase domains (Kageyama et al. 2018b). The left panel shows the expression results from the low versus high salinity analysis with genotype as a fixed effect. The right panel shows the expression results from the analysis with the RUV method, where genotype was part of the unwanted variation. See Table 1 for details of the pooling for the low versus high analysis. Significant differential expression is denoted with a star in the upper right corner. Text in brackets is the enzyme commission (EC) number.

Using protein BLAST searches against the predicted proteomes for *C. nana* and *F. cylindrus*, we identified putative *C. cryptica* homologs for each of the relevant enzymes in the DMSP biosynthetic pathway. In total, we found 12 genes from the de novo assembly of strain CCMP332 with high similarity to the candidate genes. Of these, only a putative homolog of the *C. nana* candidate methyltransferase gene (73% amino acid identity) was differentially expressed, upregulation at higher salinities (LFCshrink = 1.2; Padj = $2.25e^{-4}$; Fig. 5, Fig. S3). This was consistent with previous observations for *C. nana* where DMPS concentration was elevated, but only the methyltransferase step was upregulated in the synthesis pathway (Kageyama et al. 2018a). BLAST searches also identified similar genes in *Thalassiosira oceanica* (74% identity to C. cryptica) and the raphid pennate diatoms *Fragilariopsis*, *Phaeodactylum* and *Fistulifera* (38%–58% identity), further supporting that DMSP might be a

widespread osmoprotectant across diatoms, including both psychrophilic and temperate species.

3.5.3 Taurine as a potentially widespread osmoprotectant in diatoms.

Salinity stress experiments in *Pseudonitzschia pungens* have shown that the concentration of intracellular taurine (2-aminoethanesulfonic acid) changes rapidly in response to sudden changes in ambient salinity (Jackson et al. 1992). In *P. pungens*, exposure of cells acclimated to high salinity (S = 48) to hypoosmotic conditions (S = 15) results in a nearly ninefold decrease in the level of intracellular taurine (Jackson et al. 1992). These findings led us to investigate whether genes involved in taurine metabolism were differentially expressed C*. cryptica*. We found that a transcript, whose best KEGG annotation was gamma-glutamyl-transpeptidase/glutathione hydrolase (EC:2.3.2.2, 3.4.19.13), was repressed at higher salinities (LFCshrink = 3.3; Padj = $3.20e^{-4}$; Fig. 6, Fig. S4). This enzyme catalyses the synthesis of 5-glutamyl taurine from a peptide and taurine, reducing the abundance of free taurine in the cell. The repression of this gene at higher salinity might also indicate the reciprocal, that levels of free taurine are increased to counteract hyperosmotic conditions well into the acclimation period.

We investigated five other genes involved in taurine and hypotaurine metabolism as defined in the corresponding KEGG pathway (rn00430). We found contigs that match glutamate decarboxylase (EC:4.1.1.15), an enzyme involved in hypotaurine synthesis upstream of taurine, whose abundance we predicted should increase at higher salinity, and taurine dioxygenase

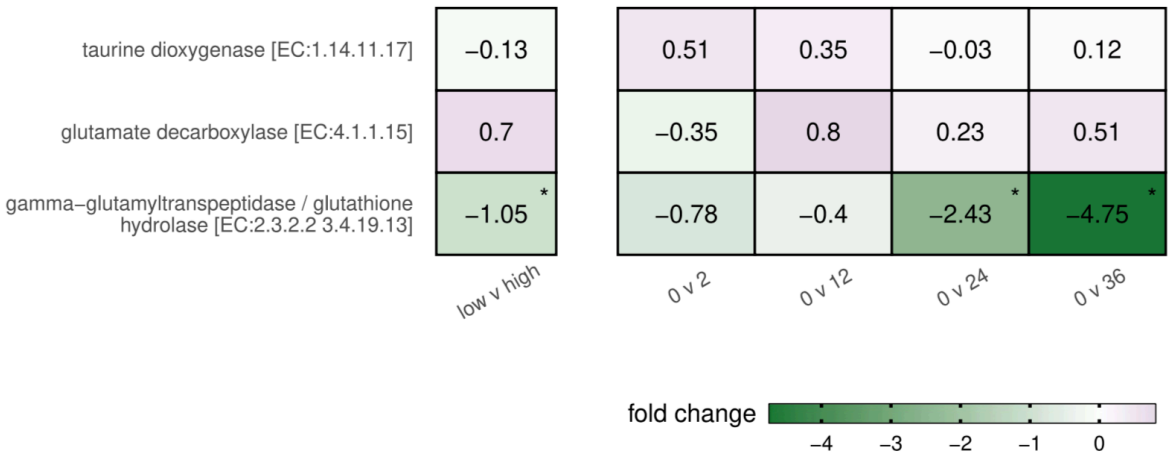


Fig. 3.6. Expression of genes from the taurine biosynthesis pathway. Shown is the ratio of transcripts per million (TPM) of a treatment against the TPM at salinity 0. The left panel shows the expression results from the low versus high salinity analysis with genotype as a fixed effect. The right panel shows the expression results from the analysis with the RUV method, where genotype was part of the unwanted variation. See Table 1 for details of the pooling for the low versus high analysis. Significant differential expression is denoted with a star in the upper right corner. Text in brackets is the enzyme commission (EC) number.

(EC:1.14.11.17), an enzyme that catalyses the breakdown of taurine into aminoacetaldehyde and sulphite, whose expression we predicted should decrease at higher salinity. The observed expression patterns were consistent with our predictions, though the effect sizes were small and tests for differential expression were not significant (Fig. 6, Fig. S4). Overall, although to our knowledge taurine has not been previously implicated in the salinity response of *C. cryptica*, our results suggest that taurine might serve as an osmoprotectant in this species. Our long-term experiments may have captured the end of the transcriptional taurine response, further highlighting the need for additional metabolomic, proteomic and transcriptomic studies across timescales ranging from the short- (minutes to hours) to long-term (days to months), acclimated response. Given the phylogenetic distance between *Cyclotella* and *Pseudonitzschia*, it is possible that regulation of the intracellular concentration of taurine is an important osmoregulatory mechanism that is conserved across diatoms.

3.5.4 Membrane transport systems.

In addition to recruitment of osmolyte synthetic pathways, several diatom species, including *C. cryptica*, are capable of scavenging compatible solutes from the environment both

for growth and salinity stress tolerance (Liu and Hellebust 1974, Petrou and Nielsen 2018, Torstensson et al. 2019). Although our growth media did not include organic compounds, our cultures were not axenic, so it was possible that *C. cryptica* cells could have scavenged compatible solutes released from dead bacterial or diatom cells. Using the best PFAM targets for genes in the CCMP332 transcriptome assembly, we examined differential expression patterns of transporters for proline, glycine betaine and small amino acids. In addition to these organic osmolyte transport systems, we searched for transporters of small cations, expecting the downregulation of Na^+ channels and the upregulation of K^+ channels at higher salinities, as the preferential accumulation of potassium ions over sodium ions might be linked to osmoprotection (Fujii et al. 1995). Finally, we used PFAM-derived annotations to examine the expression of probable aquaporins, gated membrane channels that facilitate the transport of water and small non-polar molecules and have active roles in osmoregulation in microbes (Kruse et al. 2006, Petrova et al. 2013, Matsui et al. 2018).

3.5.5 Glycine betaine and amino acids imported for osmoprotection.

We detected two differentially expressed transporters for proline and glycine betaine. These genes, one from the betaine-choline-carnitine transporter (BCCT) family and the other from the major facilitator superfamily (MFS), differ in both mechanism of transport and transcriptomic response to salinity stress. The BCCT family transporter (TC:2.A.15.1.3) was determined to symport choline alongside H^+ in the *Pseudomonas syringae* bacterium, with a homolog in the related species *Pseudomonas aeruginosa* permitting the uptake of glycine betaine (Chen and Beattie 2008). This transporter was 2.6- fold upregulated (LFCshrink; Padj = $1.07e^{-12}$) in high versus low salinity (Fig. 7, Fig. S5). Given that the choline-glycine betaine pathway genes were not induced at higher salinities, and choline import is negligible relative to glycine betaine uptake by sea ice diatoms (Torstensson et al. 2019), we presume that the BCCT transporter in *C. cryptica* functions primarily in glycine betaine import.

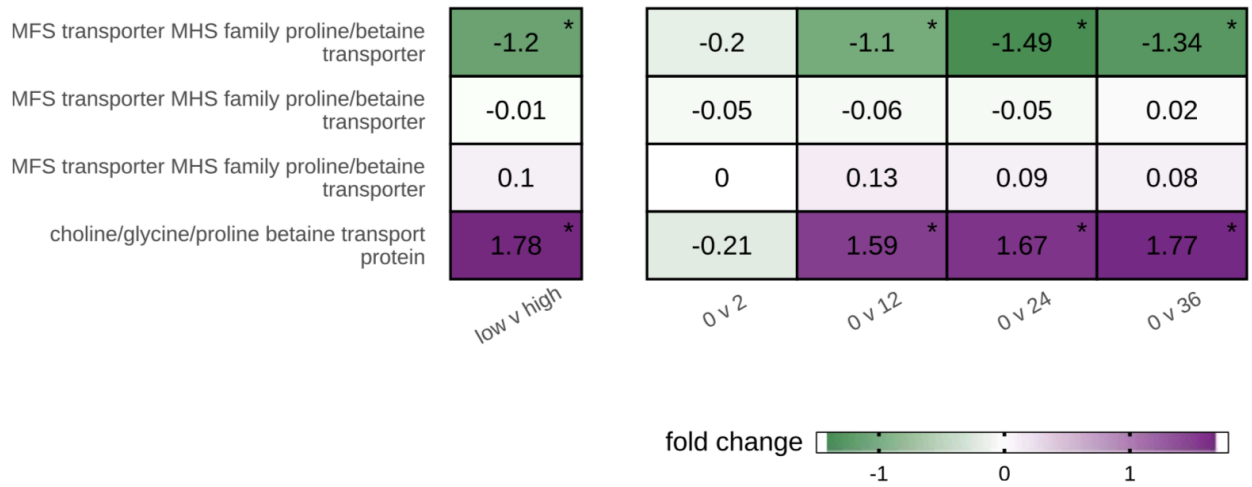


Fig. 3.7. Expression of glycine betaine transport genes. Shown is the ratio of transcripts per million (TPM) of a treatment against the TPM at salinity 0. The left panel shows the expression results from the low versus high salinity analysis with genotype as a fixed effect. The right panel shows the expression results from the analysis with the RUV method, where genotype was part of the unwanted variation. See Table 1 for details of the pooling for the low versus high analysis. Significant differential expression is denoted with a star in the upper right corner. Text in brackets is the enzyme commission (EC) number.

Three genes from our Trinity assembly showed high sequence similarity to an MFS transporter (TC:2.A.1.6.4), each of which had a different expression pattern across salinity treatments (Fig. 7, Fig. S5). Only one of the three genes displayed a consistent expression pattern across all four *C. cryptica* strains, specifically downregulation at higher salinity (LFCshrink = 1.6; Padj = $1.08e^{-12}$). This MFS transporter was identified and characterized in *E. coli* as a symporter of H^+ primarily with proline or glycine betaine, though it could also transport similar molecules, including taurine, pipecolate and dimethylglycine (Keates et al. 2010). In *E. coli*, this transporter is an osmosensor induced by increased osmotic pressure at higher salinity (Keates et al. 2010), which is the opposite of our findings in *C. cryptica*. Given previous work on the uptake of exogenous compatible solutes, including amino-acids, during salinity stress (Liu and Hellebust 1974, Petrou and Nielsen 2018, Torstensson et al. 2019), the apparent downregulation of this transporter at high salinity was unexpected, especially given its important role as an osmoregulator in various bacterial species (Peter et al. 1998, Keates et al. 2010).

3.5.6 Generic amino acid transport.

Based on PFAM annotations, many Trinity genes were determined to contain a solute carrier (SLC) family domain for amino acid transport. Of these, four were significantly differentially expressed, with one having an LFC > 2 (Figs. S6 and S7). The transcripts for this gene indicate a 2.7-fold upregulation (LFCshrink; Padj = $3.96e^{-20}$) at high salinity, as expected given the known roles of amino acids in osmoprotection (Liu and Hellebust 1974, Mazzarelli et al. 2015). According to its Transporter Classification Database annotation (TC:2.A.18.8), this gene likely encodes a proton-assisted amino acid transporter (PAT) and, based on the function of such transporters in other organisms (e.g., *Acyrthosiphon pisum*, *Drosophila melanogaster*, and *Homo sapiens*), might transport small amino acids such as glycine, alanine and proline (Goberdhan et al. 2005, Thwaites and Anderson 2011, Feng et al. 2019). Additionally, it is possible that this transporter has a lower affinity for other molecules, such as taurine and beta-alanine, as has been found for the *H. sapiens* PAT1 transporter (Anderson et al. 2009).

3.5.7 Cation transport depends on electrochemical gradients and co-transported compounds.

Sodium: Nine Na^+ transporter genes were significantly differentially expressed, eight matching the SLC family of transporters and one mapping to a Na^+:K^+ antiporter ATPase (discussed with other potassium transporters below). The functions of several of these Na^+ transporters have been experimentally validated in other organisms, so we examined how those functions might relate to the salinity response in *C. cryptica*. Two Na^+ transporter genes exhibited upregulation under increased salinity (Figs. S8 and S9). The first of these is a sodium-dependent phosphate co-transporter from SLC family 34 (TC:2.A.58.1), which has been identified as a symporter of sodium and inorganic phosphate in rats and flounder, and as an Na^+:Pi co-transporter for the import of organic phosphate in *Vibrio cholerae* (Forster et al. 1999, Lebens et al. 2002). Contrary to expectations, this gene was upregulated at high salinity (LFCshrink = 6.7; Padj = $2.37e^{-37}$). However, it is possible that this transporter is not directly

involved in osmoregulation, but rather indirectly by using the stronger Na^+ gradient caused by higher salinity to facilitate phosphate import. Thus, when the Na^+ gradient is not sufficient for phosphate transport, such as at S=0 or S=2, the transporter is not expressed. The second upregulated Trinity gene is the SLC family–sodium and hydrogen antiporter (TC:2.A.36). This transporter is active in both yeast and plants to facilitate the efflux of Na^+, and confers salinity tolerance in *Zygosaccharomyces rouxii* and in *Arabidopsis thaliana* when over-expressed (Iwaki et al. 1998, Apse et al. 1999, Dutta and Fliegel 2018). The induction of this transporter at higher salinity (LFCshrink = 1.4; Padj = $3.94e^{-8}$) may be related to the need for increased export of Na^+ against the strong Na^+ gradient at high salinity.

Three Na^+ transporter genes displayed an inverse relationship between transcript abundance and salinity (Figs. S8 and S9). Two of these, downregulated 4.3-fold (LFCshrink; Padj = $8.34e^{-17}$) and 0.6-fold (LFCshrink; Padj = $5.68e^{-8}$), matched the SLC family 13 of dicarboxylate transporters (TC:2.A.47.1) that is conserved across taxonomic kingdoms (Hall and Pajor 2005). This transporter symports Na^+ and dicarboxylates such as malate, succinate and fumarate in species ranging from *Homo sapiens* to *Staphylococcus aureus* (Fei et al. 2003, Hall and Pajor 2005, Bergeron et al. 2013). If they perform similar functions in diatoms, their repression at high salinity in C. cryptica might serve as a mechanism for preventing intracellular Na^+ build-up. The third transporter downregulated at higher salinity (LFCshrink = 0.3; P = $2.12e^{-3}$) was a Na^+ and amino acid co-transporter (TC:2.A.18.6) involved in the sequestration of amino acids within the vacuole of *Schizosaccharomyces pombe* (Chardwiriyapreecha et al. 2010). If this transporter has similar localization and function in *C. cryptica*, its downregulation at higher salinity should lead to a lower export rate of amino acids to the vacuole, and is consistent with the strategy of accumulation of amino acids as cytosolic osmoprotectants.

Potassium: Several K^+ transporters exhibited expression patterns consistent with known osmotic stress responses in other organisms. One of these (EC.7.2.2.13), a subunit of an active Na^+:K^+ antiporter found in both prokaryotes (Chan et al. 2010) and eukaryotes (Kristensen and

Juel 2010), was downregulated 1.1-fold (LFCshrink; P = $2.26e^{-4}$) at high salinity treatments (Figs. S10 and S11). As freshwaters and low salinity habitats have a lower concentration of K^+ ions, active transport may be required to pull the exogenous K^+ into the cell against a steep natural gradient in order to maintain transmembrane potential. Conversely, active transport might not be necessary for cells grown in higher salinity media where K^+ is not limiting.

Three passive K^+ transporters had expression patterns that also reflected the availability of exogenous K^+ (Figs. S10 and S11). One K^+ uptake channel (TC:2.A.38.1) from the Trk-Ktr-Kdp family, a group of K^+ transporters that play a role in osmoregulation in bacteria (Epstein 2003, Gries et al. 2013), was induced (LFCshrink = 0.4; Padj = $3.79e^{-3}$) in higher salinity treatments. The remaining two K^+ transporters (TC:1.A.2.1.2, high-affinity, and TC:1.A.1.4.1, low-affinity) are classified as inwardly rectifying K^+ channels (biased towards uptake rather than export) from the Kir family. The high-affinity transporter exhibited a positive correlation with increasing salinity (LFCshrink = 0.4; Padj = $1.97e^{-3}$), suggesting that K^+ might act as an osmoprotectant in C. cryptica, as was previously found in *Chaetoceros muelleri* (Fujii et al. 1995).

Two separate genes were best characterized as AKT1 low-affinity K^+ channels (TC:1.A.1.4.1) but showed opposing expression patterns (Figs. S10 and S11), making their potential function in osmoregulation in *C. cryptica* difficult to interpret. AKT1 channels are most effective at intermediate K^+ concentrations, and in *Arabidopsis* certain isoforms have been found to act as “potassium batteries” for other transmembrane transporters (Sandmann et al. 2011). They may have similar roles following post-transcriptional regulation in *C. cryptica*.

Aquaporins: Aquaporins are gated channels whose function is primarily regulated through a variety of post-transcriptional mechanisms. In *Arabidopsis thalina*, gating is triggered by pH fluctuation, closing the channel when environmental pH rises (Tournaire-Roux et al. 2003). Porins in *Beta vulgaris* are controlled by divalent cation-gating, with channel function inhibited by the binding of Ca^{2+} (Alleva et al. 2006). The gating of aquaporins in *Chara corallina*

is controlled by osmolyte concentrations (Ye et al. 2004) and by membrane tension in *Saccharomyces cerevisiae* (Soveral et al. 2008). A notable exception is the AQ2Y aquaporin in yeast, whose expression is inhibited by the high-osmolarity glycerol system (Soveral et al. 2010). Given the primarily post-translational functional regulation of aquaporin channels, it is unsurprising that we did not find evidence of differential expression of aquaporin genes in *C. cryptica* (Figs. S12 and S13).

3.6 Conclusions

We used RNA-seq to study patterns of gene expression in four different genotypes of the euryhaline diatom, *C. cryptica*, following a long-term acclimation trial to salinity levels that spanned the natural range of fresh to oceanic water. All four genotypes grew across the entire salinity range, although optimal salinities for growth differed between strains from New England and Italy. We observed substantial among-genotype variation in gene expression, which we accounted for in our statistical analyses resulting with inferred patterns of gene expression that were robust to differences among genotypes. Although our experiment was not designed to assess genotype-by-environment interactions, our results hint at the potential importance of intraspecific variation in understanding the osmotic stress response in diatoms.

We focused on genes involved in the synthesis and membrane transport of organic osmolytes and monovalent cations—key components of the cellular response to osmotic stress. Our results showed that in long-term acclimation to salinity, *C. cryptica* modulates the concentrations of glycine betaine, DMSP and taurine. Interestingly, we found no evidence for transcriptional regulation of the synthesis of proline, despite it being one of the most frequently reported osmolytes in short-term osmotic stress studies in diatoms. Without protein and metabolite data, our results are not entirely conclusive, but this discrepancy could suggest a temporal sequence in the mechanisms recruited to cope with osmotic stress. Expression profiles of osmolyte membrane transport systems revealed enhancement of amino acid, glycine

betaine and possibly DMSP uptake. Expression patterns of sodium and potassium transporters were consistent with expectations given the electrochemical gradients and co-transported molecules, with a general tendency for increased accumulation of potassium and increased export of sodium at higher salinity. Overall, our results highlight the transcriptional regulation of a diverse set of osmoregulatory mechanisms present in *C. cryptica* and set the stage for further examination of these patterns through studies of (i) fine-scale temporal variation in the expression of key synthetic pathway and transport genes, both at the RNA and protein level, (ii) metabolomic studies to validate patterns inferred from RNA-seq data and (iii) genetic and biochemical validation of gene functions and osmolyte concentrations.

3.7 Acknowledgements

We thank Yufei Li and Kameila Nedd for help during the acclimation trial. This work was supported by the NSF (grant no. DEB-1651087, AJA) and by a grant from the Simons Foundation (403249, AJA).

3.8 References

Alleva, K., Niemietz, C. M., Sutka, M., Maurel, C., Parisi, M., Tyer- man, S. D. & Amodeo, G. 2006. Plasma membrane of Beta vulgaris storage root shows high water channel activity regulated by cytoplasmic pH and a dual range of calcium concentrations. J. Exp. Bot. 57:609–21.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–10.

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.

Anderson, C. M. H., Howard, A., Walters, J. R. F., Ganapathy, V. & Thwaites, D. T. 2009. Taurine uptake across the human intestinal brush-border membrane is via two transporters: H+-coupled PAT1 (SLC36A1) and Na+- and Cl(-)-dependent TauT (SLC6A6). J. Physiol. 587:731–44.

Apse, M. P., Aharon, G. S., Snedden, W. A. & Blumwald, E. 1999. Salt tolerance conferred by overexpression of a vacuolar Na+/H+ antiport in Arabidopsis. Science 285:1256–8.

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S. & Ogata, H. 2019. KofamKOALA: KEGG ortho- log assignment based on profile HMM and adaptive score threshold. Bioinformatics 36:2251–2.

Bergeron, M. J., Clemencon, B., Hediger, M. A. & Markovich, D. 2013. SLC13 family of Na^+-coupled di- and tri-carboxy- late/sulfate transporters. Mol. Aspects Med. 34:299–312.

Bolger, A. M., Lohse, M. & Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–20.

Brand, L. E. 1984. The salinity tolerance of forty-six marine phy- toplankton isolates. Deep Sea Res. 18:543–56.

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Eigh, N. D., Kuo, T. H., Davis, F. G. & Bateman, J. 2017. A tissue-mapped Axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Rep. 18:762–76.

Buchfink, B., Xie, C. & Huson, D. H. 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12:59–60. Burg, M. B. & Ferraris, J. D. 2008. Intracellular organic osmolytes: function and regulation. J. Biol. Chem. 283:7309–13.

Bussard, A., Corre, E., Hubas, C., Duvernois-Berthet, E., Le Corguille, G., Jourdren, L., Coulpier, F., Claquin, P. & Lopez, P. J. 2017. Physiological adjustments and transcriptome reprogramming are involved in the acclimation to salinity gradients in diatoms. Environ. Microbiol. 19:909–25.

Chan, H., Babayan, V., Blyumin, E., Gandhi, C., Hak, K., Harake, D., Kumar, K., Lee, P., Li, T. T., Liu, H. Y. & Lo, T. C. T. 2010. The p-type ATPase superfamily. J. Mol. Microbiol. Biotechnol. 19:5–104.

Chardwiriyapreecha, S., Mukaiyama, H., Sekito, T., Iwaki, T., Takegawa, K. & Kakinuma, Y. 2010. Avt5p is required for vacuolar uptake of amino acids in the fission yeast Schizosaccharomyces pombe. FEBS Lett. 584:2339–45.

Chen, B., Wan, C., Mehmood, M. A., Chang, J. S., Bai, F. & Zhao, X. 2017. Manipulating environmental stresses and stress tolerance of microalgae for enhanced production of lipids and value-added products–a review. Bioresour. Technol. 244:1198– 206.

Chen, C. & Beattie, G. A. 2008. Pseudomonas syringae BetT is a low- affinity choline transporter that is responsible for superior osmoprotection by choline over glycine betaine. J. Bacteriol. 190:2717–25.

Cheng, R. L., Feng, J., Zhang, B. X., Huang, Y., Cheng, J. & Zhang, C. X. 2014. Transcriptome and gene expression analysis of an oleaginous diatom under different salinity conditions. Bioenergy Res. 7:192–205.

Coupel, P., Ruiz-Pino, D., Sicre, M. A., Chen, J. F., Lee, S. H., Schiffrine, N., Li, H. L. & Gascard, J. C. 2015. The impact of freshening on phytoplankton production in the Pacific Arctic Ocean. Prog. Oceanogr. 131:113–25.

Dickson, D. M. J. & Kirst, G. O. 1987. Osmotic adjustment in marine eukaryotic algae: The role of inorganic ions, quaternary ammonium, tertiary sulphonium and carbohydrate solutes. I. Diatoms and a rhodophyte. New Phytol. 106:645–55.

d'Ippolito, G., Sardo, A., Paris, D., Vella, F. M., Adelfi, M. G., Botte, P., Gallo, C. & Fontana, A. 2015. Potential of lipid metabolism in marine diatoms for biofuel production. Biotechnol. Biofuels. 8:28.

Dutta, D. & Fliegel, L. 2018. Structure and function of yeast and fungal Na+ /H+ antiporters. IUBMB Life 70:23–31.

Edwards, K. F., Thomas, M. K., Klausmeier, C. A. & Litchman, E. 2015. Light and growth in marine phytoplankton: allometric, taxonomic, and environmental variation. Limnol. Oceanogr. 60:540–52.

Emms, D. M. & Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157.

