The ISME Journal (2021) 15:562–576
https://doi.org/10.1038/s41396-020-00797-7

ARTICLE

Mating type specific transcriptomic response to sex inducing pheromone in the pennate diatom *Seminavis robusta*

Gust Bilcke[1,2,3,4] · Koen Van den Berge[4,5] · Sam De Decker[1] · Eli Bonneure[6] · Nicole Poulsen[7] · Petra Bulankova[2,3] · Cristina Maria Osuna-Cruz[2,3,8] · Jack Dickenson[9,10] · Koen Sabbe[1] · Georg Pohnert[11] · Klaas Vandepoele[2,3,8] · Sven Mangelinckx[6] · Lieven Clement[4,8] · Lieven De Veylder[2,3] · Wim Vyverman[1]

Received: 6 July 2020 / Revised: 10 September 2020 / Accepted: 24 September 2020 / Published online: 7 October 2020

Abstract

Sexual reproduction is a fundamental phase in the life cycle of most diatoms. Despite its role as a source of genetic variation, it is rarely reported in natural circumstances and its molecular foundations remain largely unknown. Here, we integrate independent transcriptomic datasets to prioritize genes responding to sex inducing pheromones (SIPs) in the pennate diatom *Seminavis robusta*. We observe marked gene expression changes associated with SIP treatment in both mating types, including an inhibition of S phase progression, chloroplast division, mitosis, and cell wall formation. Meanwhile, meiotic genes are upregulated in response to SIP, including a sexually induced diatom specific cyclin. Our data further suggest an important role for reactive oxygen species, energy metabolism, and cGMP signaling during the early stages of sexual reproduction. In addition, we identify several genes with a mating type specific response to SIP, and link their expression pattern with physiological specialization, such as the production of the attraction pheromone diproline in mating type − (MT−) and mate-searching behavior in mating type + (MT+). Combined, our results provide a model for early sexual reproduction in pennate diatoms and significantly expand the suite of target genes to detect sexual reproduction events in natural diatom populations.

Introduction

Sexual reproduction is a virtually universal feature in the life cycle of eukaryotic organisms and a wealth of reproductive strategies have evolved across different phyla [1]. Likewise, sexual reproduction is an essential phase in the diplontic life cycle of most diatoms, an extraordinarily diverse group of microalgae that play an

These authors contributed equally: Gust Bilcke, Koen Van den Berge

Supplementary information The online version of this article (https://doi.org/10.1038/s41396-020-00797-7) contains supplementary material, which is available to authorized users.

✉ Wim Vyverman
wim.vyverman@ugent.be

1 Protistology and Aquatic Ecology, Department of Biology, Ghent University, 9000 Ghent, Belgium

2 Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium

3 VIB Center for Plant Systems Biology, 9052 Ghent, Belgium

4 Department of Applied Mathematics, Computer Science and Statistics, Ghent University, 9000 Ghent, Belgium

5 Department of Statistics, University of California, Berkeley, Berkeley, CA, USA

6 SynBioC, Department of Green Chemistry and Technology, Ghent University, Coupure Links 653, 9000 Ghent, Belgium

7 B CUBE Center for Molecular Bioengineering, Technical University of Dresden, Tatzberg 41, 01307 Dresden, Germany

8 Bioinformatics Institute Ghent, Ghent University, Technologiepark 71, 9052 Ghent, Belgium

9 Marine Biological Association, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK

10 School of Ocean and Earth Science, University of Southampton, Southampton, UK

11 Bioorganic Analytics, Institute for Inorganic and Analytical Chemistry, Friedrich Schiller University Jena, Lessingstr. 8, 07743 Jena, Germany

important role in primary production and biogeochemical cycling in the oceans [2, 3]. Their unique life cycle is characterized by cell size reduction during vegetative growth [4]. Cells become sexually active once their size is below a species-specific sexual size threshold (SST). Sexual reproduction restores the maximum cell size by expansion of the zygote to form large auxospores that release an initial cell [4]. Although the conservation of meiotic genes and population genetic data on sexual homologous recombination [5–7] suggest that sexual reproduction occurs in natural diatom populations, reports on sexual events remain scarce and are usually restricted to field sites with high frequency monitoring [8–11]. Successful crossing of diatoms in laboratory conditions, however, revealed a diverse range of life cycle strategies with unique features for centric, araphid pennate, and raphid pennate diatoms [4, 12].

