LUND UNIVERSITY

Origin and the evolution of diatoms through the integration of paleontology and phylogenetics

Brylka, Karolina

2024

LITHOLUND THESES 41

Origin and the evolution of diatoms through the integration of paleontology and phylogenetics

KAROLINA BRYŁKA
LITHOSPHERE AND BIOSPHERE SCIENCE | DEPARTMENT OF GEOLOGY | LUND UNIVERSITY 2024

LITHOLUND THESIS 41

Origin and the evolution of diatoms through the integration of paleontology and phylogenetics

Karolina Bryłka

LUND
UNIVERSITY

Lithosphere and Biosphere Science
Department of Geology

DOCTORAL DISSERTATION
by due permission of the Faculty of Science, Lund University, Sweden.
To be defended at Pangea, Geocentrum II, Sölvegatan 12. Date 03.05.2024 and time 13:15.

Faculty opponent
Bánk Beszteri
University of Duisburg-Essen

Cover illustration front by Karolina Bryłka

Cover illustration back by Rebecca A. Pickering

Faculty of Science
Department of Geology

ISBN: Litholund theses 978-91-87847-82-0 (print)

ISBN: Litholund theses 978-91-87847-83-7 (pdf)

ISSN: 1651-6648

Printed in Sweden by Media-Tryck, Lund University, Lund 2024

Organization
LUND UNIVERSITY
Department of Geology
Sölvegatan 12
SE-22362, LUND
Sweden

Author: Karolina Bryłka

Document name

DOCTORAL DISSERTATION

Date of issue: May 3, 2024

Sponsoring organization

Title and subtitle: Origin and the evolution of diatoms through the integration of paleontology and phylogenetics.

Abstract

Diatoms, the prominent photosynthetic eucaryotes, have inhabited the world's oceans for at least the past 120 Ma since their first appearance in the Lower Cretaceous. There are also records of older diatoms, from the Jurassic dating to ca. 172 Ma and ca. 165 Ma, however these are poorly documented. The predicted origin time of diatoms using evolutionary relationships (molecular phylogenetics) yields an earlier date of origin of 200 Ma. These dates point towards some expected gaps in the fossil record, which may bias our understanding of early diatom evolution. Diatoms influence major geochemical cycles and sustain oceanic ecosystems in the modern ocean; hence, it is essential to learn about their past. To study diatom evolution across the Mesozoic and Cenozoic time periods, scientists use two approaches: a traditional approach based on fossil record interpretation, and the more recently developed approach, which evolves around molecular and phylogenetic studies. This Ph.D. project is a combination of both approaches.

This thesis has the following objectives: (1) search for older diatom microfossils than previously described in the lower Cretaceous and Jurassic sediments, (2) revisit the oldest published diatom microfossils to establish their reliability, (3) analyze existing fossil information and progress the use of paleontology in phylogenetic studies, (4) reconstruct molecular evolution of environmentally responsive and an ecologically important gene family in diatoms. These objectives were addressed in four subprojects.

An extensive search for Lower Cretaceous and Jurassic diatoms discovered no new fossils. Instead, scarce sponge spicules and radiolarians were observed at several study sites and exhibited a high degree of dissolution and alteration. This finding suggests that diagenetic processes biased our observations, and potentially caused the complete absence of diatoms. The study of the oldest diatoms of the Lower and Middle Jurassic age revealed these fossils were not diatoms but most likely calcareous nannofossils and testate amoebae, respectively. This discovery extended the gap between the oldest fossils and estimated origin time to 80 million years. The lack of diatoms in ancient sediments inspired further research on Cretaceous diatoms. To identify trends in distribution, diversity, and emergence of genera in time and space we compiled the Cretaceous Diatom Database. We identified extant diatom genera as far back as 100 million years and compiled a list of well documented fossils for future use as calibration points in molecular clocks. We also identified areas for future taxonomil work. Overall, based on the lower Cretaceous biogeographic dispersal and the morphological diversity of the oldest diatoms, together with diagenetic evidence from radiolarians and sponge spicules, we suggest that older diatoms than so far described are yet to be discovered.

We used molecular tools to unravel the evolution of diatoms, specifically genes responsible for silicon transport (SITs). This study focused on a diatom clade called Thalassiosirales, which has abundant representatives in the marine and freshwater realms-environments with varying levels of nutrients including dissolved silicon. Previous studies have shown that Thalassiosirales, both marine and freshwater, have multiple SITs which likely differ in their affinity and capacity for transport, moreover marine and freshwater diatoms exhibit differences in silicon metabolism. So far little evidence supports adaptive evolution on the sequence level, where diatom SITs have been shown to evolve predominantly under strong purifying selection. Advances in genome sequencing and updated codon models allowed us to formulate new questions and improve previous inferences. We showed extensive and ongoing history of gene duplication and loss. Furthermore, our data suggest the optimization of gene expression has played a central role in shaping sequence evolution and gene family dynamics of SITs, diatoms continuously balance gene dosage and expression to optimize silicon transport across major environmental gradients.

This Ph.D. project was a multidisciplinary approach to study the evolution of diatoms. For certain studies, such as the evolution of SITs, molecular means were applicable, but to draw broad conclusions on the evolution of diatoms, it is crucial to combine the fossil record and phylogenetic studies.

Key words: diatoms, evolution, Mesozoic, fossil record, molecular evolution, silicon transporter proteins, gene duplication

Classification system and/or index terms (if any)

Supplementary bibliographical information | Language: English

ISSN and key title: 1651-6648 LITHOLUND thesis | ISBN: 978-91-87847-82-0 (print) 978-91-87847-83-7 (pdf)

Recipient's notes | Number of pages: 142 | Price

Security classification

Signature Date 2024-03-18

“In one drop of water are found all the secrets of all the oceans; in one aspect of You are found all the aspects of existence.”

Kahlil Gibran

Contents

List of publications

This thesis is based on four publications listed below, which have been appended to this thesis.

Paper I

Bryłka, K., Alverson, A.J., Pickering, R.A., Richoz, S., and Conley, D.J., 2023. **Uncertainties surrounding the oldest fossil record of diatoms.** *Scientific Reports*, *13*, 8047. https://doi.org/10.1038/s41598-023-35078-8

Paper II

Bryłka, K., Pinseel, E., Roberts, W.R., Ruck, E.C., Conley, D.J., and Alverson, A.J., 2023. **Gene duplication, shifting selection, and dosage balance of silicon transporter proteins in marine and freshwater diatoms.** *Genome Biology and Evolution*, *15*(12), p. evad212. https://doi.org/10.1093/gbe/evad212

Paper III

Bryłka, K., Alverson, A.J., Richoz, S., and Conley, D.J. **Looking for the oldest diatoms.** Under review in Marine Micropaleontology.

Paper IV

Bryłka, K., Alverson, A.J., Ashworth, M.P., and Conley, D.J**. The Cretaceous Diatom Database: A tool for investigating early diatom evolution.** Under review in Journal of Phycology.

Acknowledgements

First and foremost, I would like to thank Daniel. **Daniel**, you gave me this incredible opportunity to do the Ph.D. at one of the best universities in the world. Through this experience, your supervision, scientific discussions and life-related conversations that we shared I have grown not only as a scientist but also as a person. This chance that you have given me will forever change my life. Who I am as a young scientist today is largely thanks to you.

Further I would like to thank **Sylvain** and **Andy**. Supervision that I have received throughout my Ph.D. was top notch! You were always there for me to help, discuss and cheer me up. Your expertise progressed my project and resulted in four publications that I am really proud of. **Sylvain**, you have put tremendous work into collecting all the material I needed for the diatom search and made sure I understood the geological part of my project. **Andy**, you have inspired me and gave me this incredible chance to work with your group on something that, let's be honest, I had no idea about.

Thank you: **Elizabeth, Eveline, Matt**, **Rebecca**, **Wade**, for joining my projects and being there for me when I had millions of questions. Your help tremendously improved my papers and progressed my understanding of the field. You were all incredibly fast and responsive and I would have never thought that all my papers will be published/submitted. Alongside I would like to thank **Jakub Witkowski**, **David Harwood,** and **David Williams**. You discussed my project with me along the way and helped me understand what I was dealing with.

Thanks to all **collogues and friends from the department** for creating friendly environment. It made my time here special and joyful. Thank you: **Helena F.** and **Johan**, for answering all the questions I had about the Ph.D and making sure all the organizational aspects are in place. Thanks to **Gert** for keeping the electronic world run and to **Kansli** for all the help I have received.

Thanks to all **PhD students** for time spend on fikas, lunches, dinners and parties. Thank you: **Rebecca**, **Kristin** and **Sylvain**, for going with me to the field in Germany. I had a fantastic time! **Rebecca**–my work bestie! Thank you for being there for me at every step of the way, when I was happy, when I was panicking and crying and when I was quitting. Also, my first paper–it wouldn't have happened without you. **Kasia** and **Pati** aka polish mushroom pickers in Sweden (inside joke) you made my last months so joyful! Sharing an office with you makes coming to work so easy. **Kasia** thank you for reading my thesis and finding the ridiculous amount of missing commas. Thanks to all of the office mates and friends I had during the past 5 years: **Aaron, Cindy, Franzi, Ingrid, Josefin, Rosine**, **Tjördis**, **Petra** and **Will**.

I wouldn't be where I am without my wonderful fiancée **Ethan,** humorously a fellow diatomist. You are my biggest supporter; you believe in me and you always keep my spirit up. You also had a great impact on all my research and you truly helped me shape my projects though endless conversations, discussions and brain storming. Thank you with all my heart! I also want to thank our amazing dog **Kajtek**, for always wiggling his tail and keeping me in a good mood.

Last but not least. Mieszkam w Szwecji już od pięciu lat, teraz tu jest mój dom, and nosze was w sercu codziennie. Moje dziewczyny, szwedzkie okrzemki: **Dominika**, **Magda**, **Marta** i **Ula**, za każdym razem, gdy przyjeżdżam do Polski nie mogę się doczekać by was zobaczyć. Tęsknie za wami każdego dnia i jestem ogromnie wdzięczna za wasze wsparcie! **Mamo**, **Tato**, moja najwspanialsza siostro **Marto** i moje ukochane dziewczynki **Martynka** i **Marcela** dziękuję za wsparcie mojej decyzji w wyprowadzce i wyborze kariery naukowej, dla nas wszystkich było to trudne i w dalszym ciągu jest to wyzwanie. Dziękuje że we mnie wierzycie i wspieracie mnie każdego dnia!

1 Introduction

Water bodies cover more than 75% of our planet, comprising oceans, seas, lakes, rivers, and icecaps – diverse environments with huge biodiversity spread across thousands of unique niches that, through biological processes, work together and propel life. However different, these niches have something in common – thriving diatom communities.

Diatoms are unicellular, autotrophic microorganisms that constitute a prominent group of eukaryotic algae and influence major geochemical cycles of oxygen, carbon, silica, phosphorous, and nitrogen (Julius and Theriot, 2010). Because of their unique cell wall, termed a frustule, they are easily recognizable (Schmid et al., 1981). The first diatoms and their descriptive drawings were published in 1702 (Anonymous, 1702). By the end of the 19th century, a large amount of work was published, primarily on the taxonomy of diatoms. Over time, researchers gained interest in the biology and the ecology of these important, widespread organisms. With advances in ocean exploration and ocean drilling (e.g., DSDP–Deep Sea Drilling Project, ODP–Ocean Drilling Project, IODP–International Ocean Discovery Project) scientists could recover deep-sea sediment cores and study diatoms over larger timescales in the fossil record. This opened a window into studying the origin, early evolution, paleoecology, and climatic responses of diatoms over long time periods. Nevertheless, fossil evidence is only one way to study the evolution and adaptations of the organism.

Over the past 20 years, our understanding of cellular functionality, early diatom evolution, and the time of their origin has rapidly progressed with the advent of DNA sequencing with the first fully sequenced diatom genome published in 2004 (Armbrust et al., 2004). Whole genome sequencing, the development of tools such as molecular clocks, models, and methods for reconstructing the evolutionary changes in amino acid substitutions across the phylogenetic frameworks continuously progress and allow scientists to answer evolutionary questions regarding diatoms. Some of these tools require integration of the information in the fossil record and DNA analysis.

This Ph.D. addresses several questions about the evolution of diatoms from both paleontological and phylogenetic perspectives to enhance the understanding of the origin and early evolution of diatoms and the molecular evolution of specialized genes.

2 Aims of the thesis

The aim of this thesis is to strengthen the knowledge of the origin and the early evolution of diatoms through the integration of paleontology and molecular phylogenetics, and the molecular evolution of genes responsible for dissolved silicon uptake in diatoms. The main research objectives can be summarized as:

1. Search for the Mesozoic sediment deposits of Jurassic and Cretaceous age (201-66 Ma) to complement fossil information with new records of the marine diatoms,
2. Evaluation of the oldest described diatom microfossils,
3. Fossil record analysis and interpretation to aid the integration of paleontological data and DNA studies,
4. Unraveling the molecular evolution of silicon transporter proteins in diatoms across the salinity gradient.

To achieve these aims, my thesis work comprises both paleontological and molecular approaches.

3 Background

3.1 Diatom description

Diatoms are microscopic photosynthetic eukaryotes that live in every aquatic environment in both benthic and pelagic habitats (Julius and Theriot, 2010). Diatoms can exist solitary or in chains and they can attach or move

(Round et al., 1990). All diatoms share an obligate requirement for dissolved silicon (abbreviated as DSi but chemically known as orthosilicic acid with the formula $Si(OH)_4$) necessary to build their frustules. A frustule of a diatom is composed of two halves, called thecae, which fit together like a petri dish and therefore one theca is always smaller. During cell division, when two daughter cells are produced, each cell keeps one theca and grows a smaller half to fit within it. The consequence is that after each division the cells get smaller. To restore the cell size, for those cells that can no longer divide, sexual reproduction needs to occur (Round et al., 1990). Based on the shape, arrangement, type, and structure of perforation, and the presence of special structures on the silica cell wall (Fig. 1), diatoms are taxonomically divided into four major morphological groups: radial centric, bi(multi)polar centric, araphid, and raphid (Julius and Theriot, 2010). The typical size of a frustule varies between 10 μm and 100 μm (Round et al., 1990).

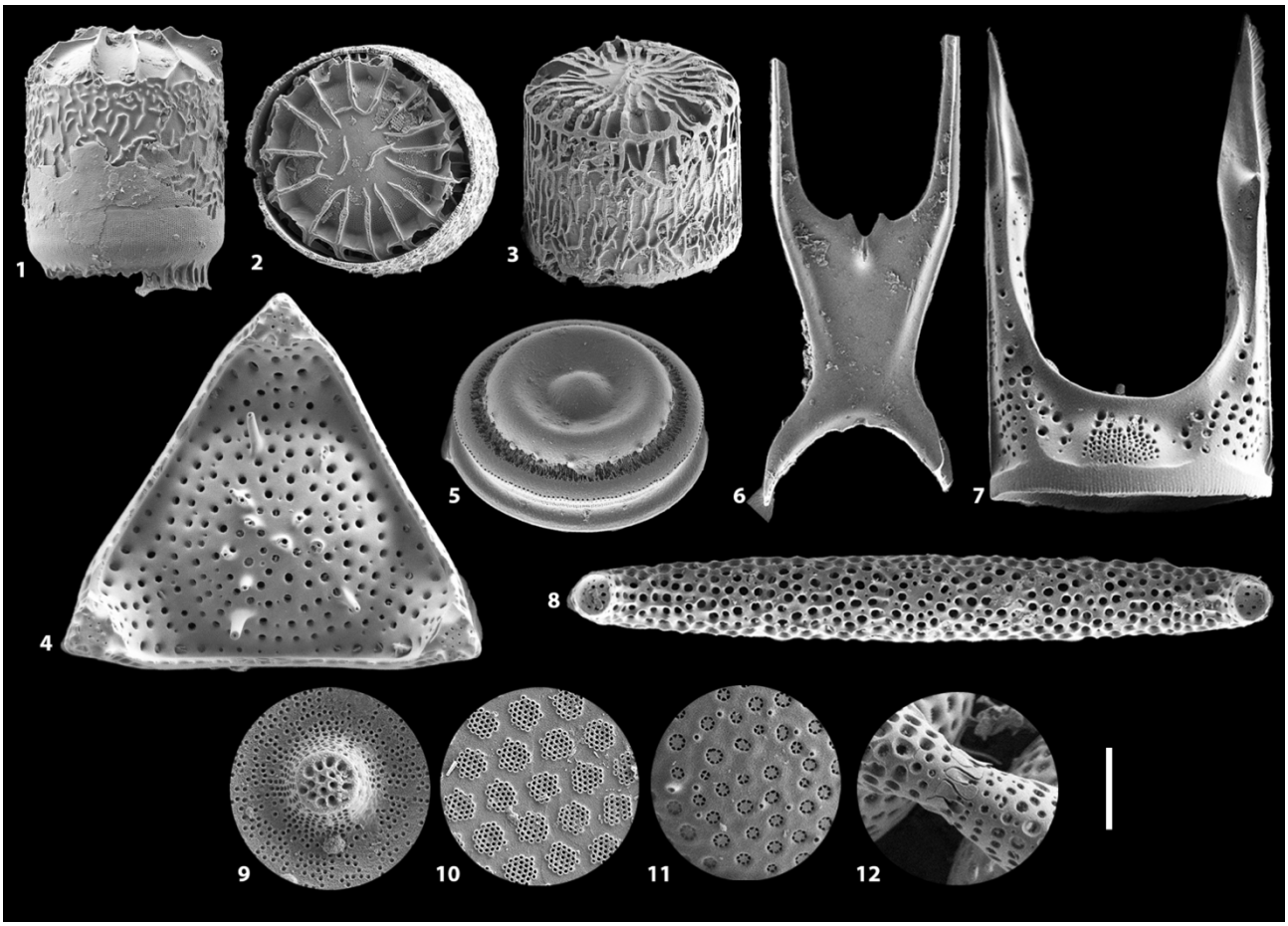

Fig. 1. Cretaceous diatoms from ODP core 693B (1–3, 6, 9,10) and ODP core 275 (4, 5, 7, 8, 11, 12). 1-3–*Calyptosporium*, 4–*Trinacria*, 5–*Pseudopodosira*, 6–*Kreagra*, 7–*Hemiaulus*, 8–*Cortinocornus*, 9–internal view of central perforate process of *Praethalassiosiropsis*, 10–internal view of areolae with vela of *Praethalassiosiropsis*, 11–external view of arealoe with cribra and rimmoportulae openings (tubes) of *Trinacria*, 12–apical interlocking processes of *Trinacria* that join adjacent frustules. Scale bar 10 μm.

Sample from ODP core 693B was provided by David M. Harwood, sample from ODP core 275 was provided by The International Ocean Discovery Program's (IODP) Gulf Coast Repository (GCR).

As diatoms photosynthesize, light, nutrients (e.g., phosphorous, nitrogen, silica), and CO_2 are required for growth and reproduction. Culture studies have shown that elements like sulphur, boron, barium, iron, manganese, and vitamin B-12 are also important for the growth of various diatom species (Lewin and Guillard, 1963).