Emms, D. M. & Kelly, S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20:238. Epstein, W. 2003. The roles and regulation of potassium in bacteria. Prog. Nucleic Acid Res. Mol. Biol. 75:293–320.

Fei, Y. J., Inoue, K. & Ganapathy, V. 2003. Structural and functional characteristics of two sodium-coupled dicarboxylate transporters (ceNaDC1 and ceNaDC2) from Caenorhabditis elegans and their relevance to life span. J. Biol. Chem. 278:6136–44.

Feng, H., Edwards, N., Anderson, C. M. H., Althaus, M., Duncan, R. P., Hsu, Y. C., Luetje, C. W., Price, D. R., Wilson, A. C. & Thwaites, D. T. 2019. Trading amino acids at the aphid-Buchnera symbiotic interface. Proc. Natl. Acad. Sci. USA 116:16003– 11.

Finn, R. D., Clements, J. & Eddy, S. R. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39:W29–37.

Forster, I. C., Loo, D. D. F. & Eskandari, S. 1999. Stoichiometry and Na+ binding cooperativity of rat and flounder renal type II Na+-Picotransporters. Am. J. Phys. 276:F644–9.

Fujii, S., Nishimoto, N., Notoya, A. & Hellebust, J. A. 1995. Growth and osmoregulation of Chaetoceros muelleri in relation to salinity. Plant Cell Physiol. 36:759–64.

Gage, D. A., Rhodes, D., Nolte, K. D., Hicks, W. A., Leustek, T., Cooper, A. J. & Hanson, A. D. 1997. A new route for synthesis of dimethylsulphoniopropionate in marine algae. Nature 387:891–4.

Goberdhan, D. C. I., Meredith, D., Boyd, C. A. R. & Wilson, C. 2005. PAT-related amino acid transporters regulate growth via a novel mechanism that does not require bulk transport of amino acids. Development 132:2365–75.

Goyer, A., Johnson, T. L., Olsen, L. J., Collakova, E., Shachar-Hill, Y., Rhodes, D. & Hanson, A. D. 2004. Characterization and metabolic function of a peroxisomal sarcosine and pipecolate oxidase from Arabidopsis. J. Biol. Chem. 279:16947–53.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. & Chen, Z. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–52.

Gries, C. M., Bose, J. L., Nuxoll, A. S., Fey, P. D. & Bayles, K. W. 2013. The Ktr potassium transport system in Staphylococcus aureus and its role in cell physiology, antimicrobial resistance and pathogenesis. Mol. Microbiol. 89:760–73.

Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V. & Mock, T. 2015. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. Plant J. 81:519–28.

Hall, J. A. & Pajor, A. M. 2005. Functional characterization of a Na(+)-coupled dicarboxylate carrier protein from Staphylococcus aureus. J. Bacteriol. 187:5189–94.

Harrison, P. F., Pattison, A. D., Powell, D. R. & Beilharz, T. H. 2019. Topconfects: a package for confident effect sizes in differential expression analysis provides a more biologically useful ranked gene list. Genome Biol. 20:67.

Hartmann, M., Zeier, T., Bernsdorff, F., Reichel-Deland, V., Kim, D.,Hohmann,M.,Scholten,N.,Schuck,S.,Br€autigam,A., Ho€lzel, T. & Ganter, C. 2018. Flavin monooxygenase-generated N-hydroxypipecolic acid is a critical element of plant systemic immunity. Cell 173:456–69.e16.

Hellebust, J. A. 1985. Mechanisms of response to salinity in halo- tolerant microalgae. In Pasternak, D. & San Pietro, A. [Eds.] Biosalinity in Action: Bioproduction with Saline Water. Springer, Dordrecht, pp. 69–81.

Iwaki, T., Higashida, Y., Tsuji, H., Tamai, Y. & Watanabe, Y. 1998. Characterization of a second gene (ZSOD22) of Na+/H+ antiporter from salt-tolerant yeast Zygosaccharomyces rouxii

and functional expression of ZSOD2 and ZSOD22 in Saccha- romyces cerevisiae. Yeast 14:1167–74.

Jackson, A. E., Ayer, S. W. & Laycock, M. V. 1992. The effect of salinity on growth and amino acid composition in the mar- ine diatom Nitzschia pungens. Can. J. Bot. 70:2198–201.

Kageyama, H., Tanaka, Y., Shibata, A., Waditee-Sirisattha, R. & Takabe, T. 2018a. Dimethylsulfoniopropionate biosynthesis in a diatom Thalassiosira pseudonana: Identification of a gene encoding MTHB-methyltransferase. Arch. Biochem. Biophys. 645:100–6.

Kageyama, H., Tanaka, Y. & Takabe, T. 2018b. Biosynthetic path- ways of glycinebetaine in Thalassiosira pseudonana; functional characterization of enzyme catalyzing three-step methylation of glycine. Plant Physiol. Biochem. 127:248–55.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. 2016a. KEGG as a reference resource for gene and pro- tein annotation. Nucleic Acids Res. 44:D457–62.

Kanehisa, M., Sato, Y. & Morishima, K. 2016b. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428:726–31.

Keates, R. A. B., Culham, D. E., Vernikovska, Y. I., Zuiani, A. J., Boggs, J. M. & Wood, J. M. 2010. Transmembrane helix I and periplasmic loop 1 of Escherichia coli ProP are involved in osmosensing and osmoprotectant transport. Biochemistry 49:8847–56.

Kettles, N. L., Kopriva, S. & Malin, G. 2014. Insights into the regulation of DMSP synthesis in the diatom Thalassiosira pseudo- nana through APR activity, proteomics and gene expression analyses on cells acclimating to changes in salinity, light and nitrogen. PLoS ONE 9:e94795.

Kirst, G. O. 1996. Osmotic adjustment in phytoplankton and macro algae. In Kiene, R. P., Visscher, P. T., Keller, M. D. & Kirst, G. O. [Eds.] Biological and Environmental Chemistry of DMSP and Related Sulfonium Compounds. Springer US, Boston, MA, pp. 121–9.

Krell, A., Beszteri, B., Dieckmann, G., Glo€ckner, G., Valentin, K. & Mock, T. 2008. A new class of ice-binding proteins discovered in a salt-stress-induced cDNA library of the psychrophilic diatom Fragilariopsis cylindrus (Bacillariophyceae). Eur. J. Phycol. 43:423–33.

Krell, A., Funck, D., Plettner, I., John, U. & Dieckmann, G. 2007. Regulation of proline metabolism under salt stress in the psychrophilic diatom Fragilariopsis cylindrus (Bacillariophyceae). J. Phycol. 43:753–62.

Kristensen, M. & Juel, C. 2010. Potassium-transporting proteins in skeletal muscle: cellular location and fibre-type differences. Acta Physiol. 198:105–23.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305:567–80.

Kruse, E., Uehlein, N. & Kaldenhoff, R. 2006. The aquaporins. Genome Biol. 7:206.

Lavoie, M., Waller, J. C., Kiene, R. P. & Levasseur, M. 2018. Polar marine diatoms likely take up a small fraction of dissolved dimethylsulfoniopropionate relative to bacteria in oligotrophic environments. Aquat. Microb. Ecol. 81:213–8.

Lebens, M., Lundquist, P., So€derlund, L., Todorovic, M. & Carlin, N. I. A. 2002. The nptA gene of Vibrio cholerae encodes a functional sodium-dependent phosphate cotransporter homologous to the type II cotransporters of eukaryotes. J. Bacteriol. 184:4466–74.

Li, W. K. W., McLaughlin, F. A., Lovejoy, C. & Carmack, E. C. 2009. Smallest algae thrive as the Arctic Ocean freshens. Science 326:539.

Litchman, E., Klausmeier, C. A. & Yoshiyama, K. 2009. Contrasting size evolution in marine and freshwater diatoms. Proc. Natl. Acad. Sci. USA 106:2665–70.

Liu, M. S. & Hellebust, J. A. 1974. Uptake of amino acids by the marine centric diatom Cyclotella cryptica. Can. J. Microbiol. 20:1109–18.

Liu, M. S. & Hellebust, J. A. 1976a. Effects of salinity and osmolarity of the medium on amino acid metabolism in Cyclotella cryptica. Can. J. Bot. 54:938–48.

Liu, M. S. & Hellebust, J. A. 1976b. Regulation of proline metabolism in the marine centric diatom Cyclotella cryptica. Can. J. Bot. 54:949–59.

Love, M. I., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.

Love, M. I., Soneson, C. & Robinson, M. D. 2019. Importing transcript abundance datasets with tximport. Available at: https:// bioconductor.org/packages/devel/bioc/vignettes/tximport/inst/doc/tximport.html (last accessed August 16, 2019).

Lyon, B. R., Bennett-Mintz, J. M., Lee, P. A., Janech, M. G. & DiTullio, G. R. 2016. Role of dimethylsulfoniopropionate as an osmoprotectant following gradual salinity shifts in the sea- ice diatom Fragilariopsis cylindrus. Environ. Chem. 13:181–94.

Lyon, B. R., Lee, P. A., Bennett, J. M., DiTullio, G. R. & Janech, M. G. 2011. Proteomic analysis of a sea-ice diatom: salinity acclimation provides new insight into the dimethylsulfoniopropionate production pathway. Plant Physiol. 157:1926–41.

Mann, D. G. 1999. Crossing the Rubicon: the effectiveness of the marine/freshwater interface as a barrier to the migration of diatom germplasm. In Mayama, S., Idei, M., & Koizumi, I. [Eds.] Proceedings of the 14th International Diatom Symposium, Koenigstein, 1999. Koeningstein: Koeltz Scientific Books, pp. 1–21.

Matsui, H., Hopkinson, B. M., Nakajima, K. & Matsuda, Y. 2018. Plasma membrane-type aquaporins from marine diatoms function as CO2/NH3 channels and provide photoprotection. Plant Physiol. 178:345–57.

Mazzarelli, C. C. M., Santos, M. R., Amorim, R. V. & Augusto, A. 2015. Effect of salinity on the metabolism and osmoregulation of selected ontogenetic stages of an Amazon

population of Macrobrachium amazonicum shrimp (Decapoda, Palae- monidae). Braz. J. Biol. 75:372–9.

McCarthy, D. J. & Smyth, G. K. 2009. Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics 25:765–71.

Nakov, T., Beaulieu, J. M. & Alverson, A. J. 2019. Diatoms diversify and turn over faster in freshwater than marine environments. Evolution 73:2497–511.

Nakov, T., Theriot, E. & Alverson, A. 2014. Using phylogeny to model cell size evolution in marine and freshwater diatoms. Limnol. Oceanogr. 59:79–86.

Parks, M. B., Wickett, N. J. & Alverson, A. J. 2017. Signal, uncertainty, and conflict in phylogenomic data for a diverse lineage of microbial eukaryotes (Diatoms, Bacillariophyta). Mol. Biol. Evol. 35:80–93.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14:417–9.

Peter, H., Weil, B., Burkovski, A., Kr€amer, R. & Morbach, S. 1998. Corynebacterium glutamicum is equipped with four secondary carriers for compatible solutes: identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier. EctP. J. Bacteriol. 180:6005–12.

Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8:785–6.

Petrou, K. & Nielsen, D. A. 2018. Uptake of dimethylsulphoniopropionate (DMSP) by the diatom Thalassiosira weissflogii: A model to investigate the cellular function of DMSP. Biogeochemistry 141:265–71.

Petrova, D. P., Khabudaev, K. V., Marchenkov, A. M., Galachyants, Y. P., Kalyuzhnaya, O. V., Zakharova, Y. R., Likhoshvai, E. V. & Grachev, M. A. 2013. Aquaporin-like protein of the diatom Synedra acus. Dokl. Biochem. Biophys. 448:5–8.

Potapova, M. 2011. Patterns of diatom distribution in relation to salinity. In Seckbach, J. & Kociolek, P. [Eds.] The Diatom World. Springer Netherlands, Dordrecht, pp. 313–32.

Reimann, B. E. F., Lewin, J. M. C. & Guillard, R. R. L. 1963. Cyclotella cryptica, a new brackish-water diatom species. Phycologia 3:75–84.

Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. 32:896–902.

Ruck, E. C., Nakov, T., Alverson, A. J. & Theriot, E. C. 2016. Phylogeny, ecology, morphological evolution, and reclassification of the diatom orders Surirellales and Rhopalodiales. Mol. Phylogenet. Evol. 103:155–71.

Sandmann, M., Skłodowski, K., Gajdanowicz, P., Michard, E., Rocha, M., Gomez-Porras, J. L., Gonzalez, W., Corr^ea, L. G. G., Ramırez-Aguilar, S. J., Cuin, T. A. & van Dongen, J. T.

2011. The K (+) battery-regulating Arabidopsis K (+) channel AKT2 is under the control of multiple post-translational steps. Plant Signal. Behav. 6:558–62.

Sayanova, O., Mimouni, V., Ulmann, L., Morant-Manceau, A., Pas- quet, V., Schoefs, B. & Napier, J. A. 2017. Modulation of lipid biosynthesis by stress in diatoms. Philos. Trans. R. Soc. Lond. B Biol. Sci. 372:20160407.

Song, L. & Florea, L. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. Gigascience 4:48.

Soveral, G., Madeira, A., Loureiro-Dias, M. C. & Moura, T. F. 2008. Membrane tension regulates water transport in yeast. Biochim. Biophys. Acta. 1778:2573–9.

Soveral, G., Prista, C., Moura, T. F. & Loureiro-Dias, M. C. 2010. Yeast water channels: an overview of orthodox aquaporins. Biol. Cell. 103:35–54.Stephens, M. 2017. False discovery rates: a new deal. Biostatistics 18:275–94.

Sunda, W., Kieber, D. J., Kiene, R. P. & Huntsman, S. 2002. An antioxidant function for DMSP and DMS in marine algae. Nature 418:317–20.

Swart, N. C., Gille, S. T., Fyfe, J. C. & Gillett, N. P. 2018. Recent Southern Ocean warming and freshening driven by greenhouse gas emissions and ozone depletion. Nat. Geosci. 11:836–41.

Tanaka, T., Maeda, Y., Veluchamy, A., Tanaka, M., Abida, H., Marechal, E., Bowler, C., Muto, M., Sunaga, Y., Tanaka, M. & Yoshino, T. 2015. Oil accumulation by the oleaginous diatom Fistulifera solaris as revealed by the genome and transcriptome. Plant Cell 27:162–76.

Thwaites, D. T. & Anderson, C. M. H. 2011. The SLC36 family of proton-coupled amino acid transporters and their potential role in drug transport. Br. J. Pharmacol. 164:1802–16.

Torstensson, A., Young, J. N., Carlson, L. T., Ingalls, A. E. & Dem- ing, J. W. 2019. Use of exogenous glycine betaine and its precursor choline as osmoprotectants in Antarctic sea-ice diatoms. J. Phycol. 55:663–75.

Tournaire-Roux, C., Sutka, M., Javot, H., Gout, E., Gerbeau, P., Luu, D. T., Bligny, R. & Maurel, C. 2003. Cytosolic pH regulates root water transport during anoxic stress through gating of aquaporins. Nature 425:393–7.

Traller, J. C., Cokus, S. J., Lopez, D. A., Gaidarenko, O., Smith, S. R., McCrow, J. P., Gallaher, S. D., Podell, S., Thompson, M., Cook, O. & Morselli, M. 2016. Genome and methylome of the oleaginous diatom Cyclotella cryptica reveal genetic flexibility toward a high lipid phenotype. Biotechnol. Biofuels. 9:258.

Wadhams, P. & Munk, W. 2004. Ocean freshening, sea level rising, sea ice melting. Geophys. Res. Lett. 31:L11311. https://d oi.org/10.1029/2004GL020039

Wood, A. M., Everroad, R. C. & Wingard, L. M. 2005. Measuring growth rates in microalgal cultures. Algal Culturing Techniques 18:269–88.

Yancey, P. H. 2005. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. J. Exp. Biol. 208:2819–30.

Ye, Q., Wiera, B. & Steudle, E. 2004. A cohesion/tension mecha- nism explains the gating of water channels (aquaporins) in Chara internodes by high concentration. J. Exp. Bot. 55:449–61.

Chapter 4 Strain-specific transcriptional responses overshadow salinity effects in a marine diatom sampled along the Baltic Sea salinity cline

Eveline Pinseel[1], Teofil Nakov[1], Koen Van den Berge[2], Kala M. Downey[1], Kathryn J. Judy[1], Olga Kourtchenko[4], Anke Kremp[5], Elizabeth C. Ruck[1], Conny Sjöqvist[5], Mats Töpel[4], Anna Godhe[4] & Andrew J. Alverson[1*]

[1] Department of Biological Sciences, University of Arkansas, Fayetteville, AR, USA

[2] Department of Statistics, University of California, Berkeley, CA, USA

[3] Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden

[4] Leibniz-Institute for Baltic Sea Research, Rostock, Germany

[5] Faculty of Science and Engineering, Åbo Akademi University, Turku, Finland

Author Contributions: Olga Kourtchenko, Anke Kremp, Conny Sjöqvist, Mats Töpel, and Anna Godhe collected sediment samples. Teofil Nakov, Kala Downey, and Elizabeth Ruck designed the study. Teofil and Elizabeth isolated the samples. Teofil, Kala, and Kathryn Judy collected the data. Teofil and Kala designed machine learning program. Eveline Pinseel, Koen Van den Berge, and Kala analysed the data.

This chapter has been published:

Pinseel E, Nakov T, Van den Berge K, Downey KM, Judy KJ, Kourtchenk O, Kremp A, Ruck EC, Sjöqvist C, Töpel M, Godhe A, Alverson AJ. Strain-specific transcriptional responses overshadow salinity effects in a marine diatom sampled along the Baltic Sea salinity cline. The ISME Journal. https://doi.org/10.1038/s41396-022-01230-x

4.1 Abstract

The salinity gradient separating marine and freshwater environments represents a major ecological divide for microbiota, yet the mechanisms by which marine microbes have adapted to and ultimately diversified in freshwater environments are poorly understood. Here, we take advantage of a natural evolutionary experiment: the colonization of the brackish Baltic Sea by the ancestrally marine diatom *Skeletonema marinoi*. To understand how diatoms respond to low salinity, we characterized transcriptomic responses of acclimated *S. marinoi* grown in a common garden. Our experiment included eight strains from source populations spanning the Baltic Sea salinity cline. Gene expression analysis revealed that low salinities induced changes in the cellular metabolism of *S. marinoi*, including upregulation of photosynthesis and storage compound biosynthesis, increased nutrient demand, and a complex response to oxidative stress. However, the strain effect overshadowed the salinity effect, as strains differed significantly in their response, both regarding the strength and the strategy (direction of gene expression) of their response. The high degree of intraspecific variation in gene expression observed here highlights an important but often overlooked source of biological variation associated with how diatoms respond to environmental change.

4.2 Introduction

The salinity gradient separating marine and freshwater environments represents one of the major ecological divides structuring microbial diversity [1]. Differences in osmotic pressure impede marine–freshwater transitions, and as a consequence, transitions are generally rare, occur on longer evolutionary timescales [2, 3], and have led to repeated bursts of diversification in freshwater environments [4]. Identifying the processes underlying marine–freshwater habitat transitions is fundamental to our understanding of lineage diversification and habitat structuring on evolutionary timescales [5], as well as short-term adaptive potential to climate change as melting ice caps, altered precipitation patterns, and changes in oceanic currents have led to freshening of large regions and local changes in the seasonal or annual cycling of salinity regimes [6, 7]. Permanent establishment of ancestrally marine organisms in freshwaters depends on the ability of individual colonists to survive the initial hypoosmotic stress, acclimate to low salinity, and ultimately adapt to their new environment [8]. Consequently, these transitions should happen gradually [4, 5], and euryhaline or brackish species that can tolerate a wide range of salinities are probably more likely to successfully cross the salinity divide. Studies focused on these taxa can provide key insights into the cellular processes that help mediating marine–freshwater transitions.

Here, we take advantage of a natural evolutionary experiment: the colonization of one of the world's largest brackish water bodies, the Baltic Sea, by the ancestrally marine diatom *Skeletonema marinoi* (Fig. 1A). Geologically, the Baltic Sea is young, with sea ice from the last glacial maximum having fully receded only ~10 000 years ago and inundation of saline waters from the adjacent North Sea occurring ~8 000 years ago [9]. Today, freshwater input from rivers and precipitation, combined with inflow of saline bottom-waters from the North Sea through the Danish straits, results in a latitudinal and vertical salinity gradient ranging from near fresh to fully marine conditions [9, 10] (Fig. 1A). The Baltic salinity gradient strongly structures aquatic biodiversity at the species and population levels [11–13], including *S. marinoi* [14–16], which is

the dominant phytoplankton species and one of the main primary producers in the area [17, 18]. Fossil evidence showed that *S. marinoi* has been present in the Baltic Sea since the marine inundation or shortly thereafter [19, 20]. Although *S. marinoi* is ancestrally marine [21], it can tolerate a wide range of salinities and is common along the entire salinity gradient, from the North Sea coast to the upper reaches of the Baltic Sea [14]. Previous work showed reduced gene flow between a high-salinity North Sea population and a low-salinity Baltic Sea population, which exhibited lower genetic diversity and optimal growth at lower salinity, consistent with local adaptation [14]. Thus, *S. marinoi* presents an excellent system for understanding how marine diatoms adapt to low salinity environments.

We combined a laboratory common garden experiment with RNA-sequencing (RNA-seq) to characterize the response of *S. marinoi* to low salinity (Fig. 1). We collected eight strains along the Baltic Sea salinity cline, acclimated them to a range of salinities, and compared gene expression between high and low salinity treatments. Natural populations of *S. marinoi* exhibit a broad range of variability in several ecophysiological traits [22–24]. The inclusion of multiple strains in our experiment allowed us to characterize variation in the salinity response as well, including which aspects of the response are shared or different among strains.

4.3 Material & Methods

4.3.1 Sample collection, experimental design, and RNA processing

We collected sediment samples from eight locations across the Baltic Sea (Fig. 1A) and stored them in the dark at 5 °C. We germinated *S. marinoi* resting cells into monoclonal cultures [25] that were kept at their native salinity (Table 1) for 12–26 months prior to the experiment. Strain identity was confirmed by sequencing the LSU rDNA gene (D1–D2 region). Before our experiment, one strain per location was acclimated to the experimental salinities for one week.

In our experiment, we grew the acclimated strains in triplicate at three salinities (8, 16 and 24), a design that included both biological (eight strains) and technical replication (three

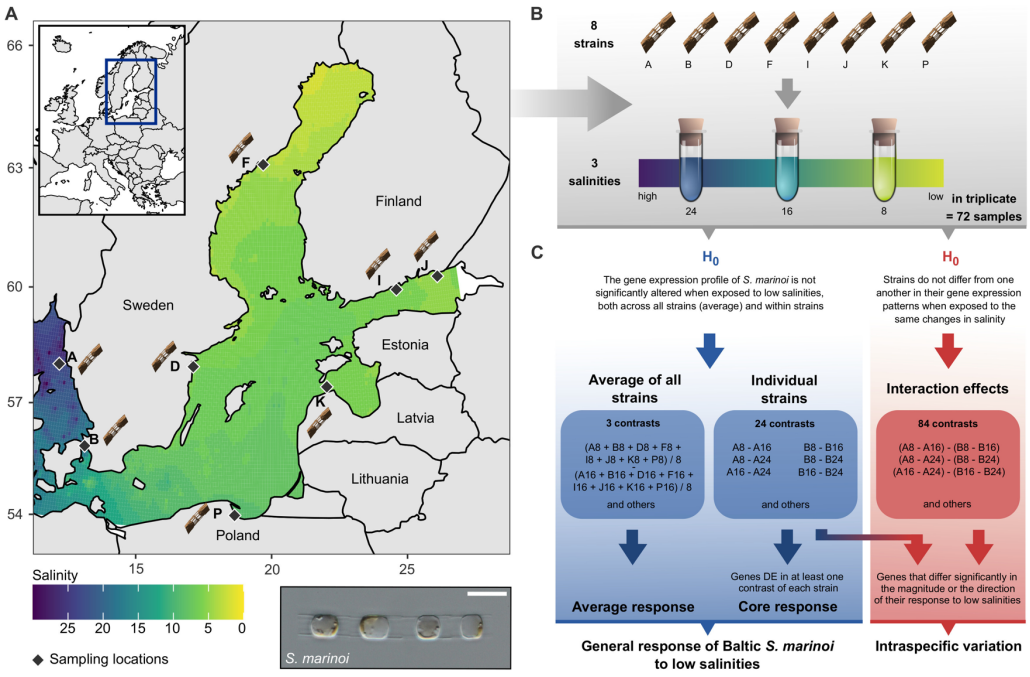


Fig. 4.1. Experimental design. a Field sampling. Natural salinity gradient in the Baltic Sea based on salinity measurements from surface samples (0–10 m depth) and interpolated across the Baltic Sea for the period 1990-2020. Salinity measurements were downloaded from ICES (ICES Dataset on Ocean Hydrography, 2020. ICES, Copenhagen) and Sharkweb (https://sharkweb.smhi.se/hamta-data/). Diamonds identify sampling locations for *S. marinoi*. The inset figure on the top left shows the general geographic area in which the Baltic Sea is located. The bottom right figure shows a light micrograph of a *S. marinoi* culture (scale bar = 10 µm). **b Laboratory experiment.** Experimental design of the laboratory experiment carried out in this study. Eight strains of *S. marinoi* were exposed to three salinity treatments (8, 16 and 24) in triplicate, resulting in 72 RNA-seq libraries. **c Statistical analyses.** Overview of the null hypotheses and contrasts tested in this study. Our experimental design allowed characterization of the general response of acclimated *S. marinoi* to low salinities as well as intraspecific variation. The lower blue arrows indicate which data were incorporated in the average and core responses, which together were used to define the general response of *S. marinoi*. Genes with significant interaction effects were subdivided in two categories using logFC values of the individual strains (blue-red gradient arrow), distinguishing genes that differed significantly in either the magnitude or direction of their response to low salinities. The first category includes genes that were DE in one strain but not the others, or that were DE in multiple strains but with significant differences in logFC values in the same direction. Genes of the second category were significantly upregulated in some strains, whereas they were significantly downregulated in other strains.

replicates/strain) (Fig. 1B). During the experiment, strains were regularly reinoculated to maintain exponential growth, growth rates were monitored via chlorophyll *a* fluorescence, and starting from day 11 cells were harvested for RNA-sequencing. For each strain, two harvests were pooled to obtain sufficient RNA for sequencing. We mapped quality-controlled and

trimmed RNA-seq reads against the reference genome of *S. marinoi* strain RO5AC v.1.1 with STAR [26], followed by gene-level read quantification with HTSeq [27]. We obtained functional annotations for all genes with InterProScan, KofamKOALA, and BLAST+ searches against Swissprot/Uniprot [28–30]. We detected orthologs of *S. marinoi* genes in other diatom genomes with OrthoFinder [31] and predicted protein targeting with MitoProt, HECTAR, SignalP, ASAFind, and TargetP [32–36]. The Supplementary Methods contain full details on the experimental design and the analyses.

