

UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

Increased aluminum exposure induces widespread changes in silicon, carbon, and nitrogen metabolism in *Entomoneis vertebralis*

Permalink

<https://escholarship.org/uc/item/7cr6h1q3>

Journal

BMC Genomics, 26(1)

ISSN

1471-2164

Authors

Ragunathan, Ramya

Purdy, Hugh M

Seppala, Susanna

et al.

Publication Date

2025

DOI

10.1186/s12864-025-12106-7

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

RESEARCH

Open Access



Increased aluminum exposure induces widespread changes in silicon, carbon, and nitrogen metabolism in *Entomoneis vertebralis*

Ramya Ragunathan¹, Hugh M. Purdy^{1†}, Susanna Seppala^{1†}, Hosu Gwak¹, Sara Calhoun⁴, Benjamin S. Twining³, Igor V. Grigoriev^{4,5}, Bradley F. Chmelka¹, Mark A. Brzezinski^{2*} and Michelle A. O'Malley^{1,6*}

Abstract

Background Diatoms are a class of algae that play an essential role in global ecology and produce valuable chemicals. They are known for forming intricate nanostructured silica cell walls (frustules). The introduction of non-siliceous elements like aluminum into diatoms induces properties such as a lower dissolution rate of the frustule, increasing the specific surface area of the frustule and enhancing metabolism. Previous studies have focused primarily on characterizing physiological impacts, leaving the genetic response(s) to non-siliceous elements largely unexplored.

Results This study investigates the transcriptional response of the pennate diatom, *Entomoneis vertebralis* to dissolved aluminum. Our findings reveal that in the presence of added 10 μ M aluminum, biogenic silica content of the cell wall increases approximately twofold along with significant changes to core metabolism. An increase in transcription of genes encoding nitrate transporters has been observed despite the apparent downregulation of nitrogen assimilation pathways. Additionally, increased transcription of genes involved in carbon fixation were noted. Amino acid and protein motif analyses identified proteins that were differentially regulated shared amino acid compositions and motifs characteristic to silicification-associated genes. A unique structure-based analysis pipeline revealed that some of these proteins have a conserved structural core while being diverse in sequence, which could be features associated with biosilicification.

Conclusion Differential expression transcriptomics has provided insight into diatom metabolism when exposed to aluminum, highlighting potential targets for metabolic engineering. Furthermore, we identified potential silicification-associated genes using tools based on structure and amino acid composition, advancing our understanding of diatom silicification.

[†]Hugh M. Purdy and Susanna Seppala contributed equally to this work.

*Correspondence:
Mark A. Brzezinski
markbrzezinski@ucsb.edu
Michelle A. O'Malley
momalley@ucsb.edu

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Keywords Algae, Diatom, Biomineralization, Aluminum, Metabolism, Structure-based similarity, Transcriptomics

Introduction

Diatoms are a large group of unicellular photosynthetic eukaryotes that contribute 20% of Earth's global primary production of organic matter, comparable to the total amount fixed carbon by all the terrestrial forests combined [1]. Diatom metabolism plays an important role in the cycling of many macronutrient elements, such as carbon, nitrogen, silicon, oxygen and phosphorus [2]. In addition, diatoms are known to produce valuable products such as lipids [3] and they synthesize complex, genetically encoded, hierarchical patterns of amorphous silica in their cell walls under environmentally benign conditions [4].

The silicified diatom cell wall, known as the “frustule”, exhibits a broad variety of morphologies [5], yet the molecular mechanisms underlying the formation of bio-silica structures remain elusive. This growing area of research is partly motivated by the potential to develop biological or bio-inspired methods that produce silica structures using sustainable conditions or inexpensive production platforms [6, 7]. Mechanistic and biosynthetic insights gained from studying the diatom's unique metabolism could guide the engineering of diatoms to synthesize valuable products or customized silica-based structures for diverse applications, including drug delivery, membrane separations, and waste management [8, 9]. These efforts would be aided by an improved understanding of the response of diatom core metabolism to manipulations that affect silica deposition and structure formation to enable processes that achieve the desired response while maintaining cell viability and growth.

Attempts have been made to modify the diatom's frustule by incorporating different non-silicon elements, such as titanium or germanium, to induce modifications with desirable properties [10, 11]. Among these, adding aluminum to frustules has resulted in favourable material properties, such as resistance to hydrolysis and higher specific area [12, 13], in addition to providing beneficial metabolic effects, such as enhancing the carbon fixation process [14]. There is some evidence that this enhancement is due to the ability of aluminum to form complexes that facilitate the conversion of specific ions, such as iron, into bioactive Fe^{2+} states [15] that are readily consumed by diatoms. Additionally, it has been demonstrated that aluminum increases the activity of alkaline phosphatases, enabling enhanced use of organic phosphorus, including dead biomass, which is typically not consumed by diatoms [16] and that could impact bio-silicification. Overall, these findings imply that aluminum influences an array of metabolic and biosynthetic processes in diatoms, including nutrient uptake and utilization of macronutrients,

and could have significant implications for understanding the bio-silicification process and the intricate interplay between elements in marine ecosystems.

Aluminum incorporation experiments have been conducted in various diatom species [13, 17–20], revealing the ability of aluminum to alter silica content or morphology. Aluminum-induced changes might differ based on whether they are centric or pennate diatoms; however, it is subject to further verification. The pennate diatoms *Navicula salinarum* [20] and *Craspedostauros* sp. [13] showed increased biosilicification in the presence of aluminum. Similar studies on the centric diatom, such as *Thalassiosira weissflogii* showed no morphological change, but phosphate metabolism was affected [16]. The centric diatom *Thalassiosira pseudonana* also exhibited no morphological change in response to aluminum, but an increase in the specific area of the silicified cell wall was reported [19]. Further insights into the possible links between silicon metabolism and aluminum come from studies of higher plants that produce silica nanospheres to alleviate aluminum toxicity [21, 22]. There has been some evidence that adding aluminum causes modifications in frustule morphology, as discussed, which varies among different classes of diatoms; however, further studies in different types of diatoms are required to understand the species-specific impacts.

Understanding how metals such as aluminum are incorporated into diatom frustules and their impacts on silica content or morphology is currently limited by incomplete understanding of silica biosynthesis. Extensive research investigating the organic components associated with bio-silicification has led to the discovery of many protein families involved in this process, such as silaffins [23], silacidin [24], cingulin [25], and silicanin [26], through the analysis of transcriptomic changes during building the silica cell wall or by conducting proteomics on extracted frustules. However, none of these studies captured the transcriptomic changes that a diatom undergoes as its frustule morphology is altered in response to an environmental cue [5, 27]. It should be noted that several of these protein families, particularly the silaffins and silacidins, although characterized in the well-studied diatom *T. pseudonana*, are not conserved even among diatoms from the same genus [28]. As previously discussed, aluminum has induced frustule-associated morphological change in certain diatoms, which may affect the expression of silicification-related genes. Therefore, merging an understanding of transcriptomic shifts and corresponding biosilicification changes in diatoms exposed to aluminum could implicate genes involved in the morphological changes.

How the incorporation of non-siliceous elements, such as aluminum, affects the processes used by diatoms to construct their intricate silica-based cell wall remains unknown. Still, it likely involves many genes and their constituent protein products [5] from various metabolic pathways involving nutrient uptake and silica polymerization. An understanding of these impacts would help identify potential targets to metabolically engineer frustule morphology and, at the same time, elucidate biosilicification mechanisms and processes. An increase in bio-silicification in the presence of aluminum has been observed, particularly in pennate diatoms as mentioned previously. Here we examine the effect of aluminum on *Entomoneis vertebralis*, a newly sequenced pennate diatom, was used in this study to investigate the impact on its bio-silicification and overall metabolism.

In the present study, the impact of aluminum on the bio-silicification process(es), as well as its impact on core metabolism, are examined. Exposing a newly sequenced pennate diatom *Entomoneis vertebralis* to aluminum increased the amount of silica per cell. The transcriptomic response revealed altered carbon and nitrogen metabolism in the presence of aluminum and its impact on the expression of genes involved in silicification, paving the way to identify metabolic targets to induce desirable properties and valuable products.

Results

Entomoneis vertebralis has a unique cell wall morphology, and the addition of aluminum leads to an increase in silicification

The morphology of *Entomoneis vertebralis* frustule was examined using scanning electron microscopy (SEM) and transmission electron microscopy (TEM), as shown in Fig. 1. SEM characterization (Fig. 1A) revealed key structural features, including a raphe, valves, and girdle bands. TEM characterization (Fig. 1B) provided a top view of a broken valve, highlighting the raphe's ring-like structures. The ring-like silica structure can be further seen when the raphe was disengaged from the rest of the frustule by sonication, as shown in Fig. 1D. These ring-like formations appeared to be the primary silicified regions of the valve, as confirmed by EDX-TEM analysis (Fig. 1C). In contrast, the rest of the valve was likely composed predominantly of organic matter, differing from the silica-rich composition typically observed in other silicified diatoms.

To understand this silicified frustule morphology, *E. vertebralis* was exposed to varying concentrations of aluminum [29]. Biogenic silica content was measured to understand the impact of increased aluminum on diatom silicification [30].

As shown in Fig. 2a, the silica content of *E. vertebralis* cells increased almost twofold in cultures that were

exposed to 10 μ M aluminum compared to the control, which was not supplemented with aluminum. Concentrations above added 5 μ M aluminum did not lead to additional increase in silicification. ICP-MS revealed some amount of aluminum is present in diatoms from control cultures (Fig. 2c), suggesting potential aluminum background levels from the salts, that are part of the artificial seawater formula. Hereafter, any mention of aluminum concentration refers to the amount of aluminum added to the culture medium, if not specified. Transmission electron microscopy (TEM) revealed that the increase in silica content was accompanied by changes in frustule morphology, whereby with 10 μ M aluminum, the frustule developed larger (From ~750 nm to ~950 nm) ring-like silicified structures within the raphe (Fig. 2b), which is the dominant silicified feature of *E. vertebralis*. Despite the increase in biogenic silica content with added aluminum, there was no noticeable change in the aluminum/silicon molar ratio across different concentrations of added aluminum to cultures, as shown in Fig. 2c.

RNA-seq analysis of *E. vertebralis* exposed to high aluminum levels reveals significant alterations in primary metabolism

RNA-seq was conducted utilizing the *Entomoneis vertebralis* reference annotation from JGI Phycocosm [31] to analyze the transcriptional response of this diatom towards supplemental aluminum. The dataset utilizing the transcript reads from samples exposed to different concentrations of added aluminum (5 μ M and 10 μ M) and a control (0 μ M) indicated transcription of 44,665 genes based on the StringTie-generated [32] gene count matrix. However, not all StringTie-generated transcripts were present in the reference genome annotated through JGI Phycocosm [31], suggesting the presence of novel transcript structures not identified by the reference transcriptome. The results from the StringTie-generated gene count matrix, as shown in Fig. 3, indicate that more genes are differentially expressed between 0 μ M and 10 μ M aluminum cultures, compared to 5 and 10 μ M aluminum cultures, in addition to new de novo transcripts observed in this study. The differentially expressed genes between 0 μ M and 10 μ M aluminum consisted of 1650 sequences, of which 818 had some predicted annotation (GO, InterPro, HMMER, Pfam, or KEGG, Supplementary Information). Analysis of the annotations indicated that aluminum affected gene expression across a diverse set of metabolic processes.