Characteristic for pennate diatoms, sexual reproduction is initiated by the interaction of sexually mature vegetative cells (gametangia) from compatible mating types [4]. Whereas passive physical forces are thought to steer cell–cell interaction in planktonic species [13], most benthic raphid diatoms actively move toward a partner of the compatible mating type [4]. Recently, experimental evidence has shown that certain pennate diatoms deploy sex pheromones to recognize and localize a suitable partner [14–17]. Furthermore, a multistage pheromone cascade was discovered in the raphid pennate diatom *Seminavis robusta* (Fig. 1), emphasizing the largely unexplored complexity of mate localization and recognition in diatoms [16]. Previous studies have briefly addressed the transcriptomic response to sex inducing pheromones (SIPs) in pennate diatoms [16, 17], but a detailed overview and timing of expression is currently lacking. Importantly, while the previously discovered multistage pheromone cascade suggests a mating type specific response to SIPs, it is unknown how this is reflected at the molecular level.

Here, we use RNA-sequencing (RNA-seq) to provide a detailed description of the response to SIPs in *S. robusta*. Responses in gene expression were compared between a newly generated time-series RNA-seq dataset for mating type + (MT+) and existing datasets for the compatible mating type − (MT−). We relate physiological changes resulting from a G1 phase arrest to differentially expressed genes throughout the cell cycle and show that SIP− increases motility of the attracted MT+. To tackle technical challenges in comparing gene expression between datasets, a workflow is introduced to integrate RNA-seq datasets. This approach allowed the identification of key genes exhibiting either mating type specific or shared responses to SIP. These key genes include a sexually induced cyclin and highlight the importance of reactive oxygen species (ROS), energy metabolism, and ubiquitination in the mating process.

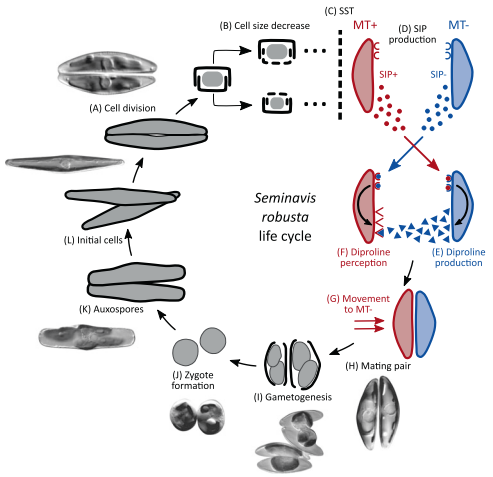

Fig. 1 The life cycle of *Seminavis robusta*. The *S. robusta* life cycle is diplontic and consists of long periods of vegetative division alternating with short periods of sexual reproduction. **a** The average cell size decreases with every mitotic division. **b** Transverse view of a vegetative cell showing the mechanism of cell size decrease. **c** When populations pass the sexual size threshold (SST) at a cell size of about 50 µm, cells become capable of sexual reproduction. **d** Mating type + (MT+) and mating type − (MT−) start producing sex inducing pheromones called SIP+ and SIP−, respectively. SIP induces a cell cycle arrest in the compatible mating type. **e** In response to SIP+, MT− secretes an attraction pheromone: the diketopiperazine diproline, while (**f**) MT+ becomes sensitive to diproline and glides toward the diproline source. **g, h** Diproline signaling leads to mate finding and pair formation. **i, j** Each partner produces two gametes that fuse with the gametes of the compatible mating type to form zygotes. Finally, auxospores (**k**) will enlarge and release an initial cell of the original cell size (**l**). This figure was modified from Moeys et al. (2016) and Gillard et al. (2012) and some microscopic pictures were obtained from Chepurnov et al. (2002) with permission.