Table 4.1. Details of the *S. marinoi* strains used in this study. The column labelled 'Year' indicates the collection year of the sediment samples (see 'Collection ID': letters between brackets refer to sampling localities in Fig. 1) from which *S. marinoi* strains (see 'Culture ID' and 'Strain ID') were germinated. All strains are publicly available from the BCCM/DCG diatom culture collection (https://bccm.belspo.be/about-us/bccm-dcg) under the DCG accession numbers listed in the table. GenBank accession numbers refer to LSU D1–D2 rDNA gene sequences used for strain identification. The salinity values indicate the salinity of the natural sample from which the respective strains were isolated ('Original salinity') and in which they were maintained prior to the experiment ('Culture salinity'). GPS coordinates indicated with an asterisk (*) represent approximate sampling locations. NA = not available.

Collection ID	Culture ID	Strain ID	DCG accession	GenBank	Country	GPS (N/E)	Collector	Year	Isolation date	Culture medium	Original salinity	Culture salinity
AJA304 (A)	AJA304-05	A.2.21b	DCG 1232	OM112317	Sweden	58.02868/11.13738 (*)	A.Godhe	2014	2017-03-28	L1	15-33	24
AJA305 (B)	AJA305-18	B.2.19b	DCG 1236	OM112318	Sweden	55.97744/12.69058	A.Godhe	2010	2017-04-07	ASW	12-15	16
AJA332 (D)	AJA332-09	D.1.27a	DCG 1238	OM112319	Sweden	58.33200/16.70583	A.Godhe, B.Andersson	2017	2018-05-14	WC + salt	8-9	8
AJA328 (F)	AJA328-03	F.1.2a	DCG 1237	OM112320	Sweden	63.65317/18.95200 (*)	A.Godhe	NA	2018-03-15	WC + salt	~8	8
AJA311 (I)	AJA311-04	I.3.11a	DCG 1233	OM112321	Finland	60.18000/25.50700	A.Kremp	2015	2017-03-09	WC + salt	5-6	5
AJA313 (J)	AJA313-31	J.3.42b	DCG 1235	OM112322	Finland	60.38964/27.37518 (*)	A.Kremp	2016	2018-04-22	WC + salt	4-5	8
AJA318 (K)	AJA318-23	K.3.3a	DCG 1234	OM112323	Estonia	57.81670/22.28330	S.Sildever	2016	2018-05-23	WC + salt	~8	8
AJA333 (P)	AJA333-06	P.2.6a	DCG 1239	OM112324	Poland	54.44778/18.57611	A.Witkowski	2018	2018-05-16	WC + salt	5-7	5

4.3.2 Hypothesis testing and GO enrichment

We tested two sets of null hypotheses, using edgeR [37] and stageR [38] (Fig. 1C). The first set tested whether gene expression was different across the salinity gradient for each strain separately and for all strains together, using 27 contrasts (Fig. 1C). Compared to solely testing the average salinity effect, simultaneously accounting for the individual strains increases the power to find differentially expressed (DE) genes, as the strain effect incorporates variability that would otherwise be unaccounted for. The second set of hypotheses tested for an interaction

effect between strain and salinity, i.e., whether there are strain-specific responses to changes in salinity. Here, we defined 84 contrasts, testing each pairwise combination of strains within all three salinity combinations (Fig. 1C). We tested the two sets of hypotheses separately using stageR's stage-wise testing procedure, thus controlling the gene-level false discovery rate (FDR) within each set at 5 % [38, 39].

Gene ontology (GO) enrichment was done with TopGO (Over-representation Analysis, ORA) [40] and CAMERA (Gene Set Enrichment Analysis, GSEA) [41]. For ORA, we performed separate GO enrichment for genes that are up- or downregulated in low salinities in each strain or average response. For GSEA, we performed GO enrichment on each contrast of the individual strains and average effects. Redundant GO terms were removed with REVIGO [42].

4.4 Results

4.4.1 Response of Baltic *S. marinoi* to low salinity

Growth reaction norms showed no differences among strains and salinities, indicating that all strains grew well across a wide salinity range (Fig. 2). These growth rates are within the range observed in previous work [14]. However, we did not find lower growth optima in strains from low salinity environments compared to those from high salinities, as previously reported for *S. marinoi* [14].

RNA-seq reads of all strains mapped equally well to the reference genome. In the combined average- and strain-specific response, 7 905 of the 22 440 predicted genes in the *S. marinoi* genome were DE using a 5 % FDR (Suppl. Fig. 1). Of the DE genes, 1 652 received no functional annotation. The 5 343 DE genes in the three contrasts of the average response was greater than that of any individual strain (Fig. 3A, Suppl. Figs 1–5), which is the result of combining data across all eight strains (8x3 replicates/salinity). Consequently, the average response allowed us to detect more DE genes, including those with small effect sizes, and shows the benefit of including biological replicates on top of the standard three technical

replicates used in many transcriptome studies. For example, whereas the total number of DE genes within individual strains is comparable with a strain of the diatom *Thalassiosira weissflogii* under changing salinity, it is much larger in the average response [43].

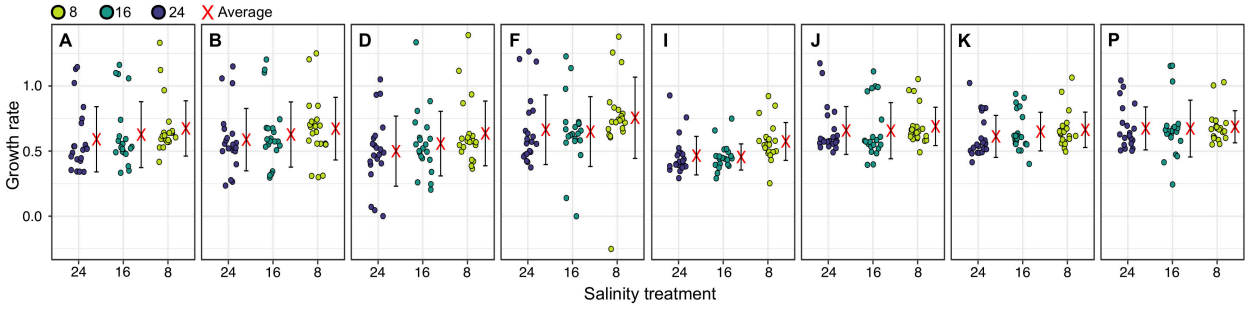


Fig. 4.2. Growth response of Baltic *S. marinoi* in low salinities. Growth rates of the eight *S. marinoi* strains examined in this study at three different salinities. The letters in the individual panels correspond with the sampling locations in Table 1 ('Collection ID') and Fig. 1A. Each point represents a single estimate of the slope of the natural logarithm of in vivo relative fluorescence against time for each sequential transfer, using a horizontal jitter of points to avoid overplotting.

The 8–24 contrasts consistently showed the most DE genes, and the least generally were found in the 16–24 contrast (Fig. 3A, Suppl. Figs 1-3). Thus, the largest drop in salinity (24 to 8), and the shift to lowest salinity (16 to 8), elicited the greatest transcriptomic responses. The number of up- and downregulated genes was comparable within contrasts (Fig. 3A, Suppl. Figs 1–3). However, when only considering the top-100 genes based on P-value or logFC for each contrast, substantially more genes were upregulated in low salinities (Fig. 3B). This indicates that genes with the strongest evidence for DE or the largest effect sizes were more likely to be upregulated in low salinities. Similarly, CAMERA GO enrichment found substantially more enriched GO terms that were upregulated in low salinities (Fig. 3C).

Next, we report on specific genes and pathways that are DE in low salinities. Unless otherwise noted, we focus on the 8–24 contrast of the average response because it represents the strongest response to salinity in our dataset and GO enrichment showed that despite presence of uniquely DE genes in each salinity contrast (Suppl. Figs 2-3), most of the same processes are enriched in the three salinity contrasts of the average response.

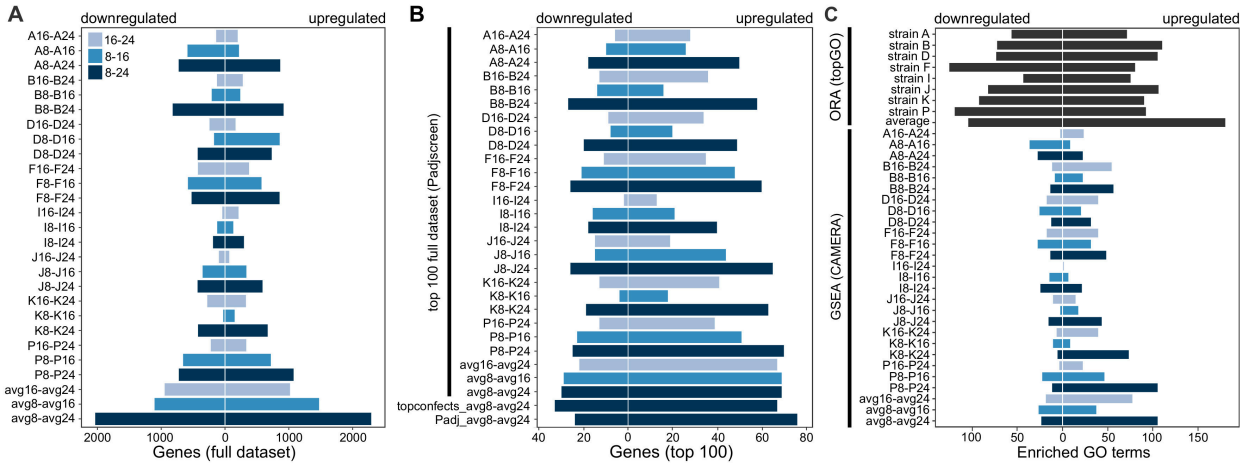


Fig. 4.3. Transcriptome response of Baltic *S. marinoi* to low salinities. a Number of DE genes at a 5 % FDR-level in the average response and the individual strains. The number of DE genes is indicated separately for each contrast, distinguishing between genes that are up- or downregulated. **b** Direction of DE in the top-100 genes of the average response and individual strains as selected by P-value or logFC. For each contrast in the average and individual strains (vertical black bar), the direction of DE is indicated for the top-100 genes selected by stageR's FDR-adjusted P-value of the global null hypothesis (Padjscreen). Thus, although a gene can have a high P-value on a dataset-wide level, it is not necessarily DE in each individual contrast. In addition, we show the top-100 genes selected by logFC (topconfects [100]) and the contrast-specific 5 % FDR-controlled P-value (Padj) for the 8-24 contrast of the average effects, as this contrast showed the greatest number of DE genes in **a**. **c** Number of enriched GO terms for the ORA and GSEA analyses. The number of up- and downregulated GSEA GO terms represents the output classification by CAMERA. The number of enriched GO terms includes Biological Process, Molecular Function and Cellular Component GO terms, prior to removal of redundant GO terms by REVIGO.

Metabolic changes in low salinities. In low salinities, *S. marinoi* experienced significant (i) upregulation of genes involved in photosynthesis, Calvin cycle, chlorophyll biosynthesis and glycolysis/gluconeogenesis, including phosphoenolpyruvate carboxylase (PEPC), and (ii) downregulation of genes involved in protein ubiquitination, proteolysis, and aerobic respiration (Fig. 4, Suppl. Figs 6-9). Most genes involved in the mitochondrial electron transport chain and the TCA cycle, including transcription factor *bZIP14* which regulates the TCA cycle [44], were slightly downregulated (Suppl. Fig. 9A-B). Biosynthesis of fatty acids and the polysaccharide chrysolaminarin (β-1,3/β-1,6-glucan) was upregulated, whereas fatty acid degradation was downregulated (Fig. 4, Suppl. Figs 6, 10A).

Genes involved in tRNA-aminoacylation, translational elongation factors, ribosomal proteins, and protein refolding were upregulated, and many genes associated with cell division were downregulated (Fig. 4, Suppl. Figs 6, 11). Two genes coding for a sulfolipid biosynthesis protein and a glycosyltransferase (involved in thylakoid membranes and membrane stability, respectively) were upregulated, suggestive of changes in membrane composition. Several transcription factors were also upregulated, including a putative heat stress transcription factor involved in DNA binding of heat shock promoter elements. Two genes coding for an extracellular subtilisin-like serine protease were upregulated, as also observed in diatoms in response to copper deficiency [45]. Finally, although activation of transposable elements has been linked to the diatom stress response, including *S. marinoi* [46–48], most genes involved in transposon activity (transposase, retrovirus-related Pol poly-protein) were downregulated or not DE (Fig. 4, Suppl. Fig. 6).

Response to oxidative stress. Multiple mechanisms to deal with reactive oxygen species (ROS) were upregulated in low salinities. This response included genes involved in the xanthophyll cycle, glutathione metabolism, ascorbate peroxidases, catalases, peroxiredoxin, and polyamine biosynthesis from ornithine via ornithine decarboxylase (Suppl. Figs 8B, 9C-D, 10B). Carotenoids for the xanthophyll cycle were likely produced primarily through the non-mevalonate pathway (Suppl. Fig. 8B). Several other genes involved in ROS elimination, such as the gene coding for superoxide dismutase (SOD), were either downregulated or not DE (Suppl. Fig. 10B).

Transmembrane transport and nitrogen metabolism. Transmembrane transporters for amino acids, polyamines, pyruvate, and essential nutrients such as nitrogen, phosphorus, molybdate, and sulfate were upregulated in low salinities (Suppl. Figs 9D, 12). *Nrt* nitrite/nitrate transporters were highly upregulated, and to a lesser extent also transporters for urea and ammonia. Most of

the imported nitrogen is probably directed to the chloroplast, where nitrogen assimilation through ferredoxin-nitrite reductase and GSII-GOGAT$_{(Fd)}$ [49] was upregulated (Suppl. Fig. 9D). In parallel, the anabolic part of the urea cycle was upregulated, including carbamoyl phosphate synthase (Suppl. Figs 9C-D), suggestive of increased recycling of ammonia and biosynthesis of arginino-succinate or arginine. In contrast, silicic acid transporters were downregulated, and this response was most evident in the 16–24 salinity contrast (Suppl. Fig. 12).

Osmotic stress response. Our data suggest that *S. marinoi* responded to differences in osmotic pressure by adjusting intracellular osmolyte concentrations to hypoosmotic conditions. Although the dimethylsulfoniopropionate (DMSP) pathway remains poorly characterized in diatoms, *S. marinoi*'s homolog of *TpMMT*, a methyltransferase that catalyzes a key reaction in DMSP biosynthesis [50], was strongly downregulated (Suppl. Fig. 10B). In addition, breakdown of the osmolyte taurine via taurine dioxygenase was upregulated (Suppl. Fig. 10B). The magnitude of DMSP downregulation in *S. marinoi* was similar to that of another euryhaline diatom, *Cyclotella cryptica*, grown in comparable salinities, whereas the effect sizes for taurine were larger than reported for *C. cryptica* [51]. Putative *BADH* and *CDH* genes involved in the biosynthesis of the osmolyte betaine from choline [52] were not DE. A homolog of the *Thalassiosira pseudonana* gene *TpGSDMT*, involved in the biosynthesis of betaine from glycine [52], was significantly downregulated (Suppl. Fig. 10B). Genes involved in proline metabolism [53] showed inconsistent expression patterns, being up- or downregulated, or not DE (Suppl. Fig. 9D).

Responses to osmotic stress also included shifts in cation import and export, such as sodium and potassium [51, 54]. Here, most potassium and sodium channels were either upregulated or not DE, and two detected aquaporins had opposite expression patterns (Suppl. Fig. 12). Several transporters for potassium or unknown cations/solutes were DE in all strains, often with large effect sizes (Fig. 5).

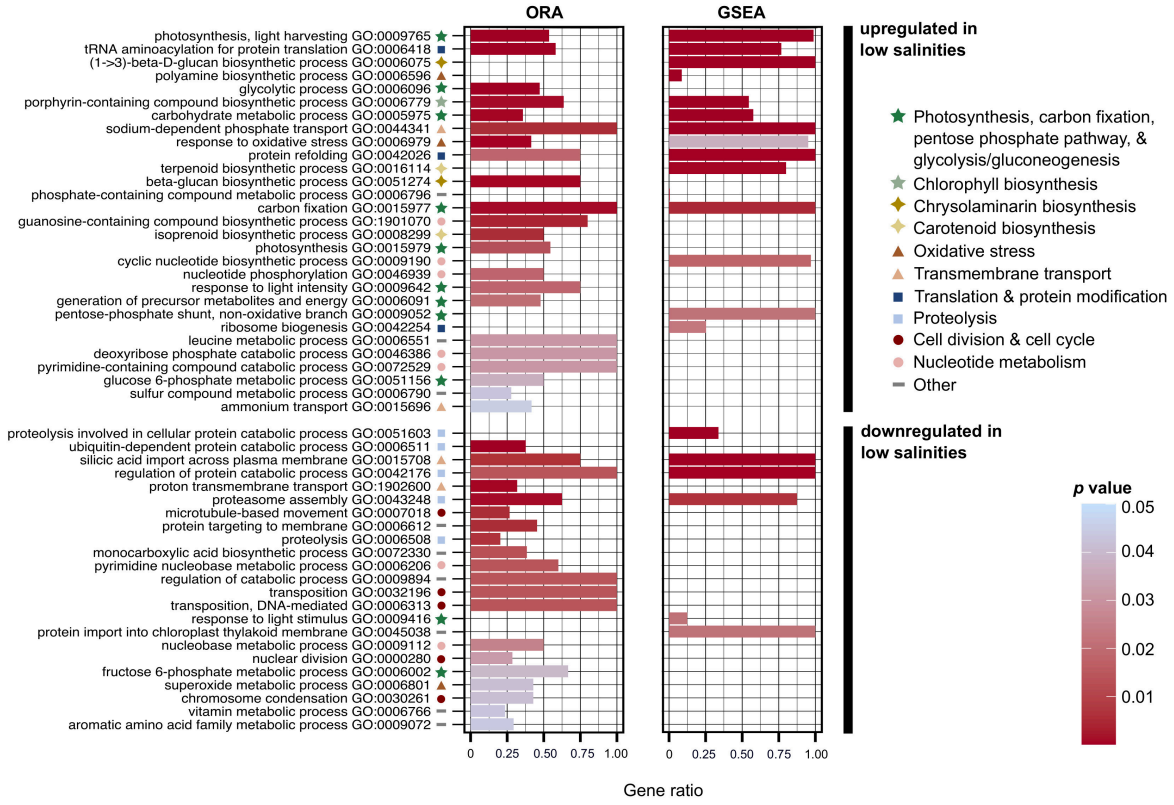


Fig 4.4. GO enrichment on the average response of *S. marinoi* to low salinities: Biological Process. The results of two types of GO enrichment analyses are shown: ORA (in topGO, Fisher's exact test, *elim* algorithm) and GSEA (in CAMERA), after removal of redundant terms by REVIGO. For ORA, we classified the total set of DE genes in the average response into two categories, distinguishing between genes that are up- or downregulated in low salinities, regardless of salinity contrast (see Supplementary Methods for details). For CAMERA, we performed GSEA analyses on each individual contrast separately, showing only the 8-24 contrast in this figure. Barplot height indicates the proportion of genes that are DE with a given GO term to the total number of genes with this GO term in the genome of *S. marinoi*. The barplots are colored according to P-value. Within the set of up- and downregulated genes, the GO terms are ranked from lowest to highest P-value, using the lowest of two P-values from ORA or GSEA. Symbols indicate major categories of cellular processes to which a GO term belongs. Only Biological Process GO terms are shown.

4.4.2 Strain-specific data reveal intraspecific variation and a conserved core response to low salinity

All previous results were based on the average response (Fig. 1C). However, when we take the responses of individual strains into account, it becomes clear that the strain effect in our dataset exceeded the salinity effect. Strains differed substantially in their responses to low

salinities, which was evidenced by a multidimensional scaling plot and poisson-distance heatmap in which samples clustered primarily by strain rather than salinity (Fig. 6). In fact, when combining data of all three salinity contrasts, 1 791 genes were uniquely DE in only one strain (and not DE in the average response), and 1

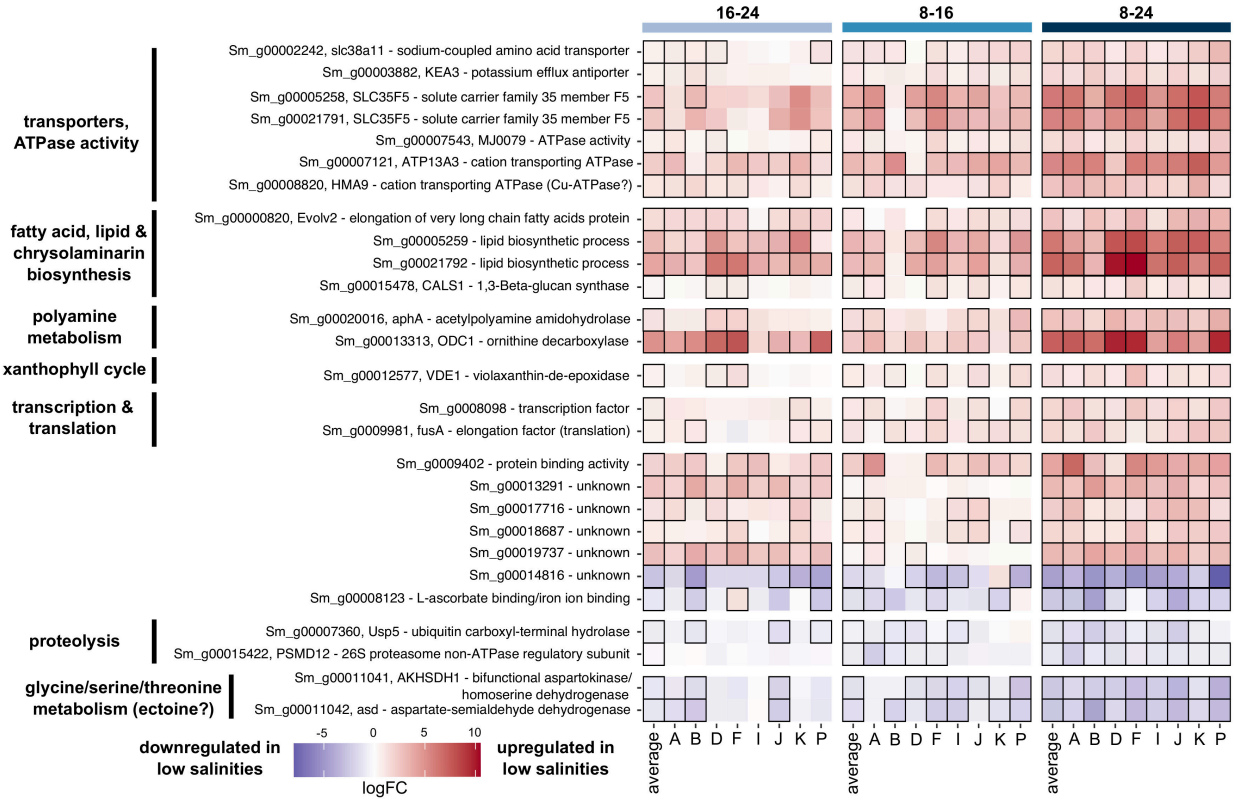


Fig. 4.5. Set of genes that are DE in at least one contrast of each strain: the core response. The heatmap shows logFC values for the individual strains and average response of the 27 core response genes. The three salinity combinations are indicated on top of the figure. Contrasts that were significant are outlined in black. Row names specify gene names and functional annotations based on Swissprot/Uniprot and/or GO terms. When DE, all genes are consistently up- or downregulated in low salinities in each strain, except for gene *Sm_g00008123*. In the 8-16 contrast, genes *Sm_g00007543* and *Sm_g00005259* are not DE for strains I and K, but appear DE in the figure due to colored edge lines from neighboring squares. Similarly for the 16-24 contrast, *Sm_g00005259* is not DE for strains A and F.

628 genes were uniquely DE in one strain and the average response. The number of uniquely DE genes per strain ranged from 103 to 317 genes, and 951 genes were DE only in the average response (Suppl. Fig. 4A). A similar pattern emerged when examining each salinity contrast separately (Suppl. Fig. 5). The high number of genes that are DE only in the average response

or in one strain plus the average response is due to the higher statistical power provided by combining data of all strains together in the average response.

