DNA Research, 2022, **29**, 1–12
https://doi.org/10.1093/dnares/dsac037
Advance access publication 5 October 2022
Research article

OXFORD

Research article

Rapid transcriptomic and physiological changes in the freshwater pennate diatom *Mayamaea pseudoterrestris* in response to copper exposure

Shigekatsu Suzuki[1,*], Shuhei Ota[1], Takahiro Yamagishi[2], Akihiro Tuji[3], Haruyo Yamaguchi[1], and Masanobu Kawachi[1]

[1]Biodiversity Division, National Institute for Environmental Studies, Tsukuba, Japan
[2]Health and Environmental Risk Division, National Institute for Environmental Studies, Tsukuba, Japan
[3]Department of Botany, National Museum of Nature and Science, Tsukuba, Japan
[*]Corresponding author: Tel. +81 29 850 2204. Fax. +81 29 850 2587. Email: suzuki.shigekatsu@nies.go.jp

Abstract

Diatoms function as major primary producers, accumulating large amounts of biomass in most aquatic environments. Given their rapid responses to changes in environmental conditions, diatoms are used for the biological monitoring of water quality and for performing ecotoxicological tests in aquatic ecosystems. However, the molecular basis for their toxicity to chemical compounds remains largely unknown. Here, we sequenced the genome of a freshwater diatom, *Mayamaea pseudoterrestris* NIES-4280, which has been proposed as an alternative strain of *Navicula pelliculosa* UTEX 664 for performing the Organisation for Economic Co-operation and Development ecotoxicological test. This study shows that *M. pseudoterrestris* has a small genome and carries the lowest number of genes among freshwater diatoms. The gene content of *M. pseudoterrestris* is similar to that of the model marine diatom, *Phaeodactylum tricornutum*. Genes related to cell motility, polysaccharide metabolism, oxidative stress alleviation, intracellular calcium signalling, and reactive compound detoxification showed rapid changes in their expression patterns in response to copper exposure. Active gliding motility was observed in response to copper addition, and copper exposure decreased intracellular calcium concentration. These findings enhance our understanding of the environmental adaptation of diatoms, and elucidate the molecular basis of toxicity of chemical compounds in algae.

Key words: Mayamaea pseudoterrestris, ecotoxicological test, metal toxicity, gliding motility, oxidative stress

1. Introduction

Diatoms are one of the most diverse groups of algae and are the main primary producers in freshwater, brackish water, and seawater environments.[1–3] Diatoms account for up to 20–40% of the carbon fixation in the ocean. In freshwater, diatoms inhabit standing water (ponds and lakes) and flowing water (rivers and shallows), and exhibit planktonic and benthic lifestyles.[4] Diatoms, with different sizes and shapes of silica cell frustules,[5] are estimated to consist of more than 100,000 potential species, including undiscovered and unidentified taxa.[6]

Because each diatom species responds rapidly to specific environmental conditions, diatoms have been used to monitor water quality in rivers, shallows, and lakes for a long time.[7] To monitor and assess water pollution, changes in the composition of a microbial communities have been observed.[8–10] The diatom community is affected by various environmental conditions, such as trophic and nutrient conditions,[9] and atmospheric conditions, such as light and temperature.[11] Indices using diatom communities have been proposed to detect metal pollution.[12,13] Therefore, understanding the physiological responses of diatoms to metal exposure is important, as these are the driving forces behind changes in the diatom species composition. The physiological responses of diatoms to heavy metal exposure have been reported in several studies[14–18]; however, most of these studies have focused on specific metabolic pathways, such as scavenging systems of reactive oxygen species (ROS).

Gonçalves *et al.*[15] revealed the comprehensive metabolic responses of a heavy metal-tolerant freshwater diatom, *Tabellaria flocculosa*, to zinc (Zn) using metabolomic approaches. For example, under Zn exposure, the activity of ROS scavenging enzymes, including superoxide dismutase (SOD), catalase (CAT), and glutathione-*S*-transferase (GST), significantly increased, and glutathione was mainly found to exist in its reduced form (GSH). Zn exposure also induced a change in the amount of attached exopolysaccharides (EPSs) and frustulins in *T. flocculosa*. The amount of EPSs decreased at low Zn concentrations but increased at high Zn concentrations, and the amount of frustulins increased at high Zn concentrations. The authors inferred that diatoms chelate heavy metals using EPSs and frustulins to reduce metal toxicity. In contrast to the increase in EPSs, the sucrose levels were reduced under Zn exposure. However, the genetic basis of the

Received 10 May 2022; Revised 27 September 2022; Accepted 3 October 2022

metabolic response of freshwater diatoms to metal exposure remains largely unknown.

Most pennate diatoms exhibit a smooth gliding motility.[19] Although the detailed molecular mechanisms for gliding remain poorly understood, except for relationships between actin and myosin,[19–21] heavy metal pollution affects the motility patterns of natural pennate diatoms.[22] An increase of copper (Cu) and Zn induces erratic motility, and the motility pattern has been proposed as a potential biomonitoring tool to evaluate the water quality. Therefore, the transcriptional response of pennate diatoms under heavy metal exposure can provide key information for elucidating the detailed mechanism for cell motility of pennate diatoms.

Given the importance of diatoms in understanding the impact of chemical compounds on primary producers in aquatic environments, the diatom strain UTEX 664 (*Navicula pelliculosa*) has been recommended for ecotoxicological analyses by the Organisation for Economic Co-operation and Development (OECD; TG201).[23] Ecotoxicological test is used to evaluate the effects of chemical compounds, including heavy metals, on algal growth in water. However, this strain is currently extinct in UTEX and other official culture collections. Therefore, a freshwater species, *Mayamaea pseudoterrestris* NIES-4280 (= UTEX 661), which is thought to be closely related to UTEX 664, has been proposed as an alternative strain for ecotoxicological test.[24] NIES-4280 was originally a terrestrial alga isolated from soil, together with UTEX 664; however, both of these strains can be cultivated in freshwater. Connolly *et al.*[25] showed that a freshwater diatom strain related to NIES-4280 and UTEX 664 had a smaller genome, which led us to speculate that the metabolic responses of this strain to heavy metals may be simpler in freshwater diatoms. In this study, we sequenced the genome and transcriptome of *M. pseudoterrestris* to elucidate its metabolic responses to heavy metals, especially Cu, and to determine the genomic basis of this response. High concentrations of Cu are toxic and are used for coating ships[26] and as agricultural pesticides[27] in natural environments; thus, Cu is a potential pollutant.

2. Materials and methods

2.1. Diatom cultivation

Mayamaea pseudoterrestris strain NIES-4280 (= UTEX 661) was obtained from the National Institute for Environmental Studies (NIES, Japan). The strain was maintained in CSi medium[28] at 20°C under approximately 80 µmol photons/m²/s light intensity and a 12-h light/12-h dark cycle.

2.2. DNA extraction and genome sequencing

To perform short-read sequencing, the NIES-4280 strain was grown in 100 ml of CSi medium for several days. The cells were collected via gentle centrifugation (3,000 × *g* for 3 min) and stored at −80°C freezer for 10 min. DNA was extracted from frozen cells using the MagAttract HMW DNA Kit (Qiagen, Düsseldorf, Germany), according to the manufacturer's protocol, and then concentrated using the Genomic DNA Clean & Concentrator Kit (Zymo Research, Irvine, CA, USA). To construct the library, approximately 380 ng of DNA was sheared using a Covaris M220 Focused-ultrasonicator (Covaris, Woburn, MA, USA) to a fragment size of approximately 550 bp. DNA libraries were then constructed using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA), and sequenced on Illumina MiSeq and HiSeqX platforms (Illumina, San Diego, CA, USA) to generate 300 and 150 bp paired-end reads, respectively. A total of 1.8 Gb (59× coverage) and 3.5 Gb (114× coverage) data were generated from the MiSeq and HiSeqX platforms, respectively (Supplementary Table S1). The sequence reads were deposited to the DNA Database of Japan (DDBJ)/GenBank/European Nucleotide Archive (ENA) under accession numbers DRR374606 and DRR374607.

To perform long-read sequencing, the NIES-4280 strain was grown in approximately 400 ml of CSi medium for 1 week. The cells were collected via gentle centrifugation and lysed in 1 ml of sodium dodecyl sulphate (SDS) buffer (2% SDS, 50 mM Tris–HCl [pH 8.0], 50 mM EDTA, 150 mM NaCl, and 0.5% 2-mercaptoethanol). The cells were then incubated at 50°C for 2 h with gentle rotation. Then, 200 µl of 5 M potassium acetate was added to the mixture, and the cell debris was removed via centrifugation at 20,000 × *g* for 2 min at 4°C. After the pellet was removed, the DNA was purified using phenol:chloroform:isoamyl alcohol (25:24:1) and chloroform:isoamyl alcohol (24:1) mixtures. Subsequently, DNA was extracted from the aqueous phase via ethanol precipitation and treated with 1 µl of RNase Cocktail Enzyme Mix (ThermoFisher Scientific, Waltham, MA, USA) at 37°C for 5 min, to remove the RNA. Subsequently, the DNA was purified using the Genomic DNA Clean & Concentrator Kit. Short DNA fragments were removed using 0.5 volume of Ampure XP (Beckman Coulter, Brea, CA, USA). The DNA library for Nanopore sequencing was constructed from 1.6 µg of DNA using the Ligation Sequencing Kit (SQK-LSK109; Oxford Nanopore Technologies, Oxford, UK), according to the 'Genomic DNA by Ligation (SQK-LSK109) version GDE_9063_v109_revT_14Aug2019' protocol. Sequencing was performed using MK1C (Oxford Nanopore Technologies), generating 1.6 Gb data (52× coverage; read N50 length: 8.6 kb) (Supplementary Table S1). The sequence reads were deposited to DDBJ/GenBank/ENA under accession number DRR374608.