Biochemical assays indicated that there was a significant increase in silicification between diatom cultures without supplemental aluminum and those exposed to 5 μ M or 10 μ M of aluminum. However, there was no significant difference in silicification between cultures that were treated with 5 μ M aluminum and cultures that were treated with 10 μ M aluminum (Fig. 2). To focus on

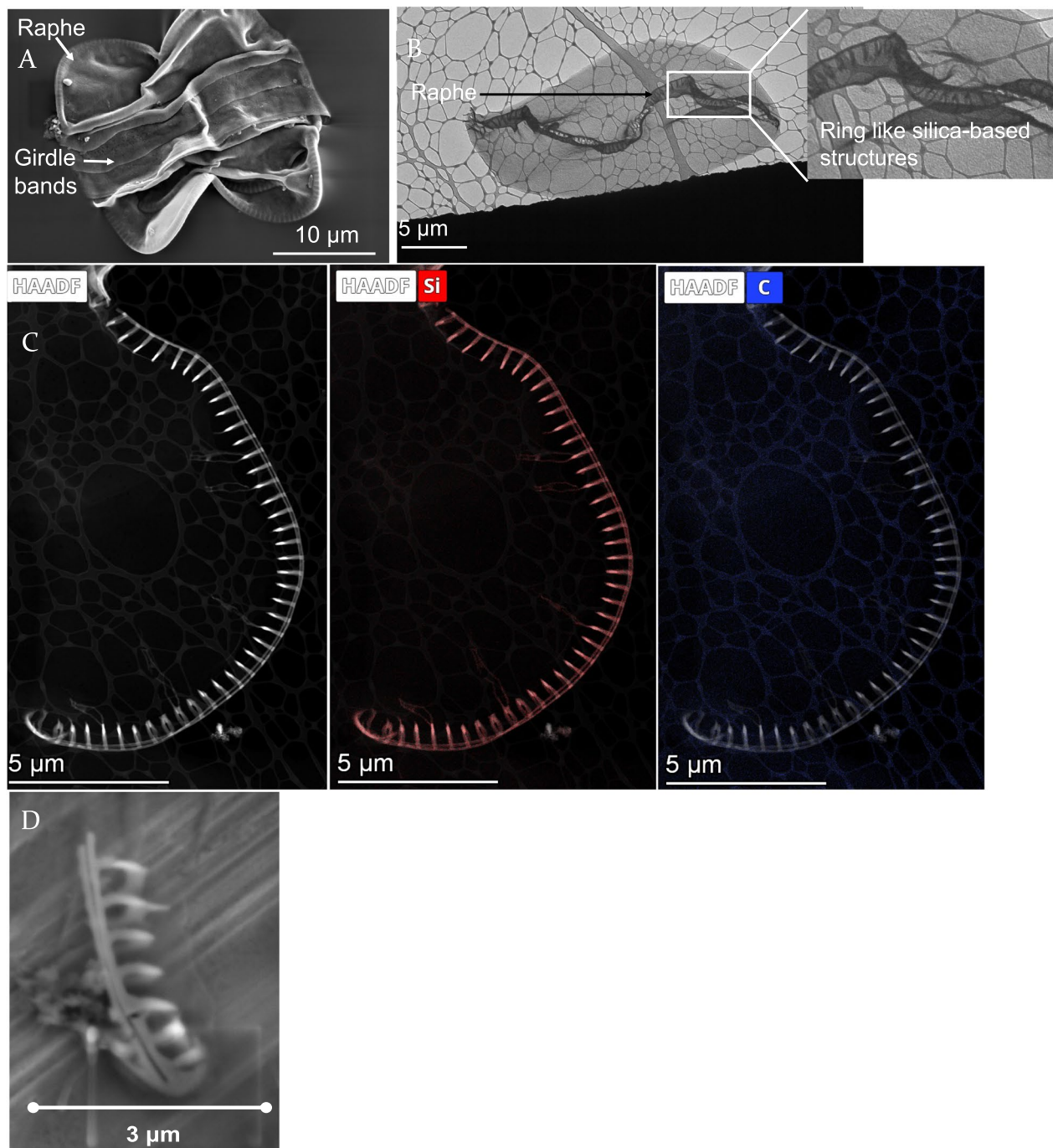


Fig. 1 **A.** SEM image of *Entomoneis vertebralis* depicting the overall frustule morphology, with ring-like structures in the raphe (outermost border as depicted in the picture). **B.** Top view of H_2O_2 -cleaned frustule of *Entomoneis vertebralis* as seen in TEM, showing that the raphe majorly consists of ring-like structures. **C.** TEM-EDX on *Entomoneis vertebralis*, the composition of carbon and silica, shows that most of the silica was concentrated on the protruding circles and the rest of the cell wall is likely composed of carbon rather than silica. **D.** SEM image of a broken raphe showing the circular silica rings protruding from the raphe

identifying responses associated with changes in silicification, we used these differences in silica content to isolate transcriptomic changes related to shifts in silicification. To identify genes involved in silicification, transcripts that were differentially expressed in 0 μM and

10 μM aluminum, but were not differentially expressed between the 5 μM and 10 μM aluminum treatments, were analyzed. Enzymes participating in polyamine synthesis were upregulated in the presence of added aluminum. Among these was Protein ID 999264, which

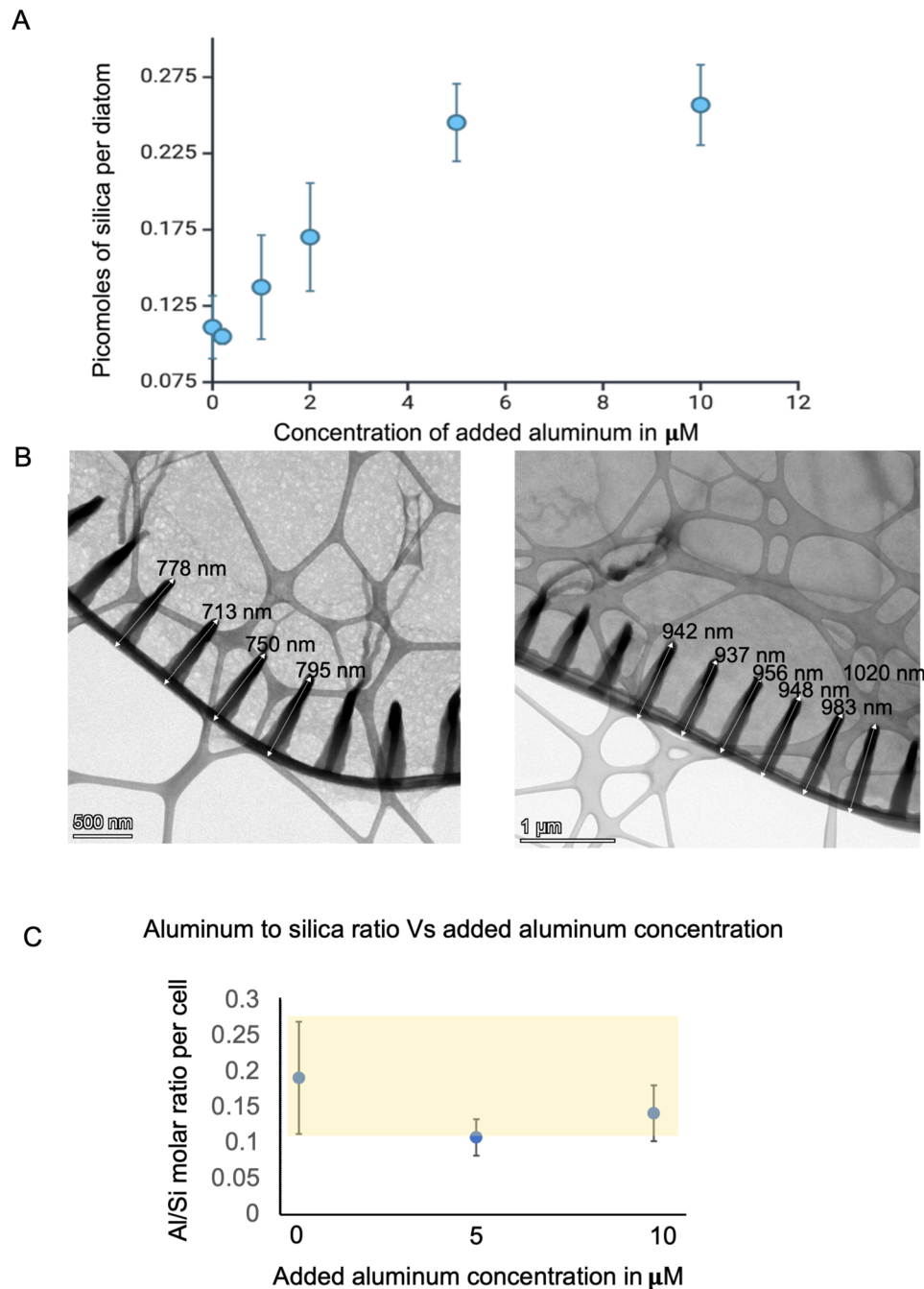


Fig. 2 **A.** Effect of aluminum on biosilicification of *Entomoneis vertebralis*. As the amount of aluminum added was increased, the amount of silica per cell increased, however it saturated after the 5 μM aluminum scenario as indicated in the shaded box. **B.** TEM images of no aluminum (right) versus 10 μM aluminum exposed diatoms (left). Enlarged circles protruding from the raphe relative to the condition without aluminum were observed. **C.** Al/Si molar ratios for different concentrations of aluminum added to the media shows that there was no significant change in Al/Si ratio as depicted in the yellow box. Here aluminum refers to amount of aluminum present inside the cell, measured using ICP-MS

encodes a putative homospermidine synthase. Protein ID 1041712 encodes what appears to be a fusion protein that contains a spermidine synthase domain as well as a SAM decarboxylase domain [33]. It has been suggested that these fusion proteins are unique to diatoms and that they are involved in synthesizing long-chain polyamines

[33]. Furthermore, several serine proteases (subtilase family) were downregulated in higher aluminum conditions (Protein IDs 1050279, 1087708, and 1087724). Various ubiquitin ligases also appeared to be significantly downregulated (Protein IDs 1003455, 1070178, 626300, 814710, 960722, 970818, and 1073132), while some

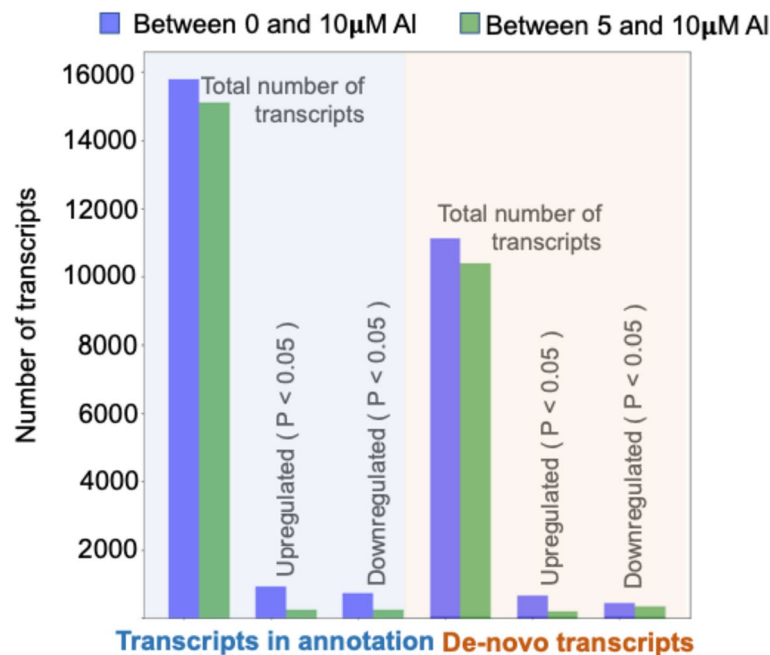


Fig. 3 RNA-seq data characterizing the statistics of genes that were differentially expressed in *Entomoneis vertebralis* between 0 and 10µM aluminum (blue) and 5 and 10 µM aluminum (green), in both the classes, i.e. transcripts that are aligned to the genome (denoted in blue) and de-novo transcripts discovered using StringTie (denoted in red). “Total number of transcripts” indicates the total number of transcripts that were expressed with cpm > 0.5 in that case. Significantly upregulated and downregulated genes were chosen based on the adjusted p-value < 0.05. The adjusted p-value was calculated using edgeR (Benjamini-Hochberg (BH) method). There were more genes differentially regulated between 0 and 10 µM relative to the 5 and 10 µM scenario