We defined a core response to low salinities by selecting genes that are DE in at least one contrast of each strain, which resulted in a set of 27 shared genes that are DE in each of the eight strains (Figs 1C, 5). Obtaining this set of shared genes required subsetting the full set of DE genes, so the 5 % FDR could not be guaranteed for these 27 genes. However, these core-response genes were characterized by a combination of high logFC and low P-values (Fig. 5, Suppl. Fig. 1), thus providing strong evidence for DE in each strain. These genes are also among the top DE genes: 13 overlapped with the top-25 DE genes as ranked by stageR's FDR-adjusted P-value of the global null hypothesis (Padjscreen), 22 were part of the top 100, and all were detected within the top-225 genes. Core-response genes upregulated in low salinities were involved in key processes previously identified in the average response, including transport of amino acids and cations, biosynthesis of fatty acids, lipids and chrysolaminarin,

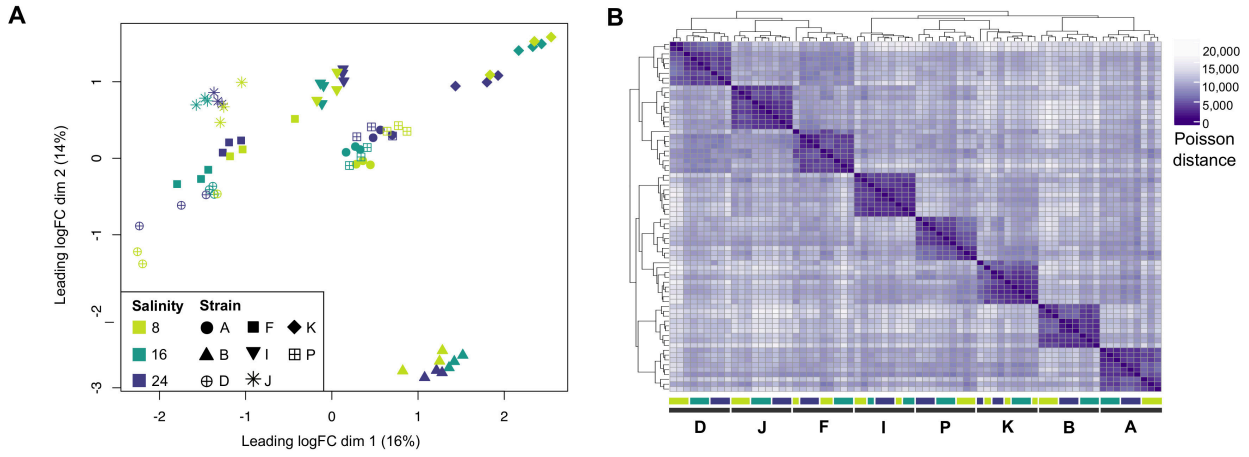


Fig. 4.6. Intraspecific variation in the response of Baltic *S. marinoi* to low salinities. a Multidimensional scaling (MDS) plot, showing that samples cluster primarily by strain rather than salinity. Distances between the samples are based on logFC changes in the top-500 genes, selecting the top-500 genes separately for each pairwise comparison between the samples. **b** Poisson-distance heatmap of the full dataset. Colored bars below the heatmap indicate the position of samples belonging to different strains and salinities (Fig. 1A), showing that samples of different strains cluster together.

polyamine metabolism/biosynthesis, the xanthophyll cycle, and transcription/translation (Fig. 5). By contrast, core-response genes that were downregulated in low salinities were involved in proteolysis or the glycine/threonine/serine pathway (Fig. 5, Suppl. Fig. 13). Seven core-response genes had unknown functions (Fig. 5).

4.4.3 Interaction effects reveal differences among strains in their response to low salinity

A total of 3 857 genes showed significantly different expression patterns between strains with a 5 % FDR (interaction effects, Fig. 1C). Of these, 2 820 differed between strains in the magnitude of their response to low salinities, whereas far fewer (1 037) differed in the direction of their response. However, 92 of the top-100 genes with interaction effects (ranked by stageR's Padjscreen) differed in the direction of their response (Suppl. Fig. 14). Thus, although more genes showed differences in the magnitude of DE, those with differences in direction of DE dominated the top interaction-effect genes.

The two classes of interaction-effect genes were enriched for different processes (Fig. 7, Suppl. Fig. 15). Genes that differed significantly across strains in magnitude were enriched for many of the processes identified in the average response, including photosynthesis, glycolysis, and the biosynthesis of chlorophyll, carotenoids, and fatty acids. By contrast, the gene set that differed significantly between strains in the direction of their response was enriched for transcription regulation, peroxidase activity, aerobic respiration and urea transmembrane transport, and contained genes involved in inositol metabolism, cell wall and calcium-binding messenger proteins, and heat shock proteins/chaperones (Suppl. Figs 14-15). Both classes were enriched for genes involved in translation, cell cycle progression, mitosis, and meiosis. For example, two genes coding for meiotic recombination protein SPO11-2 were part of the top-100 interaction effects (Suppl. Fig. 14). Nevertheless, depending on the strain and the salinity contrast, their effect sizes were 3-50 times smaller than reported during sexual reproduction in *S. marinoi* [55].

4.5 Discussion

4.5.1 The average and core response of *S. marinoi* to low salinities

Taken together, our data show that exposure to low salinities triggers a stronger response compared to high salinities (Figs 3B-C), and suggest that the ancestrally marine diatom *S. marinoi* reprograms its metabolism by upregulating several pathways to function in low salinities. Different numbers of significantly up- and downregulated genes between low and high salinities were also detected in another euryhaline diatom, *T. weissflogii* [43]. Here, analysis of the average and core responses in *S. marinoi* suggested that in low salinities photosynthesis and carbon fixation are upregulated, and there is less protein recycling. This contrasts with carbon fixation in *T. weissflogii* which was not impacted by low salinities [43]. However, like in *T. weissflogii*, we observed upregulation of PEPC in low salinities [43]. PEPC has multiple functions, including supplying oxaloacetate to the TCA cycle, which is, however, slightly downregulated in *S. marinoi*. In some diatoms, PEPC appears to be involved in the carbon concentrating mechanisms (CCMs) of a C4 mechanism similar to that of plants [56, 57].

Our expression data suggest PEPC might play a similar role in *S. marinoi*. Upregulation of this gene could reflect an increased need to dissipate energy and/or increase CO_2 concentrations near Rubisco to compensate for a potential decrease in the availability of dissolved inorganic carbon and/or Na^+-dependent HCO_3^- transport in low salinities, as was suggested for *T. weissflogii* [43]. Alternatively, given that the Calvin cycle is also upregulated in low salinities, upregulation of PEPC might contribute to a net increase of carbon fixation in low salinities.

Glycolysis/gluconeogenesis is upregulated in low salinities. Protein targeting suggests both pathways are compartmentalized across the chloroplasts, cytosol, and mitochondria (Suppl. Figs 7E-F), presumably allowing them to run simultaneously to supply precursors for biosynthesis of both fatty acids and polysaccharides [58]. Indeed, genes involved

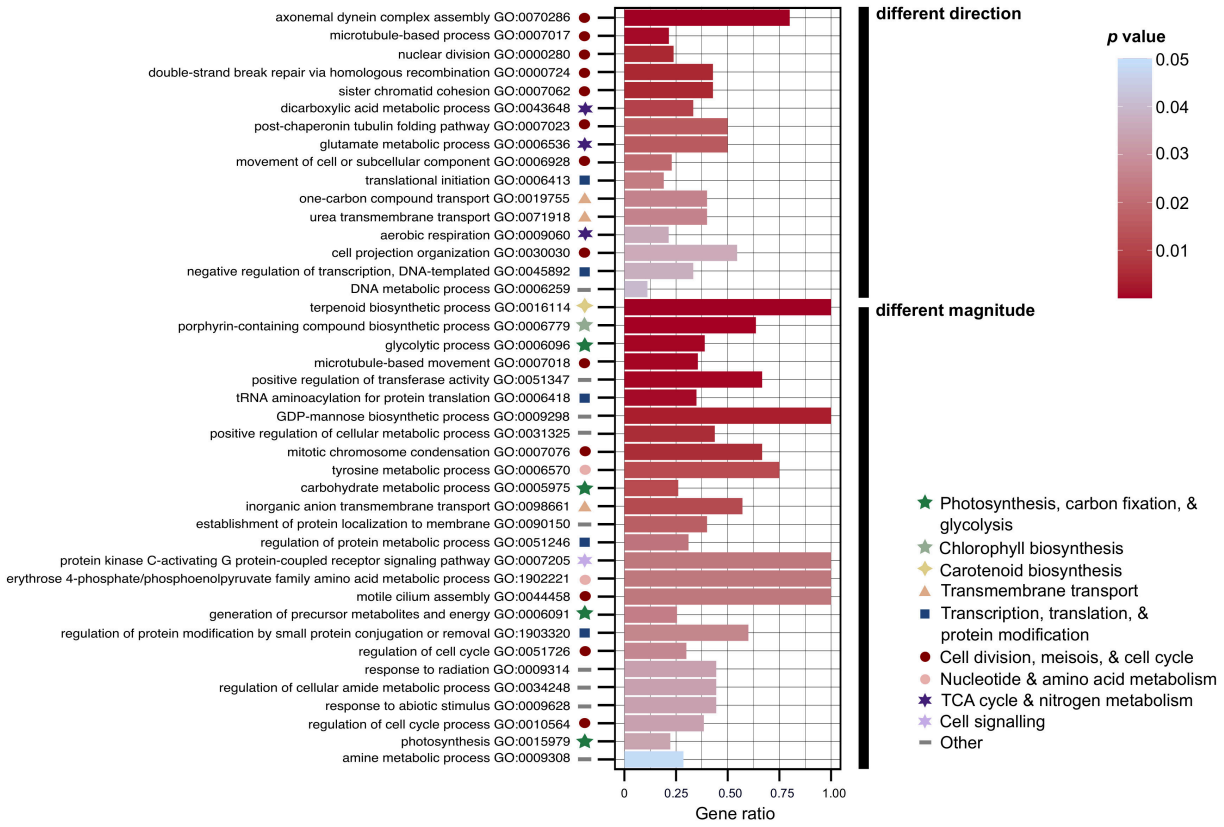


Fig. 4.7. GO enrichment of the interaction effects: Biological Process. The barplot visualizes the significant GO terms retrieved by ORA (topGO, Fisher's exact test, *elim* algorithm) after removal of redundant GO terms by REVIGO. Two sets of GO enrichment were carried out which distinguished between genes that differ significantly between strains in the direction or magnitude of their response to low salinities. Barplot height indicates the proportion of genes that are DE with a given GO term to the total number of genes with this GO term in the genome of *S. marinoi*. The barplots are colored, and the GO terms ranked, according to P-value. Symbols indicate major categories of cellular processes to which a GO term belongs. Only Biological Process GO terms are shown.

in biosynthesis of fatty acids and chrysolaminarin, an important storage polysaccharide in diatoms [59], were upregulated, including four of the core-response genes. In addition, a BASS2-like pyruvate transporter was upregulated, suggestive of increased transport of pyruvate from the cytosol to the chloroplast where it serves as precursor for fatty acid biosynthesis [60]. Diatoms are known to accumulate storage compounds in unfavorable growth conditions, and to modify the fatty acid and lipid composition of their membranes in response to osmotic changes, which alters membrane permeability and fluidity under salinity stress [61–63]. Upregulation of

these genes thus suggests that low salinities represent suboptimal growth conditions for *S. marinoi*.

The hypothesis that low salinities are suboptimal is further supported by expression data that suggest a decrease in nuclear division and silicic acid uptake in low salinities, consistent with a decrease in cell division. Growth rates in the euryhaline diatom *T. weissflogii* also decreased in lower salinity [64], and in the marine diatom *Chaetoceros*, low salinity was found to negatively affect silicon metabolism [65]. Paradoxically, decreased mitosis was not reflected in our growth data measured from relative chlorophyll *a* fluorescence, which showed approximately equal growth rates across salinities for all strains, with even slightly higher (but not significant) rates in low salinities (Fig. 2). However, upregulation of the chlorophyll biosynthesis pathway in low salinities, despite constant light levels, suggests that a decreased growth rate could have been masked by an increase in per-cell chlorophyll content. Such increase in chlorophyll content under moderate hypersalinity stress was previously detected in green algae and is thought to drive elevated photosynthesis [66, 67]. Our data suggest that *S. marinoi* adopts a similar response to low salinity. Given that major salinity stress in algae usually results in a decreased chlorophyll content [68] and less photosynthesis [69], our data indicate that although low salinities are not optimal for *S. marinoi*, when acclimated the diatom is not severely stressed in these conditions. Furthermore, this observation might have important consequences for similar experiments that use chlorophyll *a* as a proxy for growth. Further research is necessary to unravel the link between fluorescence, chlorophyll content, and salinity in *S. marinoi*, given that many factors can influence fluorescence measurements [70].

The response of *S. marinoi* to potential oxidative stress experienced in low salinities is complex. On the one hand, our data suggest that proteins are repaired at higher rates in low salinities, which might reflect an increase in damage caused by oxidative stress [71]. In addition, both the xanthophyll cycle and polyamine biosynthesis were strongly upregulated in low salinities. The former plays a critical role in protection from oxidative stress due to excess light,

but also from ROS generated by other stressors, such as salinity [72]. Polyamines function in abiotic stress responses in land plants, including salinity stress, by increasing antioxidant enzyme activity, triggering the stress signal transduction chain, and serving an osmolyte function [73]. In diatoms, polyamines are known to increase in response to both heat and salinity stress [74, 75], and our data suggest a similar role in salinity acclimation. Violaxanthin-de-epoxidase (xanthophyll cycle), and two genes involved in polyamine biosynthesis belonged to the core response, underscoring the highly conserved nature of this response. On the other hand, several other genes involved in ROS elimination were slightly downregulated, or not DE at all, in low salinities, including the SOD gene which is a first line of defense against ROS in land plants and macroalgae [76, 77]. This might indicate that *S. marinoi* was not acutely stressed, but instead reached an adaptive state of long-term ROS management allowing for survival and growth in suboptimal conditions.

Given the approximately equal growth rates across salinities and transcriptomic evidence consistent with decreased cell division in low salinities, the observation of possibly increased nutrient transport and nitrogen assimilation in low salinities suggests a higher per cell nutrient demand in low salinities. Differences in nutrient uptake and nitrogen assimilation between different salinities have been previously reported in microalgae [78, 79], and could reflect increased biosynthesis of, for example, nitrogen-rich compounds such as polyamines and amino acids, which are essential for the stress response and protein biosynthesis, respectively [61, 80]. Indeed, genes involved in protein biosynthesis are upregulated in low salinities. Previously, higher protein content in lower salinities was observed in the euryhaline diatom *T. weissflogii* [81]. Furthermore, upregulation of amino acid and polyamine transporters pointed to increased demands for compounds essential for cell functioning and/or osmoregulation. This suggests that acclimated *S. marinoi* require more energy and resources to maintain homeostasis in low salinities.

Several of the processes described above were DE in each strain. This core response encompassed 27 genes involved in key processes such as ROS elimination, storage compound biosynthesis, proteolysis, and transmembrane transport. It included one probable transcription factor (*Sm_g00008098*), which is a promising target to unravel the role of gene expression regulation in the salinity response. Increasing the number of technical replicates would likely enlarge the set of core-response genes, as higher replicate numbers improve detection of DE genes, especially those with small effect sizes [82]. Our set of core-response genes is, consequently, not exhaustive but gives a first indication of which genes and processes are likely to be part of a conserved and possibly ancestral response to low salinity in *S. marinoi*.

4.5.2 Osmoregulation in *S. marinoi*

Diatoms produce a variety of osmolytes, small organic molecules that mitigate hyperosmotic stress typical of marine environments [50–53]. Consequently, a decrease in salinity should trigger a drop in osmolyte biosynthesis. Indeed, the expression pattern in low salinities is consistent a decrease in biosynthesis of DMSP, taurine, and possibly betaine. Two core-response genes that were strongly downregulated in all strains could be involved in the biosynthesis of another osmolyte, ectoine. These genes encode a bifunctional aspartokinase/homoserine dehydrogenase (*Sm_g00011041*) and an aspartate-semialdehyde dehydrogenase (*Sm_g00011042*) (Fig. 5, Suppl. Fig. 13). Both are involved in the early steps of the glycine/threonine/serine pathway and convert aspartate into aspartate-semialdehyde and/or homoserine. The *S. marinoi* genome contains several other homologs of both genes. When DE, these homologs show opposite expression patterns to the aforementioned genes: they are upregulated in low salinities, following the expression pattern of other genes in this pathway (Suppl. Fig. 13). Peptide targeting predictions revealed that this pathway is compartmentalized across the chloroplasts, cytoplasm, and mitochondria, presumably allowing *S. marinoi* to run opposite reactions simultaneously while avoiding futile cycles (Suppl. Fig. 13). Given their expression patterns and compartmentalization, *Sm_g00011041* and *Sm_g00011042* are likely

not involved in conventional amino acid biosynthesis. Instead, one of their products, aspartate-semialdehyde, is a known precursor for ectoine, an osmolyte common in bacteria [83]. Elevated levels of aspartate-semialdehyde dehydrogenase have been detected in bacteria occupying high salinities [84]. Recently, marine diatoms were found to both biosynthesize ectoine and import ectoine of bacterial origin [85]. Several *S. marinoi* genes may be homologous to bacterial ectoine genes (*ectA*, *ectB*, *ectC*) that convert aspartate-semialdehyde to ectoine. However, low sequence similarity (maximum 47.8 %), and lack of downregulation in low salinities, raises doubt about whether those genes are responsible for ectoine biosynthesis in *S. marinoi*. Furthermore, all putative homologs received annotations different from ectoine-related genes in Swissprot/Uniprot. It is possible that diatoms have other unknown genes involved in ectoine biosynthesis, or alternatively, diatoms might provide ectoine precursors (e.g., aspartate-semialdehyde) to extracellular bacteria that synthesize and return ectoine to the diatom. Such metabolite exchanges have been shown to occur in diatom–bacteria interactions [86]. Our expression data are consistent with both scenarios and suggest ectoine might be an important osmolyte in *S. marinoi*.

4.5.3 Incorporating multiple strains reveals intraspecific variation

The above observations were based on the average and core response of all eight strains, which revealed the general response of Baltic *S. marinoi* to low salinities. However, when taking data of individual strains into account, we observed substantial intraspecific variation in gene expression. Many of the processes identified in the average response differed significantly among strains in magnitude, indicating that strains vary in the strength of their salinity response. This becomes clearer when examining differential expression in different pathways of individual strains (Suppl. Figs 7-13). Whereas many pathways are almost entirely DE in the average response, this is often not the case for individual strains, which have fewer DE genes than the average response, and differ from one another in which genes are DE as well as the strength of DE (measured as logFC). For example, whereas four strains significantly

upregulated almost the entire chlorophyll biosynthesis pathway, none or only a few genes in this pathway are DE in the other strains (Suppl. Fig. 8A). Similarly, four and five strains significantly upregulated most genes involved in chrysolaminarin and fatty acid biosynthesis, respectively, whereas the other strains have only few genes DE in the same pathways (Suppl. Fig. 10A).

A second set of genes differed among strains in the direction of their response. Thus, strains deviated in their strategies to cope with low salinity. This included both cell wall and calcium-binding messenger proteins as well as heat shock proteins/chaperones. The latter are known to help mitigate elevated salt stress in sea-ice diatoms [87], and our data suggest they play a role also in acclimation to low salinity, although this role is variable across strains. Altogether, these data highlight intraspecific differences in how salinity stress affects cell functioning, including cell-signaling pathways. For example, Ca^{2+}-signaling is involved in osmotic sensing in diatoms [88], suggesting strains differ in how they respond to osmotic stress.

The interaction effects included several DE genes related to the cell cycle. Diatoms have an unusual cell cycle that involves progressive cell size reduction through mitotic cell divisions until cell size drops below a species-specific ‘sexual size threshold’ (SST) at which point the diatom can undergo sexual reproduction with a partner cell (allomixis), usually in response to an environmental trigger [89]. In *S. marinoi*, sexual reproduction in cells below the SST can be induced by shifts to higher salinity [90]. Because cultures were shifted to experimental salinities one week before the experiment, any sexual reproduction that occurred at this time was long finished upon RNA harvesting [55, 90]. However, when below the SST, *S. marinoi* cells can restore their maximum cell size, through an auxospore-like stage that is not contingent upon a salinity shift and might involve autogamy, apomixis, or vegetative cell enlargement [89, 90]. Although the genes involved in this process are unknown, they likely include much of the cell cycle genes identified as DE in the interaction effects, suggesting salinity impacts size restoration differently in different strains. This could happen through a direct impact on size restoration (cultures in suboptimal salinity might redirect more energy to maintaining

homeostasis or growth, and less to size restoration, or *vice versa*), or through an indirect impact on growth rates (higher or lower growth rates in different salinities might result in a different proportions of cells under the SST, resulting in more or less size restoration). To rule out that size restoration alone was responsible for the interaction effects, we removed from our dataset all *S. marinoi* sex-induced genes identified by [55], as well as genes with GO terms related to cell cycle/mitosis/meiosis (6 218 genes, 1 797 of which are DE in the interaction effects). Upon removal, samples still clustered principally by strain, not salinity (Suppl. Fig. 16), indicating size restoration is not driving the interaction effects. The high growth rates throughout the experiment confirm this (Fig. 2) [91].

Intraspecific variation in the response to low salinity likely allowed *S. marinoi* to colonize and grow throughout the Baltic Sea. Diatoms can harbor high levels of genotypic and phenotypic variation [22–24, 92–96], and *S. marinoi* even shows trait variation between individual cells of the same clone [23]. The variation in gene expression shown here is a natural extension of these observations. Our study design did not allow testing whether the intraspecific variation is related to the natural salinities at which the different strains occur, as this would require sampling of multiple strains within populations. Nevertheless, visual comparison of gene expression patterns did not show consistent differences across low- (D, F, I, J, K, P) and high-salinity (A, B) populations (Suppl. Figs 7–13), nor did those populations cluster separately (Fig. 6). This suggests that if signals of local adaptation along the Baltic salinity cline [14] are due in part to differences in gene expression between high- and low-salinity populations, then those differences are subtle. In any case, substantial intraspecific variation in gene expression in *S. marinoi* exists and is likely critical to its survival, acclimation, and adaptation to a dynamic environment such as the Baltic Sea, where in addition to salinity, marked gradients and seasonal fluctuations in nutrients and temperature also occur [24, 97]. The variation in gene expression observed here increases the chance that at least some cells can survive rapidly fluctuating, potentially adverse, conditions in the short term. Assuming some of this variation is

heritable, variable gene expression might also enable long-term evolutionary adaptation by providing targets for natural selection [8, 98].

4.6 Conclusion

Our study design, in which transcriptome data from eight strains were combined into a single analysis, allowed for a holistic view of the response of *S. marinoi* to low salinities in the Baltic Sea, the world's largest brackish water body. Transcriptome studies often include technical replicates of a single strain, but an increasing number of studies [48, 51] show that experiments without biological replicates are unlikely to be generalizable, as different strains can exhibit markedly different patterns of gene expression. Here, inclusion of both technical and biological replicates allowed us to characterize both conserved and variable responses to low salinity.

We found that when *S. marinoi* experiences long-term exposure to low salinities that mimic the natural Baltic Sea salinity gradient, the diatom is not severely stressed but experiences elevated energy and nutrient demands, increases photosynthesis and storage compound biosynthesis, and deploys a complex response to oxidative stress. This response likely allowed the ancestrally marine *S. marinoi* to grow successfully in low salinity environments and become one of the dominant primary producers in the Baltic Sea. Our analyses revealed substantial intraspecific variability in the response of *S. marinoi* to low salinities, highlighting an important source of biological variation in diatoms. Metatranscriptomics offers a powerful approach for identifying community- and species-level responses to other natural gradients in the ocean [99]. Similar studies of the Baltic Sea would provide valuable corroboration of the results from our controlled laboratory experiment. Altogether, our data indicate variable gene expression plays an important role in how diatoms respond and adapt to environmental change.

4.7 Data availability

RNA-seq data are available from the Sequence Read Archive (NCBI) under project number PRJNA772794. The *S. marinoi* reference genome (v.1.1) used for read mapping, gene-level count data, and the scripts needed to reproduce all analyses and figures are available from Zenodo (doi 10.5281/zenodo.5266588). All *S. marinoi* strains are publicly available from the BCCM/DCG diatom culture collection (https://bccm.belspo.be/about-us/bccm-dcg) under accession numbers DCG 1232–1239. LSU D1–D2 rDNA gene sequences used for strain identification are available from GenBank (OM112317–OM112324).

4.8 Acknowledgements

This work was supported by two grants from the Simons Foundation (725407, EP and 403249, AJA) and a grant from the National Science Foundation (1651087, AJA). EP also benefited from postdoctoral fellowships from Fulbright Belgium and the Belgian American Educational Foundation (B.A.E.F.). K.V.d.B. is a FWO junior postdoc fellow (project 1246220N) and previously benefited from a B.A.E.F. grant. We are grateful to Sirje Sildever (Tallinn University of Technology, Estonia), Björn Andersson (University of Gothenburg, Sweden), Andrzej Witkowski (University of Szczecin, Poland), Jörg Dutz (Leibniz Institute for Baltic Sea Research Warnemuende, Germany), Justyna Kobos (University of Gdansk, Poland), and Anu Vehmaa (University of Helsinki, Finland) for sample collection. We thank Wade Roberts (University of Arkansas in Fayetteville, USA) and Quinten Bafort and Gust Bilcke (Ghent University, Belgium) for providing advice on the data analysis.

4.9 References

1. Lozupone CA, Knight R. Global patterns in bacterial diversity. *Proc Natl Acad Sci U S A* 2007; **104**: 11436–11440.

2. Logares R, Bråte J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, Rengefors K. Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol* 2009; **17**: 414–422.

3. Cavalier-Smith T. Megaphylogeny, cell body plans, adaptive zones: causes and timing of eukaryote basal radiations. *J Eukaryot Microbiol* 2009; **56**: 26–33.