Diatoms evolved several physiological responses to stress: formation of a resting stage and/or resting cell in periods of nutrient limitation (Kuwata and Takahashi, 1990; Jewson et al., 2008) or reducing phosphorous and iron demand under nutrient-replete conditions (Marchetti et al., 2006; Van Mooy et al., 2009). Diatoms can survive without light, which disables photosynthesis, by fermentation or respiration of exogenous sugars (Lewin and Lewin, 1960). Diatoms are primarily autotrophic, but there are several known species of diatoms living as symbionts within Foraminiferas (Lee and Correia, 2005) and Dinoflagellates (Tamura et al., 2005). Developed strategies and flexibility are what make diatoms so widespread and successful.

Diatom production is responsible for 40% of the primary production in the ocean (Granum et al., 2005) and 20% of global primary production (Tréguer and De La Rocha, 2013). These microalgae carry out one-fifth of photosynthesis on Earth (Nelson et al., 1995). Diatoms contribute about 40% of particulate organic carbon export (Jin et al., 2006), which can either reach the mesopelagic layer (~1000m) or deeper into the bathypelagic layer also called the "CO_2 sequestration layer" (Tréguer et al., 2017). Carbon reaching the bathypelagic layer is removed from the atmosphere for at least 100 years, and ultimately, some of the carbon will reach the sea floor and be buried (Tréguer et al., 2017). In addition, high-energy lipid reserves make diatoms nutritionally a vital food source for heterotrophs (Ahlgren et al., 1990).

There are specific regions of the oceans worldwide where diatom abundance is exceptionally high. Those regions include the Southern Ocean, the North Pacific Ocean, and the Eastern Indian Ocean, as well as upwelling areas and estuarine environments (Marinov et al., 2010). These water masses are both DSi and nutrient-rich, many are upwelling areas and, therefore, are some of the most fertile ecosystems in the world (Renaudie, 2016). As the ocean is undersaturated with regards to DSi (Garcia et al., 2013), the preservation of opaline frustules will occur only when the export flux is greater than the dissolution rate both in the water column and in the sediment (Tréguer and De La Rocha, 2013). The highly productive regions are also the most probable places to form diatom-rich sediments and the primary deposits of fossil marine-rich diatom sediments (Renaudie, 2016). Diatoms appeared in the fossil record in the late Lower Cretaceous. The oldest diatom fossil record is dated in the Aptian (ca. 121 Ma) from deposits in Australia and is represented by moderately preserved, diverse communities (Nikolaev et al., 2001b). The oldest best-preserved fossils come from the Weddell Sea and are dated to be Albian (113-107 Ma) (Gersonde and Harwood, 1990; Harwood and Gersonde, 1990). There have been other fossils described from Albian, but these samples show chemical alteration (i.e., pyritization) and are difficult to identify and properly classify (Jousé, 1949; Wall, 1975; Geroch, 1978; Strel'Nikova and Martirosjan, 1981; Foucault et al., 1986). Rothpletz (Rothpletz, 1896, 1900) described three fossil species dating from the Jurassic (182 Ma and ca. 164 Ma), however, the work presented in this thesis overturned these fossils (see Paper I) (Bryłka et al., 2023).

There are several different hypotheses about the origin of the diatoms. Round and Crawford (1981) surmised that diatoms must have diverged from a naked photosynthetic cell that acquired a siliceous coating, they called it the Ur-diatom. This proto-diatom presumably appeared in a shallow, benthic habitat of unknown salinity. Mann (1989) puts diatoms in a close relationship with Parmales (a group called *Bolidophycea*), a siliceous marine picoplankton. This suggests the origin of the diatoms to have occurred in a marginal pelagic environment. Based on morphological similarities, Nikolaev and Harwood (2000a) assigned diatoms in close relation to ancient Xanthophyceae, occupying shallow seas. Molecular analysis supports the relationship

between diatoms and Parmales, implying that diatoms must have originated in the marine environment (Ichinomiya et al., 2016). DNA studies on organellar genes in diatoms revealed that diatoms emerged from a serial secondary endosymbiosis involving red and green algae and a heterotrophic host and bacterial exosymbiont (Prihoda et al., 2012) resulting in a red algal-derived chloroplast empowered by green algal proteins and mitochondrion derived from the non photosynthetic exosymbiont (secondary plastid) (Prihoda et al., 2012).

From 100 Ma, the abundance of the diatoms increased in the geological record, with two peaks during the Cenozoic era at 34 Ma (Eocene-Oligocene boundary) and at 20 Ma (lower Miocene), which reflect massive global fluxes of DSi into the ocean connected with enhanced continental weathering (Renaudie et al., 2018). The prominent radiation over geological time enabled diatoms to dominate modern aquatic ecosystems, impacting the global carbon, nitrogen, and silica cycles.

3.2 Diatoms and silicon

Silicon dioxide [SiO_2] (also known as silica) is the seventh most abundant element in the universe and the second most abundant element on the Earth, comprising about 25% of the Earth's crust. The silicon budget in the World's oceans is controlled by biological, geological, and chemical processes, such as weathering, hydrothermal activity, freshwater and groundwater inputs, and biomineralization (Yool and Tyrrell, 2003).

Biosilicification (the formation of biological structures from silica) is at the center of the global silica cycle. Silica biomineralization occurs in several clades across the Tree of Life, including stramenopiles (e.g., diatoms, chrysophytes), rhizarians (e.g., radiolarians), opisthokonts (e.g., sponges), testate amoeba and land plants, which use this element to form protective cell coverings (Marron et al., 2016). As diatoms dominate in many regions of today's oceans, they highly influence the global silica cycle (Benoiston et al., 2017). Every silicon atom that enters the ocean is recycled about 17 times by biological uptake followed by dissolution before being buried in seafloor sediments (Tréguer et al., 2021), which is the major loss of silicon in the world's oceans today (Tréguer and De La Rocha, 2013). Diatoms take up dissolved silicon (DSi) in the form of orthosilicic acid $Si(OH)_4$, which varies widely in concentration across the ocean. Ocean surface water DSi concentrations are generally low (≈10 μM Si) but reach higher concentrations (≈ 80 μM Si) in some areas, such as the Southern Ocean (Frings et al., 2016). The freshwater realm is characterized by higher DSi values, on the order of 85–100 μM in many lakes and rivers (Frings et al., 2014, 2016).

Differences in DSi concentrations potentially impact the silicon metabolism of marine and freshwater diatoms. Freshwater diatoms have, on average, an order of magnitude more silicon per cell than marine diatoms (Conley et al., 1989). Marine diatoms have lower half-saturation constants (concentration at which half of the maximum intake rate is reached) parameters (K_s and K_μ) than freshwater species (Martin–Jézéquel et al., 2000), which suggests marine diatoms have a greater enzymatic affinity for DSi, which is expected in environments where silicon is scarce (Alverson, 2007). In laboratory experiments, the model diatom *Cyclotella nana* (formerly *T. pseudonana*; Alverson et al., 2011) grew faster and had less silicon when grown in marine versus freshwater media, regardless of DSi concentration, suggesting an interaction between growth, salinity, and silicon availability (Olsen and Paasche, 1986). DSi uptake occurs during the S phase of the cell cycle (DNA synthesis phase), so rapid growth and shorter cell cycles allow less time for DSi uptake and deposition, resulting in thinner frustules (Thamatrakoln et al., 2006). The differences in silicon metabolism between marine and freshwater environments have been attributed to a combination of factors, including DSi, salinity, pH, ionic composition, and osmotic pressure (Olsen and Paasche, 1986; Alverson, 2007).

At high DSi concentrations (>30 μM Si), diatoms are able to acquire DSi passively through diffusion (Hildebrand et al., 1997),

whereas at low and potentially growth-limiting concentrations, DSi is actively imported by a family of silicon transporter proteins (SITs) (Hildebrand et al., 1997; Durkin et al., 2016; Marron et al., 2016). SITs are sodium/silicic acid symporters in the stoichiometry of 1:1, $Si(OH)_4$:Na^+ (Knight et al., 2016). SIT coding genes are not exclusively in diatoms, they appear in other silicifying lineages and calcifying organisms (Marron et al., 2016) and in picocyanobacteria (Baines et al., 2012), although their structure differs across these groups. SITs are localized in the plasma membrane and are built of transmembrane helixes (domains) and conserved amino acid residues (motifs) (Fig. 2). Diatom SITs possess 10 transmembrane domains, as do some coccolithophores and choanoflagellates (10-TMD), which is double the size compared to other eukaryotic lineages with SITs (Marron et al., 2016). SITs possessing 5-TMDs are called SIT-Ls and were recognized in coccolithophores, radiolarians, dinoflagellates, and foraminiferas (Marron et al., 2016).

Conserved motifs have a consistent position on transmembrane helixes, and their presence creates an environment necessary for transmembrane transport (Hildebrand, 2008). 10-TMDs SITs were suggested to result from duplication and fusion of presumably ancestral 5-TMDs SIT-Ls (Marron et al., 2016). Hildebrand (2008) proposed that a double set of motifs in SITs enables both external DSi binding and the release of DSi into the cell. Hence, the role of 5-TMDs with one set of motifs of SIT-Ls in transporting and binding silicon is unknown (Durkin et al., 2016).

It has been hypothesized that SITs may have arisen initially to prevent Si toxicity in the high DSi levels in Precambrian Oceans (Marron et al., 2016). Reduction in DSi concentrations throughout geological time led to widespread losses of SIT and SIT-Ls, resulting in today's distribution of genes across the eukaryotic tree of life (Marron et al., 2016). Structural differences between SIT and SIT-Ls support higher efficiency of diatom DSi uptake compared to radiolarians, the biggest diatom competitor for DSi in today's oceans.

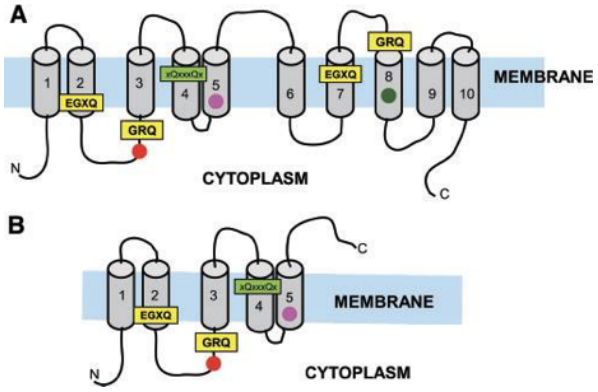

Fig. 2. Visualization of SITs for A. diatoms and B. foraminifera from Marron et al. (2016). Transmembrane domains are represented by gray cylinders, functionally important motifs are labeled in yellow, green rectangles and circles represent conserved residues. N and C are N-terminal and C-terminal amino acid residues of the protein chain.

DSi uptake in diatoms mediated by SITs enables DSi transport from the extracellular to the intracellular environment, however, the intracellular pathway is not yet known. Thamatrakoln and Hildebrand (2008) proposed the involvement of an intracellular bounding component, which would transport DSi into intracellular pools (silicon deposition vesicle) that store DSi at concentrations much higher than the polymerization level. This storage allows a continuous uptake without immediate usage, which sustains cell requirements during nutrient starvation periods (Martin–Jézéquel et al., 2000). Analysis of SIT sequences from various diatom species revealed structural differences of these proteins between taxa of different morphological groups (Thamatrakoln et al., 2006; Durkin et al., 2016).

SIT coding genes form a family with multiple gene copies in the genome (Thamatrakoln et al., 2006). The number of gene copies for SITs ranges from 3–5 in three model diatoms–*Cyclotella nana, Cylindrotheca fusiformis*, and *Phaeodactylum tricornutum*–with phylogenetic studies highlighting a dynamic history of mostly recent, taxon-specific gene duplications and losses (Alverson, 2007; Thamatrakoln and Hildebrand, 2007). Multiple SIT genes are often differently expressed through the cell cycle, suggesting sub-functionalization, allowing for

control of the timing, affinity, and capacity for DSi transport (Hildebrand et al., 1998; Martin-Jézéquel et al., 2000; Thamatrakoln et al., 2006).

4 Molecular approaches

During the past two decades, a large part of diatom research has been focused on culturing and DNA sequencing studies. Through these means scientists have been exploring adaptations and interactions and cellular activity of diatoms. DNA sequences of multiple species may provide phylogenetic frameworks based on which various evolutionary questions and hypothesis on molecular evolution and the origin of diatoms can be formulated. The hypotheses tested in this Ph.D. project were developed using such frameworks.

Box. 1. Simplified definitions of terms used in this section.

DNA-deoxyribonucleic acid, carries genetic information for the development and functioning of an organism

Nucleotide-basic building block of the DNA. Nucleotides of the DNA are adenine (A), cytosine (C), guanine (G) and thymine (T)

Amino Acid-molecules used by organisms to build proteins; one amino acid is composed of three nucleotides (three nucleotides=codon); several combinations can decode one amino acid: for example, Alanine amino acid can be decoded by the following codons: GCG, GCA, GCC

Alleles-variation of the same sequence of nucleotides at the same locus on DNA, e.g., each human has two alleles for eye color

Gene-short section of DNA; genes can code for a trait (for example eye color) or contain the instruction to produce proteins

Protein-large molecules composed of amino acids; proteins perform work in cells and are required for the cells' structure, function, and regulation

Mutation-change in the DNA sequence

Phylogeny-representation of the evolutionary history and relationships between groups of organisms inferred from DNA sequences

Molecular evolution-process of change in the sequence composition of DNA or proteins

Molecular clock-the technique that uses mutation rate to deduce the time of when two forms diverged in the past

Relaxed molecular clock-model of molecular clock which allows the evolutionary mutation rate to vary across organisms

Lineage-series of organisms connected by a continuous line of descent from ancestor to descendant

Crown group-a collection of species composed of the living organisms and their most recent common ancestor

Insertion-type of mutations in which one or more nucleotides are added to DNA segment

Deletion-type of mutation in which one or more nucleotides are removed from DNA segment

Gene conversion-allele at one locus (recipient) is changed by copying sequence from another locus (donor); recipient and donor can be different alleles of the same gene

Gene duplication-production of one or more copies of a region of the DNA that contains a gene and its insertion in another region

Natural selection-heritable traits that are beneficial for the organism become more common in a population over time; selective pressure (evolutionary force), which drives natural selection, can be, e.g., environmental such as competition

Synonymous (dS) and nonsynonymous (dN) substitution-dS is a type of mutation that does not change the amino acid, dN is a type of mutation that changes the amino acid

Purifying (negative) selection-removal of deleterious mutations that decrease the fitness of an organism

Paralog-gene that arose via gene duplication (orthologs arise via speciation)

Positive selection-type of natural selection that promotes a spread of beneficial alleles

Relaxed selection-occurs when natural selection is reduced or eliminated; both positive and purifying selection are weakened
Pseudogenization-accumulation of deleterious alleles resulting in a loss of gene functionality

To study early diatom evolution, scientists apply the so-called "molecular clock". The molecular clock is a figurative term describing a technique that uses a mutation rate of biomolecules to estimate the time when two or more taxa (e.g., diatom genera) diversified (Tiley et al., 2020). To calculate the molecular clock, we estimate the number of nucleotide or amino acid substitutions in the genetic sequences of organisms and the time when these organisms first occurred in the fossil record (Fig. 3). This combination allows us to calculate how long it took for these sequences to diverge from the identical sequence representing the last shared common ancestor (Ho and Duchêne, 2014). A molecular clock works best with a large array of sequenced genes from one organism and extensive fossil information on the first and, ideally, the next occurrence of the organism or the sister organism. The molecular clock allows for the time calibration of phylogenies (which depict the lines of evolutionary descent), and these are called time-calibrated phylogenies.

There are known issues with the molecular clock methods. For example: evolutionary rate variation, where evolution might have been rapid during the organism radiation creating a large genetic distance in short time periods (Lee and Ho, 2016), this however is impossible to address while calculating the molecular clock. Further issues are rate heterogeneity among genes (Lee and Ho, 2016), saturation, where multiple sites might have undergone multiple substitution (Lozano-Fernandez, 2022) which is undetectable and the use of erroneous calibration points (Angelis and Dos Reis, 2015). Although methods such as a relaxed molecular clock were introduced, which allows for irregular substitution rates across organisms (Tiley et al., 2020), many of the above issues still lack the solution and may bias inferences of molecular clocks.

Molecular clocks and phylogenies bring insight into early diatom evolution by calculating the timing of the origin and diversification of morphological groups and various diatom lineages. The most recent and most robust study inferring the age of the crown group of diatoms and diversification of taxa and lineages was published by Nakov et al. (2018) (Fig. 4). This phylogeny was based on 11 genes and 1151 taxa across all major morphological groups (radial centric, bi(multi)polar centric, pennate araphid, and pennate raphid). This study inferred that diatoms originated ca. 200-190 Ma, near the Triassic-Jurassic boundary. Furthermore, we can identify the timing of the emergence of morphological groups and relationships among taxa and lineages. The first diatoms to emerge were those of radial centric symmetry, followed by pennate araphid and raphid diatoms (Nakov et al., 2018).

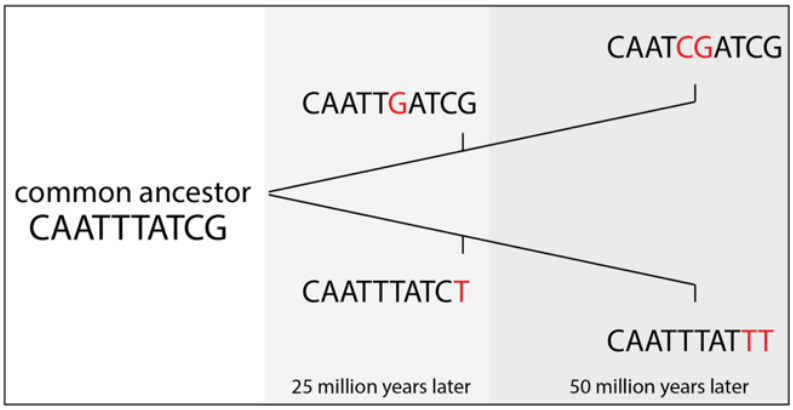

Fig. 3. Model of the molecular clock. With time organisms diverge by accumulation of nucleotide substitutions designated by red letters, longer the time the higher the divergence from the common ancestor.