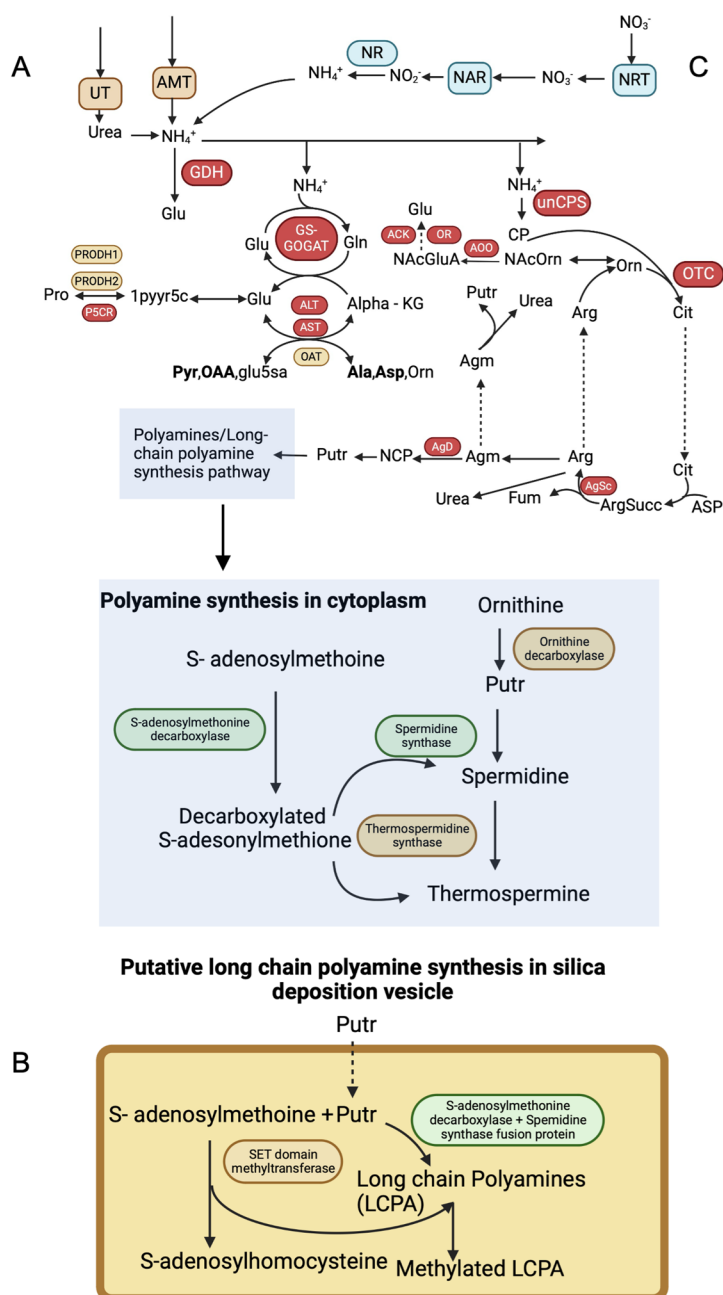
were upregulated (Protein IDs 960,989 and 1027804). Overall, an apparent downregulation of protein degradation mechanisms was observed in higher aluminum conditions. Moreover, silicon transporters (Protein IDs 912858, 1020251, 1010885, and 8915) were weakly downregulated. dAnk-like proteins (Protein IDs 1044363 and 1003123), which are believed to regulate phase separation between the organic matter and silica, were upregulated. Inorganic phosphate transporters (Protein IDs 1036788, 883475, 915403, 915394, 1389917, 902258, 846407, 1015189, 784627, and 1070080) were downregulated.

As certain metabolic processes were affected by the presence of aluminum, we investigated its effect on core metabolism, such as the nitrogen/ammonium assimilation and carbon metabolism pathways. The nitrogen metabolism pathway, as shown in Fig. 4, was constructed using the framework outlined by Smith et al. 2019 [34]. While not all genes within this pathway exhibited differential regulation, some genes involved in ammonium assimilation were downregulated.

Ammonium assimilation, as shown in Fig. 4a, primarily occurs through three distinct pathways: direct reduction of ammonium to glutamate by a glutamate dehydrogenase; the GS-GOGAT cycle, which involves glutamate synthase and results in the production of alanine, asparagine, and ornithine; and the urea cycle, which generates fumarate, putrescine, and urea; as outlined by

Smith et al. [34]. These pathways can be predicted to be localized to mitochondria, chloroplasts, and other cell compartments (genes without mitochondria/chloroplast targeting sequences), as shown in Fig. 4c. In the current transcriptomics analysis, most of these transcripts involved in nitrogen metabolism lacked clear mitochondrial or chloroplast targeting sequences, suggesting that these proteins are not targeted to either organelle, as shown in Fig. 4. Notably, the differentially expressed glutamate synthase appeared to lack mitochondria or chloroplast targeting sequences and was determined to be unique to *E. vertebralis*, as it does not share similarity to other known diatom glutamate synthases from MCL clustering. Moreover, despite the upregulation of nitrate transporters and a nitrate reductase in the presence of aluminum, a paradoxical downregulation was observed in the ammonium assimilation pathway, as shown in Fig. 4a. Differential expression of genes possibly involved in polyamine synthesis was also observed, as shown in Fig. 4b. Furthermore, a significant downregulation was noted in the transmembrane amino acid transporter family 2.A.18.6.2.

Figure 5 provides an overview of the differential regulation of diatom genes related to carbon metabolism in the presence of aluminum. Notably, there was a pronounced upregulation of genes associated with the Calvin cycle, aligning with the concurrent upregulation of bicarbonate



Name	Symbol	Location	LogFC (p<0.05)
Ammonium transporter	AMT	-	-3.373,-1.4126,-0.879,-0.837,-0.286,
Urea transporter	UT	-	-0.589
Glutamate dehydrogenase	GDH	Signal peptide and mitochondria	-0.927,-0.68,-0.687
Glutamine synthase	GSIII-GOGAT	Cytosol	-1.157,-1.135
Alanine aminotransferase	ALT	Cytosol	-0.465
Aspartate aminotransferase	AST	Cytosol	-0.852
Ornithine aminotransferase	OAT		-
Pyrroline-5-carboxylate reductase	P5CR	Mitochondria	-0.531
Proline dehydrogenase	PRODH1		-
Proline dehydrogenase	PRODH2		-
Carbamate kinase	unCPS	Cytosol	-1.1875
Ornithine carbamoyltransferase	OTC	Cytosol	-0.5168
Acetylornithine aminotransferase	AOO	Chloroplast	-0.689
N-acetyl-gamma-glutamyl-phosphate reductase.	OR	Signal peptide	-0.9106
Amino-acid N-acetyltransferase	ACK	Cytosol	-1.076
Argininosuccinate lyase	AgSc	Cytosol	-0.98
Agmatine deiminase	AgD	Cytosol	-0.395
S-adenosylmethoine decarboxylase + Spermidine synthase fusion protein	-	Signal peptide	0.771
S-adenosylmethonine decarboxylase	-	Cytosol	0.3185
Spermidine synthase	-	Cytosol	0.4523
Saccharopine dehydrogenase /Homospermedine synthase	-	Signal Peptide	1.523
Nitrite transporter	NAR		1.733
Nitrate transporter	NRT		6.964,2.1326,0.6745,0.05973,0.3823
Ferredoxin--nitrite reductase	NR	Chloroplast	2.933

Fig. 4 A Putative mapped nitrogen metabolism pathway in the diatom *Entomonesis vertebralis* and transcriptional responses of this pathway when aluminum was added. Red indicates the downregulation of an enzyme, and green represents the upregulation of an enzyme in response to aluminum. An overall increase in nitrogen uptake can be seen by mapping the transcriptional response of nitrate and nitrite transporters along with upregulation of nitrite reductase, indicating more ammonium is being produced in presence of aluminum. However, downregulation of 3 ammonium pathways, i.e. conversion into glutamate by GDH, the GS-GOGAT cycle, and the urea cycle, in addition to the upregulation of polyamine synthesis and long chain polyamine synthesis, can be noted. Glu – glutamate, Gln – glutamine, Ala – Alanine, Asp – Aspartate, Pyr – Pyruvate, OAA – Oxaloacetate, glu5a – gflutamate-5-semialdehyde, NCP-N – Carbamoyl-putrescine, NACOrn – Acetylornithine, Orn – Ornithine, NACGluA – N-acetylglutamate semialdehyde, Arg – Arginine, Cit – Citrulline, CP – Carbamoyl phosphate, Putr – Putrescine, Agm – Agmatine, 1pyr5c – pyrroline-5-carboxylate, Pro – proline, Alpha-KG – Alpha-Ketoglutaric acid, Fum – Fumarate. **B** Putative long chain polyamine synthesis pathway, illustrating the fusion protein that was upregulated in the presence of aluminum. **C** Log-fold changes of expression of genes involved in this pathway in 10 μ M aluminum compared to 0 μ M aluminum

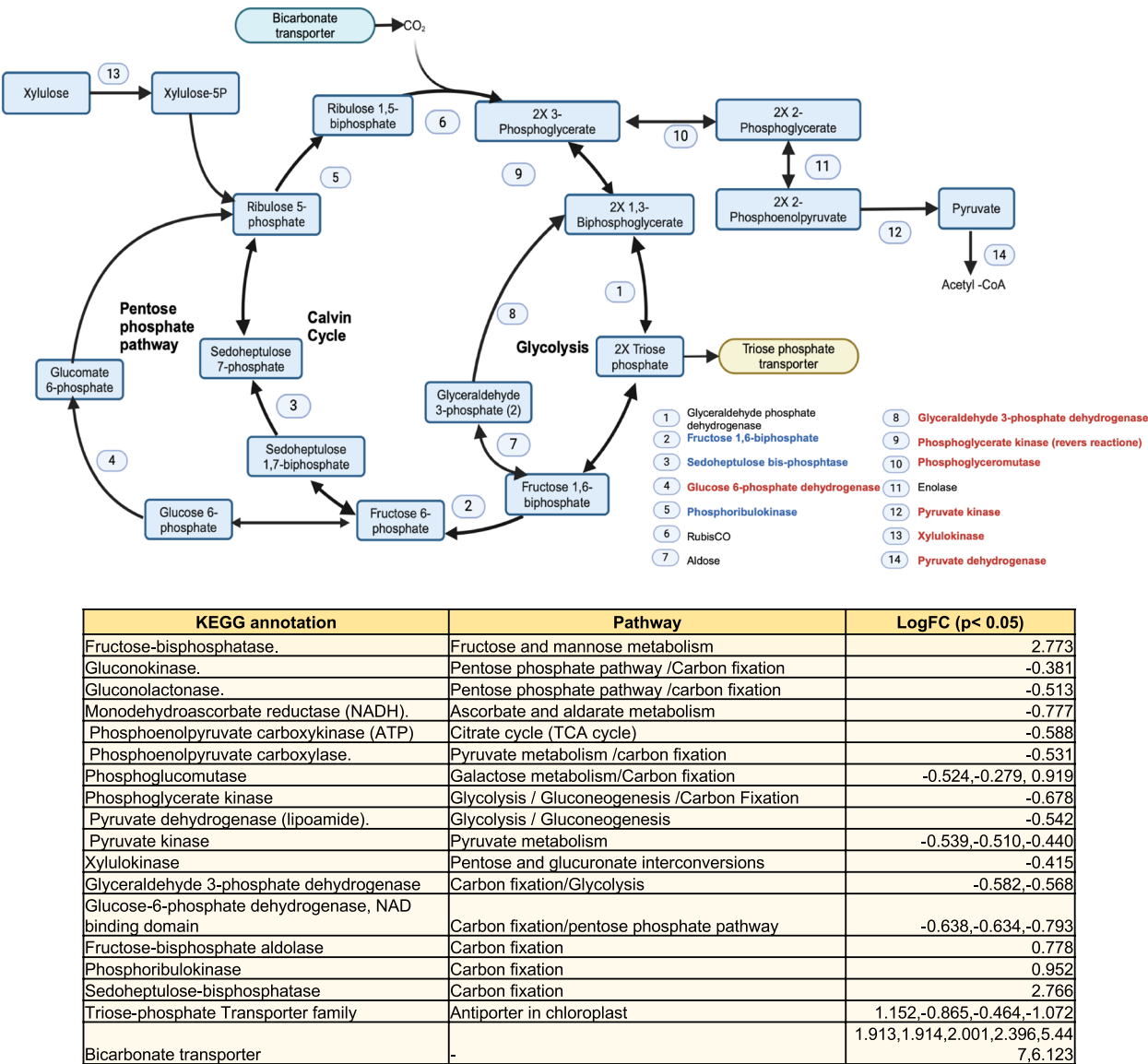


Fig. 5 A Response of carbon metabolism of *E. vertebris*, particularly the Calvin cycle, to the presence of aluminum. An overall upregulation of genes involved in the Calvin cycle was observed. Genes involved in the cycle such as fructose 1,6-biphosphate, sedoheptulose bis-phosphate, and phosphoribulokinase were upregulated. In addition, increased transcription of bicarbonate transporters was observed, indicating increased carbon fixation. However, downregulation was observed of pathways involved in glycolysis and the pentose phosphate pathway, which are ATP and NADH generating pathways, respectively. **B** LogFC of genes involved in carbon metabolism in 10 μ M aluminum compared to 0 μ M aluminum

transporters when aluminum was present. This suggests an overall augmentation of carbon fixation (increased carbon uptake per cell) processes during aluminum supplementation. However, a subtle downregulation was observed in genes linked to the ATP-generating phase of glycolysis, which contrasts with the heightened energy-intensive carbon fixation.