Material and methods

Culture conditions

Seminavis robusta strains were obtained from the BCCM/DCG diatom culture collection at Ghent University (http://bccm.belspo.be/about-us/bccm-dcg, see Supplementary Table 1 for an overview of used strains) and were grown in sterile filtered natural sea water from the North Sea enriched with Guillard's F/2 solution, except for the experiments involving MT+ motility and MT− flow cytometry where cultures were grown in artificial sea water with F/2 solution. Prior to the experiments, cultures were made axenic by adding 500 mg/L penicillin, 500 mg/L ampicillin, 100 mg/L streptomycin and 50 mg/L gentamycin to the medium for one week before the experiment. Cultures were grown at 18 °C in 12 h:12 h light:dark cycles under cool white fluorescent lamps unless stated otherwise.

Preparation of SIP containing filtrate

MT− cultures (strain 85B) and MT+ cultures (strain D6) below the SST were cultured in 150 mL cell culture flasks for 1 week. When cultures reached the late-exponential growth phase the medium was vacuum filtered using a Stericup with pore size of 0.22 µm (Merck GmbH, Germany) in order to obtain a filtrate containing SIP− and SIP+, respectively. The potency of the filtrate was assessed using a cytokinetic arrest and a diproline attraction assay (see Supplementary Methods). The SIP filtrate was used for the RNA-seq experiment, for cell cycle analysis using flow cytometry and for assessing its effect on motility of MT+ (Supplementary Methods).

RNA-seq experimental setup, data analysis and functional interpretation

RNA-seq data generated in this study representing the response of MT+ (strain 85A) to SIP− filtrate was complemented with existing data on the response of MT− to a chromatographic fraction containing SIP+ (time points 15 min, 1 h, 3 h) [16] and SIP+ filtrate (time point 10 h) [18]. The experimental setup for the two MT− RNA-seq datasets are described in their respective publications [16, 18]. For the novel MT+ dataset, we harvested control cultures in the dark (time point 0 h) and further harvested dark-synchronized cultures consisting of three control replicates and three SIP− treated replicates at five time points: 15 min, 1 h, 3 h, 6 h and 9 h. Details about the experimental setup are described in Supplementary Methods.

All three RNA-seq datasets were mapped to gene models from the *S. robusta* genome v1.0 (available at https://bioinformatics.psb.ugent.be/orcae/overview/Semro) [19] using Salmon v0.9.1 [20] (Supplementary Fig. 1A) and differential expression (DE) analysis was performed using negative binomial models implemented in the edgeR package [21]. Details about the model, design matrix, and contrasts of interest are described in Supplemenatry Methods. Functional annotation for all genes was derived using an ensemble of three methods: InterProScan [22], AnnoMine [23] and eggNOG-mapper [24]. Gene families were computed by clustering protein sequences with TRIBE-MCL [25]. Specific details on the construction of functional annotation are described in Supplementary Methods.

To identify biologically relevant DE genes, we used two complementary approaches, both of which are based on the same statistical model. First, we used the results from the conventional DE analysis to identify genes and biological processes involved in sexual reproduction, based on the genes' functional annotation and the current literature. Second, we developed a novel integrative workflow that allowed us to compare the response to SIP in different RNA-seq datasets from both mating types, by testing against a fold change cut-off. We restricted the comparison to the time points that are available for both mating types (15 min, 1 h and 3 h). Genes discovered by the integrative analysis represent key genes with a shared versus mating type specific response to SIP. Deriving both analyses from the same statistical model implies that genes discovered by the more stringent fold-change analysis are also discovered in a conventional analysis. Three sets of genes were defined: genes responding to SIP in both mating types (SRBs: "SIP Responsive in Both mating types"), genes with a MT− specific response (SRMs: "SIP Responsive in mating type Minus") and genes with a MT+ specific response (SRPs, "SIP Responsive in mating type Plus"). Details about the integrative workflow can be found in Supplementary Methods.