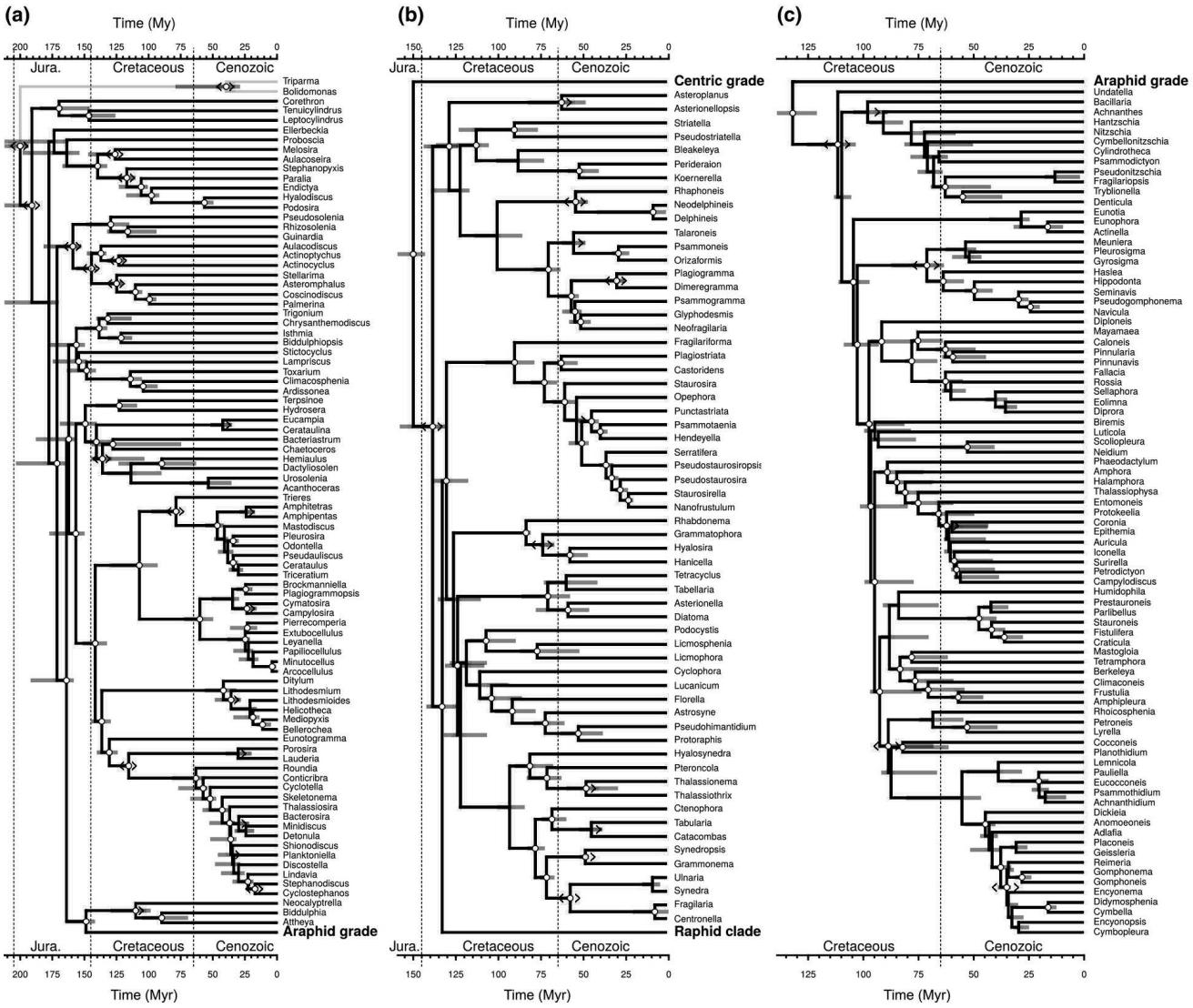

Fig. 4. Phylogenetic tree of diatoms from Nakov et al., (2018).

Because DNA degrades with time (Hofreiter et al., 2001), we cannot study the evolutionary relationships like lines of descent among the diatom genera and species in the Cretaceous. The diatom frustule preserved in the sediments brings an insight into morphological changes throughout geological time, conveying some concepts on the evolution of characteristics but not a direct answer on the relationships or when diatoms evolved. Moreover, the fossil record is incomplete, and in the case of the diatoms, only about 3% of the living communities are preserved in sediments (Tréguer et al. 1995). The fossil record does not always align with molecular phylogenies, causing dissociation when interpreting the timing of the origin and the evolution of traits. However, these two means should be combined when drawing conclusions about early evolution of diatoms, since both represent different pieces of the same evolutionary history. In Paper IV we show how these two means can be integrated.

Phylogenetic reconstructions are possible due to the continuous mutational change of alleles. DNA sequences continuously undergo nucleotide substitutions, insertions, deletions, gene conversions, duplication, and loss. The consequence is that a mutant sequence or gene may produce a new morphological or physiological character or new function, which may be further fixed in the descending population and conserved or mutated again, creating genetic distance between species (Nei and Kumar, 2000). Particularly important in this project is gene evolution driven by gene duplication. Genes are in constant flux, being gained through duplication and occasionally retained or more frequently lost from a genome (Hakes et al., 2007). Gene duplication provides new genetic material for mutation drift and natural selection to act upon, the result of which may be a new gene function (Zhang, 2003). The natural selection acting on genes can change after duplication i.e., by accumulation of deleterious mutation not removed by selection leading to pseudogenization, by strong purifying selection against mutations preventing duplicated genes from diverging, and by the accumulation of positive substitutions leading to novel functions (Fig. 5) (Zhang, 2003).

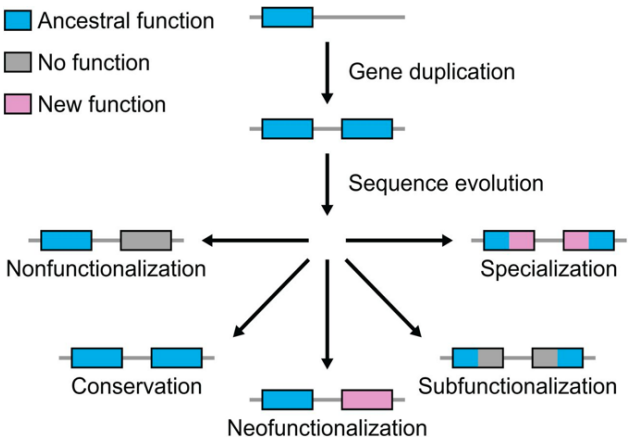

Fig. 5. Evolutionary trajectories of duplicated genes. Evolution may result in the loss of one functional copy by nonfunctionalization (pseudogenization), or retention of two functional copies by either conservation where both copies perform ancestral function, neofunctionalization where one copy performs new function, subfunctionalization where both copies perform ancestral function with reduced capacity, or specialization where both copies simultaneously perform ancestral function with reduced capacity and a new function. Figure from deGiorgio and Assis (2021).

The type of selection duplicated genes undergo depends largely on the functional benefits that multiple paralogs (duplicates) provide (Zhang, 2003), and the nature of the selection may be imposed by environmental changes (Hoffmann and Willi, 2008), e.g., habitat shift (Mikołajewski et al., 2016). Studies indicate that genes associated with environmental responses tend to evolve more quickly and undergo more rapid rates of loss and duplication (Clark et al., 2007). By example, in *Prochlorococcus*, marine cyanobacteria, genes involved in light responses are evolving quicker than core genes (those responsible for the housekeeping functions and primary metabolism) (Kettler et al., 2007). A recent study on hemoglobin genes across vertebrates shows that adapting to different oxygen levels required numerous gene duplications, both whole genome duplication and small-scale duplication, and loss events. Overall, hemoglobin genes were found to evolve under purifying selection, with exception to few members which varied in the rates of evolution

following the genome duplication. This suggests the adaptation to the environment was more constrained on one paralog than the other. Lastly, the number of gene copies of paralogs varied across lineages, suggesting gene duplication after habitat shift promoted the adaptation to different oxygen levels (Mao et al., 2023).

Another study on the Toll-like receptors gene family (TLR_1), involved in pathogen defense across vertebrates, revealed a complex history of gene duplication and inferred that TLR_1 are mostly subjected to purifying selection with varying numbers of sites under positive selection across paralogs in the lineage suggesting different selective constraint acting on gene duplicates (Huang et al., 2011). The above examples highlight the role of gene duplication and selection pressure in the adaptation to a changing environment. Duplicated genes form gene families and vary in the number of members (Zhang, 2003), e.g., in higher plants the aquaporin family constitutes from 30 to more than 70 diverse members (Deokar and Tar'an, 2016). However, some gene families are much smaller, like the silicon transporter protein (SITs) gene family, constituting only a few members, for example in diatoms (Hildebrand et al., 1998).

To study the dynamics and fate of the duplicated genes and selection pressure acting upon paralogs, many models and methods were introduced over the years. These models measure rates of nonsynonymous (dN, changes the amino acid) and synonymous (dS, does not change the amino acid) nucleotide substitutions (Fig. 6) to estimate the relative impacts of positive, negative (purifying), or relaxed selection (Yang, 2002). Each codon is fitted into three classes of ω (ω=*dN/dS*): ω<1, ω=1, ω>1, which indicate purifying, neutral, and positive selection, respectively (Yang, 2002). Depending on the selection pressure acting on the paralog, it is possible to determine whether its function was conserved by purifying selection or optimized by positive selection.

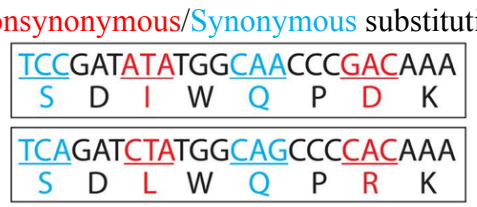

Fig. 6. Examples of non-synonymous and synonymous substitutions. Non-synonymous substitutions lead to amino acid replacement, in this case from I to L (Isoleucine to Leucine) and D to R (Aspartic Acid to Arginine). Figure from Luo (2015).

Further, transcript and protein levels can be measured to study the functionality of the paralogs to determine the relative contribution to assigned function. Genes in a genome are expressed through the process of transcription and translation. In the transcription process, information from the DNA is transferred to messenger RNA (mRNA). DNA serves as a template to complementary base paring and an enzyme termed RNA polymerase catalyzes the formation of the pre-mRNA molecule, further processed to mature mRNA also called a transcript. Mature mRNA is translated into proteins in the organelle termed ribosome, in this process genetic code in mRNA is translated into amino acids a chain of which forms a protein (Fig. 7) (Clancy et al., 2008). Sequencing mRNA is a standard method to measure and compare levels of gene expression and determine the relative contribution to the assigned function (Pertea, et al., 2016). In general, the more a gene is transcribed, the more protein will be made.

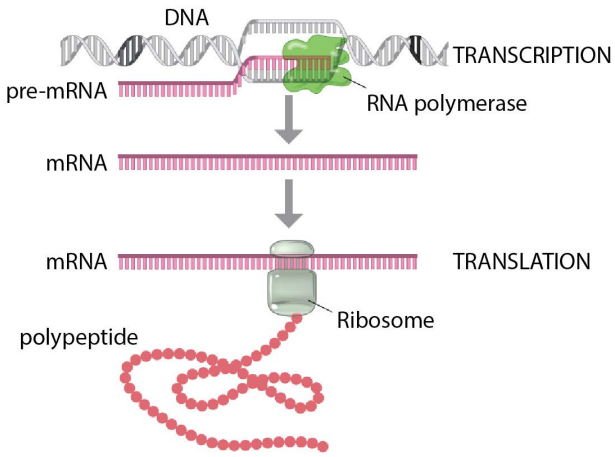

Fig. 7. Gene expression process through transcription and translation. Polypeptide=chain of amino acid=protein. Figure from Clancy et al. (2008).

Reconstruction of the duplication patterns, selection acting upon duplicated genes and measuring transcript levels of prologs is another way to reconstruct the evolutionary dynamics of the organism. It is particularly useful for functional genes spread across populations occupying various niches like silicon transporter proteins of diatoms shown in Paper II.

5 Diatoms in sediments

On average, in today's oceans, the burial of diatoms represents around 3% of the living community, demonstrating that most frustules dissolve while sinking or at the ocean's bottom (Tréguer et al., 1995). The physicochemical dissolution of the frustule in the water column depends on DSi concentration, temperature, pH, salinity, pressure, and frustule characteristics (e.g., thickness, reactive surface area, and presence of delicate features) (Passow et al., 2011). In sediments, dissolution is controlled by various processes, including the presence of lithogenic minerals, aging processes, and DSi in the pore waters (Loucaides et al., 2012).

5.1 Search for diatoms in the Mesozoic

While looking for the oldest diatoms in the geologic record, a few caveats need to be considered. Theoretically, there might not have been diatoms before 120 Ma. However, molecular clock estimates (discussed in Section 6), the diversity of the oldest described assemblages and wide dispersal of the lower Cretaceous diatoms suggest otherwise. Another possibility is that the oldest diatoms may have had a different frustule. Diatoms are closely related to Parmales, another group of marine picoplankton that build cell coverings of small plates (Ichinomiya et al., 2016). Therefore, the first diatoms could have been misidentified, or the plates could have dispersed after cell death. A diatom frustule is uniformly constructed across taxa consisting of two valves that match together like a petri dish (Round et al., 1990). Therefore, it seems unlikely that diatoms switched from one frustule type to another. An early assumption states that the original diatoms did not have a cell covering or the covering was primitive, non-siliceous (i.e., Ur-diatom hypothesis), therefore, the cells were not preserved (Round and Crawford, 1981). However, capacity to build cell covering is likely an inherited trait, therefore was present already in ancestral diatoms (Kooistra et al., 2003). We also must consider that marine deposits of the pre-Cretaceous age that might have contained diatoms were eroded, consumed by subduction or diagenetically altered or destroyed.

In the modern ocean, diatoms are the most abundant in upwelling zones and high latitudes (Marinov et al., 2010), and therefore, examining past upwelling zones has the highest potential for fossil preservation. In addition, it is estimated that a large fraction of the current export of diatoms to sediment is through aggregates (particulate organic matter with inclusions of bacteria and phytoplankton) and fecal pellets (Ploug et al., 2008). Aggregates often terminate the diatom bloom and result in a large quantity of cells sinking together (Alldredge and Gotschalk, 1989). Fecal pellets and coprolites of the Cretaceous age were described to contain diatoms (Chin et al., 2008), hence, analysis of Mesozoic coprolites may provide new fossil communities. Previous studies suggest that diatom preservation is elevated under anoxic conditions (McMinn, 1995). Soft tissues of invertebrates buried under anoxic conditions were found in sediments as old as 380 Ma (Melendez et al., 2013). Therefore, analysis of Mesozoic sediments deposited under oxygen-depleted conditions may also act to preserve diatoms (but these environments may also cause pyritization). Some of the oldest (120-100 Ma) fossils were described from phosphatic and calcareous nodules (Forti and Schultz, 1932; Nikolaev et al., 2001b). Preservation of fossils in concretions is excellent in many localities, conserving soft tissues, bones and delicate structures (Arena, 2008; Gaines et al., 2012; Mccoy et al., 2015). Accordingly, nodules have a high potential to preserve diatom fossils.

Preservation and dissolution factors and ancient paleogeography, sediment chemistry, and

lithology should be considered while looking for potential fossil-bearing sediments to maximize the chances of discovering new diatom assemblages. All the localities investigated in this Ph.D. project are described in Paper III.

5.2 The fossil diatom record in the Cretaceous

Cretaceous diatoms were observed at 35 study sites worldwide, described in 51 scientific papers (Table 1, Fig. 8). Fifteen of 35 sites are represented by diagenetically altered frustules where the original opal was substituted with pyrite crystals. Sediments containing diatom fossils are globally distributed (Table 1), but the diversity and abundance differ, which may reflect past ecological preferences and/or preservation potential. Tropical locations such as Site 758 contain 18 different diatom species belonging to 12 distinct genera (Fourtanier, 1991). Sub-tropical sites such as Tonga Trench or the Sinai Peninsula contain 24 species in 13 genera and 35 species in 18 genera, respectively (Chambers, 1997; Zalat, 2013). The most diverse and rich communities are from higher latitudes with richness as high as 180 species in 47 genera for Canadian Arctic sites (Tapia and Harwood, 2002) and 140 species in 39 genera in Seymour Island sediments (Harwood, 1988). In the modern ocean, diatoms are the most abundant at high latitudes, cold environments and upwelling zones. Cretaceous study sites with abundant diatoms are likewise predominantly located in higher latitudes (Chambers, 1997). We can, therefore assume that early communities also preferred relatively colder water masses, with weaker stratification enhancing vertical mixing, of the higher latitudes.

Table 1. Compilation table of study sites describing Cretaceous diatoms.

Location	Age	Age Ma	Lithology	Preservation	Reference
Carpathians, Poland	early Aptian	125-113	black shales	pyritized	Geroch, 1978
Queensland Australia	Aptian	121-120	-	opaline	Dun et al., 1901
			carbonate concretions	opaline	Harper, 1977
			calcareous concretions	poor	Nikolaev et al., 2001b
			mudstone, siltstone	pyritized	Haig and Barnbaum, 1978
Hannover Germany	late Aptian	115-113	-	pyritized	Kemper, 1975
	late Aptian	115-113	-	pyritized	Benda, 1982
Stavropol, Russia	Albian	113-100	-	pyritized	Strel'Nikova and Martirosjan, 1981
Penza region, Russia	Albian	113-100	argillaceous sands	pyritized	Jousé, 1949
Hannover Germany	Albian/Aptian	113	phosphorite nodules	moderate	Forti, 1933; Rust, 1885
	Albian	113-100	phosphorite nodules	moderate	Forti and Schultz, 1932
	early Albian	113-110	glauconitic sand	pyritized	Georgi, 1978
Ligurian Alps, Italy	Albian-early Cenomanian	113-100	flysch	pyritized	Foucault et al., 1986
ODP site 693 Weddell Sea	early Albian	113-107	siliceous ooze	very good	Gersonde and Harwood, 1990; Harwood and Gersonde, 1990
Sinai Peninsula, Egypt	Albian-Maastrichtian	113-70	limestones, shales	good	Zalat, 2013
Alberta, Canada	middle Albian	107	-	pyritized	Wall, 1975
Canadian Arctic (Horton River, Eglinton Island, Ellef Ringnes Island, Ellesmere Island)	Cenomanian-Campanian	100-83	siltstones, sandstones	good	Tapia and Harwood, 2002
Sergipe Basin, Brazil	early Turonian-late Campanian	93-72	shales, calcareous mudstones	pyritized	Koutsoukos and Hart, 1990
Bohemia, Czech	Turonian	93-89	-	poor	Wiesner, 1936