4. Nakov T, Beaulieu JM, Alverson AJ. Diatoms diversify and turn over faster in freshwater than marine environments. *Evolution* 2019; **73**: 2497–2511.

5. Dittami SM, Heesch S, Olsen JL, Collén J. Transitions between marine and freshwater environments provide new clues about the origins of multicellular plants and algae. *J Phycol* 2017; **53**: 731–745.

6. Dickson B, Yashayaev I, Meincke J, Turrell B, Dye S, Holfort J. Rapid freshening of the deep North Atlantic Ocean over the past four decades. *Nature* 2002; **416**: 832–837.

7. Aretxabaleta AL, Smith KW, Kalra TS. Regime changes in global sea surface salinity trend. *J Mar Sci Eng* 2017; **5**: 57.

8. López-Maury L, Marguerat S, Bähler J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* 2008; **9**: 583–593.

9. Björck S. A review of the history of the Baltic Sea, 13.0-8.0 ka BP. *Quat Int* 1995; **27**: 19–40.

10. Krauss W. Baltic Sea Circulation. In: Steele JH (ed). *Encyclopedia of Ocean Sciences*. 2001. Academic Press, Oxford, pp 236–244.

11. Telesh I, Schubert H, Skarlato S. Life in the salinity gradient: Discovering mechanisms behind a new biodiversity pattern. *Estuar Coast Shelf Sci* 2013; **135**: 317–327.

12. Johannesson K, Le Moan A, Perini S, André C. A Darwinian Laboratory of Multiple Contact Zones. *Trends Ecol Evol* 2020; **35**: 1021–1036.

13. Olofsson M, Hagan JG, Karlson B, Gamfeldt L. Large seasonal and spatial variation in nano- and microphytoplankton diversity along a Baltic Sea-North Sea salinity gradient. *Sci Rep* 2020; **10**: 17666.

14. Sjöqvist C, Godhe A, Jonsson PR, Sundqvist L, Kremp A. Local adaptation and oceanographic connectivity patterns explain genetic differentiation of a marine diatom across the North Sea-Baltic Sea salinity gradient. *Mol Ecol* 2015; **24**: 2871–2885.

15. Busse S, Snoeijs P. Gradient responses of diatom communities in the Bothnian Bay, northern Baltic Sea. *Nova Hedwigia* 2002; **74**: 501–525.

16. Gasiūnaitė ZR, Cardoso AC, Heiskanen A-S, Henriksen P, Kauppila P, Olenina I, et al. Seasonality of coastal phytoplankton in the Baltic Sea: Influence of salinity and eutrophication. *Estuar Coast Shelf Sci* 2005; **65**: 239–252.

17. Jochem F. Distribution and importance of autotrophic ultraplankton in a boreal inshore area (Kiel Bight, Western Baltic). *Mar Ecol Prog Ser* 1989; **53**: 153–168.

18. Lange CB, Hasle GR, Syvertsen EE. Seasonal cycle of diatoms in the Skagerrak, North Atlantic, with emphasis on the period 1980–1990. *Sarsia* 1992; **77**: 173–187.

19. van Wirdum F, Andrén E, Wienholz D, Kotthoff U, Moros M, Fanget A-S, et al. Middle to Late Holocene Variations in Salinity and Primary Productivity in the Central Baltic Sea: A Multiproxy Study From the Landsort Deep. *Frontiers in Marine Science* 2019; **6**: 51.

20. Warnock J, Andrén E, Juggins S, Lewis J, Ryves DB, Andrén T, et al. A high-resolution diatom-based Middle and Late Holocene environmental history of the Little Belt region, Baltic Sea. *Boreas* 2020; **49**: 1–16.

21. Alverson AJ. Timing marine–freshwater transitions in the diatom order Thalassiosirales. *Paleobiology* 2014; **40**: 91–101.

22. Kremp A, Godhe A, Egardt J, Dupont S, Suikkanen S, Casabianca S, et al. Intraspecific variability in the response of bloom-forming marine microalgae to changed climate conditions. *Ecol Evol* 2012; **2**: 1195–1207.

23. Olofsson M, Kourtchenko O, Zetsche E-M, Marchant HK, Whitehouse MJ, Godhe A, et al. High single-cell diversity in carbon and nitrogen assimilations by a chain-forming diatom across a century. *Environ Microbiol* 2019; **21**: 142–151.

24. Olofsson M, Almén A-K, Jaatinen K, Scheinin M. Temporal escape - adaptation to eutrophication by *Skeletonema marinoi*. *FEMS Microbiol Lett* 2022.

25. Godhe A, Härnström K. Linking the planktonic and benthic habitat: genetic structure of the marine diatom *Skeletonema marinoi*. *Mol Ecol* 2010; **19**: 4478–4490.

26. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics* 2015; **51**: 11-14.

27. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; **31**: 166–169.

28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410.

29. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014; **30**: 1236–1240.

30. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 2020; **36**: 2251–2252.

31. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 2019; **20**: 238.

32. Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, et al. Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance* 2019; **2**: e201900429.

33. Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J* 2015; **81**: 519–528.

34. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**: 783–795.

35. Gschloessl B, Guermeur Y, Cock JM. HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinformatics* 2008; **9**: 393.

36. Claros MG. MitoProt, a Macintosh application for studying mitochondrial proteins. *Comput Appl Biosci* 1995; **11**: 441–447.

37. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; **26**: 139–140.

38. Van den Berge K, Soneson C, Robinson MD, Clement L. stageR: a general stage-wise method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. *Genome Biol* 2017; **18**: 151.

39. Heller R, Manduchi E, Grant GR, Ewens WJ. A flexible two-stage procedure for identifying gene sets that are differentially expressed. *Bioinformatics* 2009; **25**: 1019–1025.

40. Alexa A, Rahnenführer J. Gene set enrichment analysis with topGO. *Bioconductor Improv* 2009; **27**.

41. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res* 2012; **40**: e133.

42. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 2011; **6**: e21800.

43. Bussard A, Corre E, Hubas C, Duvernois-Berthet E, Le Corguillé G, Jourdren L, et al. Physiological adjustments and transcriptome reprogramming are involved in the acclimation to salinity gradients in diatoms. *Environ Microbiol* 2017; **19**: 909–925.

44. Matthijs M, Fabris M, Obata T, Foubert I, Franco-Zorrilla JM, Solano R, et al. The transcription factor bZIP14 regulates the TCA cycle in the diatom *Phaeodactylum tricornutum*. *EMBO J* 2017; **36**: 1559–1576.

45. Kong L, Price NM. Transcriptomes of an oceanic diatom reveal the initial and final stages of acclimation to copper deficiency. *Environ Microbiol* 2021.

46. Amato A, Sabatino V, Nylund GM, Bergkvist J, Basu S, Andersson MX, et al. Grazer-induced transcriptomic and metabolomic response of the chain-forming diatom *Skeletonema marinoi*. *ISME J* 2018; **12**: 1594–1604.

47. Maumus F, Allen AE, Mhiri C, Hu H, Jabbari K, Vardi A, et al. Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genomics* 2009; **10**: 624.

48. Pargana A, Musacchia F, Sanges R, Russo MT, Ferrante MI, Bowler C, et al. Intraspecific diversity in the cold stress response of transposable elements in the diatom *Leptocylindrus aporus*. *Genes* 2019; **11**: 9.

49. Smith SR, Dupont CL, McCarthy JK, Broddrick JT, Oborník M, Horák A, et al. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat Commun* 2019; **10**: 4552.

50. Kageyama H, Tanaka Y, Shibata A, Waditee-Sirisattha R, Takabe T. Dimethylsulfoniopropionate biosynthesis in a diatom *Thalassiosira pseudonana*: Identification of a gene encoding MTHB-methyltransferase. *Arch Biochem Biophys* 2018; **645**: 100–106.

51. Nakov T, Judy KJ, Downey KM, Ruck EC, Alverson AJ. Transcriptional Response of Osmolyte Synthetic pathways and membrane transporters in a euryhaline diatom during long-term acclimation to a salinity gradient. *J Phycol* 2020; **56**: 1712–1728.

52. Kageyama H, Tanaka Y, Takabe T. Biosynthetic pathways of glycinebetaine in *Thalassiosira pseudonana*; functional characterization of enzyme catalyzing three-step methylation of glycine. *Plant Physiol Biochem* 2018; **127**: 248–255.

53. Krell A, Funck D, Plettner I, John U, Dieckmann G. Regulation of proline metabolism under salt stress in the psychrophilic diatom *Fragilariopsis cylindrus* (Bacillariophyceae). *J Phycol* 2007; **43**: 753–762.

54. Latta LC, Weider LJ, Colbourne JK, Pfrender ME. The evolution of salinity tolerance in *Daphnia*: a functional genomics approach. *Ecol Lett* 2012; **15**: 794–802.

55. Ferrante MI, Entrambasaguas L, Johansson M, Töpel M, Kremp A, Montresor M, et al. Exploring molecular signs of sex in the marine diatom *Skeletonema marinoi*. *Genes* 2019; **10**.

56. Kroth PG. The biodiversity of carbon assimilation. *J Plant Physiol* 2015; **172**: 76–81.

57. Obata T, Fernie AR, Nunes-Nesi A. The central carbon and energy metabolism of marine diatoms. *Metabolites* 2013; **3**: 325–346.

58. Smith SR, Abbriano RM, Hildebrand M. Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. *Algal Research* 2012; **1**: 2–16.

59. Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, et al. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One* 2008; **3**: e1426.

60. Furumoto T, Yamaguchi T, Ohshima-Ichie Y, Nakamura M, Tsuchida-Iwata Y, Shimamura M, et al. A plastidial sodium-dependent pyruvate transporter. *Nature* 2011; **476**: 472–475.

61. Chen G-Q, Jiang Y, Chen F. Salt-induced alterations in lipid composition of diatom *Nitzschia laevis* (Bacillariophyceae) under heterotrophic culture condition. *J Phycol* 2008; **44**: 1309–1314.

62. Sayanova O, Mimouni V, Ulmann L, Morant-Manceau A, Pasquet V, Schoefs B, et al. Modulation of lipid biosynthesis by stress in diatoms. *Philos Trans R Soc Lond B Biol Sci* 2017; **372**: 20160407.

63. Vårum KM, Myklestad S. Effects of light, salinity and nutrient limitation on the production of β-1,3-d-glucan and exo-d-glucanase activity in *Skeletonema costatum* (Grev.) Cleve. *J Exp Mar Bio Ecol* 1984; **83**: 13–25.

64. Radchenko IG, Il'yash LV. Growth and photosynthetic activity of diatom *Thalassiosira weissflogii* at decreasing salinity. *Biology Bulletin* 2006; **33**: 242–247.

65. Adams C, Bugbee B. Enhancing lipid production of the marine diatom *Chaetoceros gracilis*: synergistic interactions of sodium chloride and silicon. *J Appl Phycol* 2014; **26**: 1351–1357.

66. Shetty P, Gitau MM, Maróti G. Salinity stress responses and adaptation mechanisms in eukaryotic green microalgae. *Cells* 2019; **8**.

67. Jacob A, Kirst GO, Wiencke C, Lehmann H. Physiological responses of the Antarctic green alga *Prasiola crispa* ssp. *antarctica* to salinity stress. *J Plant Physiol* 1991; **139**: 57–62.

68. Bazzani E, Lauritano C, Mangoni O, Bolinesi F, Saggiomo M. *Chlamydomonas* responses to salinity stress and possible biotechnological exploitation. *J Mar Sci Eng* 2021; **9**: 1242.

69. Cheng R-L, Feng J, Zhang B-X, Huang Y, Cheng J, Zhang C-X. Transcriptome and gene expression analysis of an oleaginous diatom under different salinity conditions. *Bioenergy Res* 2014; **7**: 192–205.

70. Stock W, Blommaert L, Daveloose I, Vyverman W, Sabbe K. Assessing the suitability of Imaging-PAM fluorometry for monitoring growth of benthic diatoms. *J Exp Mar Bio Ecol* 2019; **513**: 35–41.

71. Reichmann D, Voth W, Jakob U. Maintaining a healthy proteome during oxidative stress. *Mol Cell* 2018; **69**: 203–213.

72. Latowski D, Kuczyńska P, Strzałka K. Xanthophyll cycle--a mechanism protecting plants against oxidative stress. *Redox Rep* 2011; **16**: 78–90.

73. Chen D, Shao Q, Yin L, Younis A, Zheng B. Polyamine function in plants: Metabolism, regulation on development, and roles in abiotic stress responses. *Front Plant Sci* 2018; **9**: 1945.

74. Liu Q, Nishibori N, Imai I, Hollibaugh JT. Response of polyamine pools in marine phytoplankton to nutrient limitation and variation in temperature and salinity. *Mar Ecol Prog Ser* 2016; **544**: 93–105.

75. Scoccianti V, Penna A, Penna N, Magnani M. Effect of heat stress on polyamine content and protein pattern in *Skeletonema costatum*. *Mar Biol* 1995; **121**: 549–554.

76. Alscher RG, Erturk N, Heath LS. Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *J Exp Bot* 2002; **53**: 1331–1341.

77. Kumar M, Kumari P, Gupta V, Reddy CRK, Jha B. Biochemical responses of red alga *Gracilaria corticata* (Gracilariales, Rhodophyta) to salinity induced oxidative stress. *J Exp Mar Bio Ecol* 2010; **391**: 27–34.

78. von Alvensleben N, Magnusson M, Heimann K. Salinity tolerance of four freshwater microalgal species and the effects of salinity and nutrient limitation on biochemical profiles. *J Appl Phycol* 2016; **28**: 861–876.

79. Rijstenbil JW, Wijnholds JA, Sinke JJ. Implications of salinity fluctuation for growth and nitrogen metabolism of the marine diatom *Ditylum brightwellii* in comparison with *Skeletonema costatum*. *Mar Biol* 1989; **101**: 131–141.

80. Mansour MMF. Nitrogen containing compounds and adaptation of plants to salinity stress. *Biol Plant* 2000; **43**: 491–500.

81. Garcia N, Lopez Elias JA, Miranda A, Martinez Porchas M, Huerta N, Garcia A. Effect of salinity on growth and chemical composition of the diatom *Thalassiosira weissflogii* at three culture phases. *Lat Am J Aquat Res* 2012; **40**: 435–440.

82. Van den Berge K, Hembach KM, Soneson C, Tiberi S, Clement L, Love MI, et al. RNA Sequencing data: Hitchhiker's guide to expression analysis. *Annu Rev Biomed Data Sci* 2019; **2**: 139–173.

83. Vargas C, Argandoña M, Reina-Bueno M, Rodríguez-Moya J, Fernández-Aunión C, Nieto JJ. Unravelling the adaptation responses to osmotic and temperature stress in *Chromohalobacter salexigens*, a bacterium with broad salinity tolerance. *Saline Systems* 2008; **4**: 14.

84. Khmelenina VN, Sakharovskii VG, Reshetnikov AS, Trotsenko YA. Synthesis of osmoprotectants by halophilic and alkaliphilic methanotrophs. *Microbiology* . 2000. , **69**: 381–386

85. Fenizia S, Thume K, Wirgenings M, Pohnert G. Ectoine from bacterial and algal origin is a compatible solute in microalgae. *Mar Drugs* 2020; **18**: 42.

86. Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 2015; **522**: 98–101.

87. Krell A, Beszteri B, Dieckmann G, Glöckner G, Valentin K, Mock T. A new class of ice-binding proteins discovered in a salt-stress-induced cDNA library of the psychrophilic diatom *Fragilariopsis cylindrus* (Bacillariophyceae). *Eur J Phycol* 2008; **43**: 423–433.

88. Helliwell KE, Kleiner FH, Hardstaff H, Chrachri A, Gaikwad T, Salmon D, et al. Spatiotemporal patterns of intracellular Ca^{2+} signalling govern hypo-osmotic stress resilience in marine diatoms. *New Phytol* 2021; **230**: 155–170.

89. Kaczmarska I, Poulíčková A, Sato S, Edlund MB, Idei M, Watanabe T, et al. Proposals for a terminology for diatom sexual reproduction, auxospores and resting stages. *Diatom Res* 2013; **28**: 263–294.

90. Godhe A, Kremp A, Montresor M. Genetic and microscopic evidence for sexual reproduction in the centric diatom *Skeletonema marinoi*. *Protist* 2014; **165**: 401–416.

91. Annunziata R, Mele BH, Marotta P, Volpe M, Entrambasaguas L, Mager S, et al. Trade-off between sex and growth in diatoms: Molecular mechanisms and demographic implications. *Sci Adv* 2022; **8**: eabj9466.

92. Ajani PA, Petrou K, Larsson ME, Nielsen DA, Burke J, Murray SA. Phenotypic trait variability as an indication of adaptive capacity in a cosmopolitan marine diatom. *Environ Microbiol* 2021; **23**: 207–223.

93. Sjöqvist CO, Kremp A. Genetic diversity affects ecological performance and stress response of marine diatom populations. *ISME J* 2016; **10**: 2755–2766.

94. Godhe A, Rynearson T. The role of intraspecific variation in the ecological and evolutionary success of diatoms in changing environments. *Philos Trans R Soc Lond B Biol Sci* 2017; **372**: 20160399.

95. Bulankova P, Sekulić M, Jallet D, Nef C, van Oosterhout C, Delmont TO, et al. Mitotic recombination between homologous chromosomes drives genomic diversity in diatoms. *Curr Biol* 2021; **31**: 3221-3232.e9.

96. Pinseel E, Janssens SB, Verleyen E, Vanormelingen P, Kohler TJ, Biersma EM, et al. Global radiation in a rare biosphere soil diatom. *Nat Commun* 2020; **11**: 2382.

97. Savchuk OP. Large-scale nutrient dynamics in the Baltic sea, 1970–2016. *Front Mar Sci* 2018; **5**: 95.

98. Gomez-Mestre I, Jovani R. A heuristic model on the role of plasticity in adaptive evolution: plasticity increases adaptation, population viability and genetic variation. *Proc Biol Sci* 2013; **280**: 20131869.

99. Lambert BS, Groussman RD, Schatz MJ, Coesel SN, Durham BP, Alverson AJ, et al. The dynamic trophic architecture of open-ocean protist communities revealed through machine-guided metatranscriptomics. *Proc Natl Acad Sci U S A* 2022; **119**.

100. Harrison PF, Pattison AD, Powell DR, Beilharz TH. Topconfects: a package for confident effect sizes in differential expression analysis provides a more biologically useful ranked gene list. *Genome Biol* 2019; **20**: 67.

Chapter 5 The Dynamic Hypoosmotic Stress Response of a Euryhaline Diatom Reveals that the Genes Involved in the Acute Response versus Acclimation are Largely Distinct.

Kala M. Downey, Kathryn J. Judy, Eveline Pinseel, Andrew J. Alverson, Jeffrey A. Lewis

[1]Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas, USA

Author Contributions: Andrew Alverson, Jeffery Lewis, and I designed the experiment. I performed data collection and analyses. Kathryn Judy and Eveline Pinseel helped with code production. I wrote the manuscript. Andy, Jeff, and Eveline provided comments and edits.

5.1 Abstract

The salinity gradient separating marine and freshwater environments is a major ecological divide, and the mechanisms by which organisms adapt to new salinity environments are poorly understood. Diatoms, a hyperdiverse group of marine microbes, have accomplished this feat multiple times. *Cyclotella cryptica*, a euryhaline diatom, is a model for studying cell wall morphogenesis, lipid production, and notably, salinity tolerance. Indeed, *C. cryptica* tolerates a wide range of salinities and occurs naturally in both oceanic and freshwater habitats, presenting a powerful system for understanding the genomic mechanisms for mitigating and acclimating to salinity stress. To understand the dynamics of gene expression changes during acute hypo-osmotic stress, we abruptly shifted *C. cryptica* from seawater to freshwater and performed transcriptional profiling at 8 time points across 10 hours. We found dramatic remodeling of the transcriptome, with over half of the genome showing differential expression in at least one time point. Using the gene annotations from KEGG, Uniprot, and Swissprot, we identified several metabolic pathways that were differentially regulated in response to early osmotic stress. Considering the results of the enrichment analyses of the clusters and 30 min, we focused on pathways associated with energy metabolism, chitin metabolism, ribosome biosynthesis, tRNA biosynthesis, and management of reactive oxygen species. We hypothesize that the function of gene repression is to redirect translational capacity to induced transcripts. Notably, transcripts largely returned to baseline by 10 hours whereas *C. cryptica* acclimated and resumed growth within 24 hours, suggesting that gene expression dynamics may be useful for predicting acclimation. Overall, this study highlights the power of analyzing dynamic responses across multiple time scales to gain new insight into stress defense and acclimation.

5.2 Introduction

Although many aquatic organisms occur exclusively in marine or freshwaters, euryhaline species are defined as being able to tolerate a wide range of salinities. Thus, they present an excellent opportunity to understand the myriad strategies used to mitigate the effects of short and long-term salinity changes. Euryhaline organisms inhabit environments that are characterized by not only a wide range of salinities (e.g, lagoons, estuaries, salt marshes, and coastal intertidal zones), but also other fluctuating environmental conditions such as water depth, light levels, and nutrient availability (Kirst 1990, Balzano et al. 2015). Osmotic shifts in these habitats can range from freshwater to fully marine and may occur as gradual transitions or sudden shocks, occurring within minutes or hours depending on rainfall, river flow, and tidal action (Balzano et al. 2015). Climate change is expected to increase the volatility of these fluctuations (Shu et al. 2018, Pörtner et al. 2019). In addition, many marine and brackish environments will experience “freshening” due to melting ice caps and altered precipitation patterns, such as the increasing frequency of large coastal storms that inundate these habitats with freshwater from both precipitation and increased river flow (Shu et al. 2018, Pörtner et al. 2019). Thus, characterizing the molecular response to acute salinity stress—particularly hypoosmotic stress—will provide important insights into whether and how these species will respond to increasingly dynamic salinity regimes in nature.

Salinity shifts induce an immediate response in diatoms, resulting in overproduction of many secondary metabolites (Cheng et al. 2014, d'Ippolito et al. 2015, Lyon et al. 2016), allowing them to survive a temporary fluctuation or eventually adapt to prolonged change. This short-term stress response commences almost immediately following the environmental trigger, and allows diatoms and other microalgae to persist long enough to either survive a temporary fluctuation, or acclimate and eventually adapt to a prolonged change in the environment (Borowitzka 2018). On evolutionary timescales, transitions between marine and freshwater have

been an important source of species diversification in diatoms (Nakov et al. 2018) and other eukaryotes (Jamy et al. 2021).

Based mostly on responses to hypersaline stress, one of the main ways diatoms balance internal osmotic pressure is to either increase (for hypersaline stress) or decrease (for hyposaline stress) the concentrations of key osmolytes (Kirst 1990). This is accomplished either through transport of inorganic ions (e.g, potassium) or by synthesizing or degrading organic osmolytes such as dimethylsulfoniopropionate (DMSP) (Lyon et al. 2011), glycine betaine (Dickson and Kirst 1987, Kageyama et al. 2018), and proline (Liu and Hellebust 1976, Krell et al. 2008). Marine organisms experiencing hypotonic stress are subject to increased turgor, causing expansion of the plasma membrane (Kirst 1990, Van Bergeijk et al. 2003, Theseira et al. 2020). In an effort to maintain constant cell turgor, microalgae and other organisms adjust the intracellular osmotic pressure (Van Bergeijk et al. 2003). This adjustment begins with a rapid movement of water into the cell, which triggers a shift in ion and osmolyte balances as the expanding cell membrane stimulates active transporters within the membrane (Kirst 1990, Scholz and Liebezeit 2012).

Cyclotella cryptica is one of many euryhaline diatoms that occurs naturally across a wide range of salinities and has emerged as a leading model for understanding the effects of salinity stress in diatoms. The effects of salinity shifts on *C. cryptica* are profound, and include considerable remodeling of the silica cell wall with alterations to thickness as well as ornamentation (Schultz 1971). In higher salinities, the cell wall became thin and more fragile, while growth in lower salinities resulted in the thicker cell walls commonly seen in freshwater diatoms (Conley et al. 1989) and ornamentation more similar to that of *Cyclotella meneghiniana*. Acute salinity shifts can also induce gametogenesis in C. cryptica (Schultz and Trainor 1970) and other euryhaline diatoms (Godhe et al. 2014).

There is an apparent disparity between *C. cryptica*'s response to hyposaline stress immediately following exposure and mechanisms relied on following acclimation (Liu and

Hellebust 1974, Liu and Hellebust 1976, Nakov et al. 2020). When initially exposed to hyposaline conditions, metabolic evidence suggests that early on in the response, *C. cryptica* incorporated osmolyte-functioning amino acids into proteins (Liu and Hellebust 1974, Liu and Hellebust 1976). Transcriptional profiling suggests this is followed by decreasing production of glycine betaine, DMSP, and taurine in later stages of acclimation (Nakov et al. 2020). Specifically, proline was found to be an important osmolyte in response to both high and low salinity in shorter timescales (Liu and Hellebust 1974, Liu and Hellebust 1976), but not regulated at all following acclimation (Nakov et al. 2020).

We exposed a brackish strain of *C. cryptica* to freshwater and used RNA sequencing (RNA-seq) to characterize changes in gene-expression in the hours after initial exposure to freshwater conditions. Comparisons to expression-level changes in fully acclimated cells highlighted important differences between short- and long-term responses to hyposalinity in diatoms, providing valuable new insights into how diatoms mitigate environmental changes in salinity and have successfully colonized and diversified in freshwater habitats worldwide.

5.3 Materials and Methods

5.3.1 Experimental conditions. *C. cryptica* strain CCMP332 was obtained from the National Center for Marine Algae and Microbiota. The culture was maintained in artificial sea water (ASW) at its original brackish salinity of 24 practical salinity units (psu) (Nakov et al. 2020). Growth rates in ASW 24 or 0 (freshwater) were determined by inoculating 1 x 106 cells into 4 mL media and measuring cell counts at various time points over 10 h via a Benchtop B3 Series FlowCAM® cytometer (Fluid Imaging Technologies).