Data availability and reproducibility

The raw data from the new RNA-seq experiments are available at the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB35793 (https://www.ebi.ac.uk/ena/data/view/PRJEB35793). Additional genomic information concerning genes mentioned here can be found on the diatom PLAZA platform for comparative genomics (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_diatoms_01/). All scripts required to reproduce the analyses and figures reported in this paper as well as Salmon estimated count matrices and results of statistical tests are available on our GitHub repository at https://github.com/statOmics/SeminavisComparative.

Results and discussion

In this study, we generated a new time-series dataset to investigate the response of dark-synchronized MT+ cultures to SIP− filtrate (6 time points, 0–9 h, Fig. 2a) and to compare expression patterns with two existing datasets on the response of MT− to SIP+ [16, 18] (4 time points, 0–10 h, Fig. 2b). Here, we first report the results of a conventional DE analysis for each dataset separately. Next, we use these results to discover genes or biological processes related to sexual reproduction, either based on their predicted functional annotation or the current literature. Finally, we integrate the results of both the novel and publicly available datasets in an integrative analysis that aims at identifying key genes involved in either a single or both mating types.

General transcriptional response and identification of key SIP responsive genes

Multidimensional scaling plots of the RNA-seq data (Fig. 2a, b) showed that in both mating types the dark-to-