Location	Age	Age Ma	Lithology	Preservation	Reference
Villers-Sur-Mer, France	Turonian	93-89	-	poor	Deflandre, 1941
Omagari Formation Japan	uppermost Santonian–lowermost Campanian	84-82		very good	Shimada et al., 2022
Devon Island Artic Ocean	Santonian-Campanian	86-72	sandstone, mudstone, bentonite	very good	Witkowski et al., 2011
Saratov Region, Russia	early Campanian	83-80	aleurolites and weakly cemented sandstones	very good	Oreshkina et al., 2013
Westphalia, Germany	Campanian	83-72	flysch	poor	Riegraf, 1995
Negev, Israel	Campanian	83-72	phosphatized carbonates; phosphorites, cherts	moderate	Moshkovitz et al., 1983
ODP site 758 Indian Ocean	late Campanian	80-72	calcareous chalk	moderate	Fourtanier, 1991
Tonga Trench, Pacific	Campanian	83-72	calcareous and siliceous pelagic sediments	moderate	Ballance et al., 1989
Alpha Ridge, Artic	late Campanian	70-72	shales	very good	Davies and Kemp, 2016; Dell'Agnese and Clark, 1994; J. F. Barron, 1985
Alberta, Canada	Campanian	83-72	shales, claystone, silty claystone and siltstones	pyritized	Given and Wall, 1971
Seymour Island Antarctica	Campanian-Danian	83-61	mudstone, sandy siltstone, diatom ooze, concretions	moderate	Harwood, 1988
Gdansk, Poland	Campanian	83-72	sponge ' reefs	good	Schulz, 1935
DSDP site 216 Indian Ocean	Campanian	83-72	-	moderate	Gresham, 1985
DSDP site 275 Pacific	Campanian	83-72	clayey slit/silty clay	very good	Hajós and Stradner, 1975
Ural, Russia	Campanian-Maastrichtian	83-66	-	very good	Jousé, 1949
central USA (Redbird, Boulder, Pueblo)	Campanian-Maastrichtian	83-66	shales	pyritized	Bergstresser and Krebs, 1983
Konto Mountains Japan	late Campanian-early Maastrichtian	74-70	siliceous mudstone, sandstone, mudstone	poor	Takahashi et al., 1999
Rougemont, Swizterland	Maastrichtian	72-66	flysch	pyritized	Weidmann, 1964
PeeDee Formation USA	Maastrichtian	72-66	glauconitic sand, sandstones, clays	pyritized	Abbott, 1978
Emperor Canyon, USA	Maastrichtian	72-66	-	moderate	Fenner, 1982
ODP site 748 Indian Ocean	early Maastrichtian	72-70	glauconitic calcareous rocks	moderate	Nikolaev and Harwood, 2000b
Moreno Formation, USA	Maastrichtian	72-66	shales	very good	Davies and Kemp, 2016
	Undifined Maastrichtian	72-66	shales	very good	Long et al., 1946
	late Maastrichtian	68-66	shales	very good	Hanna, 1927; Nikolaev et al., 2001a
E Marlborough New Zeeland	late Maastrichtian	68-66	siliceous limestone	pyritized	Hollis et al., 1995
Worldwide (DSDP Site 216, DSDP Site 275, ODP Site 758, Alpha Ridge, Moreno Formation, Tonga Trench)	Campanian-Maastrichtian	83-56	-	-	Chambers, 1997

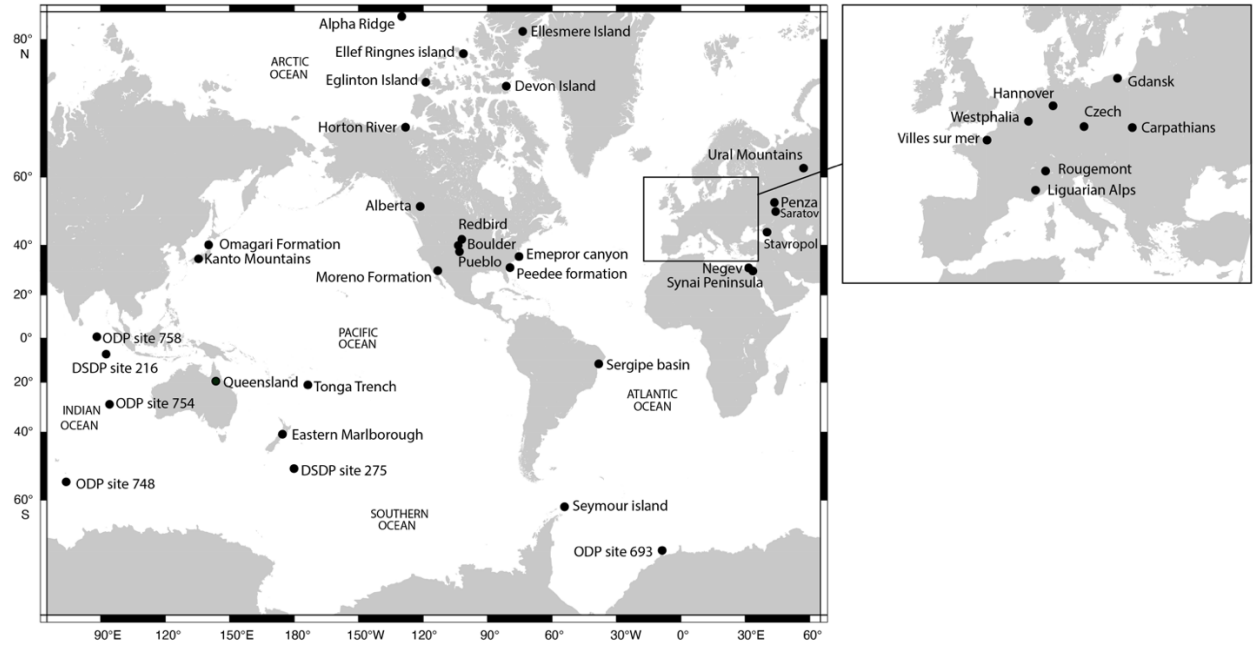

Fig. 8. Cretaceous study sites bearing diatom fossils listed in Table 1.

To understand and reconstruct the early evolution of diatoms, we need to carefully analyze and interpret what information the fossil record provides. Such analysis allows for integrating paleontology-derived data into molecular clocks and phylogenies, constraining their results. Detailed analysis of the fossil record, trends in the emergence of morphological groups and diatom genera, together with a list of taxa that can be integrated into future molecular clocks, is a subject of Paper IV presented in this thesis.

6 Materials and methods

6.1 Paper I and III

6.1.1 Study sites

To find diatom microfossils in Jurassic and Lower Cretaceous rocks we carefully selected study sites that represent environments that could have been occupied by diatoms in the past and were conductive for fossilization (see section 7.1). We collected samples in the field and received samples from collaborators and international repositories. The study sites examined represent various depositional environments and sediment types. We tested open ocean pelagic sites, shallow epicontinental seas, shelfs and deltaic setting, some of the sediments in these sites were deposited under anoxic conditions. Furthermore, we examined phosphatic nodules, coprolites, cherts and sponge mummies, alongside black shales, marls and limestones. Lastly, we sampled localities which were previously denoted as containing other siliceous microfossils such as sponge spicules and radiolarians. Compilation information and results from study sites investigated in this Ph.D. project are presented in Paper III.

6.1.2 Diatom extraction from sediments

Sediment samples, except cherts, were cleaned following common chemical digestions for diatom extraction (Trobajo and Mann, 2019). First, 5 to 300 g of each sample was weighed in a glass beaker. 10% hydrochloric acid (HCl) was added and left at room temperature for calcium salt removal. Samples remained in HCl until the reaction (fizzing) was no longer visible (up to two weeks). Samples were then rinsed three times in Milli-Q® water. Treatment continued with 35-65% nitric acid solution (HNO_3) on a hotplate (120°C) for organic matter removal. Samples remained in HNO_3 until the color of the sediment brightened (up to 4 weeks), indicating the removal of the proportion of organic matter. After HNO_3 treatment,

samples were washed in Milli-Q® water three times. The remaining organic matter was oxidized using 33% hydrogen peroxide (H_2O_2) on a hotplate (100°C) until the sample color was light gray to white. This step lasted up to 6 weeks. Samples were then cleaned three times in Milli-Q® water.

After chemical treatment, samples were dried in a freeze dryer. Dried sediment was mixed with a heavy liquid solution (sodium polytungstate of a density of 2.1 g/cm^3) to separate siliceous materials from the rest of the particles. Both floating and settled fractions were cleaned in Milli-Q® water and transferred to clean 50 ml plastic tubes. Final solutions were used for slide preparation for examination using a light microscope (LM). Two ml of suspended material were dried on a coverslip and mounted with Norland 61.

A 4–10% hydrofluoric acid solution (HF) was used to extract siliceous microfossils from cherts (De Wever et al., 2002). A few pieces of crashed material were placed in a plastic beaker, and HF solution was added in a volume suitable to cover the sample. After 30 minutes supernatant was cleaned through a filtering system, and dissolved residue was collected on a 5µm filter and transferred to a clean 50 ml plastic tube. The process was repeated four times. After chemical treatment, samples were cleaned in Milli-Q® water, and solid slides for LM examination were prepared following the same procedure described above. More slides were prepared and analyzed for samples with a larger initial volume.

Settled fractions from the heavy liquid separation were examined as a precaution that no diatoms were lost. Samples with observed siliceous macrofossils (see Paper III) were also examined on the Scanning Electron Microscope (SEM) for further analysis and elemental mapping. For the SEM analysis, 0.5 ml of the cleaned sample was pipetted on the carbon tape attached to the SEM stub and left to dry. Samples were coated with 5 nm Platinum-Palladium (Pt-Pd) powder in a Cressington sputter coater. Samples were then examined on the variable pressure Tescan Mira3 High-Resolution Schottky FE-SEM with an Oxford EDS detector housed at the Department of Geology, Lund University, Sweden. Samples were photographed at 2 kV and elemental mapping was performed at 15 kV.

6.2 Paper II

6.2.1 SITs of Thalassiosirales

Predicted SIT sequences for this Ph.D. project were derived from genomes of 37 Thalassiosirales species (Roberts et al., 2023) (25 marine and 12 freshwater) based on homology to fully assembled SIT sequences of *Cyclotella nana*. In addition, transcriptomes of four outgroups (taxa from other genera than Thalassiosirales but closely related) were assembled (*Bellerochea*, *Ditylum*, *Lithodesmium*, and *Eunotogramma)*. Species were derived from culture collections and from cultures grown in Alverson's lab at the department of Biological Sciences at the University of Arkansas in Fayetteville, USA, AR (Fig. 9).

Fig. 9. Culture collection in Alverson's lab.

Thalassiosirales were chosen for this study for several reasons: (1) they represent a clade containing both freshwater and marine taxa, (2) Thalassiosirales possess three SIT paralogs (SIT1, SIT2, SIT3) exhibiting variable levels of transcription (Thamatrakoln et al., 2006), (3) SIT paralogs are present in multiple copies across species of Thalassiosirales (Alverson, 2007, (4) SITs were shown to evolve predominantly under strong purifying selection (Alverson, 2007). As previously mentioned, freshwater and marine diatoms exhibit differences in silicon metabolism, therefore

Thalassiosirales are suitable framework to formulate questions regarding adaptations on the sequence level and underlying molecular basis for these differences. Furthermore, multiple copies of genes and varying levels of transcription suggests a history of gene duplication and varying strength of natural selection on paralogs. Advanced codon models and whole genome sequencing applied in this study may provide new perspectives on the evolution of SIT genes in diatoms.

6.2.2 Analysis

6.2.2.1 Phylogenetic analysis and history of gene duplication and loss

Derived sequences were clustered with OrthoFinder (Emms and Kelly, 2019), resulting in a single SIT orthogroup. Preliminary alignment and the phylogenetic tree consisted of 199 sequences/branches and were used to prune partial (<400 nucleotides) sequences or redundant sequences exhibiting zero/near-zero branch length. Moreover, for species represented by multiple strains, we removed strains with the shorter sequences. Further analysis of the dataset had 109 sequences (108 ingroup and one outgroup). The final set of amino acids were aligned using UPP and trimmed using trimAl (Capella-Gutiérrez et al., 2009). Nucleotide alignment was aligned by reconciling nucleotide sequences against amino acid alignment using translatorX (Abascal et al., 2010). We further used IQ-TREE (ver. 1.6.12) to obtain tree topology. More details on the process are provided in Methods of Paper III.

SIT gene tree and corresponding species tree derived from Roberts et al. (2023) were further used to infer duplication and loss events in the evolutionary history of Thalassiosirales. To this end we used reconciliation method implemented in NOTUNG software (2.9.1.5) (Chen et al., 2000), which embeds gene tree in the species tree to pinpoint duplication and loss events.

6.2.2.2 Molecular evolution

To measure nucleotide substitutions to estimate the relative impact of natural selection on the evolution of diatom SITs we used methods implemented in the HyPhy package (Kosakovsky Pond et al., 2005), operated through DataMonkey Adaptive Evolution Server (Weaver et al., 2018). We tested for purifying (FEL method), positive (MEME method) and relaxed (RELAX method) selection; for significance see section 6 and for details on methods see Paper III.

6.2.2.3 Transcription profiles of SITs

Following the results of Thamatrakoln et al. (2007) (Fig. 9) we sequenced transcriptomes from seven species (eight strains). Taxa were grown in optimal growth conditions in a Percival incubator (15 °C, 16:8 light:dark light regime, 22 μmol photons m^{-2} s^{-1} irradiance) and measurements were taken from unsynchronized cultures. Cells were grown in 24-well plates and harvested during exponential growth, based on chlorophyll *a* fluorescence. Cells were stored at -80 °C until RNA extraction using Qiagen's RNeasy Plant Mini Kit, after which RNA quality and quantity were measured using a TapeStation 2200 (Agilent), a NanoDrop 2000c (ThermoScientific), and a Qubit 2.0 (Invitrogen). Library preparation and sequencing were performed by Arbor Biosciences using the myReads RNA-seq library prep kit and Illumina NovaSeq sequencing platform (2 x 150 paired-end reads). The above measurements were performed as a part of the larger experiment of Kala Downey and Eveline Pinseel at the Department of Biological Sciences, University of Arkansas Fayetteville, USA.

6.3 Paper IV

6.3.1 Cretaceous Diatom Database

To compile first comprehensive Cretaceous Diatom Database, we assembled the data from 18 scientific papers describing Cretaceous diatoms (see Paper IV). The data collected represent Aptian-Maastrichtian assemblages (121.4–66 Ma). The Cretaceous Diatom Database consists of the following: species list with corresponding papers, species list in age and genera list with number of species in designated age (Fig. 10).

A

	Nikolaev et al., 2001	Harwood and Gersonde	Gersonde and Harwood	Forti and Schultz 1932	Zalat 2013				Tapia and Harwood 2001				Witkowski et al., 2011		Shimada et al, 2022
current name	Lower Aptian	Lower Albian	Lower Albian	Albian	Albian	Cenomanian	Turonian	Santonian-Maastrichtian	Cenomanian-Santonian	Lower Campanian	Late late	Campanian	Unclassified Campanian?	Santonian	Campanian	Uppermost Santonian-Lowermost Campanian
Gladius antiquus	x		x	x		x	x		x			x		x		x

B

	Aptian	Albian			Cenomanian		Turonian	Conician	Santonian		Campanian			Maastrychtian		
current name	lower	lower	middle	upper	lower	upper			lower	upper	lower	middle	upper	lower	middle	upper
Gladius antiquus	x	x	x	x	x	x	x	x	x	x	x	x	x			

C

	Aptian	Albian			Cenomanian		Turonian	Conician	Santonian		Campanian			Maastrychtian		
current name	lower	lower	middle	upper	lower	upper			lower	upper	lower	middle	upper	lower	middle	upper
Gladius	6	5	1	1	2	2	2	1	1	1	6	6	7			

Fig. 10. An example of the Cretaceous Diatom Database for the species: Gladius antiquus: (A) in papers and (B) in age and (C) Gladius genus showing genus species richness in age.

6.3.2 Calibration points for molecular clocks

From the Cretaceous Diatom Database, we derived a set of well documented calibration points for the molecular clock studies. To this end we analyzed a photographic documentation of taxa published under the extant genus name and taxa that were previously used as calibration nodes in molecular clocks. We divided calibration points into two groups: crown calibrations and stem calibrations. Crown calibrations represent taxa that possess all of the genus obligatory characters (synapomorphies) and stem calibrations represent taxa that exhibit some genus characters (Fig. 11). Overall, we documented 21 taxa that may serve as calibration nodes in future molecular clock studies.

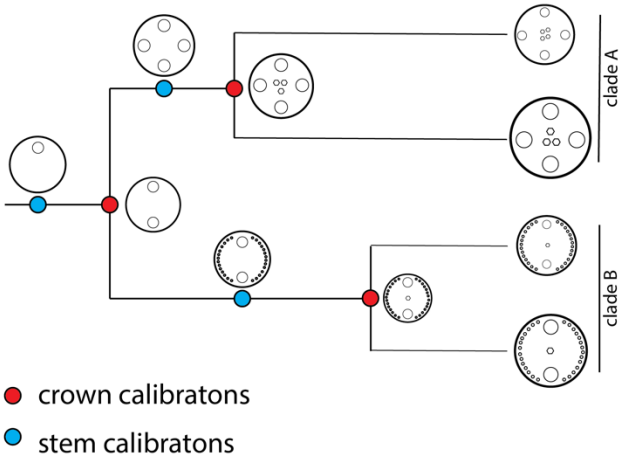

Fig. 11. Scheme representing the placement of the crown (red) and stem (blue) calibrations. Crown calibrations exhibit all the features of the terminal taxa, for clade A it's four marginal ocelli and multiple central processes, for the clade B its two marginal ocelli, single central process and ring of marginal process. Crown calibration in this case would be taxa exhibiting all characters. The stem calibration for the clade A could be a taxon with four ocelli without central processes. For the clade B it could be a taxon with two ocelli and marginal ring of processes without a central process. Crown calibration for both clades could be a taxon with only two ocelli and stem could be a taxon with one ocellus. This scheme is conceptual and does not represent actual diatom genera.

7 Summary of papers

7.1 Paper I

Bryłka, K., Alverson, A.J., Pickering, R.A., Richoz, S., and Conley, D.J., 2023. ***Uncertainties surrounding the oldest fossil record of diatoms****. Scientific Reports, 13(1), p.8047.*

Diatom microfossils from the Lower and Middle Jurassic described by A. Rothepltz in 1896 and 1900, respectively, were, and still are by many, widely considered the oldest evidence of diatom occurrences in the fossil record (Rothpletz, 1896, 1900). These diatoms were assigned to the genus *Pyxidicula* and were published based on hand-drawn representations of specimens. Microscopic slides containing

these diatoms were never found, and the sediment samples from which these diatoms were extracted were either exhausted by the author or discarded and also never found. These fossils have been broadly used in scientific inferences on the early evolution of diatoms by paleontologists and molecular biologists.

During the search for the Jurassic and Lower Cretaceous diatoms, we came across a few study sites that contained microfossils that either represented diatoms, or we classified them as diatoms due to their gross morphology. The application of stringent evaluation and contamination measures suggested these fossils were not *in situ* assemblages of Jurassic and Lower Cretaceous ages or were not diatoms. However negative, these results brought suspicions about the reliability of fossils described by A. Rothpletz and induced further evaluation.

To this end, we resampled analogous sediment from a neighboring locality to the one used in the 1896 publication and discovered fossils exhibiting the same morphological features as fossils described by Rothpletz. We subsequently classified the fossils as calcareous nannofossils due to their gross morphology and the elemental composition. For the 1900 publication, we did not resample the same sediment type, but careful analysis of the drawings presented by the author and convoluted taxonomy of the *Pyxidicula* genus allowed us to classify these fossils as testate amoeba rather than diatoms.

Paper I highlights the importance of extended evaluation of observed fossils to assess their *in situ* status and proposes that the Lower and Middle Jurassic fossils described by Rothpletz represent other microfossil groups than diatoms.