Increases in diatom silicification and carbon uptake would likely require relatively more energy. Therefore, we investigated energy metabolism by characterizing whether photosynthetic genes such as some cytochrome b (Protein IDs 1058478 and 999804) and c (Protein ID

1058093) were upregulated in the presence of aluminum, which could indicate an increased electron transfer activity associated with photosynthesis. In addition, the upregulation of chlorophyll-binding proteins (Protein IDs 1050897 and 1059248) was observed, indicating an overall increase in photosynthetic activity. High upregulation of the ferritin-like diiron domain-containing protein (Protein ID 1076983) was also observed, which could be predominately used to store ferric ions mediating photosynthesis. Lastly, genes involved in lipid degradation metabolism were strongly differentially regulated in the presence of aluminum. Genes involved in

glycerolipid synthesis such as glycerone kinase (Protein ID 986559), monogalactosyldiacylglycerol synthase (Protein ID 1006071) and glycerol-3-phosphate acyltransferase (Protein ID 1009955) were upregulated, although the triacylglycerol lipase involved in this pathway was downregulated (Protein ID 1083519). Increased upregulation of glycerone kinase and differential regulation of specific triglycerol lipases (Protein IDs 1034266 and 1083519) were downregulated, and triglycerol lipases of Protein IDs 1021083 and 797622 were upregulated [35]. Hence, differential regulation of different enzymes in lipid metabolism has been observed.

Protein motif analysis identifies genes that are characterized by motifs involved in silicification

While almost 50% of the differentially regulated diatom genes have annotations based on sequence, a notable portion lacks confident annotations or remains unannotated. The unannotated genes for which expression was associated with the increase in silicification between 0 μ M and 10 μ M aluminum were analyzed for unusual enrichment of particular amino acid patterns (referred to as motifs) relative to the entire proteome using ProminTools by Skeffington et al. [36], as shown in Fig. 6. Proteins were clustered based on the degree to which a specific motif repeats [36]. In addition, clustering was restricted to proteins that display a high degree of amino acid compositional bias (fLPS p -value $> 10^{-6}$). Figure 6 shows enrichment of various motifs among selected genes that were differentially regulated in the presence of aluminum. The color intensity in the heatmap denotes enrichment of certain motif folds. Motifs involving serine and glutamic acid (Cluster 1), can be observed in Fig. 6c. Unusually high enrichment of repeating serines can be seen in Cluster 3 (Fig. 6d). This distinctive enrichment pattern led us to identify protein 1401909, which, through motif analysis, reveals a repeating motif of SEV.S.SE (Ser-Glu-Val-X-Ser-X-Ser-Glu-X). Notably, one of the most intensely enriched motifs, which is depicted by darker shades of blue, originates from Cluster 4 (Fig. 6b) and is primarily characterized by aspartic acid-based motifs. Enrichments of serine and the negatively charged amino acids such as aspartic acid and glutamic acid are observed among the unannotated differentially expressed genes.

Structure-based annotation identifies sequence-divergent proteins with a conserved structure core

Several protein families associated with diatom silicification are poorly conserved across species [28]. Furthermore, conventional functional annotation relies on sequence-based similarity employing methods like BLAST and HMMER. As a complementary approach, we leveraged structure-based predictions for diatom proteins that were differentially expressed but lack any

sequence-based annotation based on the JGI annotation pipeline [32]. We performed structure-based homology analyses to resolve potential functions of the unannotated genes identified in this study. Structure-based clustering suggested some unannotated proteins were sequence divergent, but shared predicted structural features. As shown in Fig. 7, one of the structure-based clusters that showed the highest upregulation (Protein IDs 1069091, 1056361, 1057927, 978533 and 1056005) was predicted to fold into a conserved structured domain consisting of 5 α helices and a disordered head. The low confidence in predicting the unstructured chain might have contributed least when seeking structural similarity.

This structure-based protein class lacks sequence similarity, as shown in Fig. 7b. Employing foldseek against the CATH database revealed a predicted relatively low similarity (0.15 e-value) to certain cytoplasmic elements, such as actin and microtubules, as depicted in Fig. 7c. These sequences eventually fold into certain conserved sequences and could share similar function independent of the sequence.

Discussion

In this study, cultures of the diatom *E. vertebralis* were exposed to different concentrations of aluminum and their cellular responses were monitored by using differential transcriptomics. Previous studies have observed that aluminum enhances metabolic pathways such as carbon fixation [14, 37], which could be due to improved bioavailability of iron in the form of Fe^{2+} . Our current study supports the hypothesis that aluminum increases iron bioavailability, as we observed an upregulation of gene encoding ferritin-domain-containing proteins involved in iron storage in pennate diatoms [38]. Aluminum exposure resulted in an apparent downregulation of inorganic phosphate transporters consistent with [21], which suggests enhanced phosphate metabolism, particularly that of dissolved organic phosphate.

Additionally, we mapped responses within pathways involved in nitrogen/amino acid metabolism based on genes that were differentially regulated in the presence of aluminum. While there appeared to be an upregulation of nitrate transporters and nitrate reductase, which convert nitrates to ammonium, there was an overall transcriptional level decrease in genes involved in nitrogen assimilation. Most of the differentially regulated genes in nitrogen assimilation appeared to be lacking in chloroplasts and mitochondria targeting sequences, indicating that differentially expressed nitrogen assimilation may occur in the cytoplasm; however, additional experiments are required to verify this. The upregulation of the nitrate transporters and nitrate reductase might imply nitrogen limitation, which has been previously observed; however, we did not observe a downregulation of the Calvin cycle,

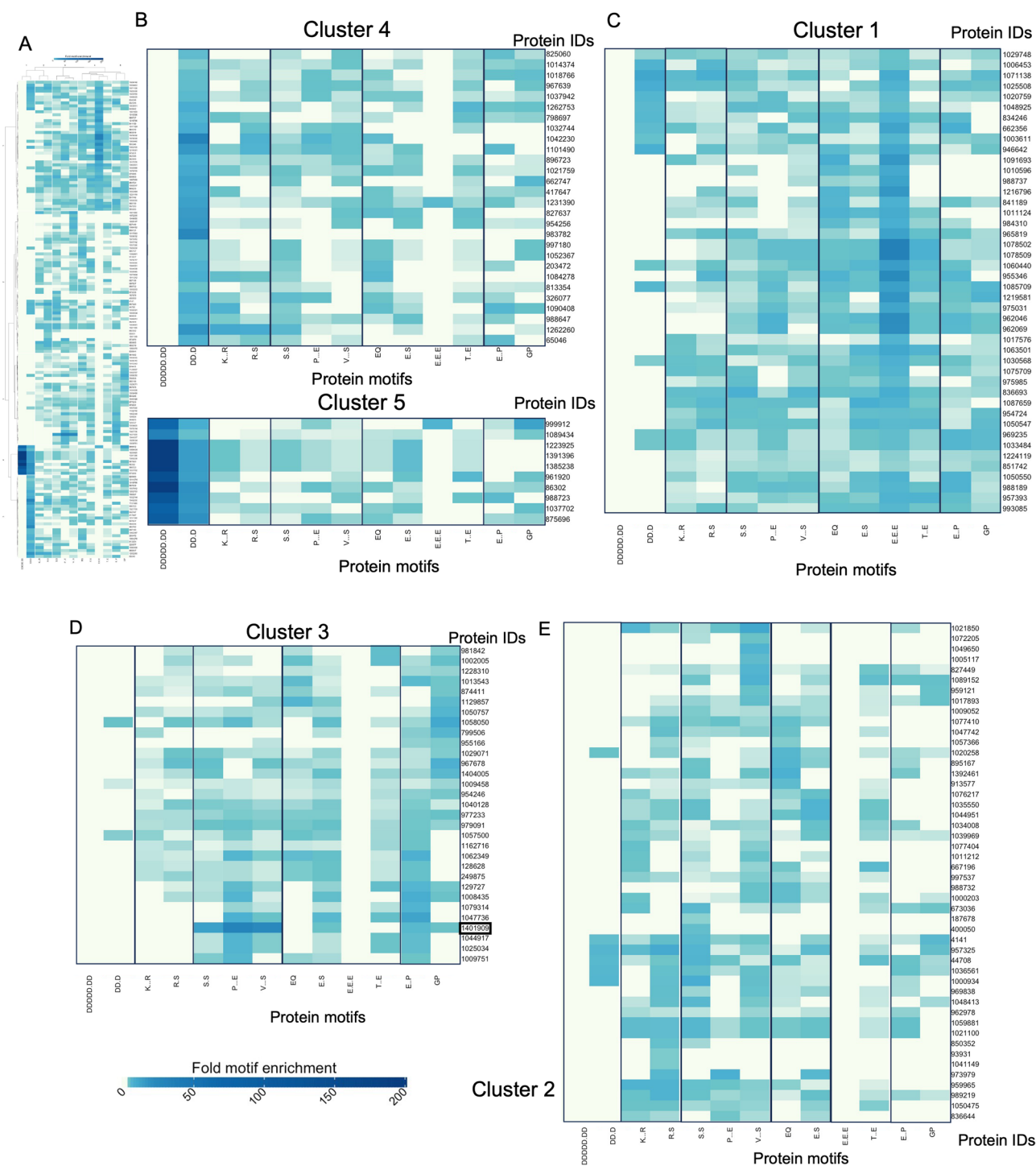


Fig. 6 Transcriptomics on *E. vertebralis* when exposed to aluminum revealed proteins with unknown annotation that are characterized by silicification-based amino acid motifs. **A** Heatmap showing genes that were differentially regulated in *E. vertebralis* when exposed to aluminum without known annotation, and clustering based on enrichment of protein motifs. The heatmap was divided into 5 major clusters, numbered based on their position in the heatmap. A protein motif finder (ProminTools) was used for this analysis. The various clusters indicate enrichment in specific motifs relative to the background, i.e. the entire predicted proteome. **B** Clusters 4 and 5 showing enrichment in aspartic acid based motifs. **C** Cluster 1 showing enrichment in glutamic acid based motifs. **D** Cluster 3 showing an unusually enriched protein in serine-based motifs that suggests its involvement in silicification. **E** Cluster 2 showing all other proteins, some of which are enriched for lysine-based motifs. The right margin indicates the protein ID based on JGI annotation. Several unannotated proteins show enrichment in silicification-associated amino acids, such as serine, aspartic acid, and glutamic acid. In particular, Protein ID 1401909, which belongs to cluster 3, is highly enriched in serine motifs