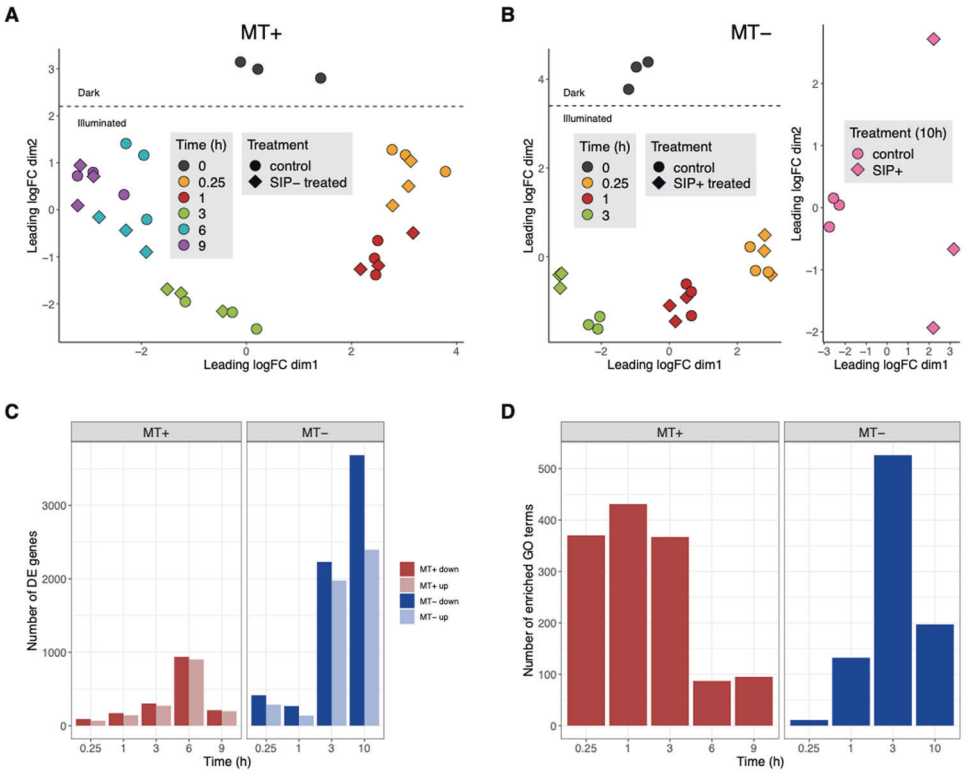

Fig. 2 Transcriptional responses induced by SIP treatment of *Seminavis robusta*. **a** Multidimensional scaling (MDS) plot for MT+ expression data (0 h–9 h), and (**b**) MDS plots of two MT− expression datasets (0 h–3 h and 10 h, respectively). Distances between samples in the MDS plot approximate the log2 fold changes of the top 500 genes. **c** Number of significant DE genes between control and SIP treated cultures for each time point in both mating types. Each dataset was analyzed on a 5% overall FDR (OFDR) level, i.e., the fraction of false positive genes over all rejected genes. The color represents mating type (MT+: red, MT−: blue) and the shade denotes direction of change (dark: downregulated after SIP treatment, light: upregulated after SIP treatment). **d** Number of significantly enriched GO terms on a 5% FDR level for each time point. The color represents mating type (MT+: red, MT−: blue). A high number of GO terms are discovered in the early time points for MT+, in comparison to the number of DE genes.

light transition and time since illumination were the major drivers of gene expression change throughout the experiments. However, as time progresses, the effect of SIP becomes more pronounced. This is supported by DE analysis showing that the number of significant genes increased markedly at later time points (Fig. 2c). Overall, most genes are significantly DE in only a single time point on a 5% overall FDR (OFDR) level (Supplementary Fig. 1B). Combined, on a 5% OFDR level, 4037 genes are DE in response to SIP treatment for MT+, while 5486 genes are found to be DE in MT− in the first 3 h and 6079 genes after 10 h. The stronger response during the first 3 time points in MT− versus MT+ may be the result of the use of a chromatographic fraction of SIP+ for MT− while MT+ cultures were treated with a SIP− containing filtrate. Gene sets for many biological processes were significantly enriched in the conventional lists of DE genes: a total of 1081 and 740 enriched biological process terms were discovered for the response to sex pheromones in MT+ and MT−, respectively (Fig. 2d, Supplemenatry Fig. 2).

We discovered key genes by developing a statistical integrative analysis workflow that is capable of testing for equivalent, i.e., non-DE, expression between conditions. Coupling equivalence testing in one mating type with DE calls in the other allowed for the discovery of key genes exhibiting mating type specific responses to SIP, while DE calls in both datasets found key genes responsive in both mating types. This workflow revealed 52 key genes responding to SIP in both mating types (SRBs), 12 genes uniquely responding in MT+ (SRPs) and 70 genes uniquely responding in MT− (SRMs) (Fig. 3a, Supplementary Figs. 3, 4, 5). Similar to the conventional DE analysis, the response of MT− was more pronounced compared to MT+, likely due to technical differences such as different protocols for pheromone administration. Remarkably, while in MT− we discovered a comparable number of down- and upregulated SRMs, we only found upregulated SRPs and SRBs, possibly indicating that sexual processes induced by SIPs are mainly driven by the induction of key genes rather than the downregulation of inhibitory genes (Fig. 3a).