7.2 Paper II

Bryłka, K., Pinseel, E., Roberts, W.R., Ruck, E.C., Conley, D.J., and Alverson, A.J., 2023. ***Gene duplication, shifting selection, and dosage balance of silicon transporter proteins in marine and freshwater diatoms.*** *Genome Biology and Evolution, 15(12), p. evad212.*

Diatoms take up DSi by the use of silicon transporter proteins (SITs) (Hildebrand et al., 1998; Durkin et al., 2016) to build siliceous covering for the cell. SITs are a small gene family consisting of only a few members per species, but exhibiting variable levels of transcription, which may mirror the capacity and affinity for transport (Thamatrakoln and Hildebrand, 2007; Sapriel et al., 2009). SITs are ubiquitous diatom proteins, present in all diatoms sequenced to date in freshwater and marine environments. Although diatoms across the salinity gradient use SITs for DSi uptake, marine diatoms have more efficient uptake kinetics and faster cell cycles and freshwater diatoms build thicker frustules (Conley et al., 1989; Martin–Jézéquel et al., 2000). So far, metabolic differences in DSi uptake and paralog-specific expression patterns were not explained by adaptive evolution on the sequence level, where SITs were shown to evolve predominantly under purifying selection (Alverson, 2007).

Paper II unravels the molecular evolution of SITs across the salinity gradient in *Thalassiosirales* by using several approaches including whole genome sequencing, gene tree reconciliation, and new methods for calculating codon substitution rates. We inferred two clades of SITs: SIT1-2 and SIT3, where SIT1-2 is present across the full phylogenetic breadth of *Thalassiosirales* with multiple copies per species, whereas SIT3 is present only in half of the species, with only one copy in most cases.

We discovered that SIT sequence diversity, copy number, and gene expression have been shaped by whole genome duplication, small-scale duplication and loss events, natural and adaptive processes, and environmental changes. Paralogs exhibiting low transcript rates (SIT3)

were frequently lost and rarely duplicated. Moreover, various codon substitution methods revealed that paralogs with low transcript rates (SIT3) and SITs of freshwater diatoms were under relaxed selection. Replete DSi levels in freshwater environments likely resulted in the relaxed selection of freshwater SITs due to lower competition and rather infrequent starvation periods. Whereas widespread loss, likely connected to low expression level, and a potentially small contribution of SIT3 to DSi uptake resulted in the relaxed selection on this paralog.

Paper II provides new insights into the evolution of environmentally responsive and ecologically important genes and highlights the role of shifts in natural selection following duplication events and habitat transitions on the complex dynamics of SITs evolution over time and across species.

7.3 Paper III

Bryłka K., Richoz S., Alverson A. J., and Conley D. J. ***Looking for the oldest diatoms.*** *Under review in Marine Micropaleontology.*

Paleontological observation of diatoms has been broadly used in variety of studies on past environments and climatic changes and how these influenced diatom communities, as well as in phylogenetic studies. The fossil record of diatoms extends back to the Lower Cretaceous, 120 Ma, and is represented by diverse, moderately preserved diatoms (Nikolaev et al., 2001b). Although, diatoms were observed in variety of niches on both hemispheres by the Upper Cretaceous, the fossil record is only densely documented from the uppermost Cretaceous and the resulting paleontological observation is scarce and limits subsequent inferences. To complement the fossil information and to find diatom deposits predating the earliest fossils we examined 33 study sites worldwide from oceanic cores and outcrops on the continents.

We recovered siliceous microfossils: sponge spicules and radiolarians from eleven study sites and no diatoms were observed. Sponge spicules and radiolarians were present in small quantities and recovered from sponge mummies, cherts, shales, concretions and coprolites: sediment types considered as good for preservation. Nevertheless, the microfossils in our samples exhibited a high degree of dissolution and intricate features of radiolarian tests were absent, leaving behind overall shape. This preservation points towards post depositional diagenetic alteration and is supported by the recrystallization of amorphous phase of opal called opal-A, known in pristine diatom frustules, radiolarian tests and sponge spicules, into a more crystalline phase of opal: opal-CT and quartz (Pickering, unpublished data). Our results suggests that diatoms might have been erased from ancient sediments we examined through diagenetic processes.

The lack of diatoms in the sediments predating 120 Ma, could suggest there were no diatoms prior to this time. However, Lower Cretaceous diatom diversity and worldwide dispersal suggests some earlier diversification and dispersal events. Overall, we suggest search for the diatoms in sediment older than 120 Ma must continue, but should be focused on unlithified sediments that experienced shallow burial and low heat as well as structures such as concretions (Barron, 1993). These sediments should be deposited in shallow and shelf environments of high latitudes.

7.4 Paper IV

Bryłka K., Ashworth M.P., Alverson A.J., and Conley D.J. ***The Cretaceous Diatom Database: a tool for investigating early diatom evolution.*** *Under review in Journal of Phycology.*

The Cretaceous period is the time of the first appearance of the diatoms in the fossil record and gives direct evidence of the age, the diversity and morphological variability in the diatom lineage. The fossil record is incomplete and often circumvented by extrapolation through the use of molecular clocks and phylogenetics. These means offer two perspectives on the early diatom evolution, which is still poorly understood.

In this article we present a first comprehensive Cretaceous Diatom Database as a tool to investigate the taxonomy, stratigraphic range and species accumulation throughout the Cretaceous. To further aid the integration of the fossil information with molecular clock studies we derive a set of well supported calibration points of extant and extinct diatom genera from the database. We selected 21 genera that were documented with high quality photographs for crown and stem calibrations. We advise all the calibrations should be used in future molecular clock studies.

Further analysis of the Cretaceous Diatom Database revealed that most diverse are extant genera. Nevertheless, many taxa in these genera exhibit little morphological variation, hence the total diversity may be overestimated. Further, a third of the genera is monotypic and often documented only with LM photographs. We suggest the above would benefit from further taxonomic work. Previous estimates points towards numbers of 300 diatom species by the end of the Cretaceous (Lazarus et al., 2014; Knoll and Follows, 2016; Jewson and Harwood, 2017). Although the number of 700 species in the Database may be overestimated, we are confident that Cretaceous diversity is substantially higher than previous estimates.

The Cretaceous diatom database and the list of diatoms strongly support calibration points and is a valuable resource for future taxonomical, palaeoecological and molecular studies.

Table 2. Author's contribution to the papers.

	Paper I	**Paper II**	**Paper III**	**Paper IV**
Planning and study design	K. Bryłka D.J. Conley	K. Bryłka, A.J. Alverson	K. Bryłka D.J. Conley	K. Bryłka A.J. Alverson D.J. Conley
Literature review	K. Bryłka	K. Bryłka	K. Bryłka	K. Bryłka M.P. Ashworth
Fieldwork and data collection	K. Bryłka, R.A. Pickering, S. Richoz K. Doering*	E. C. Ruck, W. R. Roberts	K. Bryłka, S. Richoz, R.A. Pickering*	K. Bryłka
Labwork, sample preparation	K. Bryłka	E.C. Ruck, W.R. Roberts, E. Pinseel K. Downey*	K. Bryłka I. Doverbrat* R.A. Pickering*	n/a
Analysis	K. Bryłka	K. Bryłka, E. Pinseel, W.R. Roberts, A.J. Alverson	K. Bryłka	K. Bryłka
Figures	K. Bryłka	K. Bryłka, A.J. Alverson, E. Pinseel	K. Bryłka	K. Bryłka R.A. Pickering*
Data interpretation and discussion	All authors	All authors	All authors	All authors
Lead author	K. Bryłka	K. Bryłka	K. Bryłka	K. Bryłka
Comments and editing of manuscript	All authors	All authors	All authors	All authors

**Contributions who are not co-authors*

8 Discussion

The overall motivation for this Ph.D. was to reconstruct the evolution of diatoms by joining the perspectives of paleontology and molecular phylogenetics. The primary goal was to discover older diatom microfossils than what is currently known from the geologic record and then to integrate these data into a molecular clock analysis. The failure to do so has not affected the overall research presented herein.

Through the collection of four papers, we tackled the evolution of diatoms. We addressed overall perspectives such as early evolution and the origin of marine diatoms, and more focused perspectives, e.g., the evolution of silicon transporter proteins in one diatom clade.

Paper I and Paper III present a purely paleontological approach. In the following we evaluated the oldest claimed diatoms and rejected them as reliable fossils. In the latter we present efforts in the search for Mesozoic diatoms and infer the oldest diatoms are yet to be discovered. In Paper IV we highlighted the importance of joining both perspectives and assembled a set of well documented calibration points for future molecular clocks derived from the Cretaceous Diatoms Database.

Finally in paper II we presented a purely molecular perspective in which we reconstructed the molecular evolution of silicon transporter proteins (SITs). This study shows that environmental pressure and gene duplication events shape the evolution of ecologically and metabolically important genes. This, to some extent, dissociated collection of papers shows that the evolution of diatoms can be reconstructed 1) by paleontological means, as we show in Paper I and Paper III; 2) by phylogenetic and molecular means, as we showed in Paper II; 3) but large perspectives like origin and timing of the emergence of lineages and genera can only be inferred through joining perspectives of paleontology and phylogenetics (Paper IV).

8.1 Developments in the fossil record of marine diatoms

The primary and unachieved goal of this Ph.D. was to find new fossil assemblages of diatoms in the Lower Cretaceous and Jurassic, which would predate the oldest described diatoms. However, we also examined the Upper Cretaceous. The complementing fossil information extends our knowledge of diatom dispersal, the ecology of past environments, and the diversity of the organism at a point in its geological history. These findings improve our understanding of the early evolution of diatoms and past biogeochemical cycles and allows us to apply these data to predictive models (Sims et al., 2006; Conley et al., 2017; Naidoo-Bagwell et al., 2023).

For the search of Mesozoic diatoms, we investigated 33 study sites worldwide and found no fossil diatoms. We observed other siliceous microfossils such as radiolarians and sponge spicules and these were highly degraded, pointing towards post-depositional diagenetic processes (Paper III). Due to the unsuccessful search for the diatoms in Mesozoic sediments, we analyzed and revisited the fossil record which was already documented (Paper I and Paper IV). In Paper I we examined the oldest diatoms. First, based on the morphology and elemental composition of specimens, we have established that the oldest claimed diatoms (*Pyxidicula*) from the Lower Jurassic (Rothpletz, 1896) recovered from the trace fossil *Phymaroderma* in black shales were most likely calcareous nannofossils (*Schizosphaerella*). Second, based on morphological characteristics of specimens drawn by Rothpletz, we suggest that the Middle Jurassic *Pyxidicula* is likely a testate amoeba and not a diatom (Rothpletz, 1900). Lastly, although we did not recover equivalent sediment type to Rothpletz's "*Spongiolites fellenbergii"* from Middle Jurassic, we have investigated (Paper III) sediments of equivalent age from the same locality (Palfris) in different lithology. No microfossils were observed, neither diatoms, nor testate amoeba. Through this inference, we increased the fossil record gap to 80 million years spanning from the origin time, which has been estimated at 200 Ma (Nakov et al., 2018) to the occurrence of the oldest described diatom communities at 121 Ma from Australia (Nikolaev et al., 2001b).

The main conclusions of Paper I were: 1) *Pyxidicula* was used as a calibration in the molecular clocks (e.g., Nakov et al., 2018), and this practice should be discontinued; 2) through three examples we show the relevance of stringent evaluation when describing new fossils (elemental composition analysis, exclusion due to contamination, age control). We suggest that a detailed evaluation should be an obligatory practice whilst publishing new diatom records; 3) based on the analysis of the previous literature we infer that the first diatom described in the *Pyxidicula* genus (*P. operculata* Ehrenberg) was established based on a testate amoebae specimen that exhibited some diatom characteristics (e.g., cell wall made of two pieces and ornamented surface) (Fig. 12). Regardless of the later transfers to testate amoeba (Lachmann, 1858), *Pyxidicula* has been continuously used in diatom taxonomy (Pritchard, 1861; Nakov et al., 2018). We suggest *Pyxidicula* Ehrenberg should be recognized exclusively as a testate amoeba genus.

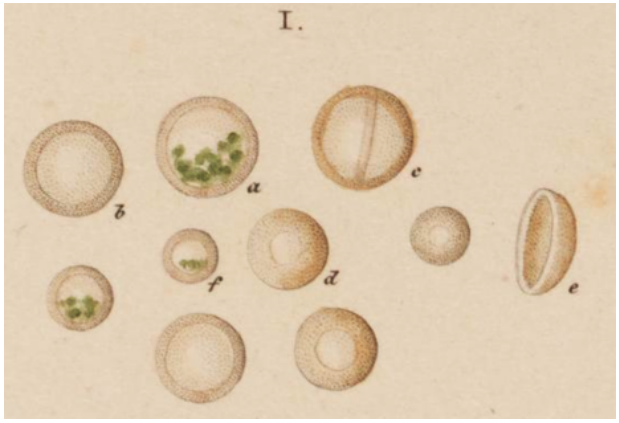

Fig. 12. *Pyxidicula operculata* Ehrenberg. Figure from Ehrenberg, 1838.

The main conclusions of Paper III are: 1) the absence of the diatoms in the fossil record of the Lower Cretaceous and Jurassic is likely an outcome of diagenesis where the original opal-A component was affected by dissolution-reprecipitation processes and conversion of opal-A to more crystalline form of opal-CT and quartz or alternatively original opal-A was substituted by pyrite crystals; 2) siliceous microfossils (sponge spicules and radiolarians) were preserved in structures such as coprolites, sponge mummies, phosphorites and in cherts which are likely connected to early cementation and were absent in the majority of the calcareous rocks we investigated; oldest diatom communities were likewise preserved in concretions; 3) Cretaceous diatom deposits with well-preserved abundant frustules were predominantly found in higher latitudes in sediments deposited in shallow and shelf environments that experienced shallow burial; 4) the future search for diatoms in geological archives should focus on sediments that experienced shallow burial and low heat, preferably deposited in higher latitudes and shallow marine environment (Barron, 1993); lithology wise clayey facies retard opal to quartz conversion (Riech and von Rad, 1979) and potentially these should be a target together with structures such as concretions. Although we have not added new species to the fossil record, our findings should have a large impact on the understanding of diatom evolution.

The fossil record gap of 80 Ma and the lack of fossils at the study sites inspired Paper IV. In Paper IV we compiled the Cretaceous Diatom Database. The database contains all the diatoms described in the Cretaceous fossil record, with corresponding stratigraphic range and species richness of genera. Through the analysis of the database, we show: 1) Cretaceous diversity is substantially higher than previously proposed; 2) Most species-rich are genera of extant representatives; 3) a third of the Cretaceous genera is monotypic; 4) by the Upper Cretaceous, representatives of all morphological groups were present. We also highlighted the areas for future taxonomical work. For example, the revision of the photographic documentation suggests that, e.g., *Triceratium* doesn't appear in the Cretaceous at all, yet 16 different species were described in the Cretaceous literature (example Fig. 13). All the documented forms of *Triceratium* lack genus obligatory feature (synapomorphy) which is apical ocellus (Round et al., 1990). We suggest these taxa should be reclassified to another multipolar genus, potentially.

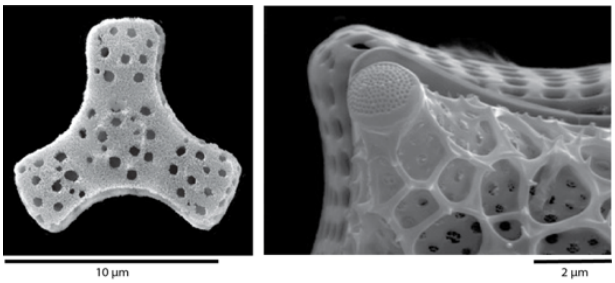

Fig. 13. Left–*Triceratium sp.* (Shimada et al., 2022), right–*Triceratium dubium* (Ashworth et al., 2013) with apical ocellus.

8.2 Calibration points for the molecular clock studies

There are several methods to construct molecular clocks and one of them requires an array of fossil calibration points to tie phylogenies to absolute ages (Tiley et al., 2020). In Paper IV we present a set of well-documented calibration points derived through the analysis of photographs provided in the Cretaceous Diatom Database. We chose the tie points based on the morphological evidence justifying the placement of the fossil in the chosen taxonomical level and accurate dating of the sediments containing these fossils (Parham et al., 2011). This set is of high value for future

molecular clock studies, since many previously used tie points were erroneous. For example, pyritised fossils which lack visible synapomorphies (Fig. 14) (Geroch, 1978) or fossils overturned due to contamination (Girard et al., 2009) should not be included as tie points. To maximise the use of the fossil information we proposed crown and stem calibrations. For the crown calibrations we chose taxa that exhibit all the genus obligatory characteristics (synapomorphies) and for the stem calibrations we chose taxa that exhibit some genus characters, and those that are present suggest clear association (Fig. 15).

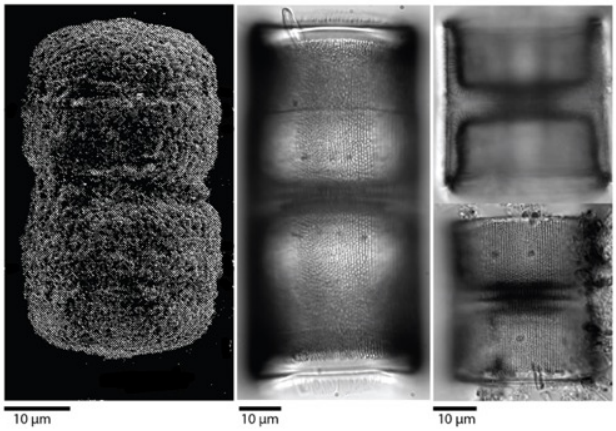

Fig. 14. Left– Cretaceous *Melosira* (Geroch 1996), middle and right–extant *Melosira species* (Spoulding and Edlund, 2008). Pyritised fossil (left) lack detailed frustule characters of *Melosira*.

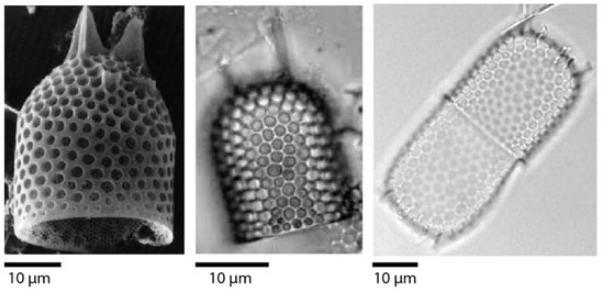

Fig. 15. Left–Cretaceous *Amblypyrgus* (Gersonde and Harwood, 1990), middle-Paleocene *Stephanopyxis* (66-56 Ma), right-extant *Stephanopyxis* (http://plankton.image.coocan.jp/).

For example, we suggest extinct *Amblypurgus* as a stem calibration for *Stephanopyxis*. *Stephanopyxis* characters are the following: domed valves; large hexagonal areolae opened by large foramina with underlying continuous sheet of silica, ring of tubular processes on the valve face and ring of marginal rimpoportulae (Round et al., 1990). Extinct *Amblypyrgus* exhibits the above characters but processes on the valve face are not tubes but spines and some valves lack spines (Fig. 15). These characters suggest *Amblypyrgus* may be ancestral to *Stephanopyxis* or may represent extinct sister lineage. In both cases it can be used as a calibration point in molecular clocks.