2.3. Genome assembly and annotation

Genome size was estimated from MiSeq reads using the GenomeScope webserver.[29] Nanopore long reads were individually assembled using Raven version 1.6.0[30] (with default settings) and NextDenovo version 2.4.0 (https://github.com/Nextomics/NextDenovo). Raw Nanopore reads were mapped to both assemblies using minimap2-2.17-r941, with the option -x map-ont.[31] Both the assemblies were polished using PEPPER-Margin-DeepVariant pipeline version 0.0.6 (https://github.com/kishwarshafin/pepper) and subsequently using NextPolish v1.2.4,[32] with Illumina reads trimmed using fastp version 0.20.0[33] with default settings. Both assemblies were merged with Raven-based assembly (query) and NextDenovo-based assembly (reference) using MAC 2.0.[34] The merged assembly was polished using POLCA[35] in MaSuRCA version 4.0.4,[36] using trimmed Illumina reads. Subsequently, because diatoms are diploid, haplotigs were purged from the merged assembly using Purge Haplotigs version 1.1.1,[37] with the following options: -l 5, -m 25, and -h 80. To purge the haplotigs, Nanopore reads were mapped to the assembly using minimap2-2.17-r941, with the option -x map-ont. Putative organellar genomes were identified using blastx with the available organellar genomes of diatoms and removed. The final assembly comprised 41 contigs and

30,627,229 bp (N50 length: 1,068 kb). The completeness of the assembly was evaluated using BUSCO 5.0.0 (with the eukaryotic dataset eukaryota_odb10),[38] and RNA-seq reads (2.2 Gb) were mapped to the assembly (described below) using hisat2 version 2.2.1.[39]

To annotate the genome, the final assembly was soft-masked using RepeatModeler version 2.0.2[40] and RepeatMasker version 4.1.1 (https://www.repeatmasker.org/). Gene models were constructed from RNA-seq reads using the funannotate pipeline (version 1.8.9; https://github.com/nextgenusfs/funannotate). The funannotate pipeline was also used for functional annotation and comparative analyses. Subcellular localization of the proteins was predicted using the HECTAR web server.[41]

2.4. Phylogenetic analysis

Phylogenetic analysis of 284 nuclear-encoded proteins was performed using *Triparma laevis* (GCA_012489335.1)[42] as an outgroup, whose assembly was annotated in this study. The repeat sequences of the assembly were soft-masked using RepeatModeler (version 2.0.1) and RepeatMasker (version 4.1.1). Gene models were predicted using the funannotate pipeline (version 1.8.0). The dataset was composed of 12 diatom species. Highly conserved orthologues were searched using reciprocal best-hit analyses with the following cutoffs: similarity >50% and High-scoring Segment Pair (HSP) coverage >50%. The dataset contained 128,950 amino acids belonging to 284 proteins shared by all species. Sequence alignment, read trimming, model testing, maximum likelihood analysis, and Bayesian analysis were performed using MAFFT v7.475,[43] trimAl v1.4.rev15,[44] ModelTest-NG version 0.1.6,[45] RAxML-NG version 1.0.3,[46] and MrBayes version 3.2.7a,[47] respectively, as described by Makita *et al.*[48]

2.5. Prediction of genes encoding G-protein-coupled receptors, G proteins, and effectors

Genes encoding G-protein-coupled receptors (GPCRs) were predicted using the GPCR-PEn web server,[49] with positive predictions obtained using both GPCRpred[50] and GPCRtm.[51] Genes encoding G proteins and effectors were predicted by BLAST search using the human-gpDB web server,[52] with an *E*-value cutoff of <1*E*−3.

2.6. Comparative transcriptome analyses under Cu exposure

Before Cu exposure, the NIES-4280 strain was precultivated in 25 ml of modified OECD medium[53] in six glass flasks for 2 days. Then, $CuSO_4$ was added to three flasks (final concentration: 1.75 µM) 3 h after illumination, while the remaining three flasks were used as controls. The concentration of Cu^{2+} corresponded to the value of half maximal effective concentration (EC50) of UTEX B673.[53] The cells were cultivated for an additional hour. Subsequently, the cells were collected via centrifugation at 3,000 × *g* for 2 min and immediately frozen using liquid nitrogen. RNA was extracted from the cells using RNeasy Plus Mini Kit (Qiagen), according to the manufacturer's protocol. To predict gene models, RNA was extracted from cells in the dark phase.

RNA libraries were constructed using the NEBNext Ultra Directional RNA Library Prep Kit (New England BioLabs), with the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England BioLabs), according to the manufacturer's instructions. The libraries were sequenced on the Illumina MiSeq platform to generate 300 bp paired-end reads (Supplementary Table S1), which were later deposited to DDBJ/GenBank/ENA under accession numbers DRR374599–DRR374605. All reads were trimmed using fastp version 0.20.0[33] (default settings) and then mapped to the genome assembly using hisat2 (default settings). The number of reads mapped per gene was counted using featureCounts v2.0.1.[54] Subsequently, 9,714 genes mapped by at least three reads in one of the replicates were extracted and used for expression analysis. Differentially expressed genes (DEGs) were identified using the likelihood ratio test and fold-change (FC) in edgeR (version 3.32.1),[55] with a false discovery rate (FDR) cutoff of <0.05 and logFC >0.5 or <−0.5. Subsequently, gene ontology (GO) enrichment analysis of the DEGs was performed using goatools (version 1.2.3)[56] with an FDR cutoff of <0.05. The FDR values were corrected using the Benjamini–Hochberg procedure.

2.7. Evaluation of NIES-4280 growth inhibition under Cu exposure

To evaluate the effect of Cu on the growth of NIES-4280, the chlorophyll fluorescence of the strain was measured using SpectraMax M2e (Molecular Devices, San Jose, CA, USA), as described previously.[53] Cells in CSi medium were transferred to modified OECD medium and incubated for 2 days. Subsequently, the cells were collected via gentle centrifugation and washed once with modified OECD medium. The cells were resuspended in modified OECD medium, and 200 µl of the cell suspension (2.8×10^6 cells/ml) was aliquoted into each well of a 96-well microplate. Then, $CuSO_4$ was added to each well (final concentration of Cu^{2+}: 0.157, 0.787, 1.57, 7.87, 15.75, and 157.48 µM), and cells were grown for 84 h. Three independent biological replicates were performed for each Cu^{2+} treatment.

2.8. Time-lapse video microscopy

The NIES-4280 cells were cultivated for a few weeks in CSi medium. The cells (2 ml) were transferred to OECD agar plates (modified OECD medium supplemented with 1% agarose) in 50 mm Petri dishes (AGC Techno Glass, Shizuoka, Japan). The cells were incubated for 6 h and washed twice with modified OECD medium. Subsequently, the cells were incubated for 2 days under the same light conditions. After incubation, the medium was changed to 2 ml of modified OECD medium with or without 2.68 µM $CuSO_4$ and the cells were incubated for 1 h under light. Observations were performed using a Nikon ECLIPSE Ni-U microscope (Nikon, Tokyo, Japan) equipped with a DS-Fi3 digital camera (Nikon). Images were taken every 10 s for 10 min.

Sixty images were stacked, converted to 8-bit colours, and colour-inverted using Fiji (Image J2 version 2.3.0).[57] Cell motility was analysed using TrackMate version 7.7.2.[58] The cells were detected using a LoG (Laplacian of Gaussian) detector with an estimated object diameter of 10 pixels and quality threshold of 2. The cells were tracked using a simple LAP tracker with default options. We defined cells with a mean speed of <3 µm/min as immotile.

2.9. Detection of changes in intracellular Ca^{2+} concentration

Changes in the intracellular Ca^{2+} concentration of NIES-4280 under Cu exposure were detected via Fluo 3-AM staining (Dojindo Laboratories, Kumamoto, Japan). Briefly, NIES-4280 cells were transferred from CSi medium to modified

OECD medium and cultured for 1 h. Cells were collected via gentle centrifugation and incubated in 500 µl of Fluo 3 medium (25 µg/ml Fluo 3-AM and 0.4% Pluronic F-127) for 1 h in the dark. Subsequently, the cells were collected via gentle centrifugation, and incubated in modified OECD medium supplemented with or without 2.68 µM $CuSO_4$ for 2 h under illumination. The green fluorescence (FL2; 543/22 nm) of the non-colonized cells was detected with an On-chip Sort (On-chip Biotechnologies, Tokyo, Japan) using 488 nm blue laser excitation.