Prior to the experiment, cells were grown in a 500 mL Erlenmeyer flask for 7 days in a Percival incubator at 15℃ and 22 µmol photons m^{-2} s^{-1} irradiance under a 16:8 h light:dark cycle to obtain sufficient biomass for the experiment. Cells were homogenized by agitating the source flask, and three 2-mL samples were used to inoculate three 1-L Erlenmeyer flasks containing

500 mL ASW 24. Growth was monitored by counting cells with a Benchtop B3 Series FlowCAM® cytometer (Fluid Imaging Technologies). Upon reaching exponential growth in ASW 24, cells from each triplicate culture were immediately exposed to freshwater conditions (ASW 0) as follows. We harvested 24 x 10^6 cells from each replicate for freshwater treatments, centrifuged at 800 rcf for 3 min at 4 ℃, decanted the supernatant, and resuspended the cells in 16 mL of ASW 0 to a final concentration of 1.5 x 10^6 cells/ml. The 16-mL of cells in freshwater were aliquoted (2 mL each) into 8 tubes containing 38 mL of ASW 0. During the experiment, cells were incubated at 15℃ and 20 µmol photons m^{-2} s^{-1} irradiance with gentle agitation using a Boekel Scientific adjustable speed wave rocker to prevent the cell settling.

Cells were collected for transcriptional profiling at seven different time points following inoculation in ASW 0: 15 min, 30 min, 1 h, 2 h, 4 h, 8 h, and 10 h. A 0 min sample was collected immediately following inoculation in ASW 0 and used as an unstressed control. Cells were harvested by centrifugation at 400 rcf for 3 min at 4℃, flash-frozen in liquid nitrogen, and stored at -80 ℃ until processed.

5.3.2 RNA extraction and library preparation. To prevent batch effects, we randomized all samples during both RNA extraction and library construction. RNA was extracted with an RNeasy Plant Mini Kit (QIAGEN, Netherlands) and quantified with a Qubit 2.0 Fluorometer (Invitrogen, USA). RNA quality was assessed using an Agilent Technologies 2200 TapeStation (Agilent Technologies, USA). All samples had an RNA integrity number > 6.5 (Supplementary Table 1). We constructed dual-indexed RNA libraries with the KAPA mRNA HyperPrep kit (KAPA Biosystems, USA) using half reaction volumes. Libraries were multiplexed and sequenced on a single lane of an Illumina HiSeq 4000 (2 x 100 bp paired-end reads) at the Biological Sciences Core Facility (University of Chicago). Details on RNA extractions, library prep, and mapping are available in Supplementary Table 1.

5.3.3 RNA-seq analysis. Quality of the raw reads was examined using FastQC v0.11.5 (Andrew 2019). Low-quality reads were removed and adapter sequences trimmed (KAPA v3 dual indices) using kTrim v1.1.0 (Sun 2020) with the following settings: -t 15 -p 33 -q 20 -s 36 -m 0.5. Reads were mapped to the reference genome of *C. cryptica* (Roberts et al. 2020) using STAR v2.7.3a with settings '--alignIntronMin 1 --alignIntronMax 22618' to account for the size distribution of annotated introns (Dobin and Gingeras 2015). Gene-level counts were quantified from uniquely mapped reads with HTSeq v0.11.3 in *union* mode (Anders et al. 2010).

Trimmed mean of M-values (TMM) normalization and differential expression analysis were conducted using the Bioconductor package edgeR v3.30.3 (Robinson et al. 2010) Only genes with at least one read count per million (CPM) in at least three samples included in the analysis. Differential expression in edgeR used the quasi-likelihood (QL) model (glmQLFit) with a group-model design that included each replicate–timepoint combination and a Benjamini-Hochberg false discovery rate (FDR) adjusted p-value cutoff of 5% (Lund et al. 2012). Each time point was contrasted against the 0 min ASW 0 treatment. To control for testing multiple hypotheses for each gene across each comparison (i.e., significant differential expression across multiple time points), we used stageR v1.14.0 (Van den Berge et al. 2017), which allowed us to control the gene-level FDR across all contrasts with a 5 % cutoff (Heller et al. 2009, Van den Berge et al. 2017).

We compared previously published RNA-seq data from *C. cryptica* CCMP332 after long-term (120 days) acclimation to ASW 0 (Nakov et al. 2020). This dataset included only one replicate for strain CCMP332, we followed recommendations by Robinson et al. (2010) for unreplicated data and used the edgeR exactTest function with an assigned dispersion of 0.16 to identify genes with differential expression when contrasting an unstressed control in ASW 24 vs. long-term acclimated growth in ASW 0. To further compensate for the lack of replication, we restricted our analyses to genes with a ≥ 1 $\log_2$-fold change cutoff.

Global similarity of gene expression patterns for each time point was assessed with metric multidimensional scaling (mMDS) of logCPM for the top 500 genes with largest standard deviations across the samples, using limma's plotMDS function (Ritchie et al. 2015). Hierarchical clustering was performed with Cluster v3.0 (Eisen et al. 1998) using uncentered Pearson's correlation as the similarity metric and centroid linkage. Gene ontology (GO) enrichment analyses were conducted using the elim algorithm and Fisher's exact test implemented in TopGO v2.36.0 (Alexa and Rahnenfuhrer 2010), with Bonferroni-corrected p-values ≤ 0.05 considered significant.

Functional gene annotations were based primarily on the published annotation of the reference genome (Roberts et al. 2020) but were augmented in some cases with NCBI-BLASTP (Altschul et al. 1990) searches against the Swissprot and Uniprot databases with a cutoff e-value of 1e-6. KEGG pathway annotations were obtained from the KofamKOALA web server on 2020-09-17 (Aramaki et al. 2020). Related GO terms were condensed using the online server REVIGO (Supek et al. 2011) based on a 0.5 similarity threshold using the SimRel algorithm.

5.4 Results and Discussion

5.4.1 Hypoosmotic stress causes slower growth *C. cryptica*. To understand the physiological and transcriptomic effects of acute hypoosmotic stress, we transferred the diatom *C. cryptica* from its native brackish (ASW 24) water—without any acclimation—to freshwater, mimicking a dispersal event or rapid environmental fluctuation. We measured cell division in the hours and days following exposure to freshwater and used hierarchical clustering to characterize patterns of gene expression associated with the short-term response to acute hypoosmotic stress. Observed growth rates were similar to previous estimates for CCMP332 (Nakov et al. 2020), i.e., *C. cryptica* divided more rapidly in its native brackish water (ASW 24) compared to the freshwater (ASW 0) treatment over the course of our 7-day experiment (Fig. 1). Growth was temporarily halted for 4 h, but recovered by 10 h with the culture reaching a new

steady-state of exponential growth by 12 h (Fig 1). Transient arrest of the cell cycle during stress is common in eukaryotes (Nitta et al. 1997, West et al. 2004, Skirycz et al. 2011, Seaton and Krishnan 2016), and likely provides time to reestablish homeostasis before resuming growth (Shin et al. 1987, Escoté et al. 2004, Clotet et al. 2006). Consistent with this hypothesis, changes in gene expression also peaked within the period of arrested growth.

5.4.2 Hypoosmotic stress causes transcriptional remodeling in C. cryptica. We used RNA-seq to measure changes in gene expression in the 10 hrs following exposure to freshwater. Of the 21,250 predicted genes in the *C. cryptica* genome, 12,939 (61%) were expressed (defined as >1 cpm in at least 3 samples), and among the expressed genes, a large 10,566 (82%) of them were differentially expressed for at least one time point. This corresponds to roughly half of all genes in the genome, highlighting the large magnitude of transcriptional remodeling that occurred in response to hyposalinity stress. A large fraction (31%) of differentially expressed genes were induced or repressed less than 1.5-fold, which we interpret as hyposalinity stress causing small but reproducible downstream effects on many aspects of *C. cryptica* physiology. For example, at 30 min there are a large number of genes involved in the oxidation reduction and regulation of active potassium channels that are upregulated to a low degree, as well as high volume of downregulated genes associated with translation and protein transport.

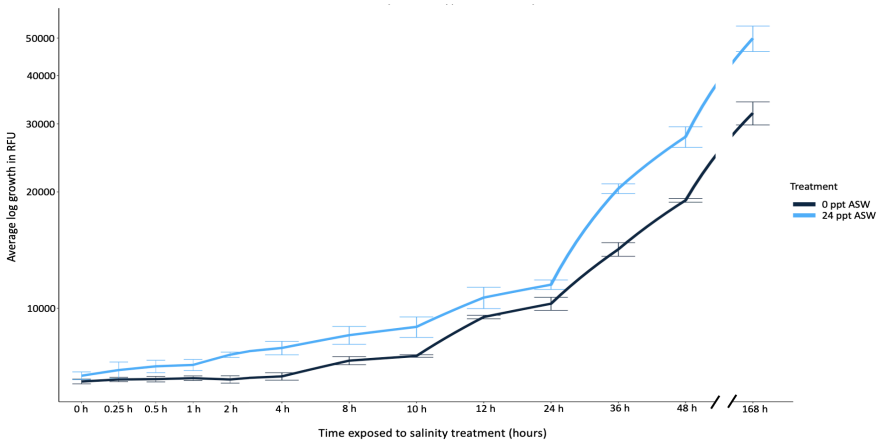
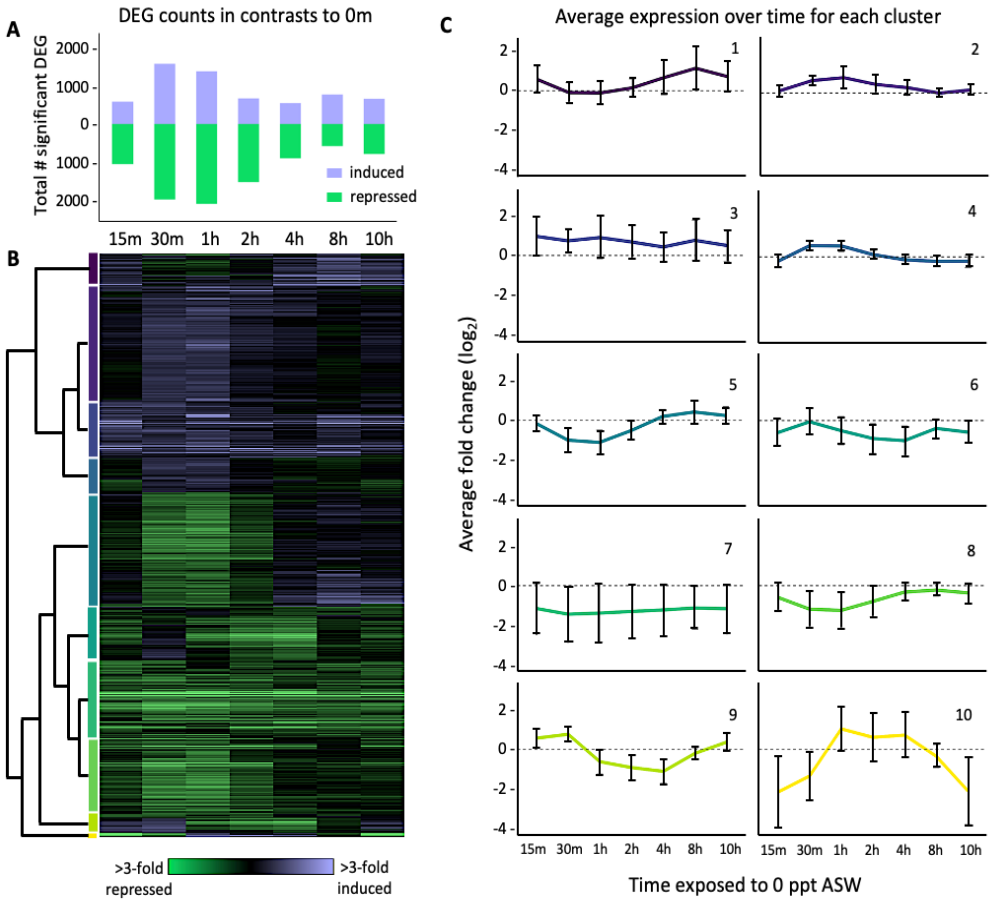


Fig 5.1. Growth of *C. cryptica* in freshwater (ASW 0) and its native brackish water (ASW 24) over a seven day period. Growth was measured from chlorophyll *a* fluorescence and is reported in relative fluorescence units (RFU).

With the rationale that genes most critical to the stress response should have consistent patterns of differential expression across multiple time points, we narrowed our focus to genes with significant differential expression in the same direction for ≥ 2 consecutive time points. 4,298 genes (41%) met these criteria and were included in all downstream analyses. We were first interested in understanding the dynamics of the transcriptional response to hypo-saline stress, and whether those dynamics are related to the effects of the stress on growth rate. The largest numbers of differentially expressed genes, and the largest magnitude of expression changes, occurred at the 30 and 60 minute time points (Fig 2), suggesting that the peak of the transcriptional response occurs within this time frame. Additionally, an MDS plot comparing the overall similarity of expression profiles at each time point showed that the 30 and 60 minute samples were the most dissimilar to the 0 minute unstressed control (Supplemental Fig 1). The peak



D

Cluster	No. genes	Enriched GO terms p < 0.01
1	243	rRNA processing^{BP}, DNA-binding transcription factor activity^{MF}
2	849	Photosynthesis (light harvesting)^{BP}, metal ion transport^{BP}, solute:proton antiporter activity^{MF}, violaxanthin de-epoxidase activity^{MF}, oxidoreductase activity^{MF}
3	410	Transcription regulation^{BP}, sulfur compound synthesis^{BP}, iron ion binding^{MF}
4	252	Cellular amino acid catabolism^{BP}, inward potassium channel activity^{MF}, oxidoreductase activity^{MF}
5	848	Ribosome biogenesis^{BP}, translation initiation^{BP}, GTP binding^{MF}
6	494	Carboxylic acid catabolism^{BP}, oxidoreductase activity^{MF}, cobalamin binding^{MF}, iron ion binding^{MF}
7	465	Macromolecule localization^{BP}, negative regulation of biological processes^{BP}, NADP binding^{MF}, endonuclease activity^{MF}
8	590	TCA cycle^{BP} , mRNA transport^{BP}, ATP synthesis coupled proton transport^{BP}, RNA binding^{MF}, GTP binding^{MF}
9	108	Carbohydrate metabolism^{BP}, chitin binding^{MF}
10	29	Photosynthesis (light harvesting)^{BP}, ribonuclease III activity^{MF}, ferric iron binding^{MF}

Fig 5.2. A) Total number of up- and downregulated differentially expressed genes at each time point. B) Heatmap and dendrogram showing the ten clusters of genes with similar expression patterns. Clusters are denoted with color-coded lines. C) Graphs depicting the average log_2-fold change of each cluster over time with error bars representing a 95% confidence interval. D) Table showing the number of differentially expressed genes for each cluster and significantly enriched GO terms (p-value < 0.01).

transcriptome remodeling at 30–60 minutes gradually returned to near pre-stressed levels by 4-10 hrs (Fig 2B, Supplemental Fig 1). That 4-10h time period coincides with the time period of resumption of growth (Fig 1), suggesting that the short-term acute transcriptional response to hypo-saline stress potentiates growth acclimation (see below).

Processes differentially regulated in response to hyposaline stress. We next sought to understand the major biological processes that were differentially regulated in response to hyposaline stress. We first focused on the major enriched GO categories ($p < 0.05$) for genes with LFC ≥ ±1 at the 30-minute time point closest to the peak of the response. Two of the top upregulated enriched functions at 30 min were related to 'DNA-binding transcription factor activity' and 'sequence specific DNA binding' (Fig 3B). Most genes belonging to these classes were either transcription factor proteins or heat shock factor proteins (Supplemental Table 2). These genes are known to respond to most stressors in many species, including humans, plants, and diatoms (Watanabe and Tanaka 2018, Strauch and Haslbeck 2018, Ghosh et al. 2018). Second, 'hydrolase activity' was enriched at 30 min, and all differentially expressed genes with this GO term were involved in carbohydrate metabolic processes and carbohydrate binding. Downregulation of 'RNA binding' activity at 30 min includes genes involved in ribosomal structure and intramolecular transferase activity (Supplemental Table 2), comprising 54 large and small ribosomal subunits, and numerous genes involved in the synthesis of pseudouridine, a protein that serves a stabilizing function in tRNA and rRNA biosynthesis (Carlile et al. 2014). In addition, the enriched 'NADP binding' GO term includes genes such as homoserine dehydrogenase which plays a role in amino acid biosynthesis (Bromke 2013), and several flavin-dependent monooxygenases that can assist with hormone production as well defense against oxidative stress (Huijbers et al. 2014).

We then shifted our focus to the genes involved in the initial stages of acclimation, signified by cell division recovering shortly after the 10 h time point. Upregulated functions

unique to 10 h include ‘protein serine/threonine kinase activity’ and ‘regulation of metabolic processes’ (Fig 3D). The latter classification contains nearly all proteins associated with regulation of cellular activity. At 10 h, a wide range of such regulatory genes were upregulated, including genes associated with rRNA processing, signal transduction, translation, and cell division (Supplemental Table 2). All three genes responsible for the enrichment of protein serine/threonine kinase activity were RIO kinases 1 and 2 (Supplemental Table 2), which are classified as regulators of rRNA processing. Additionally, RIO kinases 1 and 2 are both known kinase regulators of ribosome biogenesis and the cell cycle (LaRonde-LeBlanc and Wlodawer 2005).

Several genes downregulated at 10 h were associated with ‘iron ion binding’ activity, which includes genes related to L-ascorbic acid binding and regulation of oxidative stress. The repressed genes classified as ‘vitamin binding’ were primarily transaminases involved in catalyzing the conversion of fructose-6P to glucosamine-6P, which represent the first step in the chitin pathway (Supplemental Table 2, Fig 4). Enrichment of ‘peptide transport’ resulted from the downregulation of several solute carriers involved in metal ion binding, and enrichment of ‘nitrogen compound transporters’ was related to the downregulation of several vesicle-mediated protein transporters (Supplemental Table 2).

A distinctly different set of genes was involved at the peak response (30 min) and versus onset of acclimation (10 h) (Fig 3A,C). Specifically, the majority of differentially expressed genes with peak expression levels at the 30 min time point (922 genes) returned to expression levels near those of the initial steady state by 10 h. Furthermore, the majority of the 200 differentially expressed genes with peak expression at 10 h displayed reversed expression patterns at 30 min (i.e., genes that were upregulated at 30 min become repressed by 10 h and vice versa), indicating that very few of the genes involved in the immediate salt stress response played the same role in the early stages of acclimation. Despite these differences in gene-expression profiles and number of differentially expressed genes during peak period of stress and the final

time point, there is some overlap of differentially expressed genes between the two. Of the 1,942 repressed genes at 30 min, 381 were still significantly repressed at 10 h, while 260/1555 of the genes induced at 30 min remained so at 10 h.

There is minimal overlap between the 30 min and 10 h time points regarding GO terms, with only 'ribosome biogenesis' and 'carbohydrate binding' being shared. Specifically for ribosome biogenesis, genes with this GO term show opposite expression patterns in both time points, and enrichment of this GO term is caused by a different set of genes at both time points (Supplemental Fig 2). The initial repression of ribosome biogenesis is likely related to the lag in cell growth in the first several hours of exposure, after which ribosome biosynthesis is increased prior to resumption of cell division (Thomas 2000). In contrast, nearly 50 % of the genes driving the enrichment of carbohydrate binding activity are shared between the two time points. The majority of these genes are involved in chitin biosynthesis (Fig 4A).

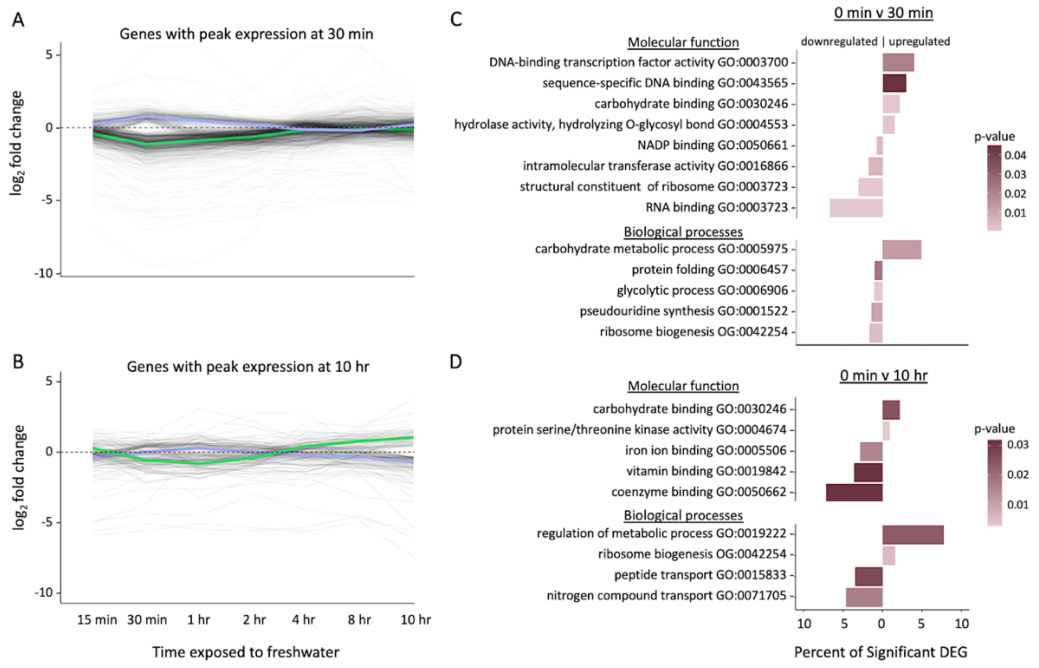


Fig 5.3. A) Plot of gene expression over time for 922 genes that have the highest magnitude of change in expression at 30 min. Solid black lines represent individual genes. Purple solid line is median expression of genes upregulated at 30 min. Green solid line is median expression of genes downregulated at 30 min. B) Plot of gene expression over time for 200 genes that have the highest change in expression at 10 h. Functional enrichments for significantly differentially expressed genes with at least a 2-fold change in expression at 30 min (C) and 10 h (D).

Cyclotella species, including *C. cryptica*, produce β-chitin, which is extruded as long crystallized threads around the margin of the cell (McLachlan and Craigie 1966, LeDuff and Rorrer 2019). In a previous study, changes in salinity triggered morphological changes of *C. cryptica* frustules (Schultz 1971), which suggests that increased expression of genes involved in chitin metabolism during the peak period of stress at 30 min to 1 h and the beginning of acclimation at 10 h could be related to salinity-induced alteration in the cytoskeletal structure of the cell.

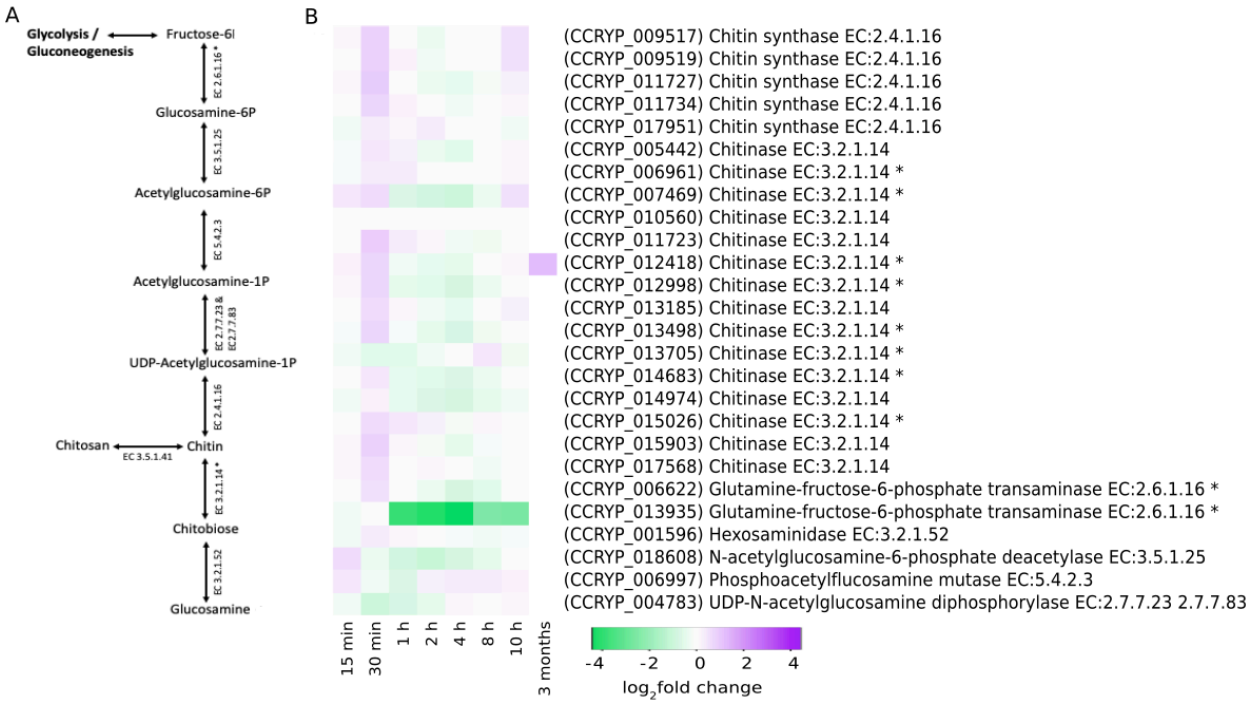


Fig 5.4. A) Hypothetical *C. cryptica* chitin biosynthesis pathway based on Traller et al. (2016). B) Heatmap of chitin biosynthesis genes. Asterisk (*) indicates significant differential expression. Gene names in parentheses.