In Paper IV we also show that many taxa published under extant genus names were either not documented with photographs or were most likely misclassified (Fig. 13). Therefore, molecular biologists wanting to retrieve accurate information from the Cretaceous literature either will not use such fossils as calibration nodes or will use them without proper assessment (which in the ideal situation should not be necessary). This can disrupt the inferences of molecular clocks as either the wrong information is used or the correct information is unused.

We advise that future taxonomical studies (see Paper I), especially on Cretaceous diatoms, should focus on the analysis of the character and discussion where the extinct fossils could be placed on time-calibrated phylogenies. Such fossils should always be documented with photographs and, when possible, SEM photographs. Overall, our dataset represents valuable resources for future evolutionary studies of modern and fossil diatoms.

8.3 Updated codon substitution methods bring new insight into the evolution of SITs

Evidence across several diverse diatoms suggests that gene duplication and functional differentiation have played important roles in the evolution of diatom SITs, with important consequences for silicon metabolism (Martin–Jézéquel et al., 2000; Durkin et al., 2016). There is less evidence for adaptive evolution at the sequence level, where diatom SITs have been shown to evolve predominantly under strong purifying selection (Thamatrakoln et al., 2006; Alverson, 2007). Previous tests on natural selection in SITs of diatoms were based on older

models implemented in PAML software, which assume a constant substitution rate across the sites and branches on the phylogeny (Kosakovsky Pond and Muse, 2005), which is unrealistic for most empirical datasets. We used HyPhy, which implements a much broader range of models that allow for more specific and nuanced hypothesis testing. A great example is the RELAX model, designed to specifically test for relaxed selection in a set of foreground branches (Wertheim et al., 2015). We used RELAX to test for relaxed selection on freshwater branches and found some support for this hypothesis. Another example is the test for episodic selection, which has increased power to detect positively selected sites amid a sea of purifying selection (Murrell et al., 2012), as is the case for SITs. These tests are not implemented in PAML, which was used by Alverson (2007) in a previous study on *Thalassiosirales* SITs.

Further advancements were achieved by whole genome sequencing in comparison to the PCR method used by Alverson (2007). PCR method failed to amplify and sequence SIT3 paralog in Thalassiosirales and therefore, the gene tree and history of gene duplication presented by Alverson (2007) was based only on SIT1-2, missing the history of the highly divergent SIT3. Whole genome sequencing allowed us to retrieve both SIT1-2 and SIT3 sequences across a full phylogenetic breadth of *Thalassiosirales,* and we show SIT3 is widespread across genera and habitats. Furthermore, we established a divergent history of gene duplication and loss between SIT3 and SIT1-2 paralogs and inferred these emerged through whole genome duplication some 80 Ma (Parks et al., 2018). We hypothesize that following whole genome duplication, both paralogs (here ohnologs) had equal functions, and with time, SIT3 accumulated deleterious mutations, which overall decreased its functionality. Indeed, the RELAX test inferred SIT3 to be under relaxed selection. Retention of SIT3 in 17 diatoms, however, suggests it likely has a function, but this function is not crucial to diatom's fitness, and its loss has no consequences for DSi uptake.

Measured transcript levels of 7 diatoms, revealed that join expression of SIT1-2 paralogs exceed the levels of SIT3 for all but one diatom–Cyclostephanos invisitatus (Fig. 16). Although SIT3 in some cases exceeds the transcript levels of SIT1-2 paralog the definitive functional study on Thalassiosirales SITs (Thamatrakoln and Hildebrand, 2007) clearly showed that: a) SIT1 and SIT2 actively transport dissolved silicon (DSi) and b) SIT3 is expressed at levels 40–80-fold lower than SITs 1 and 2 and may not transport DSi. This functional divergence is consistent with the SIT gene phylogeny shown in Paper II, where SITs 1 and 2 comprise one large clade, and SIT3 another clade. Based on these collective findings, we summed across SITs 1 and 2 as a way to show total expression of "active" SITs.

Transcript levels of SIT3 in Cyclostephanos invisitatus presumably suggests this taxon utilizes SIT3 for DSi uptake, we also show it was under intensified selection (where both positive and purifying selection are enhanced) relative to other SIT3 paralogs. Further, we blasted (finding region of interest in the genome) SIT sequences from individual species against their genomes and found that SITs 1 and 2 of C. invisitatus were tandemly duplicated (in simple words present next to each other in the genome) unlike any other SITs in our data set. There is no supportive evidence to suggest that tandemly duplicated SIT1 and SIT2 are unable to perform the function of DSi uptake, and were substituted by SIT3. Moreover, genomes of Thalassiosirales are not fully assembled hence the above inference of tandem duplication may be erroneous. Further studies are therefore needed to unravel the DSi uptake in C. invisitatus. Unlike our measurements, such studies should measure transcript levels in synchronized cultures to determine whether SIT3 is indeed expressed at the highest levels during the S phase of the cell cycle prior to the formation of the new frustule (Thamatrakoln and Hildebrand, 2007).

In Paper II, we show the power of a genome-derived dataset and the advantage of updated codon models to uncover the evolutionary dynamics of metabolically important genes, including the history of gene duplication and loss and the strength of natural selection across paralogs and habitats.

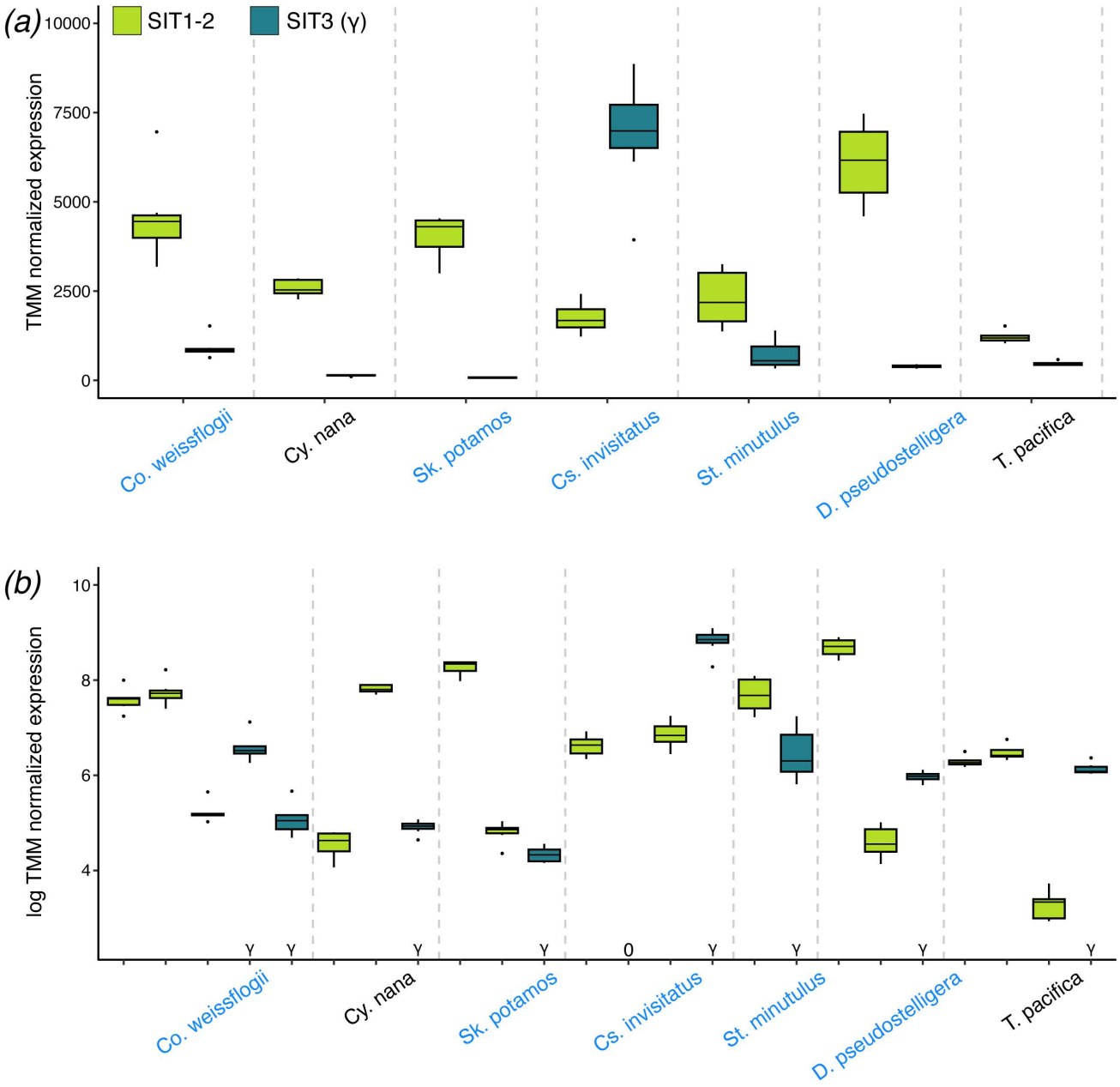

Fig. 16**.** Transcriptional profiles of divergent SIT paralogs in diatoms, with transcript levels summed across paralogs for SIT1-2 versus SIT3 (a) or transcript level per paralog (b). In panel (b), unexpressed paralogs are labeled '0' and for clarity, SIT3 paralogs are labeled 'γ'. Marine taxa are in black and freshwater taxa are in blue. Genus abbreviations: *Conticribra* (Co), *Cyclotella* (Cy), *Cyclostephanos* (Cs), *Discostella* (D), *Skeletonema* (Sk), *Stephanodiscus* (St), *Thalassiosira* (T).

Figure prepared by Eveline Pinseel.

9 Conclusions

1. The oldest diatoms previously identified in the literature from the Lower and Middle Jurassic ages are not diatoms and likely represent calcareous nannofossils and testate amoebae respectively. Evaluation of the elemental composition of the fossils using SEM-EDS and removal of contaminants should be a standard procedure when publishing new records of diatoms.

2. The absence of diatoms in the Mesozoic sediments and the poor preservation of other siliceous microfossils highlights the role of diagenesis in the ancient sediments. Conversion of opal-A to opal-CT and quartz or pyritization exhibits dissolution-reprecipitation dynamics that likely erased the original biogenic component of diatoms in study sites we have examined.

3. Worldwide dispersal of diatoms in the Early Cretaceous and the diversity of the oldest diatoms from Australia, which likely belong to several lineages, suggests prior evolutionary events. Diatoms needed time to diversify, disperse and prevail in new environments to establish communities which were further preserved in sediments. We therefore suggest that communities older than 120 Ma are yet to be discovered.

4. Further search for Mesozoic diatoms must continue, but should focus on unlithified sediments that experienced shallow burial. Such sediments should be deposited in shallow seas or shelf environments of higher latitudes. Lithology wise, clayey facies and concretions, coprolites, cherts and sponge mummies, should be a target.

5. The Cretaceous fossil record of diatoms is burdened by taxonomical inconsistencies and a lack of extensive documentation of observed fossils through light and scanning electron microscopes. We highlight the areas where such research should be carried out.

6. The Cretaceous period is a time of diversification of many taxa that could be implemented as calibrations points in future molecular clocks. These calibrations will improve and constrain the inferences on the origin and the early evolution of diatoms.

7. The molecular evolution of silicon transporter proteins (SITs) of diatoms is complex and dynamic. Small-scale and whole-genome duplication events, adaptive processes, and environmental changes have shaped SIT sequence diversity, copy number, and gene expression. Complex evolutionary dynamics of SITs suggest optimizing the silicon transport in diatoms is a difficult or moving target.

8. Combination of whole genome sequencing and the use of updated codon substitution methods is advantageous when uncovering the evolutionary dynamics of ecologically and metabolically important genes.

9.1 Popular summary

To study the evolution of the organism, scientists use two approaches: 1) the fossil record which evolves around studying organismal remains and their traces in ancient sediments, 2) DNA sequencing which focuses on studying genes of living organisms. Both approaches may answer different evolutionary questions. The study of fossil diatoms can reveal what organism were living in a certain locality at a certain time and how they were changing morphologically through millions of years. DNA sequencing may reveal the relationships and lines of descent between different organisms and if combined with the fossil record can be applied to long time scales. In this Ph.D. both methods were applied and discussed in four papers to unravel the origin and evolution of diatoms.

Diatoms are microscopic organisms with siliceous cell coverings called a frustule and are one of the world's most prominent photosynthetic organisms that play a crucial role in today's oceans. Diatoms regulate important biogeochemical cycles (oxygen, carbon, silica) and produce fats and sugars that are nutritionally important for higher organisms. Diatoms first appeared in the fossil record some 120 million years ago (Ma) in the Lower Cretaceous, but based on evolutionary relationships of diatoms inferred from DNA sequences, it was estimated that diatoms originate ca. 200 Ma near the Triassic-Jurassic boundary. There were also diatoms described from the Jurassic (182 Ma and 165 Ma), but these diatoms have been poorly documented.

To learn more about the earliest diatoms and their evolution, we searched for the diatoms in sediments around the world, revisited the Jurassic diatoms and analysed existing literature on Cretaceous diatoms. In one project we established that the Jurassic diatoms were misidentified and represent other organisms. Furthermore, no new fossils were discovered, however based on the state of preservation of other siliceous microfossils we suggest that Cretaceous and Jurassic diatoms were dissolved through processes occurring during the transformation from sediments to rocks known as diagenesis. Analysis of the literature on Cretaceous diatoms revealed that: 1) diatoms were dispersed worldwide already in the Lower Cretaceous, 2) the oldest diatom communities found are represented by diverse morphologies suggesting prior evolutionary events, and 3) wide array of Cretaceous fossils can be implemented into DNA studies to establish origin and the evolution of diatom lineages. We suggest older diatoms than so far described are yet to be discovered.

The fourth project in this Ph.D. focused on evolution of genes encoding specialised proteins in diatoms called silicon transporter proteins (SITs) that diatoms use to uptake silicic acid from the surrounding water to build their frustules. Many factors shape the evolution of protein-coding genes. For example, their function can be optimized by adaptive mutations that change an amino acid or by altering gene expression dynamics through gene duplications or losses. Our results show that SITs have an extensive and ongoing history of gene duplication and loss. Numerous sources of data together suggest that optimization of gene expression has played a central role in shaping sequence evolution and gene family dynamics of diatom SITs.

This Ph.D. project combines perspectives of paleontology and DNA based studies. Articles in this dissertation are of high importance for future evolutionary studies of fossil and modern diatoms.

9.2 Podsumowanie popularnonaukowe

Aby zbadać ewolucję organizmów, naukowcy wykorzystują dwa sposoby: 1) zapis kopalny, który ewoluuje wokół badania szczątków organizmów i ich śladów w dawnych osadach, 2) sekwencjonowanie DNA, które koncentruje się na badaniu genów żywych organizmów. Oba sposoby mogą odpowiedzieć na różne pytania ewolucyjne, pierwszy może ujawnić, jakie organizmy żyły w określonym miejscu w określonym czasie i jak zmieniały się morfologicznie przez miliony lat. Drugi może ujawnić relacje i linie pochodzenia między różnymi organizmami, a w połączeniu z zapisem kopalnym może być aplikowany do długich skalach czasowych. W tym doktoracie obie metody zostały zastosowane i omówione w czterech artykułach w celu odkrycia pochodzenia i ewolucji okrzemek.

Okrzemki są mikroskopijnymi organizmami z krzemionkowymi pancerzykami zwanymi okrywami i są jednymi z najbardziej znanych na świecie organizmów fotosyntetyzujących, które odgrywają kluczową rolę w dzisiejszych oceanach. Okrzemki regulują ważne cykle biogeochemiczne (tlen, węgiel, krzemionka) i wytwarzają tłuszcze i cukry, które są ważne pod względem odżywczym dla organizmów wyższych. Okrzemki po raz pierwszy pojawiły się w zapisie kopalnym około 120 milionów lat temu (Ma) w dolnej Kredzie, ale na podstawie relacji ewolucyjnych wywnioskowanych z sekwencji DNA oszacowano, że okrzemki wyewoluowały ok. 200 Ma w pobliżu granicy Triasu i Jury. Istniały również okrzemki opisane z okresu Jurajskiego (182 Ma i 165 Ma), ale zostały one słabo udokumentowane.

Aby dowiedzieć się więcej o najwcześniejszych okrzemkach i ich ewolucji, poszukiwaliśmy okrzemek w osadach na całym świecie, ponownie zbadaliśmy okrzemki Jurajskie i przeanalizowaliśmy istniejącą literaturę na temat okrzemek Kredowych. W jednym projektach ustaliliśmy, że okrzemki Jurajskie zostały błędnie zidentyfikowane i reprezentują inne organizmy. Ponadto nie odkryliśmy żadnych nowych skamieniałości okrzemkowych, jednak w oparciu o stan zachowania innych mikroskamieniałości krzemionkowych sugerujemy, że okrzemki Kredowe i Jurajskie zostały rozpuszczone w wyniku procesów zachodzących podczas transformacji osadów w skały, znanych jako diageneza. Analiza literatury na temat Kredowych okrzemek wykazała, że: 1) okrzemki były rozproszone na całym świecie już w dolnej Kredzie, 2) najstarsze znalezione zespoły okrzemkowe są reprezentowane przez różnorodne morfologie sugerujące wcześniejsze zdarzenia ewolucyjne oraz 3) szeroki wachlarz skamieniałości Kredowych może być wdrożony do badań DNA w celu ustalenia pochodzenia i ewolucji okrzemek. Sugerujemy, że okrzemki starsze niż dotychczas opisane nie zostały jeszcze odkryte.

Czwarty projekt w ramach tego doktoratu koncentrował się na ewolucji genów kodujących wyspecjalizowane białka okrzemek zwane białkami transportującymi krzem (po angielsku: SIT), które okrzemki wykorzystują do pobierania kwasu krzemowego z otaczającej wody w celu budowy swoich okryw. Ewolucję genów kodujących białka kształtuje wiele czynników. Na przykład ich funkcja może zostać zoptymalizowana przez mutacje adaptacyjne, które zmieniają aminokwas lub przez zmianę dynamiki ekspresji genów poprzez duplikacje lub utratę genów. Nasze wyniki pokazują, że SIT mają rozległą i ciągłą historię duplikacji i utraty genów. Liczne źródła danych sugerują, że optymalizacja ekspresji genów odegrała kluczową rolę w kształtowaniu ewolucji sekwencji i dynamiki rodziny genów SIT okrzemek.

Ten projekt doktorancki łączy perspektywy paleontologii i badań opartych na DNA. Opracowane przez nas artykuły mają duże znaczenie dla przyszłych badań ewolucyjnych kopalnych i współczesnych okrzemek.