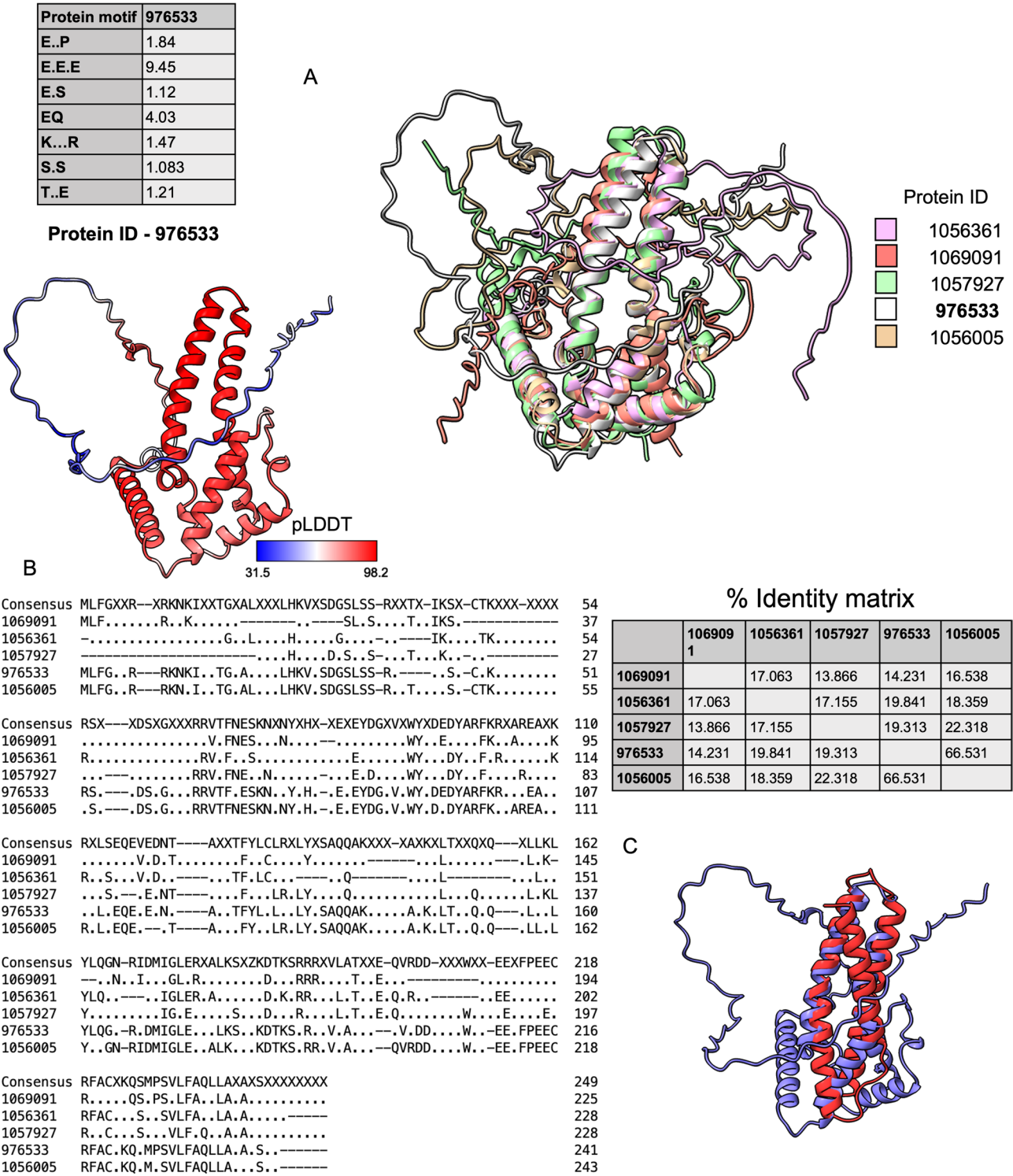


Fig. 7 Structure-based protein families were revealed using transcriptomics of *E. vertebralis* when exposed to aluminum. **A** Structural superimposition of proteins belonging to a group of proteins that share structural similarity but are divergent in sequence. All of these proteins possess a structured tail and a disordered head. However, they do not share significant sequence-based similarity, even among the structured tail. **B** Depicting sequence-based alignment of all proteins belonging in the family. They share minimal sequence identity except for two members of the cluster. **C** The first hit from a database called CATHDB, depicts the structural similarity of the structured tail with cytoplasmic elements, such as actin (red). All structures were predicted using AlphaFold (see Methods). Using structural similarity, novel protein families in response to aluminum were discovered

thus differing from the nitrogen limitation response in Bender et al. 2014 [39]. In addition, we identified a glutamate synthase unique to *E. vertebralis*, which appears to lack clear mitochondrial and chloroplast targeting sequences and is possibly localized in the cytoplasm. Furthermore, a family of transmembrane amino acid transporters (TCDB ID 2.A.18.6.2) were downregulated; these proteins could be responsible for transporting glutamate from the cytoplasm to the mitochondria. In general, a decrease in ammonium assimilation was observed; however, the genes that were differentially regulated appear to lack clear mitochondrial and chloroplast targeting sequences and is possibly localized to the cytoplasm. This nitrogen assimilation pathway may have been derived from recycling protein-derived nitrogen, which is concurrent to claims proposed by Bernard et al. 2009 [40].

The observed upregulation of genes in the Calvin cycle in the presence of aluminum aligns with previous findings, indicating enhanced carbon uptake [37]. Notably, the Calvin cycle is known to require extensive energy. A plausible response to the increase in required energy is an increase in photosynthesis, possibly attributed to heightened iron bioavailability. The concurrent upregulation of cytochromes, facilitating electron transfer mechanisms, supports this hypothesis. However, a crucial question arises regarding the fate of carbon. Contrary to expectations, the downregulation of the latter half of glycolysis, responsible for ATP synthesis, challenges the notion of glucose being the primary energy source. Intriguingly, the pronounced upregulation of lipases and glycerone kinase points towards lipid metabolism as an alternative route for carbon assimilation and energy acquisition. Upregulation of glycerone kinase suggests the upregulation of glycerol metabolism, which could be produced from certain types of lipids [41, 42]. Tanaka et al. 2015, reported that increased lipid accumulation in *Fistulifera solaris* was linked to an upregulation of the fatty acid degradation pathway and a concomitant downregulation of ammonium assimilation [43]. This leads us to hypothesize an elevated production of lipids in response to aluminum-induced stress. Our observations suggest an association between aluminum stress and lipid metabolism alterations. The upregulation of rate limiting glycerol-3-phosphate acyltransferase [44] and monogalactosyldiacylglycerol synthase, both of which are involved in glycerolipid biosynthesis, strengthens the claim that the presence of aluminum could alter lipid production.

One potential confounding variable in the above carbon and nitrogen metabolism analysis is the slight difference in cell density observed between the different culture conditions, with the addition of aluminum leading to a reduced cell count (Fig S4). However, BSI analysis shows that the total amount of silica consumed from the

media in each condition is not significantly different (see following paragraph). This reduction of cell count is proportional to the increase in silica uptake per cell (Fig. 2a) indicates that silicon is the limiting nutrient. Additionally, as f/2 media generally contains an excess of macronutrients (nitrogen and phosphate) relative to silica, and since the rate of silica depletion does not significantly differ across aluminum concentrations, the depletion of other nutrients in the culture was assumed to be similar across conditions, with concentrations still in excess at the time of sampling. However, differences in biomass density could cause changes to other media attributes such as pH, which could have an impact on nutrient uptake.

We observed that exposing *E. vertebralis* to higher aluminum concentration increases the biogenic silica content per cell and changes the frustule morphology, as observed by TEM. In addition, transcriptomic analyses suggest that the genes involved in silicification were subjected to differential regulation, which has been summarized in Fig. 8. Silicon transporters (SITs) are believed to be involved in silicic acid transport when it is deficient in the cellular environment [45].

When excess silicic acid is present in the cellular environment, some of these transporters have minimal to no roles in transport, but may instead sense extracellular silicic acid levels and allow the cell to evaluate whether it can proceed for cell division [46]. However, an estimate of the available silicic acid at the time when the cultures were harvested, based on the mass balance between the initial silicic acid concentration and the amount of biogenic silica present at the time of sampling, indicates that around 56 μM of silicic acid was available for diatoms exposed to aluminum compared to ~64 μM for those in controls (no aluminum). Both of these concentrations are greater than 30 μM , which has been previously proposed to be a threshold above which certain silica transporters are upregulated and, suggesting that cells in both treatments may rely on diffusion rather than SIT-mediated silicic acid transport. The amount of silica available per cell in aluminum-exposed diatoms was higher than that of the control, so downregulation of SIT's transcripts might be as a mechanism for sensing and evaluating the needs of silicic acid. In addition, the downregulation of silicon transporters in the presence of aluminum, despite higher levels of available silica per cell, is consistent with evidence from higher plants where genes encoding silicon transporters are downregulated in plant shoots with higher silica accumulation levels [47]. We hypothesize that to detoxify aluminum, the cell incorporates the metal into their frustules [48], and the silicon transporters could be signaling the cell for increased silica requirement by downregulating its expression. However, more studies are needed to explain this molecular mechanism of silicon transporters.

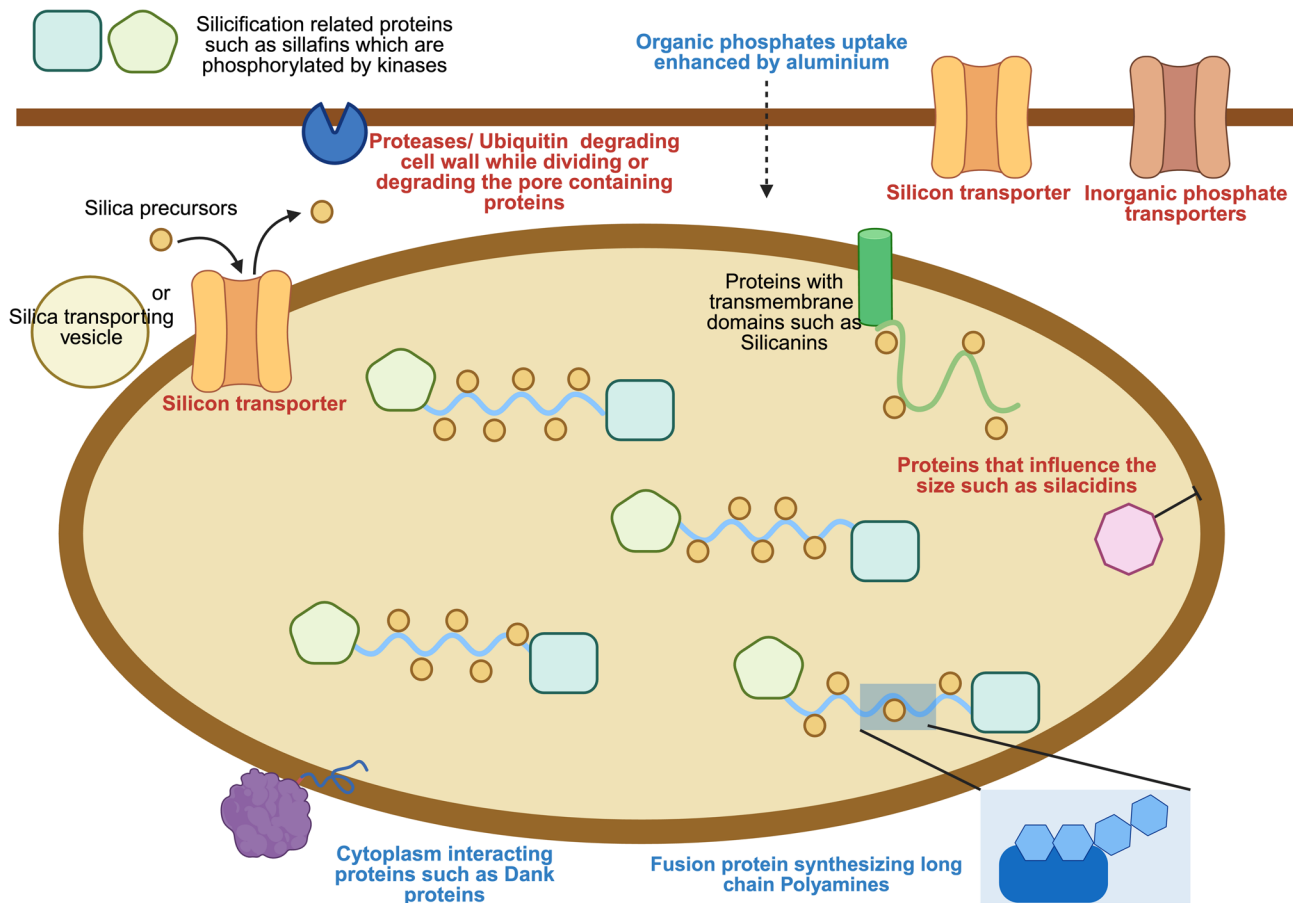


Fig. 8 Putative elements involved in diatom frustule formation or in aiding frustule formation. The model involves different pathways, such as proteolysis to break down proteins involved in pores in the cell wall; fusion proteins for synthesizing long-chain polyamine, which are engaged in silica precipitation; proteins that interact with cytoplasm, such as the dank proteins; and the proteins that are assumed to be involved in silicification, such as silacidins, or involved in transporting the silica precursors, such as the silicon transporters. Various elements are differentially expressed in the presence of aluminium, which may result in frustule morphology change. Red indicates downregulation and blue indicates upregulation in presence of aluminium