3. Results and discussion

3.1. Genomic characterization of *M. pseudoterrestris*

The genome assembly of *M. pseudoterrestris* strain NIES-4280 was 30,627,229 bp in size and contained 41 contigs with no gaps (Supplementary Table S2). The contig N50 length was 1,068 kb. However, the total number of chromosomes in this species has not been estimated yet. Based on k-mer analysis, the genome size of strain NIES-4280 was predicted to be 27.2 Mb (Supplementary Fig. S1), which was quite close to the assembly size. The k-mer spectrum showed two clear peaks, suggesting that the NIES-4280 strain is diploid with low heterogeneity (2.97%). Repeat sequences comprised at least 5.8 Mb (19.0%) of the assembly and were mainly represented by the Ty1-Copia group of long terminal repeat retrotransposons (Supplementary Table S3). The completeness of the NIES-4280 genome was 81.1%, based on the assessment of the eukaryotic dataset with BUSCO version 5.0.0, whereas that of the other species was 55.7–80.4% (Fig. 1a). Most RNA-seq reads (84.4%) were mapped to the NIES-4280 genome assembly. Additionally, 11,019 protein-coding genes were predicted in the *M. pseudoterrestris* genome (Fig. 1b). The proteins of *M. pseudoterrestris* contained 79.6% of the conserved eukaryotic proteins in the BUSCO dataset, whereas those of the other diatom species contained 81.2–53.7% of the eukaryotic proteins (Fig. 1a). Approximately 81.3% of

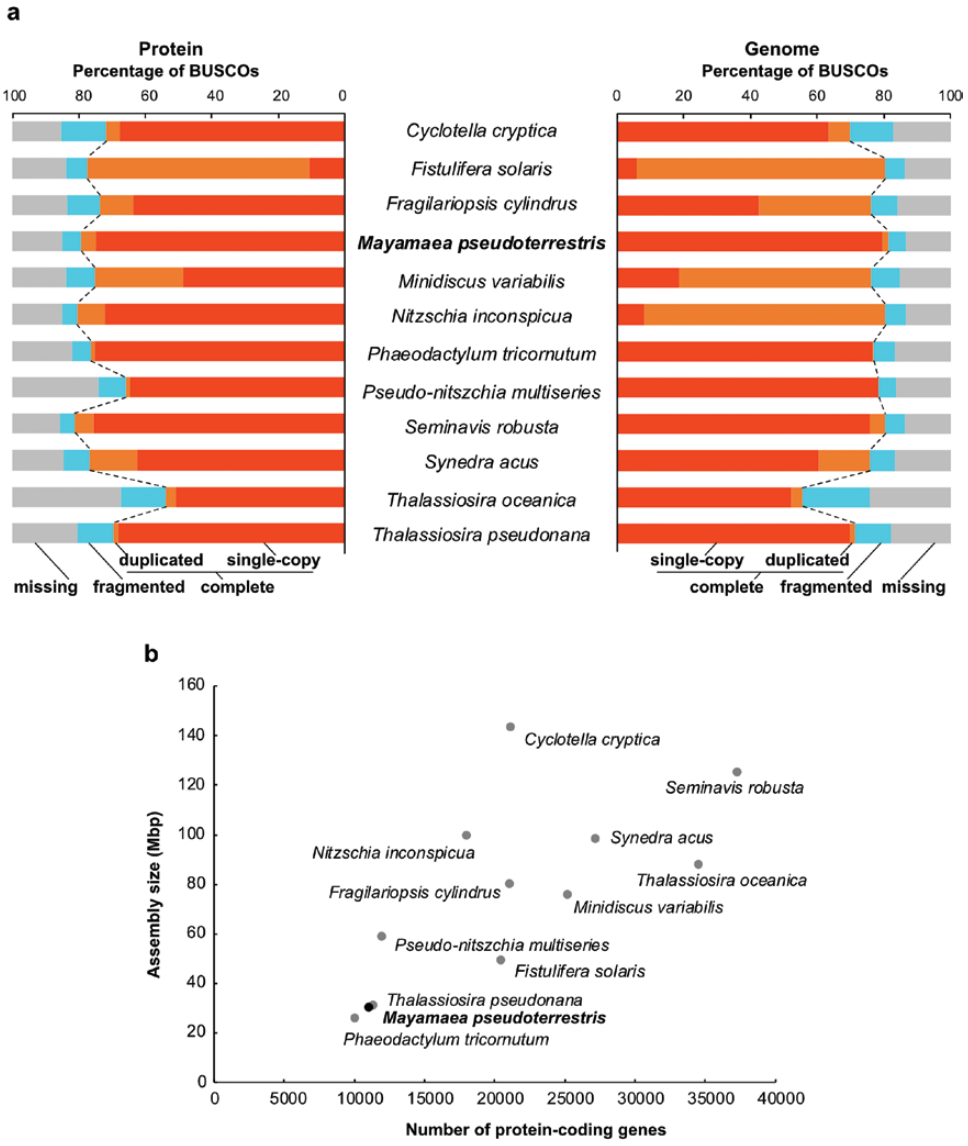

Figure 1. Quality and structural diversity of diatom genomes. (a) Assessment of genome completeness. BUSCO analyses were performed using eukaryotic protein (left) and genomic (right) datasets. (b) Plots showing the size of genome assembly and the number of protein-coding genes. The genome of *Mayamaea pseudoterrestris* was one of the smallest among diatom genomes, with compact gene contents.

RNA-seq reads were mapped to the predicted transcripts. Thus, the *M. pseudoterrestris* genome shows a high degree of completeness and highly accurate gene models.

To infer the phylogenetic relationship between *M. pseudoterrestris* and other representative diatoms with sequenced genomes, we performed phylogenetic analysis based on the amino acid sequences of 284 proteins (Fig. 2a). *Mayamaea pseudoterrestris* clustered with *Phaeodactylum tricornutum* [bootstrap value (BP) = 93], which branches from *Fisturifella solaris*. The genera belonging to the order Naviculales, including *Mayamaea*, *Phaeodactylum*, *Fisturifella*, and *Seminavis*, formed a monophyletic group.

The genome size of *M. pseudoterrestris* (30.6 Mb) was only slightly larger than that of *P. tricornutum* (26.1 Mb), which possessed the smallest genome among all diatom species evaluated in this study (Fig. 1b; Supplementary Table S2). Consistently, the number of genes in the *M. pseudoterrestris* genome was also low (11,019 genes) but slightly higher than that in the *P. tricornutum* genome (10,020 genes). Non-metric multidimensional scaling analysis based on InterProScan domains showed that *M. pseudoterrestris* is closely related to *P. tricornutum*, *Pseudo-nitschia multiseries*, and *Thalassiosira pseudonana* (Supplementary Fig. S2). These data suggest that genome structure and gene function are conserved between *M. pseudoterrestris* and *P. tricornutum*, although these species dwell in different aquatic environments (freshwater and marine, respectively).

To elucidate the evolution of the gene repertoires of representative diatoms, we performed gene family analysis using the Dollo parsimony model[59] (Fig. 2b). Based on 12 representative diatoms and *Triparma* (order Parmales), 26,986 orthogroups were found. The common ancestor of all diatoms contained 10,761 orthogroups, whereas the ancestor of the Naviculales order possessed 11,586 orthogroups. In Naviculales, the common ancestor of *P. tricornutum*, *M. pseudoterrestris*, and *Fistulifera solaris* lost 2,379 orthogroups after the branching of *Seminavis robusta*. *Phaeodactylum tricornutum*, *M. pseudoterrestris*, and *F. solaris* contained similar numbers of orthogroups (7,743, 7,818, and 8,640, respectively), and lost 1,214–1,602 orthogroups in each lineage. The *F. solaris* genome was fragmented[60]; thus, the deletion of orthogroups in the *F. solaris* lineage may be overestimated. In the *M. pseudoterrestris* lineage, 1,553 gene families were lost after branching of *P. tricornutum*. GO enrichment analysis showed that many G-protein-related GO terms were significantly decreased (P <0.05) in *M. pseudoterrestris*, including 'G-protein-coupled receptor signalling pathway', 'G-protein-coupled receptor activity', 'signalling receptor activity', 'molecular transducer activity', and 'transmembrane signalling receptor activity' (Supplementary Table S4). A search for genes involved in the G-protein-related signalling pathway led to the discovery of genes putatively encoding 64 GPCRs, 73 G proteins, and 273 effectors, fewer than those in other diatoms (Supplementary Fig. S3; Supplementary Tables S5–S7).

GPCRs are thought to potentially affect the results of ecotoxicology tests in aquatic invertebrates because many commercial pharmaceuticals target GPCRs.[61] Therefore, when performing ecotoxicological tests using *M. pseudoterrestris*, which carries only a few genes that encode GPCRs, the toxicity of pharmaceuticals to natural diatom communities may be underestimated.

3.2. Effect of Cu exposure on gene expression

We measured the EC50 value of Cu^{2+} for *M. pseudoterrestiris* after the addition of $CuSO_4$. The EC50 value for *M. pseudoterrestris* was 2.68 µM (Supplementary Fig. S4), which is similar to that of a closely related diatom, *Mayamaea terrestris* UTEX B673 (formerly known as *N. Pelliculosa*), when incubated in 1.75 µM $CuSO_4$ for 72 h.[53]

To elucidate the rapid cell responses of *M. pseudoterrestris* to Cu^{2+}, we performed RNA-seq analysis. The cells were exposed to 1.75 µM Cu^{2+} for 1 h. Overall, 426 protein-coding genes were significantly differentially expressed, of which 297 were up-regulated and 129 were down-regulated

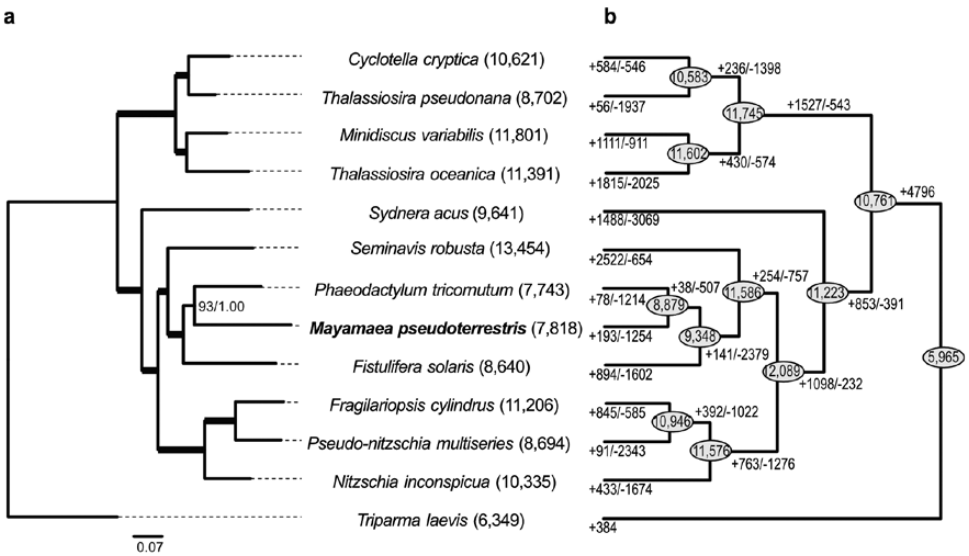

Figure 2. Phylogenetic relationships and evolution of gene families in diatoms. (a) Phylogenetic analysis based on 284 highly conserved orthologous proteins (128,950 amino acids). Non-parametric bootstrap support values (BP) and Bayesian posterior probabilities (BPPs) are shown at each node. BP and BPP values of 100 and 1.00, respectively, are indicated using bold lines. (b) Prediction of ancestral gene families in diatoms. The number of gene families in each species is indicated in parentheses. Putative gene family sets conserved in the common ancestor are outlined with circles at each node. Gains and losses of gene families are shown at each branch.