9.3 Populär sammanfattning

För att studera en organisms utveckling använder forskare två metoder: 1) fossila arkiv som bygger på studier av organismrester och deras spår i gamla sediment, 2) DNA-sekvensering som fokuserar på studier av levande organismers gener. De båda metoderna kan ge svar på olika evolutionära frågor, den första kan avslöja vilka organismer som levde på en viss plats vid en viss tidpunkt och hur de förändrats morfologiskt under miljontals år. Den andra metoden kan avslöja släktskap mellan olika organismer och hur de utvecklats och kan i kombination med fossila fynd tillämpas på långa tidsskalor. I denna doktorsavhandling har båda metoderna använts och diskuterats i fyra artiklar för att ta reda på kiselalgernas ursprung och utveckling. Kiselalger är mikroskopiska organismer med kiselhaltiga cellöverdrag som kan kallas skal. De är bland världens mest framträdande fotosyntetiserande organismer och spelar en avgörande roll i dagens hav. Kiselalger reglerar den viktiga biogeokemiska omsättningen av ämnen som syre, kol, kiseldioxid och producerar fetter och sockerarter som är näringsmässigt viktiga för högre organismer. De äldsta fossila fynden av kiselalger är från kiselalger som levde för cirka 120 miljoner år sedan (Ma) i den tidiga kritaperioden, men baserat på evolutionärt släktskap mellan kiselalger som härletts från DNA-sekvenser, har det uppskattats att kiselalger har sitt ursprung för ca. 200 Ma, nära gränsen mellan trias och jura.

Det har också funnits kiselalger beskrivna från jura (182 Ma och 165 Ma), men dessa kiselalger har varit dåligt dokumenterade. För att lära oss mer om de tidigaste kiselalgerna och deras utveckling sökte vi efter kiselalger i sediment runt om i världen, återbesökte kiselalgerna från juratiden och analyserade befintlig litteratur om kiselalger från kritatiden. I detta doktorandprojekt fastställde vi att kiselalgerna från jura var felidentifierade och i själva verket representerade andra organismer. Inga nya fossil upptäcktes, men baserat på bevarandegraden för andra kiselhaltiga mikrofossil föreslår vi att kiselalger från krita och jura löstes upp i de processer som inträffar under omvandlingen från sediment till sten, en process som benämns diagenes. En analys av litteraturen om kiselalger från krita visade att: 1) kiselalger spreds över hela världen redan under den tidiga delen av kritaperioden, 2) de äldsta kiselalgerna som hittats har morfologiska drag som tyder på tidigare evolutionär utveckling, och 3) ett brett utbud av fossil från krita kan användas i DNA-studier för att fastställa ursprung och utveckling av kiselalger. Vi föreslår att har funnits äldre kiselalger än de som hittills beskrivits och att de skulle kunna hittas framöver.

Det fjärde projektet i denna doktorandstudie fokuserade på evolutionen av gener som kodar för specialiserade proteiner i kiselalger, så kallade kiseltransportproteiner (SIT), som kiselalger använder för att ta upp kiselsyra från det omgivande vattnet och bygga sina skal. Många faktorer påverkar utvecklingen av proteinkodande gener. Till exempel kan deras funktion optimeras genom adaptiva mutationer som ändrar en aminosyra eller genom att förändra genuttryckets dynamik genom genduplikationer eller genförluster. Våra resultat visar att SIT har en omfattande och pågående historia av genduplicering och genförlust. Många datakällor tyder tillsammans på att optimering av genuttryck har spelat en central roll för att forma sekvensutveckling och genfamiljsdynamik hos kiselalgernas SIT:er.

Detta doktorsavhandling kombinerar perspektiv från paleontologi och DNA-baserade studier och resultaten är av stor betydelse för framtida evolutionära studier av fossila och moderna kiselalger.

10 References

Abascal, F., Zardoya, R., and Telford M. J. (2010). TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*. *38*, 7–13.

Abbott, W. H. (1978). Cretaceous diatoms from the Peedee formation of South Carolina. *Geologic Notes*, *22*(2), 105–108.

Ahlgren, G., Lundstedt, L., Brett, M., and Forsberg, C. (1990). Lipid composition and food quality of some freshwater phytoplankton for cladoceran zooplankters. *Journal of Plankton Research*, *12*(4), 809–818.

Alldredge, A. L., and Gotschalk, C. C. (1989). Direct blooms: characteristics, settling velocities and formation of diatom aggregates. *Deep-Sea Research. Part A, Oceanographic Research Papers*, *36*(2), 159–171.

Alverson, A. J. (2007). Strong purifying selection in the silicon transporters of marine and freshwater diatoms. *Limnology and Oceanography*, *52*(4), 1420–1429.

Alverson, A. J., Beszteri, B., Julius, M. L., and 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.

Angelis, K., and Dos Reis, M., 2015. The impact of ancestral population size and incomplete lineage sorting on Bayesian estimation of species divergence times. *Current Zoology, 61*, 874–885.

Anonymous (1702). Two Letters from a Gentleman in the Country, Relating to Mr Leuwenhoeck's Letter in Transaction, No. 283. Communicated by Mr C. *Philosophical Transactions, (1683-1775)*, 1494–1501.

Arena, D. A. (2008). Exceptional preservation of plants and invertebrates by phosphatization, Riversleigh, Australia. *Palaios*, *23*(7), 495–502.

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S., Allen, A. E., Apt, K. E., Bechner, M., Brzezinski, M. A., Chaal, B. K., Chiovitti, A., Davis, A. K., Demarest, M. S., Detter, J. C., Glavina, T., Goodstein, D., Hadi, M. Z., et al. (2004). The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. *Science*, *306*(5693), 79–86.

Ashworth, M.P., Nakov, T. and Theriot, E.C. (2013). Revisiting Ross and Sims (1971): toward a molecular phylogeny of the Biddulphiaceae and Eupodiscaceae (Bacillariophyceae). *Journal of Phycology,* 49(6), 1207-1222.

Baines, S. B., Twining, B. S., Brzezinski, M. A., Krause, J. W., Vogt, S., Assael, D., and McDaniel, H. (2012). Significant silicon accumulation by marine picocyanobacteria. *Nature Geoscience*, *5*(12), 886–891.

Ballance, P. F., Barron, J. A., Blome, C. D., Bukry, D., Cawood, P. A., Chaproniere, G. C. H., Frisch, R., Herzer, R. H., Nelson, C. S., Quinterno, P., et al. (1989). Late Cretaceous pelagic sediments, volcanic ash and biotas from near the Louisville hotspot, Pacific Plate, paleolatitude 42° S. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *71*(3-4), 281–299.

Barron, J. A. (1985). Diatom biostratigraphy of the CESAR 6 core. *Geoogical. Survey of. Canada, Paper 84-22*, 137–148.

Barron, J. A. (1993). Diatoms, in: Lipps, J.E. (Ed.), Fossil Prokaryotes and Protists. *Blackwell Scientific Publications*, 155–167.

Barron, J. A., and Washington, W. M. (1982). Cretaceous climate: A comparison of atmospheric simulations with the geologic record. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *40*(1), 103–133.

Benda, L. (1982). Die Diatomeen des späten Apt. *Geologisches Jahrbuch. Reihe A, Allgemeine Und Regionale Geologie BR Deutschland Und Nachbargebiete, Tektonik, Stratigraphie, Paläontologie*, *65*, 405–411.

Benoiston, A. S., Ibarbalz, F. M., Bittner, L., Guidi, L., Jahn, O., Dutkiewicz, S., and Bowler, C. (2017). The evolution of diatoms and their biogeochemical functions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1728).

Bergstresser, T. J., and Krebs, W. N. (1983). Late Cretaceous (Campanian-

Maastrichtian) Diatoms from the Pierre Shale, Wyoming, Colorado and Kansas. *Journal of Paleontology*, *57*(5), 883–891.

Bryłka, K., Alverson, A. J., Pickering, R. A., Richoz, S., and Conley, D. J. (2023). Uncertainties surrounding the oldest fossil record of diatoms. *Scientific Reports*, *13*(1), 8047.

Budd, G. E., and Mann, R. P. (2020). The dynamics of stem and crown groups. *Science Advances*, *6*(8), eaaz1626.

Capella-Gutiérrez, S., Silla-Martínez, J. M., Gabaldón T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*, 1972–1973.

Chambers, P. M. (1997). *Late Cretaceous and Paleocene marine diatom floras*. University of London.

Chen, K., Durand, D., Farach-Colton, M. (2000). NOTUNG: A program for dating gene duplications and optimizing gene family trees. *Journal of Computational Bioogy*. *7*, 429–447.

Chin, K., Bloch, J., Sweet, A., Tweet, J., Eberle, J., Cumbaa, S., Witkowski, J., and Harwood, D. (2008). Life in a temperate Polar Sea: a unique taphonomic window on the structure of a Late Cretaceous Arctic marine ecosystem. *Proceedings. Biological Sciences / The Royal Society*, *275*(1652), 2675–2685.

Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T. C., Kellis, M., Gelbart, W., Iyer, V. N., Pollard, D. A., Sackton, T. B., Larracuente, A. M., Singh, N. D., Abad, J. P., Abt, D. N., Adryan, B., Aguade, M., Akashi, H., and Others (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, *450*(7167), 203–218.

Conley, D. J., Kilham, S. S., and Theriot, E. (1989). Differences in silica content between marine and freshwater diatoms. *Limnology and Oceanography*, *34*(1), 205–212.

Davies, A., and Kemp, A. E. S. (2016). Late Cretaceous seasonal palaeoclimatology and diatom palaeoecology from laminated sediments. *Cretaceous Research*, *65*, 82–111.

Dobson, L., Reményi, I., Tusnády, G.E. (2015). CCTOP: A Consensus Constrained TOPology prediction web server. *Nucleic Acids Researach*, 43, W408–12.

Donoghue, P. C. (2005). Saving the stem group–a contradiction in terms? *Paleobiology, 31(4)*, 553-558.

Deflandre, G. (1941). Sur la présence de Diatomées dans certains silex creux Turoniens et sur un nouveau mode de fossilisation de ces organismes. *Comptes Rendus de l'Académie Des Sciences de Paris*, *213*, 878–880.

Dell'Agnese, D. J. and Clark, D. L. (1994). Siliceous microfossils from the warm Late Cretaceous and early Cenozoic Arctic Ocean. *Journal of Paleontology*, *68*(1), 31–47.

Deokar, A. A., and Tar'an, B. (2016). Genome-Wide Analysis of the Aquaporin Gene Family in Chickpea (Cicer arietinum L.). *Frontiers in Plant Science*, *7*, 1802.

De Wever, P., Dumitrica, P., Caulet, J.P., Nigrini, C., Caridroit, M. (2002). Radiolarians in the sedimentary record. *CRC Press*.

Dun, W. S., Rands, W. H., and David, T. W. E. (1901). Note on the Occcurence of Diatoms, Radiolaria and Infusoria in the Rolling Downs Formation (Lower Cretaceous), Queensland. *Proceedings of Linnean Society, New South Wales*, *18*, 299–309.

Durkin, C. A., Koester, J. A., Bender, S. J., and Armbrust, E. V. (2016). The evolution of silicon transporters in diatoms. *Journal of Phycology*, *52*(5), 716–731.

Ehrenberg, C. G. (1838). Die Infusionsthierchen als vollkommene Organismen. Ein Blick in das tiefere organische Leben der Natur. *Leipzig, L. Voss*.

Emms, D.M. and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol*ogy, *20*, 238.

Lachmann, C. (1858). Etudes sur les Infusoires et les Rhizopodes. *Mémoires de l'Institut National Genevois, 5(6)*, 1859.

Fenner, J. (1982). Cretaceous diatoms off New Jersey. *7th Int. Symp. Living and Fossil*.

Forti, A. (1933). Contribuzioni diatomologiche XIY-Schulziella nov. nom. Dallas Hanna et

Forti. *Atti R. 1st. Veneto Sci. Lett. Arti*, *2*(92), 1279–1283.

Forti, A., and Schultz, P. (1932). *Erste Mitteilung über Diatomeen aus dem hannoverschen Gault*.

Foucault, A., Servant-Vildary, S., Nianqiao Fang, S., Fang, N., and Powichrowski, L. (1986). Un des plus anciens gisements de diatomées découvert dans l'Albien-Cénomanien inférieur des Alpes ligures (Italie). Remarques sur l'apparition de ces algues. *Comptes Rendus de l'Académie Des Sciences. Série 2, Mécanique, Physique, Chimie, Sciences de L'univers, Sciences de La Terre*, *303*(5), 397–402.

Fourtanier, E. (1991). Diatom biostratigraphy of Equatorial Indian Qcean Site 758. *Proc. ODP, Sci. Results*, *121*, 189–208.

Frings, P. J., Clymans, W., Fontorbe, G., De La Rocha, C. L., and Conley, D. J. (2016). The continental Si cycle and its impact on the ocean Si isotope budget. *Chemical Geology*, *425*, 12–36.

Frings, P. J., Clymans, W., Jeppesen, E., Lauridsen, T. L., Struyf, E., and Conley, D. J. (2014). Lack of steady-state in the global biogeochemical Si cycle: Emerging evidence from lake Si sequestration. *Biogeochemistry*, *117*(2), 255–277.

Gaines, R. R., Briggs, D. E. G., Orr, P. J., and Van Roy, P. (2012). Preservation of giant anomalocaridids in silica-chlorite concretions from the Early Ordovician of Morocco. *Palaios*, *27*(5), 317–325.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Baranova, O. K., Zweng, M. M., Reagan, J. R., and Johnson, D. R. (2013). *World ocean atlas 2013. Volume 4, Dissolved inorganic nutrients (phosphate, nitrate, silicate)*. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data and Information Service.

Georgi, K. H. (1978). Mikrofaunistisch-lithologische Untersuchungen der Hilssandstein-Region (Apt/Alb) im Raum Salzgitter-Goslar. *Mitteilungen aus dem Geologischen Institut der Universität Hannover*, *13*(13), 5–112.

Geroch, S. (1978). Lower Cretaceous diatoms in the Polish Carpathians. *Annales Societatis Geologorum Poloniae*, *48*(3-4), 283–295.

Gersonde, R., and Harwood, D. M. (1990). 25. Lower Cretaceous diatoms from ODP Leg 113 Site 693 (Weddell Sea). Part 1: Vegetative cells. *Proc ODP Sci Res*, *113*, 365–402.

Girard, V., Néraudeau, D., Breton, G., Martin, S. S., and Martin, J. P. S. (2009). Contamination of Amber Samples by Recent Microorganisms and Remediation Evidenced by Mid-Cretaceous Amber of France. *Geomicrobiology Journal*, *26*(1), 21–30.

Given, M. M. and Wall, J. H. (1971). Microfauna from the upper Cretaceous Bearpaw Formation of south-central Alberta. *Bulletin of Canadian Petroleum Geology*, *19*(2), 502–544.

Granum, E., Raven, J. A., and Leegood, R. C. (2005). How do marine diatoms fix 10 billion tonnes of inorganic carbon per year? *Canadian Journal of Botany. Journal Canadien de Botanique*, *83*(7), 898–908.

Gresham, C. W. (1985). Cretaceous and Paleocene siliceous phytoplankton assemblages from DSDP sites 216,214 and 208 in the Pacific and Indian Oceans. *University of Wisconsin–Madison*.

Haig, D. W., and Barnbaum, D. (1978). Early Cretaceous microfossils from the type Wallumbilla Formation, Surat Basin, Queensland. *Alcheringa: An Australasian Journal of Palaeontology*, *2*(2), 159–178.

Hajós, M. and Stradner, H. (1975). Late cretaceous archaeomonadaceae, diatomaceae, and silicoflagellatae from the south Pacific Ocean, deep sea drilling project, leg 29, site 275. *Initial Reports of the Deep Sea Drilling Project*, *29*, 913–1009.

Hakes, L., Pinney, J. W., Lovell, S. C., Oliver, S. G., and Robertson, D. L. (2007). All duplicates are not equal: the difference between small-scale and genome duplication. *Genome Biology*, *8*, 1–13.

Hanna, G. D. (1927). Cretaceous diatoms from California. *California Academy of Sciences, 13.* 5–49.

Harper, H. E. (1977). A lower Cretaceous (Aptian) diatom flora from Australia. *Nova Hedwigia*, *54*, 411–412.

Harwood, D. M. (1988). Upper Cretaceous and lower Paleocene diatom and silicoflagellate biostratigraphy of Seymour Island, eastern Antarctic Peninsula. *Geological Society of America Memoirs*, *169*, 55–130.

Harwood, D. M., and Gersonde, R. (1990). 26. Lower Cretaceous diatoms from ODP Leg 113 Site 693 (Weddell Sea). Part 2: resting spores, chrysophycean cysts, an endoskeletal dinoflagellate, and notes on the origin of diatoms. *Scientific Results Ocean Drilling Program*, *113*, 403–425.

Hildebrand, M. (2008). Diatoms, biomineralization processes, and genomics. *Chemical Reviews*, *108*(11), 4855–4874.

Hildebrand, M., Dahlin, K., and Volcani, B. E. (1998). Characterization of a silicon transporter gene family in Cylindrotheca fusiformis: sequences, expression analysis, and identification of homologs in other diatoms. *Molecular and General Genetics*, *260*(5), 480–486.

Hildebrand, M., Volcani, B. E., Gassmann, W., and Schroeder, J. I. (1997). A gene family of silicon transporters. *Nature*, *385*(6618), 688–689.

Hoffmann, A. A., and Willi, Y. (2008). Detecting genetic responses to environmental change. *Nature Reviews. Genetics*, *9*(6), 421–432.

Hollis, C. J., Rodgers, K. A., and Parker, R. J. (1995). Siliceous plankton bloom in the earliest Tertiary of Marlborough, New Zealand. *Geology*, *23*(9), 835–838.

Ho, S. Y. W., and Duchêne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. *Molecular Ecology*, *23*(24), 5947–5965.

Huang, Y., Temperley, N. D., Ren, L., Smith, J., Li, N., and Burt, D. W. (2011). Molecular evolution of the vertebrate TLR1 gene family--a complex history of gene duplication, gene conversion, positive selection and co-evolution. *BMC Evolutionary Biology*, *11*, 149.

Ichinomiya, M., Dos Santos, A. L., Gourvil, P., Yoshikawa, S., Kamiya, M., Ohki, K., Audic, S., de Vargas, C., Noël, M.-H., Vaulot, D., and Kuwata, A. (2016). Diversity and oceanic distribution of the Parmales (Bolidophyceae), a picoplanktonic group closely related to diatoms. *The ISME Journal*, *10*(10), 2419–2434.

Jewson, D. H. and Harwood, D. M. (2017). Diatom life cycles and ecology in the Cretaceous. *Journal of phycology* 53: 616–628.

Jewson, D. H., Granin, N. G., Zhdanov, A. A., Gorbunova, L. A., Bondarenko, N. A., and Gnatovsky, R. Y. (2008). Resting stages and ecology of the planktonic diatom Aulacoseira skvortzowii in Lake Baikal. *Limnology and Oceanography*, *53*(3), 1125–1136.

Jin, X., Gruber, N., Dunne, J. P., Sarmiento, J. L., and Armstrong, R. A. (2006). Diagnosing the contribution of phytoplankton functional groups to the production and export of particulate organic carbon, $CaCO_3$, and opal from global nutrient and alkalinity distributions. *Global Biogeochemical Cycles*, *20*(2).

Jousé, A. P. (1949). Diatoms from Mesozoic deposits. *Diatomeen Analysiz. Gosdeolizdat, Leningrad, Pg*.