Silacidins and silaffins, which have been implicated in diatom silicification, can undergo proteolytic cleavage [24, 49]. As previously reported by Shrestha et al. [50], protein degradation is crucial for pore formation and rearrangement of cell wall components during valve synthesis [50]. Downregulation of most of the ubiquitin-ligases [50] and certain proteases could be an indication of decreased protein degradation in the presence of aluminium. Proteins and long-chain polyamines together form the organic phase that could control the silica polymerization via liquid–liquid phase separation, as proposed by Sumper 2002 [51]. The size of the peptides that form this organic phase could play an essential role in controlling the size of the cell and pores. In addition to these peptides, dAnk proteins [52] were recently identified as key players in pore formation by regulating the phase separation of the organic phase. The upregulation of dAnk proteins could indicate that the cell controls this phase separation process to mediate silicification. Although strong upregulation of dAnk proteins and

protein degradation mechanisms was seen, this strain of *Entomoneis verbralis* did not exhibit much pore-type morphology. Moreover, upregulation of a fusion protein, which has been hypothesized to be involved in the synthesis of long-chain polyamines, was upregulated in the presence of aluminium, suggesting that increases in long-chain polyamine production could also contribute to increases in silicification [53]. In conclusion, although differential regulation of conventional genes involved in silicification was not observed in this study, differential regulation was observed for other mechanisms related to silicification, such as protein degradation, synthesis of long-chain polyamines, and organic phase control. Further research is required to improve the mechanistic understanding of how these processes affect the degree of diatom silicification.

Conducting RNA-seq on aluminium-exposed diatoms provided opportunities to identify genes involved in morphological changes or differences in the diatom's frustule. *E. verbralis* is only significantly silicified in the region



Fig. 9 Protein ID- 1401909, which has unusual enrichment of serine-based motifs from Fig. 6. By further analysis, this protein has different gene models, additional experiments are required to verify which gene model is correct. Different gene models of the protein of interest putatively involved in silicification (1401909 Vs 1389153) and their comparison with Tp Silicidin is shown in this figure. Both gene models have SEVS.SE as a repeating motif, which is similar to a known silicification-associated protein (Silicidin) in *T. pseudonana*, which is known to control the size of the cell. Alternative gene model unusually repeating lysine and proline, which is known characteristics of silicification associated proteins

of the raphe in the form of circular rings (Fig. 1), making it a good candidate for linking the differential expression of genes to specific phenotypic changes or differences in raphe morphology. The identification of genes involved in silicification is challenging because it appears that, based on sequence alone, these genes are poorly conserved among and within species and are poorly characterized [28]. We hypothesize that genes that lack sequence conservation may still share conserved structural elements. Using structure prediction, a structure-based family of proteins was identified that shares a structured tail but does not share much sequence similarity. Some of the silicification-associated proteins, such as the dAnk-like proteins, have some cytoplasm-interacting domains that are structured and simultaneously have similar unstructured tails that could mediate silicification. A similar pattern is observed in this structure-based protein class, which shares some structural similarities with known cytoplasmic protein domains such as microtubules, albeit with low confidence. As one of the proteins in this cluster shares weak sequence similarity to an uncharacterized protein that is present in other sequenced pennate diatoms and is upregulated in cells that have increased silicification, it may be a pennate-specific silicification-associated protein. We speculate that this protein cluster could be involved in establishing a connection between the silica deposition vesicle and cytoplasmic elements that could guide the cell to synthesizing an asymmetric pennate-specific frustule.

Many of the unannotated genes identified in this study appear to be enriched in serine-containing amino acid motifs. Patterns of serine have been associated with silicification [54], but little is known regarding the function of these repeats. However, it is known that serine and tyrosine can undergo post-translational modifications where the hydroxyl group is modified with a negatively charged phosphate. It has been reported that silicification-associated proteins that lack phosphate modifications do not catalyze silica precipitation [54]. We speculate that serine-containing protein motifs might have a putative role in catalyzing silicification. As shown in Fig. 9, we identified protein ID 1401909, which appears unusually enriched in serine moieties. Upon further analyses of this gene product, it has a highly repetitive SEV.S.SE motif. The putative gene model contains additional repetitive PT and K...K motifs, which are known to be associated with silicification [55]. This protein sequence shares similar repeating patterns of serine as seen in silacidin from *T. pseudonana* [56], which is also enriched in a similar motif SEDS. SE; however, the protein from *E. vertebralis* does not share any sequence similarity to the *T. pseudonana* silacidin. Previously, downregulated silacidin was reported to correlate with increased cell size in *T. pseudonana* [56, 57], which is consistent with what we observe, as this particular gene is highly downregulated in the presence of 10 μ M aluminum, which causes larger rings to protrude out of the raphe in *Entomoneis vertebralis*. Exposing the *E. vertebralis* to higher aluminum

concentrations—caused increased size of the rings associated with the raphe, which could account for the increased silicification caused by aluminum.

The artificial seawater is not completely lacking aluminum, the control cultures also contain some amounts of aluminum. Quantitative analyses suggest that the molar ratio of aluminum/silicon in *E.vertebralis* remained approximately constant across varying aluminum concentrations, mirroring trends observed previously in a different diatom strain [37]. This suggests that the observed increase in silicification and incorporation of aluminum could be a way to alleviate aluminum stress and toxicity, as has been observed in plants that embed aluminum in silica nanospheres [22]. Additionally, the overall increase in dissolved organic phosphates, carbon, nitrogen, and iron caused an increase in general metabolism, which may account for enhanced bio-silicification.

Conclusion

Overall, our results characterize and correlate the diatom's transcriptional responses and frustule morphology changes when exposed to aluminum. The transcriptional response to aluminum provides new insights into the different metabolic responses of diatoms to aluminum. In addition to altered frustule properties, the impact on core metabolism has been shown to have some beneficial effects such as increased carbon fixation, which will help us engineer diatoms to produce valuable products and understand their ecological importance. In combination with state-of-the-art computational tools, such as structure-based similarity and protein motif analyses of differentially expressed genes, we have identified a set of proteins that could be directly or indirectly impact diatom silicification. These results advance our understanding of the factors influencing diatom metabolism, silicification, and production of other biosynthetic compounds, paving the way for the identification of new metabolic targets for engineering or controlling biosynthetic processes.

Materials and methods

Cultivation of diatoms

Entomoneis vertebralis (UTEX LB 3214) was received from the UTEX Culture Collection of Algae and cultivated in half salinity artificial seawater medium (~15 psu) supplemented with f/2 additives (ASW-f/2) in 14:12 day light cycle [29]. Prior inoculation, media was pH balanced to 8.2. Precautions were taken to minimize the alterations of the f/2 trace metals and aluminum concentrations during preparation of the media. Specifically, all additives—nitrate, phosphate, silicate, trace metals, vitamins, and aluminum—were filter-sterilized rather than being autoclaved. This strain may also go by *Entomoneis* sp. (11IV18-2A), based on their isolate number.

Aluminum chloride hexahydrate (99.8%) was added to the media to achieve added Aluminum concentrations of 0.2 μ M, 1 μ M, 2 μ M, 5 μ M, and 10 μ M. Three biological replicates were grown. A control culture was grown without Aluminum addition. Culture flasks were shaken manually every single day to assure proper mixing of the medium. Cell counts were taken manually using hemacytometer (HS-3510). After 22 days, cells (9 ml culture) were harvested through centrifugation at 3000 g for 10 min to isolate RNA. After centrifugation, an equal amount of RNA later was added and removed using centrifugation (3000 g for 10 min). The samples were flash-frozen in liquid nitrogen and stored at -80°C . For Biogenic silica analysis, cells (0.25 ml) were collected on a 0.6 μ m polycarbonate filter through vacuum filtration and frozen at -20°C for further analysis.

RNA isolation and data analysis

Total RNA was isolated using a Zymo Quick-DNA Fungal/Bacterial Miniprep Kit. The cells stored in RNA later were lysed in a bead beater (30 s bead beating, 2 min ice for 4 cycles) followed by the instructions in the kit. RNA was eluted in 50 μ l of RNase-free water and then re-eluted through the column to concentrate the RNA and stored at -80°C until further analyses. RNA concentration was checked on the Invitrogen Qubit 2.0 fluorometer. Quality was evaluated with an Agilent 2200 TapeStation or 2100 Bioanalyzer – all samples had a RIN or RINe score above 7. The NEBNext Ultra II Directional kit was employed to prepare RNA for sequencing, following the manufacturer's instructions. RNA was sequenced on a NextSeq 500 platform using a 75 Cycle High Output run on single end mode. Raw reads were analyzed for quality using FastQC v 0.12.1 [58], based on that, trimming and further cleaning was designed. Raw reads were trimmed, and adapter sequences were removed using fastp (setting -q 15) [59]. The reference genome was accessed through the JGI portal, PhycoCosm [32] under the name of *Entomoneis* sp. UTEX 11IV18-2A FL. Reads were aligned to the reference using HISAT2 (-p 12 -dta mode) [60]. Output files from HISAT2 (SAM files) were sorted and indexed using SAMtools v. 1.7 to BAM files [61]. Further, transcripts were assembled using StringTie v2.2.1 (settings -p 20) [31]. The assembled transcripts generated by StringTie were merged in order to account for de-novo transcripts and used as a reference genome to calculate the gene count matrix. Genes with CPM>0.5 in at least one condition were considered for further analysis. Further, the gene count matrix was analyzed using edgeR [62] for differential gene expression analysis. Voom [63] was used to calculate the mean-variance trend, which was further used by limma [64] to create contrasts between different conditions. For further

analysis, genes that are differentially regulated with a p-value less than 0.05, were considered.

Sequence-based annotations were imported through JGI PhycoCosm [32], which included SignalP [65], TMHMM [66], InterPro [67], Uniprot [68], KEGG [69] and KOG [70]. KEGG and KOG annotations were used to explore nitrogen and carbon pathways. Chloroplast/mitochondria localization prediction was done using HECTAR [71]. For structure-based analysis, the protein structures of the genes that did not have any annotation (from above) were predicted using AlphaFold 2.3.1 [72]. Further the AlphaFold [72] code was modified to run on using parallel computing. AlphaFold provides, 5 possible models for a given protein sequence, Model 0 which has the highest confidence among all the predicted models, for each protein was chosen for further analysis. To predict functionality of predicted protein structures and cluster them based on structure, Foldseek was employed [73]. Foldseek was installed on local computing cluster using the AlphaFold proteome and the Protein Data Bank (PDB) as a reference database. In addition to it, a database of AlphaFold predicted known silicification genes. For structural similarity, Foldseek with the easy-search option was used, and for creating structure-based clusters, the easy-cluster option was used with $c = -0.9$ setting. The per-residue local distance difference test (pLDDT) confidence scores for the protein structure models were retrieved from the B-factor field of the coordinate section of the output pdb file, and the final structures were visualized and aligned using UCSF ChimeraX. Protein Motif analysis was conducted using ProminTools [36], with FLPS p-value $> 10^{-6}$, using the predicted protein in the genome as the background proteins. Protein clustering performed by JGI computed based on the TRIBE-MCL clustering method [74] was used to identify genes specific to diatoms by comparing them with related Ochrophyta.