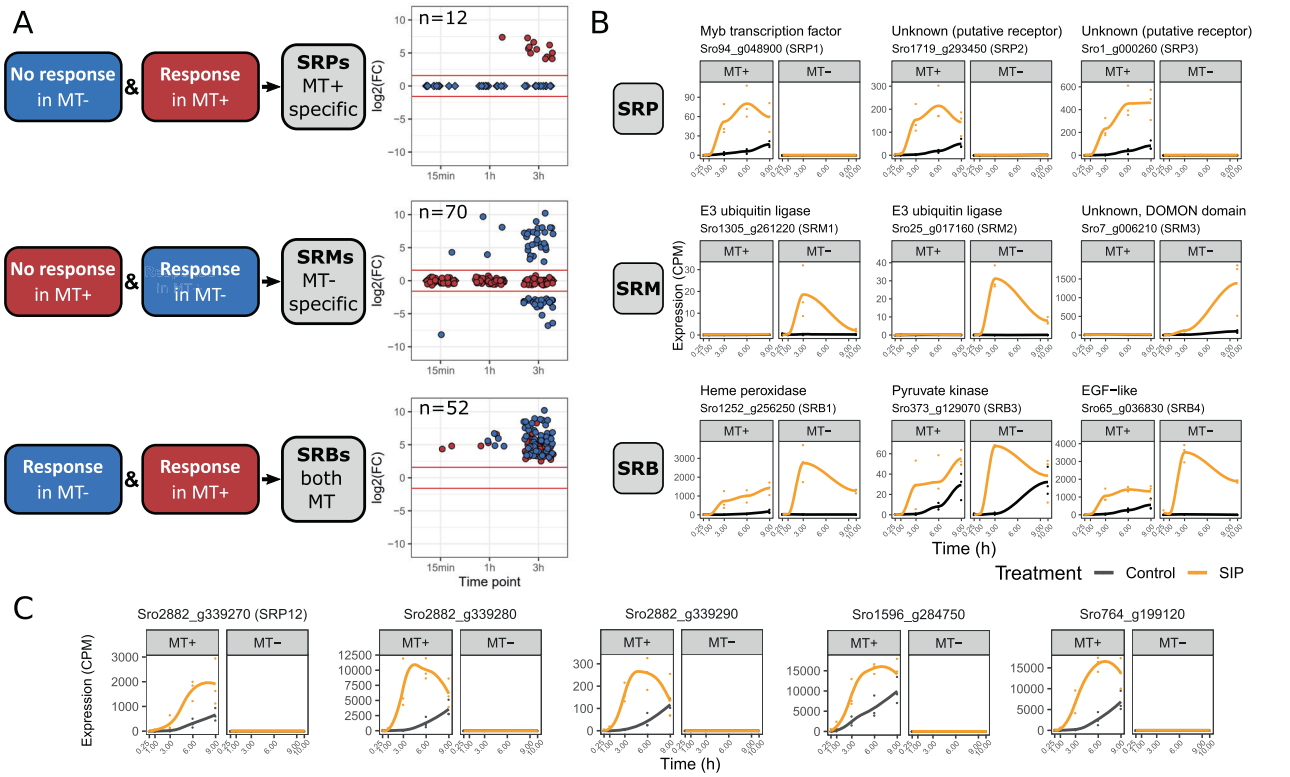

Fig. 3 Visualization and main results of the integrative workflow. a Schematic representation of the integrative workflow indicating how SIP responsive genes with a shared response (SRBs) or mating type specific response (SRPs, SRMs) were discovered. Non-responsive genes consist of genes that are equivalently expressed after SIP treatment versus control, or that are very lowly/not expressed (filtered). A log fold change (LFC) cutoff of ±log(3) was used to define responsive (differentially expressed) genes and equivalent genes. At the right side of the panel, log2 fold changes of SRMs, SRPs and SRBs in both mating types are plotted. Each gene is plotted for the time point at which they are differentially expressed. Genes which were not expressed ("filtered") in the non-responsive mating type are plotted as diamonds. The red horizontal lines represent the fold change cutoff used to determine equivalence and differential expression. The number of discovered genes is indicated in the top left corner of each plot. **b** Expression of a selection of SIP responsive genes (SRMs, SRPs, SRBs). For each gene, counts per million (CPM) are plotted as a function of time for both mating types. The data points correspond to gene expression of the replicates in each time point and the solid line represents the mean. Data points and lines are colored according to condition, i.e., black for control condition and orange for SIP treatment. **c** Expression of the five genes belonging to the gene family of SRP12 (*Sro2882_g339270*). Data are presented in the plots in the same manner as in (**b**).