The expression dynamics of metabolic pathways during hyposaline stress. While the overall peak of the hyposaline response occurred between 30-60 minutes, hierarchical clustering revealed groups of genes with distinct sub-dynamics that deviated from the group average. We manually collapsed the gene-tree dendrogram into 10 clusters with distinct expression profiles

(Fig 2B, C). GO enrichment of hierarchically clustered genes revealed that, in several cases, groups enriched for similar functional processes displayed similar patterns of differential expression (Fig 2B, C, D). For example, clusters enriched for 'ion transporter activity' and 'oxidoreductase activity' (Fig 2; clusters 2, 4, and 6) were generally upregulated at 30 min to 1 h and downregulated from 2 h onward. Conversely, clusters enriched for genes involved in transcription, translation, and ribosome biogenesis (Fig 2; clusters 1, 5, and 8) displayed lower expression from 15 min to 1 h followed by higher]expression from 2 h to 10 h (Supplemental Fig 2). In addition, cluster 8, which showed decreased expression from 15 to 4 h, was enriched for two major energy metabolism processes: the tricarboxylic acid (TCA) cycle and ATP synthesis-coupled proton transport. Notably, the expression patterns of all 6 of these clusters (i.e., 1, 2, 4, 5, 6, and 8) coincide with the lag in cell division from 0–4 h (Fig 1). Like other bacteria and eukaryotes (Warner 1999, Sharfstein et al. 2007, Rojas et al. 2014), diatoms also halt cell division in the early stages of osmotic shock until ionic and osmotic equilibrium are restored. During such conditions, energy is likely redirected to cellular processes essential in acute stress management, rather than routine transcription, translation and other typical cellular processes (Albert et al. 2019).

Despite only being enriched at the peak period of hyposalinity stress response, membrane transporters and channels were highly active throughout the 10 h of freshwater exposure, suggesting a continuous role in the initial acclimation of *C. cryptica* to freshwater conditions. Throughout the experiment, the electrochemical gradient was restored and subsequently maintained through differential regulation of numerous amino acid and ion transporters. This included five potassium channels, three potassium antiporters, three sodium/hydrogen exchangers, and two general solute:proton antiporters (Fig 5). Although none were significantly differentially expressed at two time points, we also examined regulation of aquaporins as these allow for rapid water and gas exchange in many organisms (Jahn et al. 2004, Butler et al. 2020). We identified several likely aquaporins, but only one was strongly

induced at 30 min. This may be because mitigation of the initial rapid influx of water was also addressed by continued maintenance of the electrochemical gradient.

Like many other species, diatoms utilize low-molecular weight osmolytes such as DMSP, taurine, glycine-betaine, and proline to balance osmotic pressure in response to changes in salinity in both the short and long term. Previous studies have found DMSP involved in osmotic stress responses in both hypo- and hypersaline environments in multiple diatom species (Liu and Hellebust 1976, Trevena et al. 2000, Lyon et al. 2011, Nakov et al. 2020). Of the four dimethylglycine methyltransferases found in the *C. cryptica* genome, none of the four were a match for the *Thalassiosira* methyltransferase differentially expressed in *C. cryptica* post-acclimation to salinity change (Supplemental Table 2, Nakov et al. 2020). Two of the methyltransferases were differentially expressed, displaying sharply contrasting profiles with one strongly downregulated for all 10 h of freshwater exposure and the other only mildly induced at 30 min and 1 h, then downregulated at all other time points (Fig 5). These methyltransferases catalyze reactions in both the DMSP and glycine-betaine metabolic pathways, which are known to be repressed in hyposaline conditions for *Skeletonema costatum*, *C. cryptica*, *Thalassiosira weissflogii*, and *S. marinoi* (Speeckaert et al. 2019, Nakov et al. 2020, Theresiera et al. 2020, Pinseel et al. In review).

Genes involved in taurine biosynthesis showed a similar pattern to that of the methyltransferases (Fig 5). Gamma-glutamyltransferase, which catalyzes a reaction that reduces free taurine in the cytosol, was downregulated during the peak period of salinity stress. Glutamate decarboxylase, which acts as the catalyst for the one-way synthesis of hypotaurine from 3-sulfino-L-alanine and taurine from L-cysteate, is heavily downregulated for the entire experiment. Meanwhile, cysteine sulfinic acid decarboxylase facilitates the same reaction and was mildly upregulated at all time points but 10 h. The combination of highly repressed genes and mildly induced genes in the pathways of these three compounds during the hyposaline treatments suggests *C. cryptica* heavily regulated intracellular levels of these osmolytes but

refrained from completely depleting them. *C. cryptica* is also known to decrease free cytosolic proline concentrations in the initial stages of hypoosmotic stress and carbon-labeling found it was incorporated into proteins, rather than being expelled into surrounding media (Liu and Hellebust 1976, Liu and Hellebust 1974). We found transcript evidence of that regulation here as *C. cryptica* ceases the synthesis of proline in response to hyposaline stress. Proline synthesis was actively repressed as the limiting steps of glutamate conversion from L-glutamyl-phosphate to L-glutamate 5-semialdehyde, and the breakdown of peptides into proline are both downregulated at 15 min to 2 h (Fig 5). In contrast, proline was not differentially regulated in the long-term experiment (Nakov et al. 2020). This could be due to the important role of proline in amino acid synthesis, making it a costly molecule to maintain in the cytosol long-term (Liu and Hellebust 1974).

Oxidoreductase activity was enriched in clusters 2, 4, and 6 (Fig 2D), giving the initial impression that *C. cryptica* experienced increased oxidative stress during the peak period of low salinity stress, which may have also contributed to the growth lag experienced by the cells in freshwater. However, on closer inspection, there is no clear indication for a strong cellular response to oxidative stress. Although a number of genes involved in reactive oxygen species scavenging, such as superoxide dismutase and violaxanthin de-epoxidase, are upregulated, many others, such as several glutaredoxins and thioredoxins, are not differentially regulated at all (Supplemental Fig 3).

Within the first hour of freshwater exposure, *C. cryptica* redirected the majority of cellular effort towards mitigating ionic stress via a combination of *Na+/K+:H+* and *Ca+:cation* exchangers and antiporters found in the chloroplast and cell membranes (Fig 5). The expression behavior of genes associated with metal ion binding, photosynthesis (light harvesting), and oxidoreductase activity (Supplemental Fig 3) suggests that the cell may have experienced increased oxidative stress within the first hour as well. However, without experimental verification of the presence of reactive oxygen species, this cannot be confirmed

as the activity of genes involved in oxidative stress response are often correlated with the activity of osmotic stress response genes (Mager et al. 2000).

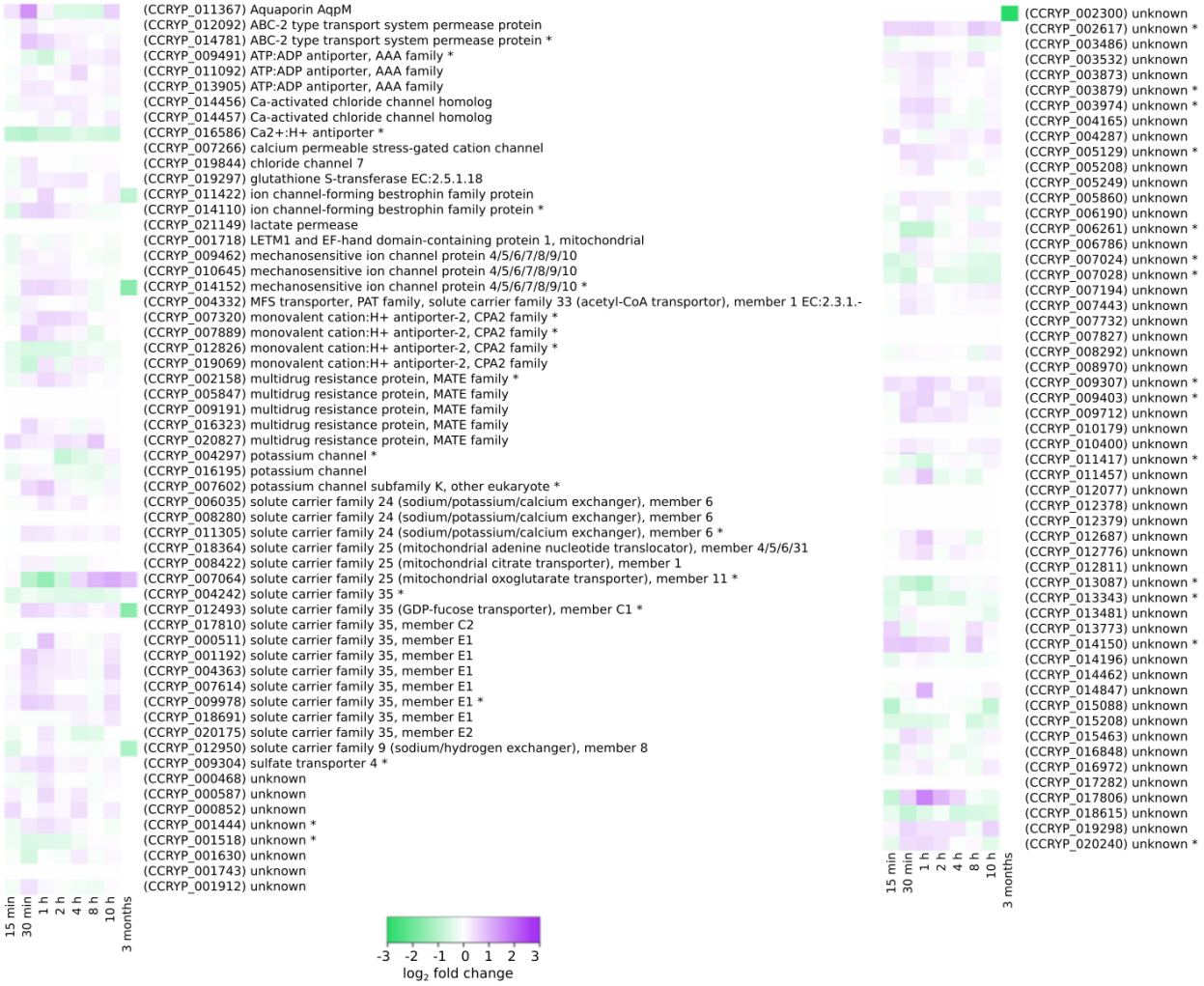


Fig 5.5. Transporters, channels, aquaporins, and osmolyte biosynthesis genes. Asterisk (*) indicates significant differential expression. Gene names in parentheses.

The cell halted energy production during the first hour. This was evidenced by downregulation of the TCA cycle, ATP synthesis-coupled proton transport in the photosynthetic electron transport chain, and seven of the eight ATPase complex subunits (Fig 6A-C). The majority of the genes involved in the glycolysis/gluconeogenesis pathway were downregulated at the peak period of salinity stress, including the primary enzymes responsible for the production of ATP (Fig 7A, B). In the absence of ATP production, transcription, translation, and

cell division were temporarily disrupted until the cell was able to achieve homeostasis again. As previously mentioned, the repression of ribosome biogenesis is closely related to the disruption of cell division as many organisms possess a narrow ribosome concentration within the cell (Warner 1999).

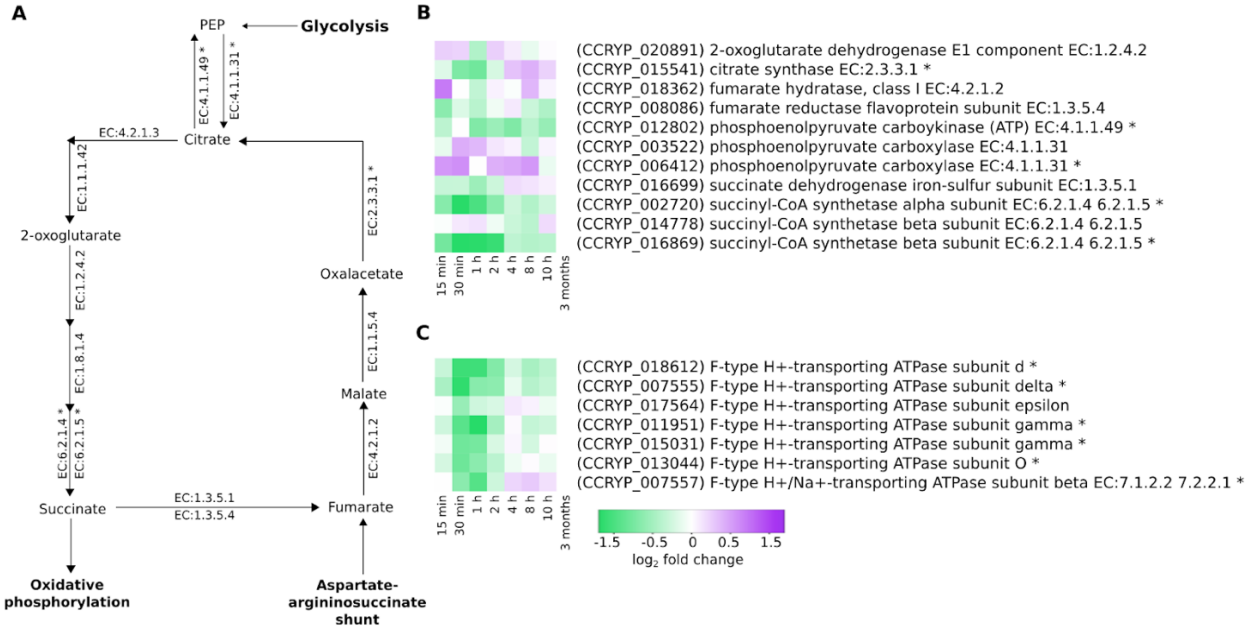


Fig 5.6. A) TCA cycle. B) Genes in the TCA cycle. C) Components of the ATPase synthase complex. Asterisks (*) denote significant differentially expressed genes. Gene names are in parentheses.

A closer look at the glycolytic/gluconeogenesis pathway shows the significant upregulated of 2 copies of fructose 1-6 bisphosphate, which catalyzes a unidirectional reaction that produces fructose 6-phosphate in gluconeogenesis, implying gluconeogenesis is upregulated. However, fructose 6-phosphate is also a necessary component in both chitin biosynthesis and the Calvin cycle (Fig 5A, 8A). Given that all differentially expressed genes in the Calvin cycle are also found in glycolysis, both processes are likely repressed at the peak point of salinity stress (Fig 8B). Though most enzymes in the chitin biosynthetic pathway are briefly upregulated in the first 30 min of hypoosmotic stress, there is a subsequent downregulation until 10 h. Thus, the usage of fructose 6-phosphate in this pathway may not

account for the upregulation of fructose 1,6-bisphosphatase for the duration of the experiment. It is possible that the cell is stockpiling fructose-6P in the cytosol, perhaps until the cell acclimates to freshwater or the salinity increases, for conversion to glucose-6-phosphate and entry into the pentose phosphate pathway to produce the nucleotides necessary for cell cycle transitions (Abbriano et al. 2018).

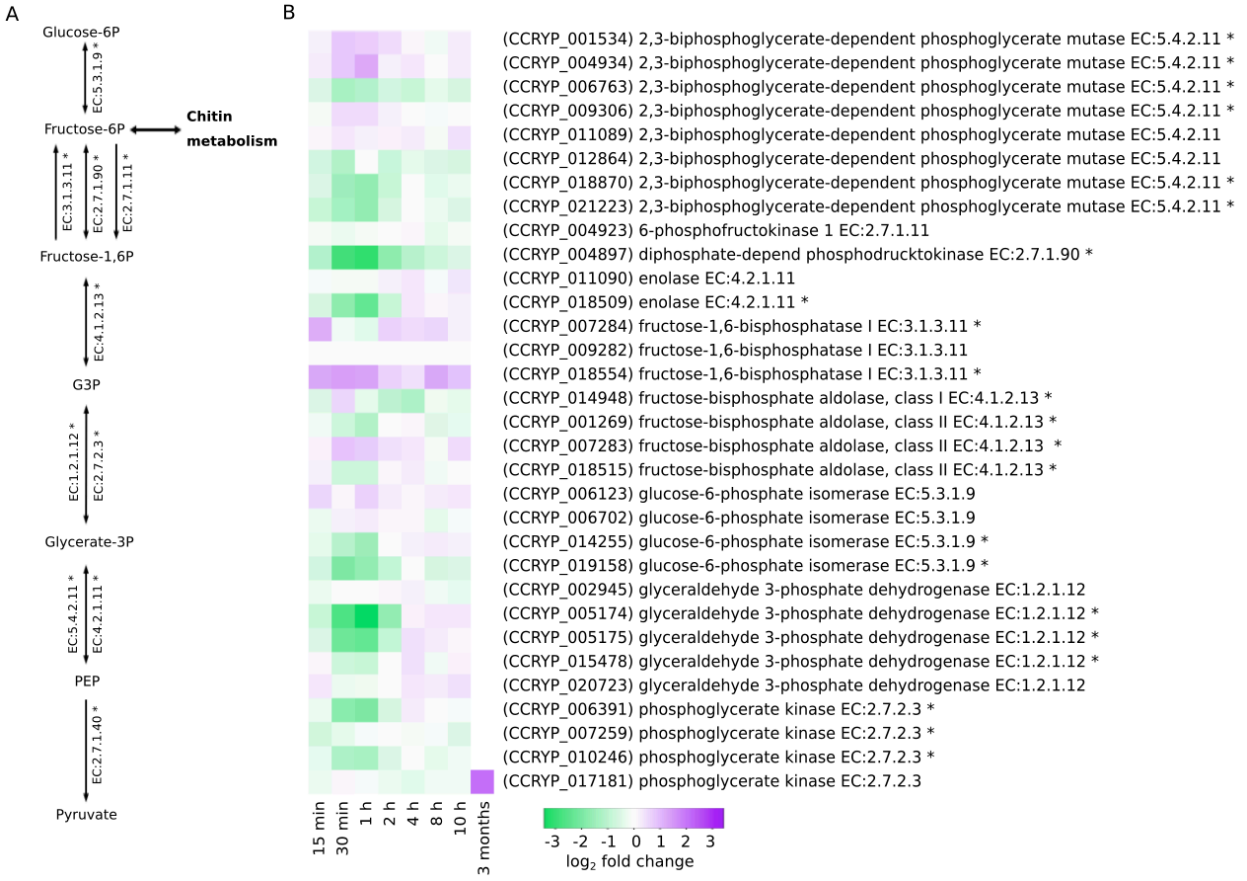


Fig 5.7. A) Glycolytic pathway. B) Heatmap of glycolysis/gluconeogenesis genes. Asterisk (*) indicates significant differential expression. Gene names are in parentheses.

We looked into paralogs because the pattern seen in the glycolytic pathway, the Calvin-Benson cycle, and chitin biosynthesis was that if there were multiple copies of a gene, only a few are differentially regulated in response to salt stress as is a common trend in many eukaryotes (Hericourt et al. 2016, Espinola et al. 2018, Noree et al. 2019, Savic et al. 2019). To determine if the pattern we were seeing was significant, we compared the number of paralogs

that occurred in a pair to artificially assigned pairs from genes with only a single copy. We found that roughly 42% of the time artificial pairs had a significant and a nonsignificant gene. In contrast, only 13% of actual paralog pairs had one significant and one nonsignificant gene.

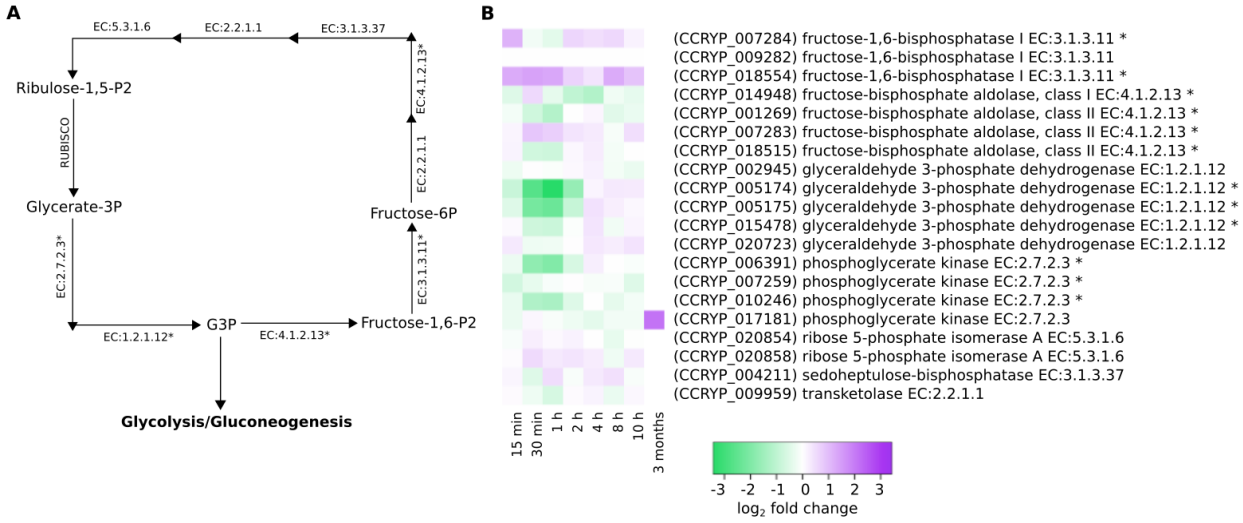


Fig 5.8. A) Calvin-Benson cycle. B) Heatmap of genes in Calvin-Benson cycle. Asterisk (*) indicates significant differential expression. Gene name is in parentheses.

Comparison of the short-term hyposalinity response to long-term acclimation. To understand whether the genes that were important for the acute hyposaline response were also important to long-term acclimation, we re-analyzed previously published RNA-seq data from Nakov et al. That dataset included the same *C. cryptica* strain used in this study (CCMP332) grown in freshwater for 120 days.

Notably, there is no overlap of enriched terms between datasets for long-term acclimation and short-term stress experiments. Of the 1,220 differentially expressed genes in the long-term dataset that met the $\log_2$ fold change cutoff of an absolute value ≥1, only 179 were also differentially expressed in the short-term stress experiment. While this is a highly significant overlap, it accounts for only 5.6% of the significantly differentially expressed genes we investigated. While enrichment analysis of the unreplicated data should be taken with a grain of salt, some patterns are worth noting. Phosphorelay signal transduction is repressed and linked

to several strongly repressed genes involved in histidine kinase pathways, which are used by both prokaryotic and eukaryotic organisms to detect and respond to a broad range of environmental cues (Abriata et al. 2017). Enriched upregulated terms are related to DNA replication and methylation as well as mRNA splicing, which may be an indication that cells in the culture are preparing for cell division. Taking into account genes that are significantly differentially expressed at the peak short-term stress response and the long term data, there is a concentration of genes involved in photosynthetic processes and the catalysis of small ubiquitin-like proteins, while repressed genes are primarily associated with carbohydrate binding. Unfortunately, most of these shared genes are currently uncharacterized so little can be determined about the relevance of their presence in both the short- and long-term study datasets. This result is of particular interest when we consider how stress results are typically interpreted. In many studies only one timescale is taken into consideration and generalizations are made based on those results. the results that we see here highlight the disparity between genes required for prolonged exposure and persistence versus those work required to mitigate immediate stress factors.

5.5 References

Abbriano, Raffaela, Nurcan Vardar, Daniel Yee, and Mark Hildebrand. 2018. Manipulation of a Glycolytic Regulator Alters Growth and Carbon Partitioning in the Marine Diatom Thalassiosira Pseudonana. Algal Research 32 (June): 250–58.

Abriata, Luciano A., Daniela Albanesi, Matteo Dal Peraro, and Diego de Mendoza. 2017. Signal Sensing and Transduction by Histidine Kinases as Unveiled through Studies on a Temperature Sensor. Accounts of Chemical Research 50 (6): 1359–66.

Albert, Benjamin, Isabelle C. Kos-Braun, Anthony K. Henras, Christophe Dez, Maria Paula Rueda, Xu Zhang, Olivier Gadal, Martin Kos, and David Shore. 2019. A Ribosome Assembly Stress Response Regulates Transcription to Maintain Proteome Homeostasis. eLife 8 (May). https://doi.org/10.7554/eLife.45002.

Alexa, Adrian, and Jorg Rahnenfuhrer. 2010. topGO: Enrichment Analysis for Gene Ontology. R Package Version 2 (0): 2010.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic Local Alignment Search Tool. Journal of Molecular Biology 215 (3): 403–10.

Anders, Simon, P. T. Pyl, and W. Huber. 2010. HTSeq: Analysing High-Throughput Sequencing Data with Python.

Andrew, S. 2019. FastQC: A Quality Control Tool for High Throughput Sequence Data 2010 [Available from: Http://www. Bioinformatics. Babraham. Ac. Uk/projects/fastqc]. Accessed.

Aramaki, Takuya, Romain Blanc-Mathieu, Hisashi Endo, Koichi Ohkubo, Minoru Kanehisa, Susumu Goto, and Hiroyuki Ogata. 2020. KofamKOALA: KEGG Ortholog Assignment Based on Profile HMM and Adaptive Score Threshold. Bioinformatics 36 (7): 2251–52.

Balzano, S., E. Abs, and S. C. Leterme. 2015. Protist Diversity along a Salinity Gradient in a Coastal Lagoon. Aquatic Microbial Ecology: International Journal 74 (3): 263–77.

Borowitzka, Michael A. 2018. The ‘stress’ Concept in Microalgal Biology—homeostasis, Acclimation and Adaptation. Journal of Applied Phycology 30 (5): 2815–25.

Bromke, Mariusz A. 2013. Amino Acid Biosynthesis Pathways in Diatoms. Metabolites 3 (2): 294–311.

Butler, Thomas, Rahul Vijay Kapoore, and Seetharaman Vaidyanathan. 2020. Phaeodactylum Tricornutum: A Diatom Cell Factory. Trends in Biotechnology 38 (6): 606–22.