(Supplementary Fig. S5). Next, we performed GO enrichment analysis of the DEGs to determine their general cell responses (Supplementary Table S8). A total of 20 GO terms were enriched among the up-regulated genes. Overall, genes related to intracellular signal transduction were up-regulated.

3.3. Cu exposure induces cell motility

Exposure to Cu significantly increased the expression levels of genes encoding cytoskeleton-related proteins such as myosin and actin, which are related to gliding motility.[19–21] The genome of *M. pseudoterrestris* harbours 1 *actin* gene (MP_007637-T1) and 10 *myosin* genes. Among the 10 *myosin* genes, 3 were significantly up-regulated upon Cu treatment (Fig. 3a), suggesting that these myosin genes play a role in the gliding mechanism of *M. pseudoterrestris* (Fig. 3b). Phylogenetic analysis of the myosin proteins of diatoms (Fig. 4) showed that all myosin proteins of *M. pseudoterrestris* clustered with ten known clades of *P. tricornutum* myosin (Myo A–I and myosin 29).[62] The up-regulated *myosin* genes encode MyoA (MP_005044-T1), MyoC (MP_005398-T1), and MyoE (MP_000497-T1), and these genes together form a monophyletic group. Interestingly, these clades were composed only of pennate diatoms, probably those exhibiting gliding movements. These data suggest that gliding-related *myosin* genes originated from a single gene in a common ancestor of pennate diatoms, and then underwent functional diversification.

Similar to the genes encoding actin and myosin, genes related to the regulation and maintenance of the cytoskeleton were also up-regulated, including those encoding calmodulin (CaM; MP_004473-T1 and MP_005929-T1), UNC-45 (MP_010642-T1), tropomodulin (MP_004515-T1), coronin (MP_007065-T1), and F-actin-capping protein (MP_006910-T1) (Fig. 3a and b). UNC-45 is involved in the folding of myosin[63] and CaM induces myosin movement. Tropomodulin and F-actin-capping protein stabilize the ends of actin filaments by capping their ends.[64,65] Coronin mediates the interaction between actin filaments and microtubules.[66]

In addition to the actin–myosin mechanism, transmembrane proteins and the extracellular matrix (ECM) are believed to facilitate the gliding of pennate diatoms.[21] Interestingly, we found that the gene encoding a spondin-like protein (MP_010827-T1) was up-regulated in *M. pseudoterrestris* (Fig. 3a). In animals, the spondin family is an ECM protein related to the adhesion and outgrowth of neurons.[67] The spondin-like protein of *M. pseudoterrestris* carries an N-terminal signal peptide and a

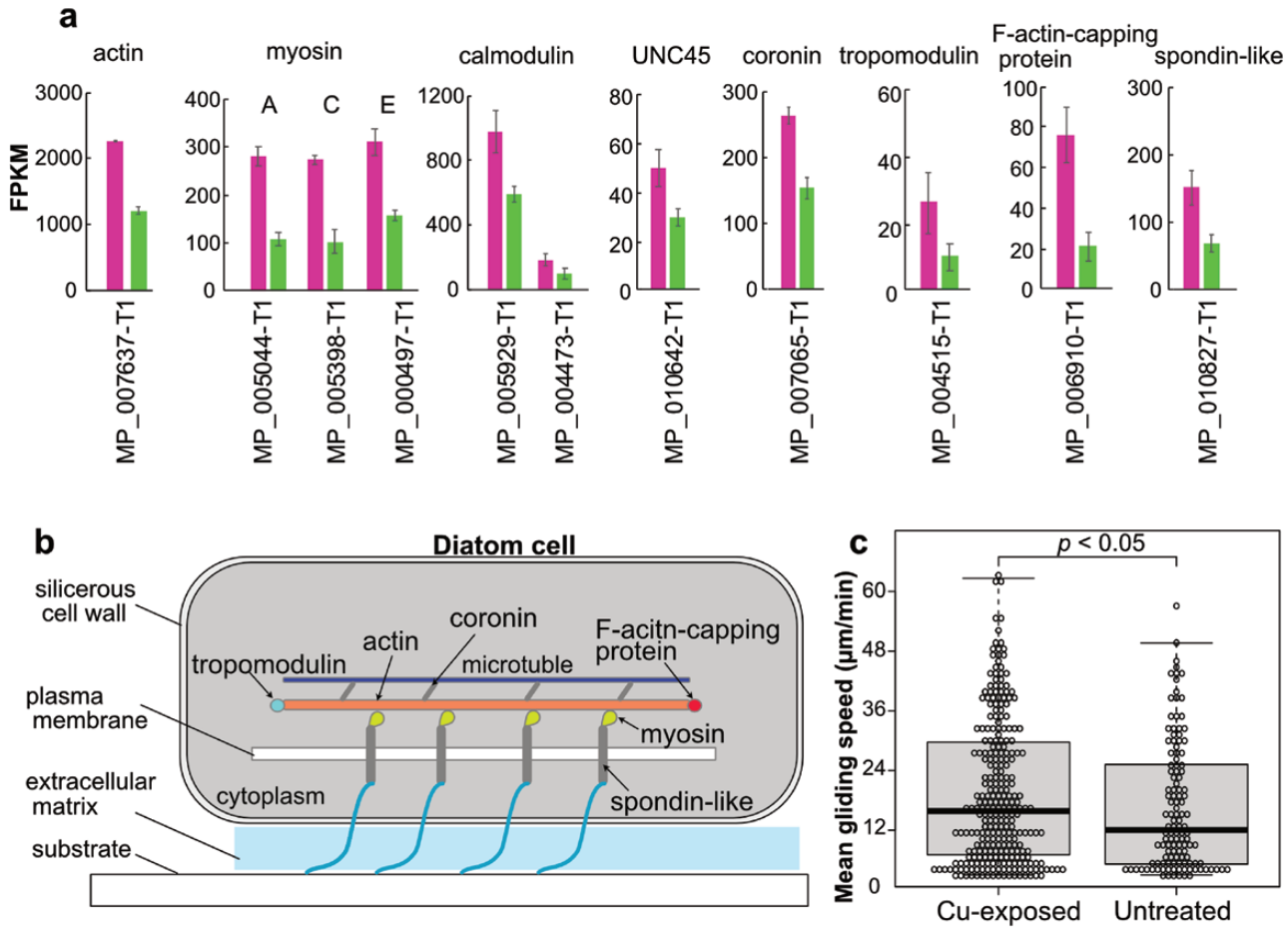

Figure 3. Gliding of diatoms in response to copper (Cu). (a) Relative expression levels of DEGs potentially related to cell motility. Gene expression was estimated as mean fragments per kilobase of exon per million reads mapped (FPKM). Left and right bars represent the gene expression levels in the Cu and control treatments, respectively. Data represent the mean ± S.D. of FPKM values. (b) Revised model showing the molecular mechanism of gliding in pennate diatoms. This model was based on DEGs and was originally conceived by Molino and Wetherbee.[21] (c) Mean gliding speed of Cu-exposed or untreated *M. pseudoterrestris*. The time-lapse images were taken every 10 s for 10 min. The difference in the speed was tested using the Wilcoxon rate sum test.

C-terminal transmembrane region, suggesting that this protein is membrane localized (Supplementary Fig. S6a). Homology-based search and phylogenetic analysis revealed that the spondin-like protein-encoding gene was present in all pennate diatoms but absent in centric diatoms (Supplementary Fig. S6b). These data suggest that spondin-like protein is a good transmembrane protein candidate that links myosin with the ECM (Fig. 3b).