Biogenic silica analysis

Biogenic silica analysis was conducted according to Krause et al. 2009 [30]. Cells were collected on 0.6 μm polycarbonate filters, dried for 2 days, and dissolved in 0.2 M NaOH. Blank tubes with filter were also included in these treatments to account for background interferences. This dissolution of silica was carried out at 95 °C for 1 h and then immediately placed in ice. Excess of NaOH was neutralized by adding 1N HCl. The dissolved sample was centrifuged to remove particulates. The supernatant was diluted, and 4 ml of 6.87 mM of ammonium molybdate (8 g ammonium molybdate in 1000 ml Nanopure water, when dissolved, 24 ml concentrated HCl were added). After 10 min, excess of ammonium molybdate that did not bind to silica was reduced by adding a 6 ml reducing agent. The reducing agent was made by combining 4 solutions: 2 ml Metol-sulfite

reagent with 1.6 ml Nanopure water, 1.2 ml 50% H₂SO₄, and 1.2 ml saturated oxalic acid for each sample. Metol-sulfite solution was made by dissolving 12 g of sodium sulfite (Na₂SO₃) in 1000 ml Nanopure water. Once dissolved 20 g para-methylaminophenol sulfate were added. The samples were kept overnight and measured for fluorescence next day using UV spectrophotometer. Similarly, standards from 0.625 to 56.25 μM silica concentrations were prepared to make a calibration curve. Once the silica concentration was known, it was used to calculate silica per cell using the cell count taken just before the filtration.

SEM and STEM on cleaned diatom frustules

50 ml of exponentially growing cultures were centrifuged (3000 g for 10 min) and further dissolved in 20 ml of 30% sodium peroxide at 60 °C for 1 h (diatoms must have turned white in color). Cleaned frustules were further washed with MilliQ water for 5 times. Further, the cleaned frustules were suspended in 70% ethanol and dried over the SEM stub. To isolate the raphe from the frustule sonication was performed for 2 min under 20% power using a Branson Tip Sonicator. The material was then sputter-coated with gold and imaged using Thermo Scientific Apreo C LoVac FEG SEM at 5 kV accelerating voltage.

Transmission electron microscopy (TEM) images were acquired using Talos F200X (Thermo Scientific) with an accelerating voltage of 200 kV. 10 μl of the diluted H₂O₂ cleaned frustules extracted from diatom cultures of different exposed aluminum concentrations were loaded on the 300-mesh lacey carbon Cu grid (Ted Pella), and the excess of water was removed using filter paper followed by drying in vacuum overnight. Scanning transmission electron microscopy (STEM) was performed on the same equipment in the high-angle annular dark-field (HAADF) mode.

ICP-MS on whole cell to measure aluminum uptake

Cells of *Entomoneis* (11IV18-2A), grown under 0, 5, and 10 μM added aluminum conditions, were washed with 10 mM EDTA (pH 4.5), were collected on a 0.6 μm Hcl cleaned polycarbonate filter and dried overnight at 60 °C. These filters were digested by adding 2 mL of a mixture of concentrated strong acids (final concentrations: 4 M HCl, 4 M HNO₃, and 4 M HF) [75]; and heated for 4 h at 110 °C in a PFA bomb. The acid digest was transferred into another bomb, ensuring complete transfer by rinsing the primary bomb with DI water. The acid digest was further dried for 2–3 h at 110–115 °C. Then, 1 ml of HNO₃/H₂O₂ mix (50% HNO₃, 15% H₂O₂, and 35% DI) was added to the dried acid digest, which was dried further for 2–3 h at 110–115 °C. After cooling down, the digest was reconstituted in 2% HNO₃ and transferred into ICP

vials. Samples were analyzed using quadrupole ICP-MS (PerkinElmer NexION 350D) with an indium internal standard and quantification by external standard curve.

Abbreviations

ICP-MS	Inductively coupled plasma mass spectrometry
HAADF	High-angle annular dark-field imaging
STEM	Scanning transmission electron microscopy
Bsi	Biological silica
CATH	Class architecture topology and homologous superfamily
BLAST	Basic local alignment search tool
HMMER	Hidden markov model
GS-GOGAT	Glutamine synthetase and Glutamine:2-oxo-glutarate aminotransferase
Pfam	Protein family database
KEGG	Kyoto encyclopedia of genes and genomes

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-12106-7>.

Supplementary Material 1.

Acknowledgements

The authors acknowledge using Biological Nanostructures Laboratory within the California NanoSystems Institute, supported by the University of California, Santa Barbara, and the University of California, Office of the President. A part of research was conducted using shared facilities of the Materials Research Science and Engineering Center at UC Santa Barbara, NSF Award No. DMR2308708. Use was made of computational facilities purchased with funds from the National Science Foundation (CNS-1725797) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Material Research Science and Engineering Center (MRSE: NSF DMR 2308708) at UC Santa Barbara. This study utilized cultures from the UTEX Culture Collection of Algae, University of Texas at Austin. In addition to that, we appreciate Dr. Matt P. Ashworth, who helped us acquire this strain and provided guidance regarding its cultivation.

Authors' contributions

RR, HP, and SS conceived and designed the study; RR conducted RNA-seq Bsi analysis, performed bioinformatics, wrote the original draft, and designed figures; HP and SS edited the draft and provided interpretations to data generated; HG conducted TEM; BT performed ICP-MS; SC and IG annotated the genome and ran AlphaFold for unannotated proteins BC, MB and MO supervised the study, provided training, and reviewed all data. All authors have read and approved the manuscript.

Funding

Research was sponsored by the U.S. Army Research Office and accomplished under cooperative agreement W911NF-19-2-0026 for the Institute for Collaborative Biotechnologies. BST was supported the U.S. National Science Foundation (OCE-2049272). The work (proposal <https://doi.org/10.46936/1.025585/60001341>) was conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility and is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Data availability

RNA-seq data generated during this study are available on NCBI SRA under BioProject ID: PRJNA1258573. AlphaFold code and structures are available on request. All the data generated during analysis is reported in supplementary information.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Chemical Engineering, University of California, Santa Barbara, CA 93106, USA

²Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, CA 93106, USA

³Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, East Boothbay, Maine 04544, USA

⁴U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, CA 94720, USA

⁵Department of Plant and Microbial Biology, University of California Berkeley, 110 Koshland Hall, Berkeley, CA 94720, USA

⁶Department of Bioengineering, University of California, Santa Barbara, CA 93106, USA

Received: 21 May 2025 / Accepted: 9 September 2025

Published online: 16 October 2025

References

- Armbrust EV. The life of diatoms in the world's oceans. *Nature*. 2009;459(7244):185–92. Available from: <https://doi.org/10.1038/nature08057>.
- Bowler C, Vardi A, Allen AE. Oceanographic and biogeochemical insights from diatom genomes. *Ann Rev Mar Sci*. 2010;2(Volume 2, 2010):333–65. Available from: <https://www.annualreviews.org/content/journals/10.1146/annurev-marine-120308-081051>.
- Yi Z, Xu M, Di X, Brynjolfsson S, Fu W. Exploring valuable lipids in diatoms. *Front Mar Sci*. 2017;4. Available from: <https://www.frontiersin.org/journals/marine-science/articles/10.3389/fmars.2017.00017>.
- Gordon R, Losic D, Tiffany MA, Nagy SS, Sterrenburg FAS. The glass menagerie: diatoms for novel applications in nanotechnology. *Trends Biotechnol*. 2009;27(2):116–27.
- Hildebrand M, Lerch SJL, Shrestha RP. Understanding diatom cell wall silicification—moving forward. *Front Mar Sci*. 2018;5(APR):125.
- Morse DE. Silicon biotechnology: harnessing biological silica production to construct new materials. *Trends Biotechnol*. 1999;17(6):230–2. Available from: <https://www.sciencedirect.com/science/article/pii/S016779999013098>.
- Rogato R, De Tommasi E. Physical, chemical, and genetic techniques for diatom frustule modification: Applications in nanotechnology. *Applied Sciences (Switzerland)*. 2020;10(23):1–23.
- Losic D, Mitchell JG, Voelcker NH. Diatomaceous lessons in nanotechnology and advanced materials. *Adv Mater*. 2009;21(29):2947–58. Available from: <https://doi.org/10.1002/adma.200803778>.
- Maher S, Maher S, Aw MS, Losic D. Diatom Silica for Biomedical Applications. In: *Diatoms: Fundamentals and Applications* [Internet]. 2019. p. 511–36. Available from: <https://doi.org/10.1002/9781119370741.ch20>.
- Maeda Y, Niwa Y, Tang H, Kisailus D, Yoshino T, Tanaka T. Development of titania-integrated silica cell walls of the titanium-resistant diatom, *fistulifera solaris*. *ACS Appl Bio Mater*. 2018;1(6):2021–9. Available from: <https://doi.org/10.1021/acsabm.8b00520>.
- Azam F, Hemmingsen BB, Volcani BE. Germanium incorporation into the silica of diatom cell walls. *Arch Mikrobiol*. 1973;92(1):11–20. <https://doi.org/10.1007/BF00409507>.
- Tian Q, Liu D, Yuan P, Li M, Yang W, Zhou J, et al. Occurrence of structural aluminum (Al) in marine diatom biological silica: visible evidence from microscopic analysis. *Ocean Science*. 2022;18(2):321–9. Available from: <https://os.copernicus.org/articles/18/321/2022/>.
- Soleimani, M., Rutten, L., Maddala, S.P. et al. Modifying the thickness, pore size, and composition of diatom frustule in *Craspedostauros* sp. with Al³⁺ ions. *Sci Rep* 10, 19498 (2020). <https://doi.org/10.1038/s41598-020-76318-5>.
- Zhou L, Liu F, Liu Q, Fortin C, Tan Y, Huang L, et al. Aluminum increases net carbon fixation by marine diatoms and decreases their decomposition: evidence for the iron–aluminum hypothesis. *Limnol Oceanogr*. 2021;66(7):2712–27. <https://doi.org/10.1002/lno.11784>.
- Zhou, L., Tan, Y., Huang, L. et al. Aluminum effects on marine phytoplankton: implications for a revised Iron Hypothesis (Iron–Aluminum Hypothesis).