Carlile, Thomas M., Maria F. Rojas-Duran, Boris Zinshteyn, Hakyung Shin, Kristen M. Bartoli, and Wendy V. Gilbert. 2014. Pseudouridine Profiling Reveals Regulated mRNA Pseudouridylation in Yeast and Human Cells. Nature 515 (7525): 143–46.

Cheng, Ruo-Lin, Jia Feng, Bing-Xin Zhang, Yun Huang, Jun Cheng, and Chuan-Xi Zhang. 2014. Transcriptome and Gene Expression Analysis of an Oleaginous Diatom Under Different Salinity Conditions. Bioenergy Research 7 (1): 192–205.

Chen, Guan-Qun, Yue Jiang, and Feng Chen. 2008. Salt-induced alterations in lipid composition of diatom nitzschia laevis (bacillariophyceae) under heterotrophic culture condition(1). Journal of Phycology 44 (5): 1309–14.

Clotet, Josep, Xavier Escoté, Miquel Àngel Adrover, Gilad Yaakov, Eloi Garí, Martí Aldea, Eulàlia de Nadal, and Francesc Posas. 2006. Phosphorylation of Hsl1 by Hog1 Leads to a G2 Arrest Essential for Cell Survival at High Osmolarity. The EMBO Journal. https://doi.org/10.1038/sj.emboj.7601095.

Conley, Daniel J., Susan S. Kilham, and Edward Theriot. 1989. Differences in Silica Content between Marine and Freshwater Diatoms. Limnology and Oceanography 34 (1): 205–12.

Dickson, D. M. J., and G. O. Kirst. 1987. Osmotic adjustment in marine eukaryotic algae: the role of inorganic ions, quaternary ammonium, tertiary sulphonium and carbohydrate solutes. New Phytologist. https://doi.org/10.1111/j.1469-8137.1987.tb00165.x.

Dobin, Alexander, and Thomas R. Gingeras. 2015. Mapping RNA-Seq Reads with STAR. Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.] 51 (September): 11.14.1–11.14.19.

Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein. 1998. Cluster Analysis and Display of Genome-Wide Expression Patterns. Proceedings of the National Academy of Sciences of the United States of America 95 (25): 14863–68.

Escoté, Xavier, Meritxell Zapater, Josep Clotet, and Francesc Posas. 2004. Hog1 Mediates Cell-Cycle Arrest in G1 Phase by the Dual Targeting of Sic1. Nature Cell Biology 6 (10): 997–1002.

Espinola, Sergio Martin, Martin Pablo Cancela, Lauís Brisolara Corrêa, and Arnaldo Zaha. 2018. Evolutionary Fates of Universal Stress Protein Paralogs in Platyhelminthes. BMC Evolutionary Biology. https://doi.org/10.1186/s12862-018-1129-x.

Ghosh, Sumit, Poulami Sarkar, Priyanka Basak, Sushweta Mahalanobish, and Parames C. Sil. 2018. Role of Heat Shock Proteins in Oxidative Stress and Stress Tolerance. Heat Shock Proteins and Stress. https://doi.org/10.1007/978-3-319-90725-3_6.

Godhe, Anna, Anke Kremp, and Marina Montresor. 2014. Genetic and Microscopic Evidence for Sexual Reproduction in the Centric Diatom Skeletonema Marinoi. Protist 165 (4): 401–16.

Heller, Ruth, Elisabetta Manduchi, Gregory R. Grant, and Warren J. Ewens. 2009. A Flexible Two-Stage Procedure for Identifying Gene Sets That Are Differentially Expressed. Bioinformatics 25 (8): 1019–25.

Héricourt, François, Françoise Chefdor, Inès Djeghdir, Mélanie Larcher, Florent Lafontaine, Vincent Courdavault, Daniel Auguin, et al. 2016. Functional Divergence of Poplar Histidine-Aspartate Kinase HK1 Paralogs in Response to Osmotic Stress. International Journal of Molecular Sciences 17 (12). https://doi.org/10.3390/ijms17122061.

Ippolito, Giuliana d', Angela Sardo, Debora Paris, Filomena Monica Vella, Maria Grazia Adelfi, Pierpaolo Botte, Carmela Gallo, and Angelo Fontana. 2015. Potential of Lipid

Metabolism in Marine Diatoms for Biofuel Production. Biotechnology for Biofuels 8 (February): 28.

Jahn, Thomas P., Anders L. B. Møller, Thomas Zeuthen, Lars M. Holm, Dan A. Klaerke, Brigitte Mohsin, Werner Kühlbrandt, and Jan K. Schjoerring. 2004. Aquaporin Homologues in Plants and Mammals Transport Ammonia. FEBS Letters 574 (1-3): 31–36.

Jamy, Mahwash, Charlie Biwer, Daniel Vaulot, Aleix Obiol, Homgmei Jing, Sari Peura, Ramon Massana, and Fabien Burki. 2021. Global Patterns and Rates of Habitat Transitions across the Eukaryotic Tree of Life. bioRxiv. https://doi.org/10.1101/2021.11.01.466765.

Kageyama, Hakuto, Yoshito Tanaka, and Teruhiro Takabe. 2018. Biosynthetic Pathways of Glycinebetaine in Thalassiosira Pseudonana; Functional Characterization of Enzyme Catalyzing Three-Step Methylation of Glycine. Plant Physiology and Biochemistry: PPB / Societe Francaise de Physiologie Vegetale 127 (June): 248–55.

Kirst, G. O. 1990. Salinity Tolerance of Eukaryotic Marine Algae. Annual Review of Plant Physiology and Plant Molecular Biology 41 (1): 21–53.

Krell, Andreas, Bánk Beszteri, Gerhard Dieckmann, Gernot Glöckner, Klaus Valentin, and Thomas Mock. 2008. A New Class of Ice-Binding Proteins Discovered in a Salt-Stress-Induced cDNA Library of the Psychrophilic Diatom Fragilariopsis Cylindrus (Bacillariophyceae). European Journal of Phycology 43 (4): 423–33.

Krell, Andreas, Dietmar Funck, Ina Plettner, Uwe John, and Gerhard Dieckmann. 2007. Regulation of Proline Metabolism under Salt Stress in the Psychrophilic Diatomfragilariopsis Cylindrus(Bacillariophyceae)1. Journal of Phycology 43 (4): 753–62.

Liu, Ming Sai, and Johan A. Hellebust. 1976. Effects of Salinity and Osmolarity of the Medium on Amino Acid Metabolism in Cyclotella Cryptica. Canadian Journal of Botany. Journal Canadien de Botanique 54 (9): 938–48.

Liu, M. S., and J. A. Hellebust. 1974a. Uptake of Amino Acids by the Marine Centric Diatom Cyclotella Cryptica. Canadian Journal of Microbiology 20 (8): 1109–18.

———. 1974b. Utilization of Amino Acids as Nitrogen Sources, and Their Effects on Nitrate Reductase in the Marine Diatom Cyclotella Cryptica. Canadian Journal of Microbiology 20 (8): 1119–25.

Lund, Steven P., Dan Nettleton, Davis J. McCarthy, and Gordon K. Smyth. 2012. Detecting Differential Expression in RNA-Sequence Data Using Quasi-Likelihood with Shrunken Dispersion Estimates. Statistical Applications in Genetics and Molecular Biology 11 (5). https://doi.org/10.1515/1544-6115.1826.

Lyon, Barbara R., Jennifer M. Bennett-Mintz, Peter A. Lee, Michael G. Janech, and Giacomo R. DiTullio. 2016. Role of Dimethylsulfoniopropionate as an Osmoprotectant Following Gradual Salinity Shifts in the Sea-Ice Diatom Fragilariopsis Cylindrus. Environmental Chemistry 13 (2): 181–94.

Lyon, Barbara R., Peter A. Lee, Jennifer M. Bennett, Giacomo R. DiTullio, and Michael G. Janech. 2011. Proteomic Analysis of a Sea-Ice Diatom: Salinity Acclimation Provides

New Insight into the Dimethylsulfoniopropionate Production Pathway. Plant Physiology 157 (4): 1926–41.

Mager, Willem H., Albertus H. de Boer, Marco H. Siderius, and Hans-Peter Voss. 2000. Cellular Responses to Oxidative and Osmotic Stress. Cell Stress & Chaperones. https://doi.org/10.1379/1466-1268(2000)005<0073:crtoao>2.0.co;2.

Nakov, Teofil, Jeremy M. Beaulieu, and Andrew J. Alverson. 2018. Insights into Global Planktonic Diatom Diversity: The Importance of Comparisons between Phylogenetically Equivalent Units That Account for Time. The ISME Journal 12 (11): 2807–10.

Nakov, Teofil, Kathryn J. Judy, Kala M. Downey, Elizabeth C. Ruck, and Andrew J. Alverson. 2020. Transcriptional Response of Osmolyte Synthetic Pathways and Membrane Transporters in a Euryhaline Diatom during Long-Term Acclimation to a Salinity Gradient. Journal of Phycology. https://onlinelibrary.wiley.com/doi/abs/10.1111/jpy.13061?casa_token=WxPO1IBOAuoAAAAA:zcF3_ZZSO99Xcn-uy8dT5HVtuy9WnS1evYKF-fIWvqHBudKtOHEHuAHjpXe_81xkpHowxkbBIu1krg.

Nitta, Makoto, Hitoshi Okamura, Shinichi Aizawa, and Masaru Yamaizumi. 1997. Heat Shock Induces Transient p53-Dependent Cell Cycle Arrest at G1/S. Oncogene. https://doi.org/10.1038/sj.onc.1201210.

Noree, Chalongrat, Naraporn Sirinonthanawech, and James E. Wilhelm. 2019. Saccharomyces Cerevisiae ASN1 and ASN2 Are Asparagine Synthetase Paralogs That Have Diverged in Their Ability to Polymerize in Response to Nutrient Stress. Scientific Reports 9 (1): 278.

Pörtner, Hans-Otto, Debra C. Roberts, Valérie Masson-Delmotte, Panmao Zhai, Melinda Tignor, Elvira Poloczanska, and N. M. Weyer. 2019. The Ocean and Cryosphere in a Changing Climate. IPCC Special Report on the Ocean and Cryosphere in a Changing Climate. https://www.ipcc.ch/site/assets/uploads/sites/3/2019/12/02_SROCC_FM_FINAL.pdf.

Rijstenbil, J. W. 2005. UV- and Salinity-Induced Oxidative Effects in the Marine Diatom Cylindrotheca Closterium during Simulated Emersion. Marine Biology. https://doi.org/10.1007/s00227-005-0015-4.

Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. Nucleic Acids Research 43 (7): e47.

Roberts, Wade R., Kala M. Downey, Elizabeth C. Ruck, Jesse C. Traller, and Andrew J. Alverson. 2020. Improved Reference Genome for Cyclotella Cryptica CCMP332, a Model for Cell Wall Morphogenesis, Salinity Adaptation, and Lipid Production in Diatoms (Bacillariophyta). Cold Spring Harbor Laboratory. https://doi.org/10.1101/2020.05.19.103069.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2010. edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. Bioinformatics 26 (1): 139–40.

Rojas, Enrique, Julie A. Theriot, and Kerwyn Casey Huang. 2014. Response of Escherichia Coli Growth Rate to Osmotic Shock. Proceedings of the National Academy of Sciences of the United States of America 111 (21): 7807–12.

Savic, Neda, Shawn P. Shortill, Misha Bilenky, Joseph M. Dobbs, David Dilworth, Martin Hirst, and Christopher J. Nelson. 2019. Histone Chaperone Paralogs Have Redundant, Cooperative, and Divergent Functions in Yeast. Genetics. https://doi.org/10.1534/genetics.119.302235.

Scholz, Bettina, and Gerd Liebezeit. 2012. Compatible Solutes in Three Marine Intertidal Microphytobenthic Wadden Sea Diatoms Exposed to Different Salinities. European Journal of Phycology 47 (4): 393–407.

Schultz, Mary E. 1971. Salinity-Related Polymorphism in the Brackish-Water Diatom Cyclotella Cryptica. Canadian Journal of Botany. Journal Canadien de Botanique 49 (8): 1285–89.

Schultz, Mary E., and Francis R. Trainor. 1970. Production of Male Gametes and Auxospores in a Polymorphic Clone of the Centric Diatom Cyclotella. Canadian Journal of Botany. Journal Canadien de Botanique 48 (5): 947–51.

Seaton, Daniel D., and J. Krishnan. 2016. Model-Based Analysis of Cell Cycle Responses to Dynamically Changing Environments. PLOS Computational Biology. https://doi.org/10.1371/journal.pcbi.1004604.

Sharfstein, Susan T., Duan Shen, Thomas R. Kiehl, and Rui Zhou. 2007. Molecular Response to Osmotic Shock. Cell Engineering. https://doi.org/10.1007/1-4020-5252-9_7.

Shin, D. Y., K. Matsumoto, H. Iida, I. Uno, and T. Ishikawa. 1987. Heat Shock Response of Saccharomyces Cerevisiae Mutants Altered in Cyclic AMP-Dependent Protein Phosphorylation. Molecular and Cellular Biology 7 (1): 244–50.

Shu, Qi, Fangli Qiao, Zhenya Song, Jiechen Zhao, and Xinfang Li. 2018. Projected Freshening of the Arctic Ocean in the 21st Century. Journal of Geophysical Research, C: Oceans 123 (12): 9232–44.

Skirycz, Aleksandra, Hannes Claeys, Stefanie De Bodt, Akira Oikawa, Shoko Shinoda, Megan Andriankaja, Katrien Maleux, et al. 2011. Pause-and-Stop: The Effects of Osmotic Stress on Cell Proliferation during Early Leaf Development in Arabidopsis and a Role for Ethylene Signaling in Cell Cycle Arrest. The Plant Cell. https://doi.org/10.1105/tpc.111.084160.

Strauch, Annika, and Martin Haslbeck. 2018. Small Heat Shock Proteins in Stress Response of Higher Eukaryotes. Heat Shock Proteins and Stress. https://doi.org/10.1007/978-3-319-90725-3_14.

Sun, Kun. 2020. Ktrim: An Extra-Fast and Accurate Adapter- and Quality-Trimmer for Sequencing Data. Bioinformatics 36 (11): 3561–62.

Supek, Fran, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PloS One 6 (7): e21800.

Theseira, Alyson M., Daniel A. Nielsen, and Katherina Petrou. 2020. Uptake of Dimethylsulphoniopropionate (DMSP) Reduces Free Reactive Oxygen Species (ROS) during Late Exponential Growth in the Diatom Thalassiosira Weissflogii Grown under Three Salinities. Marine Biology. https://doi.org/10.1007/s00227-020-03744-4.

Thomas, G. 2000. An Encore for Ribosome Biogenesis in the Control of Cell Proliferation. Nature Cell Biology 2 (5): E71–72.

Traller, Jesse C., Shawn J. Cokus, David A. Lopez, Olga Gaidarenko, Sarah R. Smith, John P. McCrow, Sean D. Gallaher, et al. 2016. Genome and Methylome of the Oleaginous Diatom Cyclotella Cryptica Reveal Genetic Flexibility toward a High Lipid Phenotype. Biotechnology for Biofuels 9 (November): 258.

Trevena, A. J., G. B. Jones, S. W. Wright, and R. L. van den Enden. 2000. Profiles of DMSP, Algal Pigments, Nutrients and Salinity in Pack Ice from Eastern Antarctica. Journal of Sea Research 43 (3): 265–73.

Tyerman, S. D., C. M. Niemietz, and H. Bramley. 2002. Plant Aquaporins: Multifunctional Water and Solute Channels with Expanding Roles. Plant, Cell & Environment 25 (2): 173–94.

Van Bergeijk, Stef A., Claar Van der Zee, and Lucas J. Stal. 2003. Uptake and Excretion of Dimethylsulphoniopropionate Is Driven by Salinity Changes in the Marine Benthic Diatom Cylindrotheca Closterium. European Journal of Phycology 38 (4): 341–49.

Van den Berge, Koen, Charlotte Soneson, Mark D. Robinson, and Lieven Clement. 2017. stageR: A General Stage-Wise Method for Controlling the Gene-Level False Discovery Rate in Differential Expression and Differential Transcript Usage. Genome Biology 18 (1): 151.

Warner, J. R. 1999. The Economics of Ribosome Biosynthesis in Yeast. Trends in Biochemical Sciences 24 (11): 437–40.

Watanabe, Yoshihisa, and Masaki Tanaka. 2018. Regulation of Autophagy by the Heat Shock Factor 1-Mediated Stress Response Pathway. Heat Shock Proteins and Stress. https://doi.org/10.1007/978-3-319-90725-3_9.

West, Gerrit, Dirk Inzé, and Gerrit T. S. Beemster. 2004. Cell Cycle Modulation in the Response of the Primary Root of Arabidopsis to Salt Stress. Plant Physiology 135 (2): 1050–58.

Chapter 6 Conclusion

6.1 Summary of results

6.1.1. Reestablishing a monophyletic *Cyclotella* clade.

The order Thalassiosirales has a very complex taxonomic history with numerous reclassifications. In particular, the two largest genera, *Thalassiosira* and *Cyclotella*, have been split into polyphyletic clades more than once with the introduction of new species and genera that share many of the defining characteristics of these groups. This is often the case when the taxonomic classification does not agree with the phylogenetic relationship, as was the case with *Spicaticribra kingstonii.* This *S. kingstonii* was the type species for this genus that was defined by many characteristics that were shared by both *Thalassiosira* and *Cyclotella.* Using phylogenetic analyses of sequences for two nuclear (*SSU* and partial *LSU* rDNA) and two plastid (*rbcL* and *psbC*) genes, we found that *Spicaticribra kingstonii* shared a most recent ancestor with *Cyclotella nana* and fell within the *Cyclotella* genus. In the interest of maintaining a monophyletic clade, we reclassified *Spicaticribra* as *Cyclotella* and in doing so also moved nine other *Spicaticribra* species, several of which were identified from fossil specimen, that had been put in this genus based on morphological features. Those that are not extinct should be collected and sequenced with the same genes to verify this placement, but species that cannot be found and placed in living culture should remain as *Cyclotella*.

6.1.2 *Cyclotella cryptica* relies on ionic and osmotic regulation after acclimation to new salinity.

Understanding how wide spread euryhaline species, such as the diatom *Cyclotella cryptica*, are able to handle the effects of short-term salinity shock and long-term growth in suboptimal conditions (determined by growth rates) will provide insight into the species distribution and diversification across the salinity barrier between marine and freshwater. We

performed a ~120 day common garden experiment growing four strains of *C. cryptica* in the full range of naturally occurring salinity from freshwater to full oceanic saltwater.

Our initial analyses of transcriptomic data from these samples determined that the variation of gene expression between the strains masked any general response to salinity. To accommodate this, we pooled the strains for each salinity treatment and conducted analyses between low and high salinity comparisons. We found that following acclimation to suboptimal salinity conditions, *C. cryptica* appears to rely on regulation of DMSP, taurine, and glycine-betaine for sustained growth in both low and high salinity conditions. Additionally, we saw increased activity of many membrane transport systems, particularly potassium and sodium transporters. These results highlighted a disparity between previously reported mechanisms, like proline regulation, during short-term responses to new salinities and after acclimation, from which we found no indication of proline regulation in any of our salinity treatments.

6.1.3 Genotypic variation has a greater impact than salinity on degree of gene expression in *Skeletonema marinoi*.

Genotypic variation has a confounding effect when interpreting gene expression behavior of multiple strains. To account for the impact of genotype, we conducted a long-term common garden experiment with eight strains, each with three technical replicates, of *S. marinoi* from different locales within the Baltic Sea. We analyzed the transcriptome data for strain-specific and average overall responses to 8, 16, and 24 ppt salinities. The number of differentially expressed genes from the averaged response of all the strains was higher than any of the strain-specific totals showing that inclusion of biological diversity improved our ability to detect differentially expression and emphasizing the value of having both technical and biological replicates in a study.

We found that while shifting from a 24 ppt to 8 ppt had the greatest effect on gene expression across all the strains, there was considerable variation between the magnitude of their responses. When looking at strain responses individually, considerably fewer genes were differentially expressed, though each strain had 100-300 unique differentially expressed genes. Only 27 genes were differentially expressed in all eight strains, suggesting a conserved response to low salinity. Both the average and core responses in *S. marinoi* suggested that in low salinities photosynthesis and carbon fixation are upregulated, and there is less protein recycling. Upregulation of genes involved in lipid and fatty acid production suggests that low salinities represent suboptimal growth conditions for *S. marinoi*, as diatoms are known to accumulate these compounds in unfavorable environments. The reduction in cell division supports this interpretation. However, the continued production of chlorophyll and upregulation of photosynthesis indicates that the lowest salinity, 8 ppt, is not a severe stress for *S. marinoi*. Though there were indicators of ROS, oxidative stress also did not appear to be severe as several key ROS scavengers like super oxide dismutase were not differentially expressed. This complex overall response likely enabled the colonization of low salinity regions of the Baltic Sea.

6.1.4 Disparity between genes regulated at different stages of stress response in *Cyclotella cryptica*.

Based on the discrepancies between the results of the post-acclimation gene expression response of *C. cryptica* to salinity change and reported short-term stress metabolic changes (Liu and Hellebust 1974, Nakov et al. 2020), there is apparent disparity in how *C. cryptica* copes with salinity stress following immediate exposure and how it acclimates. Thus we conducted a time series study determining how freshwater exposure impacts gene expression within the first 10 hours. We estimated that according to growth analyses this time frame encapsulates the peak period of salinity stress through to the initial stages of acclimation, during which ~50% of the transcriptome was remodeled.

The highest number of differentially expressed genes occurs at 30 min, indicating that this time point is closest to the peak salinity stress response for *C. cryptica*. At this time, we saw increased activity in heat shock factor proteins and, at relatively low degrees of expression but high volume, oxidoreductive processes. Conversely, transcription, translation, and amino acid biosynthesis were repressed. Much of this response was reversed by 10 hr. Regulation of typical cellular metabolic processes become upregulated and genes associated with ROS scavenging are downregulated. A comparison between the differentially expressed genes at these two time points revealed minimal overlap, suggesting that these two stages of the stress response require very different genes as the cell acclimates to the new salinity.

As part of the overall cellular response to osmotic stress, *C. cryptica* regulated production of the osmolyte proline, reducing cytosolic levels as expected in a hyposaline environment. Other common osmolytes such as DMSP, taurine, and glycine betaine did not experience any strong regulation, or were not differentially regulated at all. Additionally, Na^+ and K^+ levels were heavily regulated via various ion transporters throughout the 10 hrs of exposure. In conjunction with the results of the gene expression analyses of *C. cryptica* post-acclimation, it is clear that the cell must work continuously to maintain the osmotic gradient in suboptimal salinities. Other than this perspective, there appears to be little overlap between genes utilized in the peak stress response, initial acclimation, and post-acclimation stages. This highlights the necessity of including multiple timescales when describing a cellular stress response.

6.2 Future work

We validated the phylogenetic relationship between *Cyclotella kingstonii* and other members of *Cyclotella* to justify the reclassification of the *Spicaticribra* genus. Similar phylogenetic analyses should be performed for extant species previously classified as *Spicaticribra* to ensure they were accurately reclassified. We found that in *Cyclotella cryptica*

genotypic effects had a greater impact on variation between samples than any of our salinity stress treatments. After pooling the data to account for this, we determined that following acclimation to low and high salinity treatments DMSP, taurine, and glycine betaine were utilized as compatible solutes in most high salinity treatments. Additionally, as expected, Na+ and K+ levels within the cell were highly regulated in all treatments. At the time of this experiment, we did not have access to a high quality annotated genome for the species and a de novo transcriptome was used for determination of differential gene expression. Based on the outcome of our long-term *Skeletonema marinoi* experiment, this *C. cryptica* common garden experiment should be reproduced with proper biological and technical replication with the results determined via mapping to the improved annotated genome (Roberts et al. 2020). In *S. marinoi*, this approach enabled us to identify a core set of genes regulated by all genotypes in response to salinity change, which could be of considerable value in understanding the evolutionary history of marine-freshwater transitions in this species. We also found that our lowest salinity treatment, 8 ppt, did not elicit a severe stress response from *S. marinoi*. Future predictions of the Baltic Sea estimate that within 100 years the brackish sea will have salinity levels reminiscent of a freshwater lake (Janssen, Schrum, and Backhaus 1999). As such, the lowest salinity tolerance of *S. marinoi* should be identified and the salinity stress response in this environment determined. The initial response of *C. cryptica* to freshwater was similar to that seen in the yeast stress response: downregulate genes involved in ribosome biogenesis, transcription, and translation while upregulating genes involved in carbohydrate metabolism, redox reactions, and defense against ROS (Mager and De Kruijff 1995; Gasch et al. 2000). Here again, it would be of interest to compare this across multiple biological replicates, with technical replication, to determine how natural genetic variation affects the early stress response.

6.3 References

Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown. 2000. Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Molecular Biology of the Cell 11 (12): 4241–57.

Janssen, F., C. Schrum, and J. O. Backhaus. 1999. A Climatological Data Set of Temperature and Salinity for the Baltic Sea and the North Sea. Deutsche Hydrographische Zeitschrift. https://doi.org/10.1007/bf02933676.

Mager, W. H., and A. J. De Kruijff. 1995. Stress-Induced Transcriptional Activation. Microbiological Reviews 59 (3): 506–31.

Roberts, Wade R., Kala M. Downey, Elizabeth C. Ruck, Jesse C. Traller, and Andrew J. Alverson. 2020. Improved Reference Genome for Cyclotella Cryptica CCMP332, a Model for Cell Wall Morphogenesis, Salinity Adaptation, and Lipid Production in Diatoms (Bacillariophyta). Cold Spring Harbor Laboratory. https://doi.org/10.1101/2020.05.19.103069.