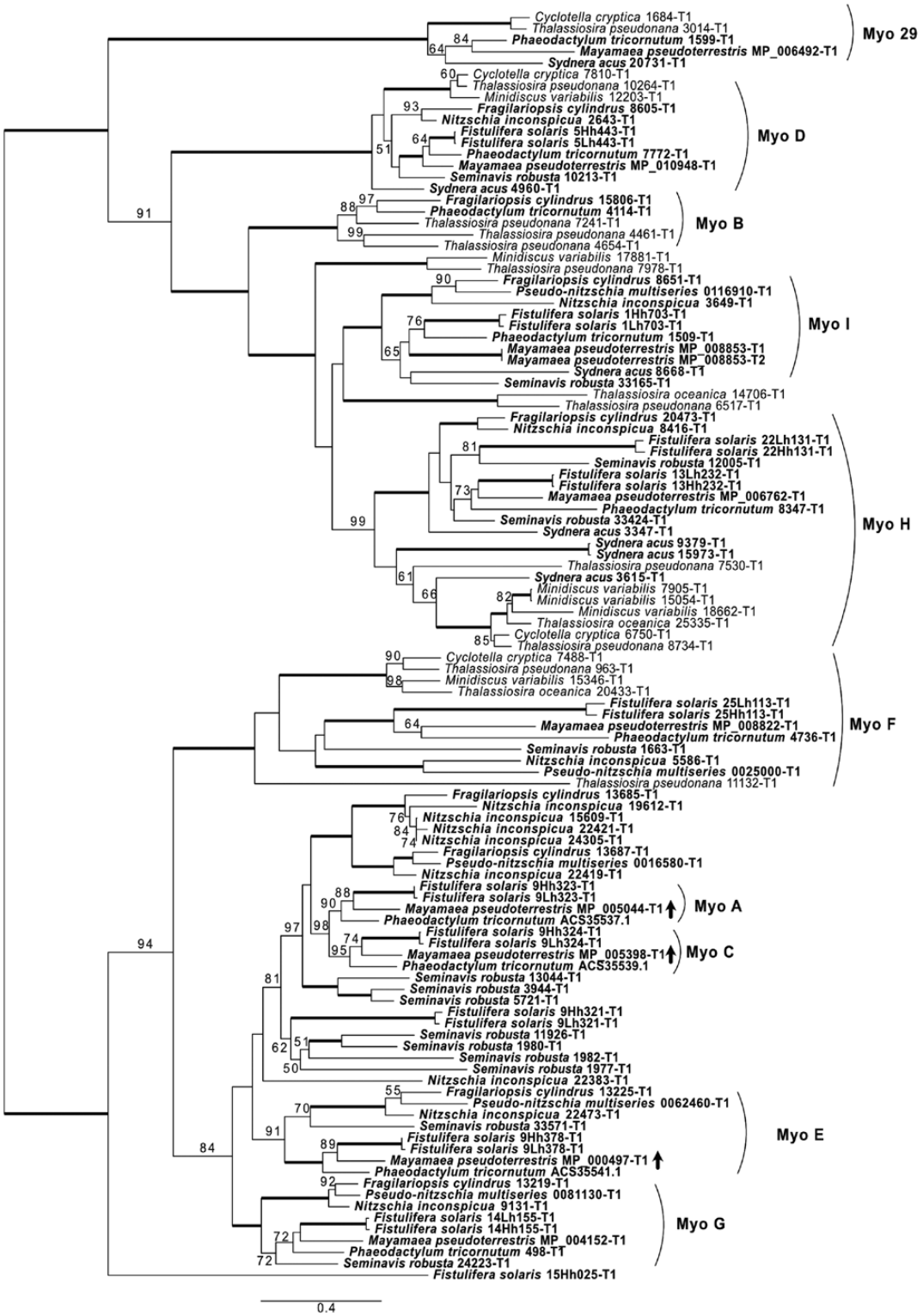

Figure 4. Phylogenetic analysis of myosin proteins in diatoms. Phylogenetic analysis was performed using 111 myosin homologues belonging to 12 diatom species with sequenced genomes. The dataset was composed of 628 amino acids. Maximum likelihood (ML) analysis was performed using the LG + I + G4 model. BP values are indicated on the branches, and bold lines represent BP = 100. BP <50 is not shown. Clades were classified according to Heintzelman and Enriquez.[62] Significantly up-regulated genes of *M. pseudoterrestris* under Cu exposure are indicated with up arrows, besides the operational taxonomic units (OTUs). Pennate diatoms are indicated with bold OTUs.

To elucidate the effect of Cu exposure on *M. pseudoterrestris* motility, we performed time-lapse observations. *Mayamaea pseudoterrestris* cells showed induction of gliding after 1 h of Cu exposure (Fig. 3c; Supplementary Videos 1 and 2). Gliding motility was observed 51.3% (Cu exposed) and 35.8% (untreated) of cells. The median motility speed of the gliding cells was 15.9 and 12.1 μm/min with and without Cu exposure, respectively, and the cells showed significantly fast gliding motility under Cu exposure (P <0.05, Wilcoxon rank sum test). These results were consistent with an increase in the expression levels of putative genes related to gliding motility.

3.4. Induction of EPS biosynthesis

Generally, diatoms produce EPSs that respond to heavy metal exposure, including Cu, to chelate metal.[68,69] In *M. pseudoterrestris*, several sugar metabolism-related genes were up-regulated under Cu exposure, including genes encoding phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, fructose-bisphosphate aldolase, transaldolase, glucokinase, and UDP-sugar pyrophosphorylase (Supplementary Fig. S7). However, the gene-encoding pyruvate kinase was down-regulated, implying that UDP-glucose is synthesized from D-glucose and glycerate-3-phosphate, which is generated by carbon fixation via photosynthesis, and not from pyruvate. Additionally, genes encoding 3-oxoacid CoA-transferase, hydroxymethylglutaryl-CoA synthase, and fumarylacetoacetate hydrolase were also up-regulated. These three genes are involved in the biosynthesis of acetoacetate, which serves as an alternate energy source under starvation conditions, suggesting that Cu exposure decreases glycolysis. The reduction in sucrose content of *Tabellaria flocculosa* under Zn exposure[15] could also be explained by a similar metabolism. Extracellular mucilage of the pennate diatom *Craspedostauros australis* is composed of glycoproteins, and their glycans are rich in xylose.[70] Although the biosynthetic pathway of xylose in diatoms remains unknown, land plants synthesize xylose from UDP-sugars.[71,72] Thus, the induction of gluconeogenesis-related genes in diatoms may imply an increase in EPS production of to chelate heavy metals.

3.5. Resistance to oxidative stress

In *M. pseudoterrestris*, the exposure to Cu significantly induced the expression of genes encoding glutathione peroxidase (GPx) and GST (Supplementary Table S9), suggesting that *M. pseudoterrestris* cells are affected by oxidative stress. ROS and Cu^{2+} can be reduced by GPx and GST, respectively, using GSH. However, Cu exposure did not significantly induce the gene expression of SOD and CAT, suggesting that these proteins are regulated at the post-transcriptional or post-translational level. Additionally, we were unable to detect the induction of genes involved in the glutathione–ascorbate cycle.

The gene-encoding DBH (dopamine beta-hydroxylase)-like mono-oxygenase was up-regulated in *M. pseudoterrestris*. DBH-like mono-oxygenase is a Cu/ascorbate-dependent mono-oxygenase family members that hydrolyses substrates with ascorbate and molecular oxygen in the endoplasmic reticulum.[73] In *M. pseudoterrestris*, we predicted signal peptides at the N-terminus of DBH-like mono-oxygenase and considered that this protein may be localized to the endoplasmic reticulum. Because DBH-like mono-oxygenase binds to Cu^{2+} in its active site, a high concentration of Cu^{2+} may induce the activity of this protein, which suggests that this protein could be used to remove ROS.

Cu exposure induces the gene-encoding spermidine synthase, which synthesizes spermidine from putrescine.[74] In the green alga *Chlorella vulgaris*, application of spermidine alleviates heavy metal-induced oxidative stress and growth inhibition, causing degradation of photosynthetic pigments, monosaccharides, and proteins.[75] This suggests a potential role for spermidine in oxidative stress resistance in *M. pseudoterrestris*.

3.6. Cu exposure decreases the intracellular Ca^{2+} concentration

The gene-encoding plasma membrane Ca^{2+} ATPase (PMCA), which is involved in Ca^{2+} efflux, was up-regulated (Fig. 5a and b). PMCA is located in the plasma membrane and exports intracellular Ca^{2+} to the extracellular space via ATP.[76] In addition to PMCA, the plasma membrane Na^{+}/Ca^{2+} exchanger (NCX) is also responsible for Ca^{2+} export; however, no significant difference in the expression of *NCX* gene was observed under Cu exposure. These data suggest that Cu exposure can induce Ca^{2+} export via PMCA, which may be useful for cellular redox homeostasis. Ca^{2+} and CaM enhance the activity of PMCA.[76] Therefore, induction of *CaM* genes can affect PMCA activity and facilitate motility (as described above) under Cu exposure. PMCA is regulated by cGMP-dependent protein kinase (PKG), and genes encoding PKG and guanyl cyclase are up-regulated in *M. pseudoterrestris* exposed to Cu.