- Biogeochemistry 139, 123–137 (2018). <https://doi.org/10.1007/s10533-018-0458-6>.
16. Zhou L, Tan Y, Huang L, Wang WX. Enhanced utilization of organic phosphorus in a marine diatom *Thalassiosira weissflogii*: a possible mechanism for aluminum effect under P limitation. *J Exp Mar Biol Ecol*. 2016;1(478):77–85.
 17. Gensemer RW. The effects of aluminum on phosphorus and silica-limited growth in *Asterionella ralfsii* var. *americana*. *SIL Proceedings*, 1922–2010. 1991;24(4):2635–9. <https://doi.org/10.1080/03680770.1989.11900042>.
 18. Gensemer RW. Role of aluminum and growth rate on changes in cell size and silica content of silica-limited populations of *asterionella ralfsii* var. *americana* (*Bacillariophyceae*)1. *J Phycol*. 1990;26(2):250–8. <https://doi.org/10.1111/j.0022-3646.1990.00250.x>. Jun 1.
 19. Köhler L, Kaskel S, Brunner E. Tailoring the surface of Aluminum-enriched Diatom Biosilica. *Chemie Ingenieur Technik*. 2023;95(11):1844–50. Available from: <https://doi.org/10.1002/cite.202300018>.
 20. Vrieling E, Poort L, Beelen T, Gieskes W. Growth and silica content of the diatoms *Thalassiosira weissflogii* and *Navicula salinarum* at different salinities and enrichments with aluminum. *Eur J Phycol*. 1999;34(3):307–16. <https://doi.org/10.1080/09670269910001736362>.
 21. Britez RM, Watanabe T, Jansen S, Reissmann CB, Osaki M. The relationship between aluminium and silicon accumulation in leaves of *Faremea marginata* (Rubiaceae). *New Phytol*. 2002;156:437–44. <https://doi.org/10.1046/j.1469-8137.2002.00531.x>.
 22. Hodson MJ, Evans DE. Aluminium–silicon interactions in higher plants: an update. *J Exp Bot*. 2020;71(21):6719–29. <https://doi.org/10.1093/jxb/eraa024>.
 23. Kröger N, Deutzmann R, Sumper M. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. *Science* (1979). 1999;286(5442):1129–32.
 24. Wenzl S, Hett R, Richthammer P, Sumper M. Silacidins: highly acidic phosphopeptides from diatom shells assist in silica precipitation in vitro. *Angew Chem Int Ed*. 2008;47(9):1729–32.
 25. Scheffell A, Poulsen N, Shian S, Kröger N. Nanopatterned protein microrings from a diatom that direct silica morphogenesis. *Proc Natl Acad Sci U S A*. 2011;108(8):3175–80.
 26. Kotzsch A, Gröger P, Pawolski D, Bomans PHH, Sommerdijk NAJM, Schlierf M, et al. Silicanin-1 is a conserved diatom membrane protein involved in silica biomineralization. *BMC Biol*. 2017;15(1):1–16. Available from: <https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-017-0400-8>. Cited 2022 Jul 11.
 27. Tesson B, Lerch SJL, Hildebrand M. Characterization of a new protein family associated with the silica deposition vesicle membrane enables genetic manipulation of Diatom Silica. *Sci Rep*. 2017;7(1):1–13. Available from: <https://www.nature.com/articles/s41598-017-13613-8>. Cited 2022 Jul 11.
 28. Skeffington AW, Gentzel M, Ohara A, Milentyev A, Heintze C, Böttcher L, et al. Shedding light on silica biomineralization by comparative analysis of the silica-associated proteomes from three diatom species. *Plant J*. 2022;110(6):1700–16. <https://doi.org/10.1111/tpj.15765>.
 29. Guillard RRL. Culture of phytoplankton for feeding marine invertebrates. In: *Culture of Marine Invertebrate Animals*: Proceedings — 1st Conference on Culture of Marine Invertebrate Animals Greenport. Smith WL, Chanley MH, editors. Boston, MA: Springer US; 1975. p. 29–60. Available from: https://doi.org/10.1007/978-1-4615-8714-9_3.
 30. Krause JW, Nelson DM, Lomas MW. Biogeochemical responses to late-winter storms in the Sargasso Sea, II: Increased rates of biogenic silica production and export. *Deep Sea Research Part I: Oceanographic Research Papers*. 2009;56(6):861–74. Available from: <https://www.sciencedirect.com/science/article/pii/S096706370900020X>.
 31. Grigoriev IV, Hayes RD, Calhoun S, Kamel B, Wang A, Ahrendt S, et al. PhycoCosm, a comparative algal genomics resource. *Nucleic Acids Res*. 2021;49(D1):D1004–11. Available from: <https://doi.org/10.1093/nar/gkaa898>.
 32. Perteu M, Perteu GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. Stringtie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–5. <https://doi.org/10.1038/nbt.3122>.
 33. Michael AJ. Molecular machines encoded by bacterially-derived multi-domain gene fusions that potentially synthesize, N-methylate and transfer long chain polyamines in diatoms. *FEBS Lett*. 2011;585(17):2627–34. Available from: <https://www.sciencedirect.com/science/article/pii/S0014579311005679>.
 34. Smith SR, Dupont CL, McCarthy JK, Broddrick JT, Obornik M, Horák A, et al. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat Commun*. 2019;10(1):4552. Available from: <https://doi.org/10.1038/s41467-019-12407-y>.
 35. Barka F, Angstenberger M, Ahrendt T, Lorenzen W, Bode HB, Büchel C. Identification of a triacylglycerol lipase in the diatom *Phaeodactylum tricornutum*. *Biochim Biophys Acta (BBA) - Mol Cell Biol Lipids*. 2016;1861(3):239–48. Available from: <https://www.sciencedirect.com/science/article/pii/S1388198115002450>.
 36. Skeffington AW, Donath A. ProminTools: Shedding light on proteins of unknown function in biomineralization with user friendly tools illustrated using mollusc shell matrix protein sequences. *bioRxiv*. 2020;2020.03.05.978718. Available from: <http://biorxiv.org/content/early/2020/03/05/2020.03.05.978718.abstract>.
 37. Liu D, Yuan P, Tian Q, Liu H, Deng L, Song Y, et al. Lake sedimentary biogenic silica from diatoms constitutes a significant global sink for aluminium. *Nat Commun*. 2019;10(1):4829. <https://doi.org/10.1038/s41467-019-12828-9>.
 38. Raven JA, Raven JA. Iron in Diatoms. In: *Diatoms: Fundamentals and Applications*. 2019. pp. 213–24. Available from: <https://doi.org/10.1002/9781119370741.ch9>.
 39. Bender S, Durkin C, Berthiaume C, Morales R, Armbrust EV. Transcriptional responses of three model diatoms to nitrate limitation of growth. *Front Mar Sci*. 2014;1. Available from: <https://www.frontiersin.org/articles/10.3389/fmars.2014.00003>.
 40. Bernard SM, Habash DZ. The importance of cytosolic glutamine synthetase in nitrogen assimilation and recycling. *New Phytol*. 2009;182(3):608–20. <https://doi.org/10.1111/j.1469-8137.2009.02823.x>.
 41. Daniel R, Stuert K, Gottschalk G. Biochemical and molecular characterization of the oxidative branch of glycerol utilization by *Citrobacter freundii*. *J Bacteriol*. 1995;177(15):4392–401. Available from: <https://doi.org/10.1128/jb.177.15.4392-4401.1995>.
 42. Leyland B, Boussiba S, Khozin-Goldberg I. A review of diatom lipid droplets. *Biology (Basel)*. 2020;9(2). Available from: <https://www.mdpi.com/2079-7737/9/2/38>.
 43. Tanaka T, Maeda Y, Veluchamy A, Tanaka M, Abida H, Maréchal E, et al. Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. *Plant Cell*. 2015;27(1):162–76. <https://doi.org/10.1105/tpc.114.135194>.
 44. Zou S, Lu Y, Ma H, Li Y, Chen G, Han D, et al. Microalgal glycerol-3-phosphate acyltransferase role in galactolipids and high-value storage lipid biosynthesis. *Plant Physiol*. 2023;192(1):426–41. <https://doi.org/10.1093/plphys/kiad091>.
 45. Thamatrakoln K, Alverson AJ, Hildebrand M. Comparative sequence analysis of diatom silicon transporters: toward a mechanistic model of silicon transport. *J Phycol*. 2006;42(4):822–34. <https://doi.org/10.1111/j.1529-8817.2006.00233.x>.
 46. P SR, Mark H. Evidence for a regulatory role of Diatom Silicon transporters in cellular silicon responses. *Eukaryot Cell*. 2014;14(1):29–40. Available from: <http://doi.org/10.1128/ec.00209-14>.
 47. Mitani-Ueno N, Yamaji N, Ma JF. High silicon accumulation in the shoot is required for down-regulating the expression of Si transporter genes in rice. *Plant Cell Physiol*. 2016;57(12):2510–8. <https://doi.org/10.1093/pccp/pcw163>.
 48. Kumar S, Rechav K, Kaplan-Ashiri I, Gal A. Imaging and quantifying homeostatic levels of intracellular silicon in diatoms. *Sci Adv*. 2024;6(42):eaaz7554. Available from: <https://doi.org/10.1126/sciadv.aaz7554>.
 49. Kroger N, Poulsen N. Biochemistry and Molecular Genetics of Silica Biomineralization in Diatoms. In: *Handbook of Biomineralization*. 2007. pp. 43–58. Available from: <https://doi.org/10.1002/9783527619443.ch3>.
 50. Shrestha RP, Tesson B, Norden-Krichmar T, Federowicz S, Hildebrand M, Allen AE. Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. *BMC Genomics*. 2012;13(1):499. <https://doi.org/10.1186/1471-2164-13-499>.
 51. Sumper M. A phase separation model for the nanopatterning of diatom biosilica. *Science* (1979). 2002;295(5564):2430–3. Available from: <https://www.science.org/doi/10.1126/science.1070026>. Cited 2024 Aug 14.
 52. Heintze C, Babenko I, Zackova Suchanova J, Skeffington A, Friedrich BM, Kröger N. The molecular basis for pore pattern morphogenesis in diatom silica. *Proc Natl Acad Sci U S A*. 2022;119(49):e2211549119. <https://doi.org/10.1073/pnas.2211549119>.
 53. Kröger N, Deutzmann R, Bergsdorf C, Sumper M. Species-specific polyamines from diatoms control silica morphology. *Proc Natl Acad Sci U S A*. 2000;97(26):14133–8. <https://doi.org/10.1073/pnas.260496497>.
 54. Kröger N. (2022). *Biomolecules Involved in Frustule Biogenesis and Function*. In: Falcietore, A., Mock, T. (eds) *The Molecular Life of Diatoms*. Springer, Cham. https://doi.org/10.1007/978-3-030-92499-7_12.
 55. Hildebrand M, Lerch SJL. Diatom silica biomineralization: parallel development of approaches and understanding. *Semin Cell Dev Biol*. 2015;46:27–35.

56. Belshaw N, Grouneva I, Aram L, Gal A, Hopes A, Mock T. Efficient gene replacement by CRISPR/Cas-mediated homologous recombination in the model diatom *Thalassiosira pseudonana*. *New Phytol*. 2023;238(1):438–52. <https://doi.org/10.1111/nph.18587>.
57. Kirkham AR, Richthammer P, Schmidt K, Wustmann M, Maeda Y, Hedrich R, et al. A role for the cell-wall protein silacidin in cell size of the diatom *Thalassiosira pseudonana*. *ISME J*. 2017;11(11):2452–64.
58. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
59. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*. 2023;2(2):e107. <https://doi.org/10.1002/imt2.107>.
60. Zhang Y, Park C, Bennett C, Thornton M, Kim D. Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Res*. 2021;31(7):1290–5. <http://genome.cshlp.org/content/31/7/1290.abstract>.
61. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008. <https://doi.org/10.1093/gigascience/giab008>.
62. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
63. Law CW, Chen Y, Shi W, Smyth GK. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15(2):R29. <https://doi.org/10.1186/gb-2014-15-2-r29>.
64. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47–e47. Available from: <https://doi.org/10.1093/nar/gkv007>.
65. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. Signalp 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol*. 2022;40(7):1023–5. <https://doi.org/10.1038/s41587-021-01156-3>.
66. Melén K, Krogh A, von Heijne G. Reliability Measures for Membrane Protein Topology Prediction Algorithms. *J Mol Biol*. 2003;327(3):735–44. Available from: <https://www.sciencedirect.com/science/article/pii/S0022283603001827>.
67. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. *Nucleic Acids Res*. 2005;33(suppl_2):W116–20. Available from: <https://doi.org/10.1093/nar/gki442>.
68. Consortium TU. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res*. 2025;53(D1):D609–17. Available from: <https://doi.org/10.1093/nar/gkae1010>.
69. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*. 2006;34(suppl_1):D354–7. Available from: <https://doi.org/10.1093/nar/gkj102>.
70. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*. 2003;4(1):41. Available from: <https://doi.org/10.1186/1471-2105-4-41>.
71. Gschloessl B, Guernsey Y, Cock JM. HECTAR: A method to predict subcellular targeting in heterokonts. *BMC Bioinformatics*. 2008;9(1):393. Available from: <https://doi.org/10.1186/1471-2105-9-393>.
72. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2>.
73. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol*. 2024;42(2):243–6. <https://doi.org/10.1038/s41587-023-01773-0>.
74. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*. 2002;30(7):1575–84. Available from: <https://doi.org/10.1093/nar/30.7.1575>.
75. Ohnemus DC, Auro ME, Sherrell RM, Lagerström M, Morton PL, Twining BS, et al. Laboratory intercomparison of marine particulate digestions including Piranha: a novel chemical method for dissolution of polyethersulfone filters. *Limnol Oceanogr Methods*. 2014. <https://doi.org/10.4319/lom.2014.12.530>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.