In what follows we will discuss in more detail the genes and pathways that are responding to SIP in both mating types or uniquely in one mating type and link these changes to physiological events in the mating process. In each section, we first discuss key genes highlighted by the integrative analysis (Fig. 3), after which we discuss results from the conventional DE analysis, focussing on selected biological processes (Fig. 4).

Responses to SIP conserved in both mating types

Integrative analysis reveals key genes responsive in both mating types

A large fraction (22/52) of SRBs, i.e., key genes with a strong response to SIP in both mating types, lack any functional annotation and homology to sequenced genomes of other diatoms (Supplementary Table 2), suggesting that the molecular mechanisms underlying early mating are highly species-specific. The remaining 30 SRBs can be linked to energy metabolism, ROS signaling and meiosis, amongst others. Pyruvate kinase (*Sro373_g129070*) and isocitrate dehydrogenase (*Sro492_g153950*), respectively involved in glycolysis and the citric acid cycle, are strongly upregulated in both mating types (Fig. 3b), suggesting an increased energy demand. Interestingly, pyruvate kinase is also upregulated during gametogenesis in the brown alga *Saccharina latissima* [26] and the parasite *Plasmodium berghei* [27]. In addition, two enzymes from the pentose phosphate pathway (PPP) are among the SRBs: transketolase (*Sro524_g159900*) and transaldolase (*Sro196_g083630*) (Supplementary Fig. 3). The PPP generates NADPH, a reductive compound needed in various metabolic reactions and involved in detoxification of ROS by regenerating glutathione [28, 29]. Furthermore, one SRB encoding a heme peroxidase (*Sro1252_g256250*) exhibited