To elucidate the effect of Cu exposure on the intracellular Ca^{2+} concentration in *M. pseudoterrestris*, we used a fluorescent Ca^{2+} indicator, Fluo-3, and performed flow cytometry (Fig. 5b). The median intensity of green fluorescence was 120.6 a.u. in untreated *M. pseudoterrestris* cells, and 82.3 a.u. in cells exposed to Cu. The number of cells with less than peak fluorescence was higher in the Cu treatment group than in the untreated control group. Without staining, a clear difference in the fluorescence peak was not observed between the untreated and Cu-treated cells. The clear decline in fluorescence intensity under Cu exposure indicated that the intracellular Ca^{2+} concentration decreased in response to Cu exposure. The intracellular concentration of Ca^{2+} is well maintained and critical for the regulation of cell responses in both the short and long terms.[77] Therefore, the rapid change in Ca^{2+} concentration observed in this study implies that the Ca^{2+}-mediated signal transduction system is disrupted in diatoms, which can be a major factors in metal toxicity. For example, the Ca^{2+}-mediated signal transduction system is related to the shear response, osmotic stress response, and response to Fe supply and reactive aldehydes in diatoms.[78,79]

3.7. Diatom-specific mechanisms of reactive dicarbonyl compound detoxification

Genes encoding two glyoxalases (Glys), glyoxalase III (GlyIII) and glyoxalase-like protein (Gly-like) were up-regulated under Cu exposure (Fig. 6a and b). Glyoxalases detoxify methylglyoxal, a reactive dicarbonyl compound generated from triosephosphate sugars.[80] The Gly system is composed of three enzymes, GlyI, GlyII, and GlyIII. Among these, GlyI and GlyII detoxify methylglyoxal to D-lactate via a two-step reaction, whereas GlyIII generates D-lactate through a one-step reaction. The GlyIII of *M. pseudoterrestris* contains a DJ-1_Pfpl domain, which is conserved in GlyIII,[81] and the Gly-like contains a glyoxalase_like domain. The induction of

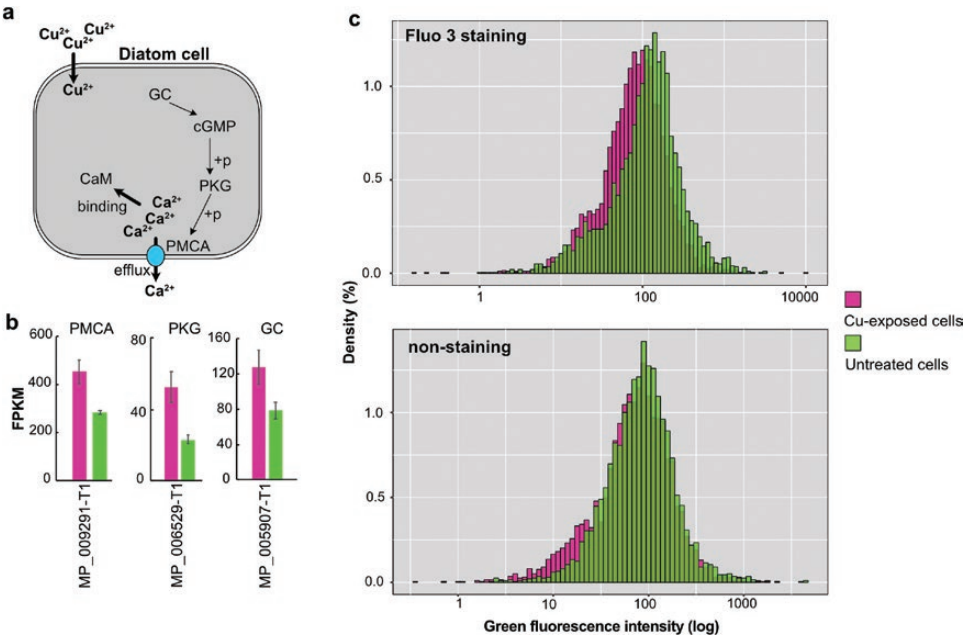

Figure 5. Calcium (Ca^{2+}) efflux in *Mayamaea pseudoterrestris* under Cu exposure. (a) A model for Ca^{2+} efflux under Cu exposure. CaM: calmodulin; GC: guanyl cyclase; PKG: cGMP-dependent protein kinase. (b) Relative expression levels of DEGs potentially related to Ca^{2+} efflux. Left and right bars indicate the gene expression levels (expressed as FPKM values) in the Cu and control treatments, respectively. Data represent the mean ± S.D. of FPKM values. (c) Frequency of cells exhibiting green fluorescence in *M. pseudoterrestris* cells. Cells stained with Fluo 3 (upper) and without staining (lower) are shown.

genes encoding Gly proteins suggests that Cu exposure leads to methylglyoxal production via oxidation, and that diatoms can detoxify methylglyoxal using different Gly enzymes.

Next, we performed phylogenetic analysis of Gly-related proteins (Fig. 6c and d). The results showed that the Gly-like protein originated from bacteria and was conserved only in diatoms among eukaryotes. Notably, centric diatoms are the only diatom species lacking this protein. In contrast, GlyIII proteins were conserved among pennate and centric diatoms. These data suggested that pennate diatoms acquired an additional *Gly-like* gene from bacteria via HGT. This functional duplication may be explained by the different lifestyles of the diatoms. Pennate diatoms adhere to substances, whereas centric diatoms can change their buoyancy and habitat vertically.[82] Therefore, pennate diatoms likely acquired an additional mechanism for detoxification of harmful oxidative compounds, in contrast to centric diatoms, which can avoid oxidative stress by vertical migration. These findings suggest that ecotoxicological tests performed using pennate diatoms can underestimate the level of toxicity caused by metals and some medicines, such as bleomycin, compared to those performed using centric diatoms. Bleomycin is an Fe-containing protein, and Gly proteins belong to the same structural family as the bleomycin resistance protein.[83]

4. Conclusion

Mayamaea pseudoterrestris NIES-4280 is a freshwater pennate diatom that has been proposed as an alternative strain of UTEX 664 for performing ecotoxicological tests. In this study, we sequenced the genome of *M. pseudoterrestris* NIES-4280, and performed RNA-seq analysis, which revealed a rapid change in gene expression in response to Cu exposure. The genome size and gene content of *M. pseudoterrestris* were compact, and *M. pseudoterrestris* shared many genes with *P. tricornutum*, a model diatom species. This suggests that *M. pseudoterrestris* may serve as a suitable model for freshwater diatoms. Gene expression analyses revealed drastic changes in the expression of genes related to cell motility, polysaccharide metabolism, oxidative stress removal, intracellular Ca^{2+} signalling, and reactive compound detoxification. Indeed, exposure to Cu causes active gliding motility and decreases the intracellular Ca^{2+} concentration in *M. pseudoterrestris*. Based on the analysis of DEGs identified under active motility, we propose a putative molecular mechanism of gliding in diatoms, which needs to be verified via genome editing of *P. tricornutum* in the future.

Our analyses also revealed the diversity of genes and gene expression patterns related to the rapid responses to Cu among diatom species. The Cu toxicity response mechanism can be greatly differentiated between marine and freshwater species, and between pennate and centric diatoms. Considering the ecological importance of diatoms in aquatic environments, ecotoxicological tests should be performed using centric diatom species as well as pennate diatom species, such as *M. pseudoterrestris*.

Funding

This work was partially supported by the National BioResource Project for Algae (Grant No. 17km0210116j0001) funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

Data availability

Mayamaea pseudoterrestris genome assembly, DNA-seq reads, and RNA-seq reads were deposited to the DNA Data Bank of Japan (DDBJ) under accession numbers BROI01000001–BROI01000041, DRR374606–DRR364608, and DRR374599–DRR374605, respectively.