Julius, M. L., and Theriot, E. C. (2010). The diatoms: a primer. In *The Diatoms: Applications for the Environmental and Earth Sciences*. Cambridge University Press.

Kemper, E. (1975). Zur biostratigraphie und palaekologie der schichtenfolge ober-apt/unter-alb im beckenzentrum noerdlich und oestlich von Hannover. Bericht der Naturhistorischen Gesellschaft zu Hannover, *0019*, 49–85.

Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G. M., Richardson, P., and Chisholm, S. W. (2007). Patterns and implications of gene gain and loss in the evolution of Prochlorococcus. *PLoS Genetics*, *3*(12), e231.

Knight, M. J., Senior, L., Nancolas, B., Ratcliffe, S., and Curnow, P. (2016). Direct evidence of the molecular basis for biological silicon transport. *Nature Communications*, *7*(1), 11926.

Knoll, A. H. and Follows, M. J. (2016). A bottom-up perspective on ecosystem change in Mesozoic oceans. *Proceedings. Biological sciences / The Royal Society* 283. royalsocietypublishing.org.

Kooistra, W. H. C. F., De Stefano, M., Mann, D.G., and Medlin, L. K., (2003). The phylogeny of the diatoms. *Progress in Molecular and Subcellular Biology*, *33*, 59-97.

Kosakovsky Pond, S. L., Frost, S. D. W., Muse S.V. (2005). HyPhy: Hypothesis testing using phylogenies. *Bioinformatics*, *21*, 676–679.

Koutsoukos, E. A. M., and Hart, M. B. (1990). Radiolarians and diatoms from the mid-Cretaceous successions of the Sergipe Basin, northeastern Brazil: Palaeoceanographic assessment. *Journal of Micropalaeontology*, *9*(1), 45–63.

Kuwata, A., and Takahashi, M. (1990). Life-form population responses of a marine planktonic diatom, *Chaetoceros pseudocurvisetus*, to oligotrophication in regionally upwelled water. *Marine Biology*, *107*(3), 503–512.

Lazarus D., Barron J., Renaudie J., Diver, P. and Türke A. (2014). Cenozoic planktonic marine diatom diversity and correlation to climate change. *PloS one* 9, e84857.

Lee, J. J., and Correia, M. (2005). Endosymbiotic diatoms from previously unsampled habitats. *Symbiosis, 38*, 251–260.

Lee, M. S. Y., Ho, S. Y. W., 2016. Molecular clocks. *Curr. Biol. 26*, R399–402.

Lewin, J. C., and Guillard, R. R. L. (1963). Diatoms. *Annual Review of Microbiology*, *17*(1), 373–414.

Lewin, J. C., and Lewin, R. A. (1960). Auxotrophy and heterotrophy in marine littoral diatoms. *Canadian Journal of Microbiology*, *6*, 127–134.

Long, J. A., Fuge, D. P., and Smith, J. (1946). Diatoms of the Moreno Shale. *Journal of Paleontology*, *20*(2), 89–118.

Loucaides, S., Van Cappellen, P., Roubeix, V., Moriceau, B., and Ragueneau, O. (2012). Controls on the Recycling and Preservation of Biogenic Silica from Biomineralization to Burial. *Silicon Chemistry*, *4*(1), 7–22.

Lozano-Fernandez, J. (2022). A Practical Guide to Design and Assess a Phylogenomic Study. *Genome Biology and Evolution*, *14*(9).

Luo, H. (2015). The use of evolutionary approaches to understand single cell genomes. *Frontiers in Microbiology*, *6*, 174.

Mann, D. (1989). The origins of the diatom and its life cycle. *The Chromophyte Algae: Problems and Perspectives*, *38*, 307–323.

Mao, Y., Peng, T., Shao, F., Zhao, Q., and Peng, Z. (2023). Molecular evolution of the hemoglobin gene family across vertebrates. *Genetica*, *151*(3), 201–213.

Marchetti, A., Maldonado, M. T., Lane, E. S., and Harrison, P. J. (2006). Iron requirements of the pennate diatom Pseudonitzschia: Comparison of oceanic (high-nitrate, low-chlorophyll waters) and coastal species. *Limnology and Oceanography*, *51*(5), 2092–2101.

Marinov, I., Doney, S. C., and Lima, I. D. (2010). Response of ocean phytoplankton community structure to climate change over the 21st century: partitioning the effects of nutrients, temperature and light. *Biogeosciences, 7*(12), 3941–3959.

Marron, A. O., Ratcliffe, S., Wheeler, G. L., Goldstein, R. E., King, N., Not, F., De Vargas, C., and Richter, D. J. (2016). The evolution of silicon transport in eukaryotes. *Molecular Biology and Evolution*, *33*(12), 3226–3248.

Martin–Jézéquel, V., Hildebrand, M., and Brzezinski, M. A. (2000). Silicon metabolism in diatoms: implications for growth. *Journal of Phycology*, *36*(5), 821–840.

Mccoy, V. E., Young, R. T., and Briggs, D. E. G. (2015). Factors controlling exceptional preservation in concrections. *Palaios*, *30*(4), 272–280.

McMinn, A. (1995). Comparison of diatom preservation between oxic and anoxic basins in Ellis Fjord, Antarctica. *Diatom Research: The Journal of the International Society for Diatom Research*, *10*(1), 145–151.

Medlin, L. K., and Kaczmarska, I. (2004). Evolution of the diatoms: V. Morphological and cytological support for the major clades

and a taxonomic revision. *Phycologia*, *43*(3), 245–270.

Melendez, I., Grice, K., Trinajstic, K., Ladjavardi, M., Greenwood, P., and Thompson, K. (2013). Biomarkers reveal the role of photic zone euxinia in exceptional fossil preservation: An organic geochemical perspective. *Geology*, *41*(2), 123–126.

Mikolajewski, D. J., Scharnweber, K., Jiang, B., Leicht, S., Mauersberger, R., and Johansson, F. (2016). Changing the habitat: the evolution of intercorrelated traits to escape from predators. *Journal of Evolutionary Biology*, *29*(7), 1394–1405.

Moshkovitz, S., Ehrlich, A., and Soudry, D. (1983). Siliceous microfossils of the Upper Cretaceous Mishash Formation, Central Negev, Israel. *Cretaceous Research*, *4*(2), 173–194.

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, *8*(7), e1002764.

Nakov, T., Beaulieu, J. M., and Alverson, A. J. (2018). Accelerated diversification is related to life history and locomotion in a hyperdiverse lineage of microbial eukaryotes (Diatoms, Bacillariophyta). *The New Phytologist*, *219*(1), 462–473.

Nei, M., and Kumar, S. (2000). Molecular Evolution and Phylogenetics. *Oxford University Press*.

Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A., and Quéguiner, B. (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*, *9*(3), 359–372.

Nikolaev, V.A. & Harwood, D.M. (2000). Diversity and system of classification of centric diatoms. In: The origin and early evolution of the diatoms: fossil, molecular and biogeographical approaches, 37-53. *Krakow: W. Szafer Institute of Botany, Polish Academy of Sciences.*

Nikolaev, V. A., and Harwood, D. M. (2000b). Morphology and taxonomic position of the Late Cretaceous diatom genus Pomphodiscus Barker and Meakin. *Micropaleontology*, *46*(2), 167–177.

Nikolaev, V. A., Kociolek, J. P., Fourtanier, E., Barron, J. A., and Harwood, D. M. (2001a). Late Cretaceous diatoms (*Bacillariophyceae*) from the Marca Shale Member of the Moreno Formation, California. *Occasional Papers of the California Academy of Sciences*, *152*, 1–119.

Nikolaev, V. L., Harwood, D. M., and Samsonov, N. I. (2001b). Early Cretaceous diatoms. *Saint Petersburg: Nauka*.

Olsen, S., and Paasche, E. (1986). Variable kinetics of silicon-limited growth in *Thalasiosira pseudonana* (Bacillariophyceae) in response to changed chemical composition of the growth medium. *British Phycological Journal*, *21*(2), 183–190.

Oreshkina, T.V., Lygina, E.A., Vozhzhova, O.A. and Ivanov, A.V. (2013). Diatoms and silicoflagellates of the Upper Cretaceous from Saratov Region: Biostratigraphy and sedimentation settings. *Stratigraphy and Geological Correlation*, *21*, 222–236.

Parham, J. F., Donoghue, P. C. J., Bell, C. J., Calway, T. D., Head, J. J., Holroyd, P. A., Inoue, J. G., Irmis, R. B., Joyce, W. G., Ksepka, D. T., Patané, J. S. L., Smith, N. D., Tarver, J. E., van Tuinen, M., Yang, Z., Angielczyk, K. D., Greenwood, J. M., Hipsley, C. A., Jacobs, L., Makovicky, P. J., Müller, J., Smith, K. T., Theodor, J. M., Warnock, R. C. M. and Benton, M. J. (2011). Best Practices for Justifying Fossil Calibrations. *Systematic Biology, 61*, 346–359.

Parks, M. B., Nakov, T., Ruck, E. C., Wickett, N. J., and Alverson, A. J. (2018). Phylogenomics reveals an extensive history of genome duplication in diatoms (Bacillariophyta). *American Journal of Botany*, *105*(3), 330–347.

Passow, U., French, M. A., and Robert, M. (2011). Biological controls on dissolution of diatom frustules during their descent to the deep ocean: Lessons learned from controlled laboratory experiments. *Deep Sea Research Part I: Oceanographic Research Papers*, *58*(12), 1147–1157.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9), 1650-1667.

Ploug, H., Iversen, M. H., and Fischer, G. (2008). Ballast, sinking velocity, and apparent diffusivity within marine snow and zooplankton fecal pellets: Implications for substrate turnover by attached bacteria. *Limnology and Oceanography*, *53*(5), 1878–1886.

Pond, S. K., and Muse, S. V. (2005). Site-to-site variation of synonymous substitution rates. *Molecular Biology and Evolution*, *22*(12), 2375–2385.

Prihoda, J., Tanaka, A., De Paula, W. B. M., Allen, J. F., Tirichine, L., and Bowler, C. (2012). Chloroplast-mitochondria cross-talk in diatoms. *Journal of Experimental Botany*, *63*(4), 1543–1557.

Pritchard, A. (1861). A History of Infusoria: Including the Dismidiaceœ and Diatomaceœ, British and Foreign. *Whitaker and Company.*

Renaudie, J. (2016). Quantifying the Cenozoic marine diatom deposition history: links to the C and Si cycles. *Biogeosciences, 13*(21), 6003–6014.

Renaudie, J., Drews, E.-L., and Böhne, S. (2018). The Paleocene record of marine diatoms in deep-sea sediments. *Mitteilungen Aus Dem Museum Fur Naturkunde in Berlin. Fossil Record*, *21*(2), 183–205.

Riech, V., von Rad, U. (1979). Silica diagenesis in the Atlantic Ocean: Diagenetic potential and transformations, in: Deep Drilling Results in the Atlantic Ocean: Continental Margins and Paleoenvironment, Maurice Ewing Series. *Washington DC: American Geophysical Union,* 315–340.

Riegraf, W. (1995). Radiolarien, diatomeen, cephalopoden und stratigraphie im pelagischen Campanium Westfalens (Oberkreide, NW-Deutschland). *Neues Jahrbuch Für Geologie Und Paläontologie-Abhandlungen*, 129–200.

Roberts, W.R., Ruck, E.C., Downey, K.M. and Alverson, A.J. (2022). Resolving marine–freshwater transitions by diatoms through a fog of gene tree discordance and hemiplasy. *bioRxiv*, 2022-08.

Rothpletz, A. (1896). Ueber die Flysch-Fucoiden und einige andere fossile Algen, sowie über liasische, Diatomeen führende Hornschwämme. *Zeitschrift der Deutschen Gesellschaft für Geowissenschaften*, 854–914.

Rothpletz, A. (1900). Nachtrag zu meinem Aufsatz über einen neuen jurassischen Hornschwamm und die darin eingeschlossenen Diatomeen. *Ztschr. Deutsch. Geol. Ges.*, 154–160.

Round, F. E., and Crawford, R. M. (1981). The lines of evolution of the Bacillariophyta. I. Origin. *Proceedings of the Royal Society of London*, *211*(1183), 237–260.

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

Rust, D. (1885). Beitrage zur Kenntniss der fossilen Radiolarien aus Gesteinen des Jura. *Palaeontographica*, *31*, 273–321.

Sapriel, G., Quinet, M., Heijde, M., Jourdren, L., Tanty, V., Luo, G., Le Crom, S., and Lopez, P. J. (2009). Genome-wide transcriptome analyses of silicon metabolism in Phaeodactylum tricornutum reveal the multilevel regulation of silicic acid transporters. *PloS One*, *4*(10), e7458.

Schmid, A. M. M., Borowitzka, M. A., and Volcani, B. E. (1981). Morphogenesis and Biochemistry of Diatom Cell Walls. In *Cytomorphogenesis in Plants* (63–97). *Vienna: Springer.*

Schulz, P. (1935). Diatomeen aus senonen Schwammgesteinen der Danziger Bucht. *Zugleich Ein Beitrag Zur Entwicklungsgeschichte Der Diatomeen, Botanisches Archiv*, *37*, 383–413.

Shimada, C., Saito-Kato, M., Jenkins, R. G., Tanaka, Y., and Hikida, Y. (2022). Late Cretaceous diatoms (Bacillariophyta) from the Teshio-Nakagawa area, Hokkaido, northern Japan: Significance for their origin and biostratigraphy. *Paleontological Research*, *26*(3), 301–313.

Spaulding, S., and Edlund, M. (2008). Melosira. In Diatoms of North America. from https://diatoms.org/genera/melosira. (Retrieved November 01, 2023)

Strel'Nikova, N. I., and Martirosjan, G. N. (1981). Lower diatom algae from Stavropol. *Viestnik LGU, Seria Biologiya*.

Takahashi, O., Kimura, M., Ishii, A., and Mayama, S. (1999). Upper Cretaceous diatoms from central Japan. *Proceedings of 14th International Diatom Symposium*, 145–155.

Tamura, M., Shimada, S., and Horiguchi, T. (2005). Galeidiniium rugatum gen. Et Sp. Nov. (*Dinophyceae*), a new Coccoid dinoflagellate with a diatom endosymbiont1. *Journal of Phycology*, *41*(3), 658–671.

Tapia, P. M., and Harwood, D. M. (2002). Upper Cretaceous Diatom Biostratigraphy of the Arctic Archipelago and Northern Continental Margin, Canada. *Micropaleontology*, *48*(4), 303–342.

Thamatrakoln, K., Alverson, A. J., and Hildebrand, M. (2006). Comparative sequence analysis of diatom silicon transporters: Toward a mechanistic model of silicon transport. *Journal of Phycology*, *42*(4), 822–834.

Thamatrakoln, K., and Hildebrand, M. (2007). Analysis of Thalassiosira pseudonana silicon transporters indicates distinct regulatory levels and transport activity through the cell cycle. *Eukaryotic Cell*, *6*(2), 271–279.

Thamatrakoln, K., and Hildebrand, M. (2008). Silicon uptake in diatoms revisited: a model for saturable and nonsaturable uptake kinetics and the role of silicon transporters. *Plant Physiology*, *146*(3), 1397–1407.

Tiley, G. P., Poelstra, J. W., Dos Reis, M., Yang, Z., and Yoder, A. D. (2020). Molecular Clocks without Rocks: New Solutions for Old Problems. *Trends in Genetics: TIG*, *36*(11), 845–856.

Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., Bittner, L., Dugdale, R., Finkel, Z., Iudicone, D., Jahn, O., Guidi, L., Lasbleiz, M., Leblanc, K., Levy, M., and Pondaven, P. (2017). Influence of diatom diversity on the ocean biological carbon pump. *Nature Geoscience*, *11*(1), 27–37.

Tréguer, P. J., and De La Rocha, C. L. (2013). The world ocean silica cycle. *Annual Review of Marine Science*, *5*, 477–501.

Tréguer, P. J., Sutton, J. N., Brzezinski, M., Charette, M. A., Devries, T., Dutkiewicz, S., Ehlert, C., Hawkings, J., Leynaert, A., Liu, S. M., Llopis Monferrer, N., López-Acosta, M., Maldonado, M., Rahman, S., Ran, L., and Rouxel, O. (2021). Reviews and syntheses: The biogeochemical cycle of silicon in the modern ocean. *Biogeosciences, 18*(4), 1269–1289.

Tréguer, P., Nelson, D. M., Van Bennekom, A. J., Demaster, D. J., Leynaert, A., and Quéguiner, B. (1995). The silica balance in the world ocean: a reestimate. *Science*, *268*(5209), 375–379.

Trobajo, R., and Mann, D. G. (2019). A rapid cleaning method for diatoms. *Diatom Research: The Journal of the International Society for Diatom Research*, *34*(2), 115–124.

Van Mooy, B. A. S., Fredricks, H. F., Pedler, B. E., Dyhrman, S. T., Karl, D. M., Koblízek, M., Lomas, M. W., Mincer, T. J., Moore, L. R., Moutin, T., Rappé, M. S., and Webb, E. A. (2009). Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature*, *458*(7234), 69–72.

Wall, J. H. (1975). Diatoms and radiolarians from the cretaceous system of Alberta: A preliminary report. *Geological Association of Canada, Special Paper*, *0013*, 391–410.

Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V. and Kosakovsky Pond, S.L., (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Molecular biology and evolution, *35(3)*, 773-777.

Weidmann, M. (1964). Présence de diatomées dans le Flysch à Helminthoïdes. *Université, Laboratoire de géologie.*

Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., and Scheffler, K. (2015). RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution*, *32*(3), 820–832.

Wiesner, H. (1936). Sur la Découverte de Diatomées et autres microfpssiles peu connus dans le crétacé supérieur de la Bohême. *Annales de Protistologie*, *5*, 151–155.

Williams, D. M., and Sims, P. A. (2023). The diatom genus Longinata Hajós (Bacillariophyta): structure, relationships and distribution. *Phytotaxa*, *591*(3), 209–219.

Witkowski, J., Harwood, D. M., and Chin, K. (2011). Taxonomic composition, paleoecology and biostratigraphy of Late Cretaceous diatoms from Devon Island, Nunavut, Canadian High Arctic. *Cretaceous Research*, *32*(3), 277–300.

Yang, Z. (2002). Inference of selection from multiple species alignments. *Current Opinion in Genetics and Development*, *12*(6), 688–694.

Yool, A., and Tyrrell, T. (2003). Role of diatoms in regulating the ocean's silicon cycle. *Global Biogeochemical Cycles*, *17*(4), 1103.

Zalat, A. A. (2013). Cretaceous diatoms biostratigraphy and taxonomy from the North-eastern Sinai, Egypt. *Micropaleontology*, *59*(2/3), 305–323.

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in Ecology and Evolution*, *18*(6), 292–29.

Printed by Media-Tryck, Lund 2024 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Lithosphere and Biosphere Science
Department of Geology
Lund University
Sölvegatan 12
SE-223 62 Lund, Sweden
Telephone +46 46 222 78 80

ISSN 1651-6648
ISBN 978-91-87847-82-0