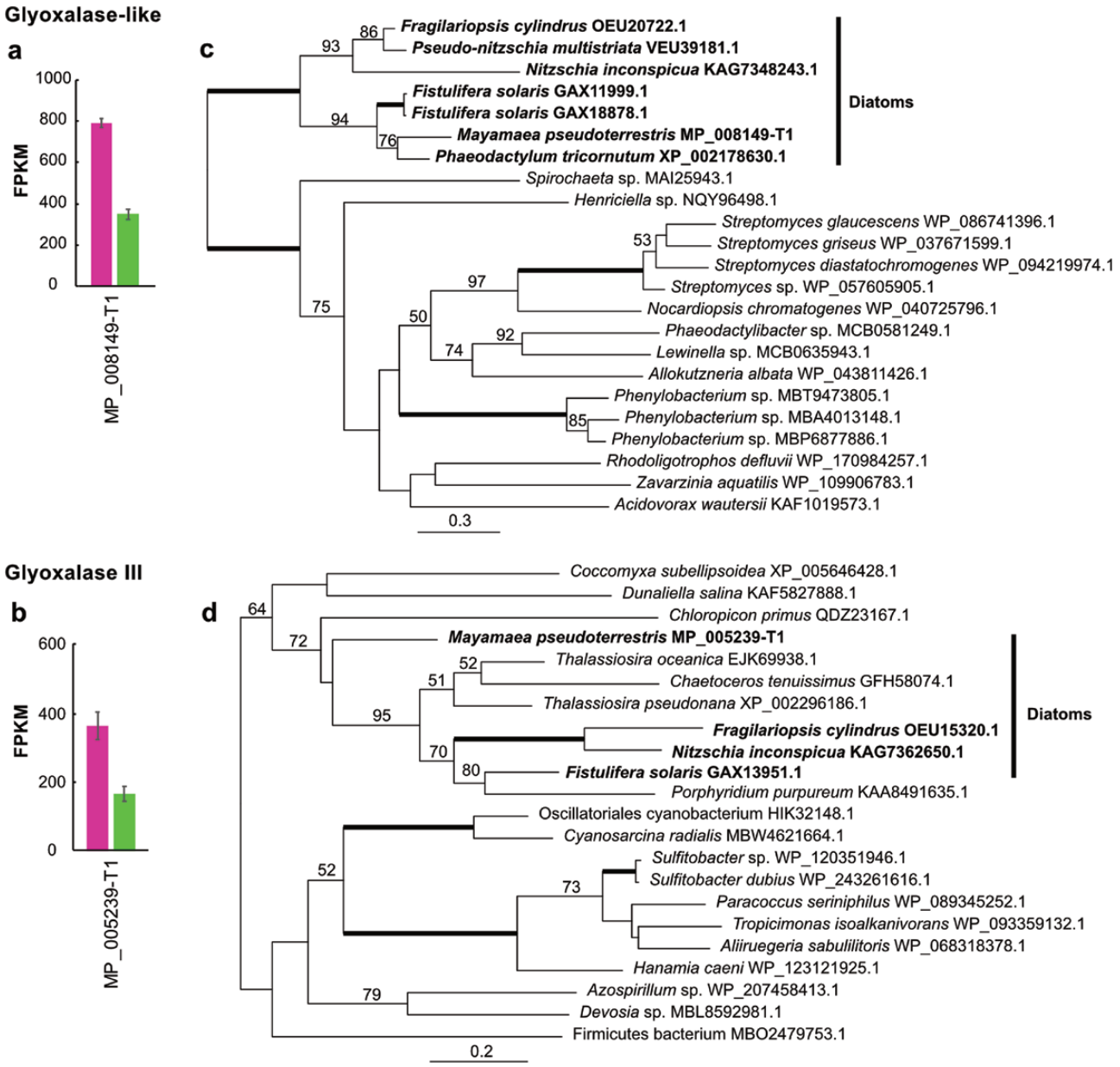

Figure 6. Diatom-specific glyoxalases (Glys). Relative expression levels of DEGs encoding Gly-like protein (a) and GlyIII (b). Left and right bars indicate the gene expression levels (FPKM) in the Cu and control treatments, respectively. Data represent the mean ± S.D. of FPKM values. Phylogenetic analysis of Gly-like (c) and GlyIII (d) proteins. In (c), the dataset was composed of 185 amino acids, and ML analysis was performed using the LG + F + I + G4 model. In (d), the dataset was composed of 187 amino acids, and ML analysis was performed using the LG + I + G4 model. BP values are indicated on the branches, and bold lines represent BP = 100. BP <50 is not shown. Bold OTUs represent the pennate diatoms.

Conflict of interest

None Declared.

Supplementary data

Supplementary data are available at *DNARES* online.

References

1. 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, *Glob. Biogeochem. Cycles*, **9**, 359–72.
2. Armbrust, E.V. 2009, The life of diatoms in the world's oceans, *Nature*, **459**, 185–92.
3. Mann, D.G. 1999, The species concept in diatoms, *Phycologia*, **38**, 437–95.
4. Sheath, R.G. and Wehr, J.D. 2015, Introduction to the freshwater algae. In: Wehr, J.D., Sheath, R.G. and Kociolek, J.P., eds. *Freshwater algae of North America*, pp. 1–11. Academic Press: Cambridge.
5. Mann, D.G. and Droop, S.J.M. 1996, Biodiversity, biogeography and conservation of diatoms. In: Kristiansen, J., ed. *Biogeography of freshwater algae*, pp. 19–32. Springer Netherlands: Dordrecht.
6. Mann, D.G. and Vanormelingen, P. 2013, An inordinate fondness? The number, distributions, and origins of diatom species, *J. Eukaryot. Microbiol.*, **60**, 414–20.
7. Venkatachalapathy, R. and Karthikeyan, P. 2015, Application of diatom-based indices for monitoring environmental quality of

riverine ecosystems: a review. In: Ramkumar, M., Kumaraswamy, K. and Mohanraj, R., eds. *Environmental management of river basin ecosystems*, pp. 593–619. Springer: Cham.
8. Kelly, M.G., Cazaubon, A., Coring, E., et al. 1998, Recommendations for the routine sampling of diatoms for water quality assessments in Europe, *J. Appl. Phycol.*, **10**, 215–24.
9. Lavoie, I., Campeau, S., Darchambeau, F., Cabana, G. and Dillon, P.J. 2008, Are diatoms good integrators of temporal variability in stream water quality? *Freshw. Biol.*, **53**, 827–41.
10. Teittinen, A., Taka, M., Ruth, O. and Soininen, J. 2015, Variation in stream diatom communities in relation to water quality and catchment variables in a boreal, urbanized region, *Sci. Total Environ.*, **530–1**, 279–89.
11. Stomp, M., Huisman, J., Mittelbach, G.G., Litchman, E. and Klausmeier, C.A. 2011, Large-scale biodiversity patterns in freshwater phytoplankton, *Ecology*, **92**, 2096–107.
12. Fernández, M.R., Martín, G., Corzo, J., et al. 2018, Design and testing of a new diatom-based index for heavy metal pollution, *Arch. Environ. Contam. Toxicol.*, **74**, 170–92.
13. Riato, L., Leira, M., Della Bella, V. and Oberholster, P.J. 2018, Development of a diatom-based multimetric index for acid mine drainage impacted depressional wetlands, *Sci. Total Environ.*, **612**, 214–22.
14. Branco, D., Lima, A., Almeida, S.F.P. and Figueira, E. 2010, Sensitivity of biochemical markers to evaluate cadmium stress in the freshwater diatom *Nitzschia palea* (Kützing) W. Smith, *Aquat. Toxicol.*, **99**, 109–17.
15. Gonçalves, S., Kahlert, M., Almeida, S.F.P. and Figueira, E. 2018, A freshwater diatom challenged by Zn: biochemical, physiological and metabolomic responses of *Tabellaria flocculosa* (Roth) Kützing, *Environ. Pollut.*, **238**, 959–71.
16. Santos, J., Almeida, S.F.P. and Figueira, E. 2013, Cadmium chelation by frustulins: a novel metal tolerance mechanism in *Nitzschia palea* (Kützing) W. Smith, *Ecotoxicology*, **22**, 166–73.
17. Bonet, B., Corcoll, N., Acuňa, V., et al. 2013, Seasonal changes in antioxidant enzyme activities of freshwater biofilms in a metal polluted Mediterranean stream, *Sci. Total Environ.*, **444**, 60–72.
18. Bonet, B., Corcoll, N. and Guasch, H. 2012, Antioxidant enzyme activities as biomarkers of Zn pollution in fluvial biofilms, *Ecotoxicol. Environ. Saf.*, **80**, 172–8.
19. Serôdio, J. 2021, Diatom motility: mechanisms, control and adaptive value In: Cohn, S., Manoylov, K. and Gordon, R., eds. *Diatom gliding motility*, pp. 159–83. Wiley: Hoboken.
20. Edgar, L.A. and Pickett-Heaps, J.D. 1984, Diatom locomotion. In: Round, F.E. and Chapman, D.J., eds. *Progress in phycological research*, pp. 47–88. Biopress Ltd.: Bristol.
21. Molino, P.J. and Wetherbee, R. 2008, The biology of biofouling diatoms and their role in the development of microbial slimes, *Biofouling*, **24**, 365–79.
22. Pandey, L.K. and Bergey, E.A. 2016, Exploring the status of motility, lipid bodies, deformities and size reduction in periphytic diatom community from chronically metal (Cu, Zn) polluted waterbodies as a biomonitoring tool, *Sci. Total Environ.*, **550**, 372–81.
23. OECD. 2011, Test No. 201: freshwater alga and cyanobacteria, growth inhibition test. In: OECD, ed. *OECD guidelines for the testing of chemicals*. OECD Publishing: Paris.
24. Tuji, A., Suzuki, S. and Yamagishi, T. 2021, Alternative diatom strains for growth inhibition tests (OECD test guideline), *Bull. Natl. Museum Nat. Sci. Ser. B*, **47**, 183–92.
25. Connolly, J.A., Oliver, M.J., Beaulieu, J.M., et al. 2008, Correlated evolution of genome size and cell volume in diatoms (Bacillariophyceae), *J. Phycol.*, **44**, 124–31.
26. Cassé, F. and Swain, G.W. 2006, The development of microfouling on four commercial antifouling coatings under static and dynamic immersion, *Int. Biodeterior. Biodegrad.*, **57**, 179–85.
27. Rabiet, M., Coquery, M., Carluer, N., Gahou, J. and Gouy, V. 2015, Transfer of metal(loid)s in a small vineyard catchment: contribution of dissolved and particulate fractions in river for contrasted hydrological conditions, *Environ. Sci. Pollut. Res.*, **22**, 19224–39.
28. Kasai, F., Kawachi, M., Erata, M., Yumoto, K. and Sato, M. 2009, NIES-collection list of strains, 8th edition, *Jpn. J. Phycol.*, **57**, 1–350.
29. Vurture, G.W., Sedlazeck, F.J., Nattestad, M., et al. 2017, GenomeScope: fast reference-free genome profiling from short reads, *Bioinformatics*, **33**, 2202–4.
30. Vaser, R. and Šikić, M. 2021, Time- and memory-efficient genome assembly with Raven, *Nat. Comput. Sci.*, **1**, 332–6.
31. Li, H. 2018, Minimap2: pairwise alignment for nucleotide sequences, *Bioinformatics*, **34**, 3094–100.
32. Hu, J., Fan, J., Sun, Z. and Liu, S. 2020, NextPolish: a fast and efficient genome polishing tool for long-read assembly, *Bioinformatics*, **36**, 2253–5.
33. Chen, S., Zhou, Y., Chen, Y. and Gu, J. 2018, fastp: an ultra-fast all-in-one FASTQ preprocessor, *Bioinformatics*, **34**, i884–90.
34. Tang, L., Li, M., Wu, F.-X., Pan, Y. and Wang, J. 2020, MAC: merging assemblies by using adjacency algebraic model and classification, *Front. Genet.*, **10**, 1–10.
35. Zimin, A.V. and Salzberg, S.L. 2020, The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies, *PLoS Comput. Biol.*, **16**, 1–8.
36. Zimin, A.V., Puiu, D., Luo, M.-C., et al. 2017, Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm, *Genome Res.*, **27**, 787–92.
37. Roach, M.J., Schmidt, S.A. and Borneman, A.R. 2018, Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies, *BMC Bioinf.*, **19**, 460.
38. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. 2015, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics*, **31**, 3210–2.
39. Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. 2019, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, *Nat. Biotechnol.*, **37**, 907–15.
40. Flynn, J.M., Hubley, R., Goubert, C., et al. 2020, RepeatModeler2 for automated genomic discovery of transposable element families, *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 9451–7.
41. Gschloessl, B., Guermeur, Y. and Cock, J.M. 2008, HECTAR: a method to predict subcellular targeting in heterokonts, *BMC Bioinf.*, **9**, 393.
42. Kuwata, A., Saitoh, K., Nakamura, Y., Ichinomiya, M. and Sato, N. 2020, Draft whole-genome sequence of *Triparma laevis* f. *inornata* (Parmales, Bolidophyceae), isolated from the Oyashio Region, Western North Pacific Ocean, *Microbiol. Resour. Announc.*, **9**, 9–10.
43. Nakamura, T., Yamada, K.D., Tomii, K. and Katoh, K. 2018, Parallelization of MAFFT for large-scale multiple sequence alignments, *Bioinformatics*, **34**, 2490–2.
44. Capella-Gutierrez, S., Silla-Martinez, J.M. and Gabaldon, T. 2009, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, *Bioinformatics*, **25**, 1972–3.
45. Darriba, D., Posada, D., Kozlov, A.M., et al. 2020, ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary, *Mol. Biol. Evol.*, **37**, 291–4.
46. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A. 2019, RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference, *Bioinformatics*, **35**, 4453–5.
47. Ronquist, F., Teslenko, M., Van Der Mark, P., et al. 2012, MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space, *Syst. Biol.*, **61**, 539–42.
48. Makita, Y., Suzuki, S., Fushimi, K., et al. 2021, Identification of a dual orange/far-red and blue light photoreceptor from an oceanic green picoplankton, *Nat. Commun.*, **12**, 3593.
49. Begum, K., Mohl, J.E., Ayivor, F., Perez, E.E. and Leung, M.-Y. 2020, GPCR-PEnDB: a database of protein sequences and derived features to facilitate prediction and classification of G protein-coupled receptors, *Database*, **2020**, 1–12.

50. Bhasin, M. and Raghava, G.P.S. 2004, GPCRpred: an SVM-based method for prediction of families and subfamilies of G-protein coupled receptors, *Nucleic Acids Res.*, **32**, W383–9.
51. Rios, S., Fernandez, M.F., Caltabiano, G., et al. 2015, GPCRtm: an amino acid substitution matrix for the transmembrane region of class A G protein-coupled receptors, *BMC Bioinf.*, **16**, 206.
52. Satagopam, V.P., Theodoropoulou, M.C., Stampolakis, C.K., et al. 2010, GPCRs, G-proteins, effectors and their interactions: human-gpDB, a database employing visualization tools and data integration techniques, *Database*, **2010**, baq019.
53. Nagai, T. and De Schamphelaere, K.A.C. 2016, The effect of binary mixtures of zinc, copper, cadmium, and nickel on the growth of the freshwater diatom *Navicula pelliculosa* and comparison with mixture toxicity model predictions, *Environ. Toxicol. Chem.*, **35**, 2765–73.
54. Liao, Y., Smyth, G.K. and Shi, W. 2014, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics*, **30**, 923–30.
55. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. 2010, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, **26**, 139–40.
56. Klopfenstein, D.V., Zhang, L., Pedersen, B.S., et al. 2018, GOATOOLS: a Python library for Gene Ontology analyses, *Sci. Rep.*, **8**, 1–17.
57. Schindelin, J., Arganda-Carreras, I., Frise, E., et al. 2012, Fiji: an open-source platform for biological-image analysis, *Nat. Methods*, **9**, 676–82.
58. Tinevez, J.-Y., Perry, N., Schindelin, J., et al. 2017, TrackMate: an open and extensible platform for single-particle tracking, *Methods*, **115**, 80–90.
59. Csurös, M. 2010, Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood, *Bioinformatics*, **26**, 1910–2.
60. Tanaka, T., Maeda, Y., Veluchamy, A., et al. 2015, Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome, *Plant Cell*, **27**, 162–76.
61. Kim, D.-H., Park, J.C. and Lee, J.-S. 2022, G protein-coupled receptors (GPCRs) in rotifers and cladocerans: potential applications in ecotoxicology, ecophysiology, comparative endocrinology, and pharmacology, *Comp. Biochem. Physiol. C: Toxicol. Pharmacol.*, **256**, 109297.
62. Heintzelman, M.B. and Enriquez, M.E. 2010, Myosin diversity in the diatom *Phaeodactylum tricornutum*., *Cytoskeleton*, **67**, 142–51.
63. Hellerschmied, D., Lehner, A., Franicevic, N., et al. 2019, Molecular features of the UNC-45 chaperone critical for binding and folding muscle myosin, *Nat. Commun.*, **10**, 4781.
64. Cooper, J.A. and Sept, D. 2008, New insights into mechanism and regulation of actin capping protein, *Int. Rev. Cell Mol. Biol.*, **267**, 183–206.
65. Rao, J.N., Madasu, Y. and Dominguez, R. 2014, Mechanism of actin filament pointed-end capping by tropomodulin, *Science*, **345**, 463–7.
66. Chan, K.T., Creed, S.J. and Bear, J.E. 2011, Unraveling the enigma: progress towards understanding the coronin family of actin regulators, *Trends Cell Biol.*, **21**, 481–8.
67. Feinstein, Y., Borrell, V., Garcia, C., et al. 1999, F-spondin and mindin: two structurally and functionally related genes expressed in the hippocampus that promote outgrowth of embryonic hippocampal neurons, *Development*, **126**, 3637–48.
68. Pistocchi, R., Guerrini, F., Boni, L. and Balboni, V. 1997, Copper toxicity and carbohydrate production in the microalgae *Cylindrotheca fusiformis* and *Gymnodinium* sp., *Eur. J. Phycol.*, **32**, 125–32.
69. Pistocchi, R., Mormile, M.A., Guerrini, F., Isani, G. and Boni, L. 2000, Increased production of extra- and intracellular metal-ligands in phytoplankton exposed to copper and cadmium, *J. Appl. Phycol.*, **12**, 469–77.
70. Chiovitti, A., Bacic, A., Burke, J. and Wetherbee, R. 2003, Heterogeneous xylose-rich glycans are associated with extracellular glycoproteins from the biofouling diatom *Craspedostauros australis* (Bacillariophyceae), *Eur. J. Phycol.*, **38**, 351–60.
71. Minic, Z., Rihouey, C., Do, C.T., Lerouge, P. and Jouanin, L. 2004, Purification and characterization of enzymes exhibiting β-D-xylosidase activities in stem tissues of *Arabidopsis*, *Plant Physiol.*, **135**, 867–78.
72. Bernal, A.J., Jensen, J.K., Harholt, J., et al. 2007, Disruption of ATCSLD5 results in reduced growth, reduced xylan and homogalacturonan synthase activity and altered xylan occurrence in *Arabidopsis*, *Plant J.*, **52**, 791–802.
73. Xin, X., Mains, R.E. and Eipper, B.A. 2004, Monooxygenase X, a member of the copper-dependent monooxygenase family localized to the endoplasmic reticulum, *J. Biol. Chem.*, **279**, 48159–67.
74. Fuell, C., Elliott, K.A., Hanfrey, C.C., Franceschetti, M. and Michael, A.J. 2010, Polyamine biosynthetic diversity in plants and algae, *Plant Physiol. Biochem.*, **48**, 513–20.
75. Piotrowska-Niczyporuk, A., Bajguz, A., Zambrzycka, E. and Godlewska-Żyłkiewicz, B. 2012, Phytohormones as regulators of heavy metal biosorption and toxicity in green alga *Chlorella vulgaris* (Chlorophyceae), *Plant Physiol. Biochem.*, **52**, 52–65.
76. Brini, M. and Carafoli, E. 2011, The plasma membrane Ca^{2+} ATPase and the plasma membrane sodium calcium exchanger cooperate in the regulation of cell calcium, *Cold Spring Harb. Perspect. Biol.*, **3**, a004168.
77. Berridge, M.J., Bootman, M.D. and Roderick, H.L. 2003, Calcium signalling: dynamics, homeostasis and remodelling, *Nat. Rev. Mol. Cell Biol.*, **4**, 517–29.
78. Falciatore, A., D'Alcalà, M.R., Croot, P. and Bowler, C. 2000, Perception of environmental signals by a marine diatom, *Science*, **288**, 2363–6.
79. Vardi, A., Formiggini, F., Casotti, R., et al. 2006, A stress surveillance system based on calcium and nitric oxide in marine diatoms, *PLoS Biol.*, **4**, e60.
80. Singla-Pareek, S.L., Kaur, C., Kumar, B., Pareek, A. and Sopory, S.K. 2020, Reassessing plant glyoxalases: large family and expanding functions, *New Phytol.*, **227**, 714–21.
81. Kumar, B., Kaur, C., Pareek, A., Sopory, S.K. and Singla-Pareek, S.L. 2021, Tracing the evolution of plant glyoxalase III enzymes for structural and functional divergence, *Antioxidants*, **10**, 648.
82. Kuwata, A. and Jewson, D.H. 2015, Ecology and evolution of marine diatoms and Parmales. In: Ohtsuka, S., Suzaki, T., Horiguchi, T., Suzuki, N. and Not, F., eds. *Marine protists: diversity and dynamics*, pp. 251–75. Springer: Tokyo, Japan.
83. Cameron, A.D., Olin, B., Ridderström, M., Mannervik, B. and Jones, T.A. 1997, Crystal structure of human glyoxalase I—evidence for gene duplication and 3D domain swapping, *EMBO J.*, **16**, 3386–95.