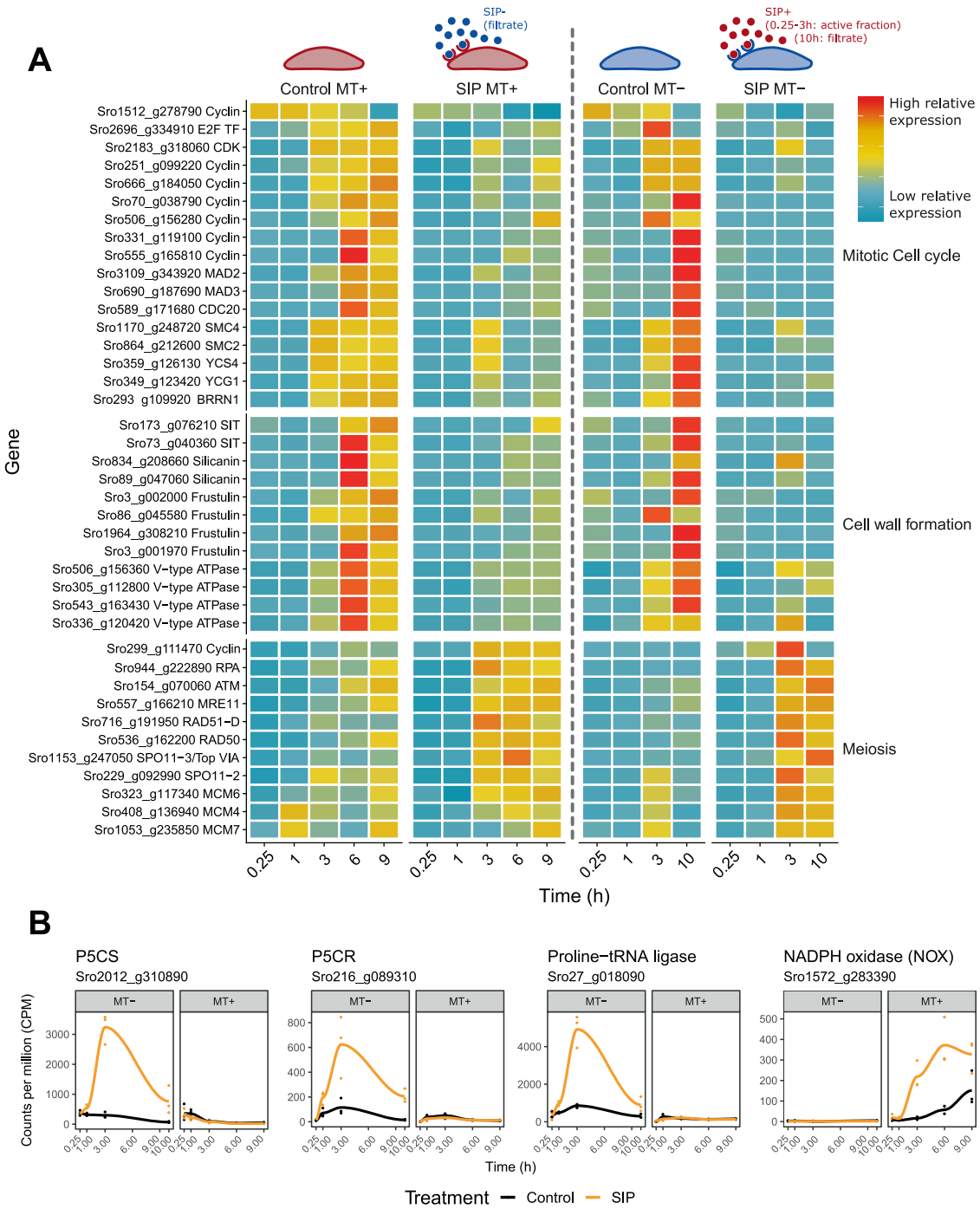

Fig. 4 Gene expression of *Seminavis robusta* genes involved in mating-related processes. a Heatmap of genes related to mitotic and meiotic cell cycle progression, which are differentially expressed (DE) in both mating types in the conventional DE analysis. Each gene is plotted for control and SIP treated conditions in both *S. robusta* mating types. Genes are specified as row names and are scaled relative to the mean expression, amounting to counts per million (CPM) standardized to zero mean and unit variance for each gene in each mating type separately. Blue indicates low expression, while red indicates high expression. **b** Expression of genes related to diproline synthesis and reactive oxygen species (ROS) production, which are significantly DE in only one mating type in the conventional DE analysis. CPM are plotted as a function of time for both mating types. The data points correspond to gene expression of the replicates in each time point while the solid line represents the mean. Data points and lines are colored according to condition, i.e., black for control condition and orange for SIP treatment. P5CS = Δ1-pyrroline-5-carboxylate synthetase; P5CR = Δ1-pyrroline-5-carboxylate reductase.

strong upregulation upon SIP treatment (Fig. 3b). Upregulation of heme peroxidases was also reported during sexual reproduction in other eukaryotes, e.g., mosquitoes (*Anopheles gambiae*) [30] and fungi [31, 32]. Heme peroxidases promote substrate oxidation in various metabolic pathways and are essential for the detoxification of ROS [33], suggesting that ROS signaling plays a role in the response to SIP, as seen in the green algae *Volvox carteri*, where high ROS levels trigger sex [34]. Finally, a highly expressed SRB encodes a transmembrane protein containing an Epidermal Growth Factor-like (EGF-like) domain (*Sro65_g036830*, Fig. 3b), with potential orthologs encoded in pennate and centric diatoms including *P. tricornutum* and *T. pseudonana* (BLASTp, E < 1e-10). EGF-like domains are generally extracellular protein modules that play a role in receptor/ligand interactions, intracellular signaling, and adhesion [35]. Membrane bound proteins containing EGF-like domains are required for gamete fusion in green algae [36] and oocyte binding by animal sperm cells [37–39]. Furthermore, multiple EGF-like repeats were discovered in sexually induced genes Sig1–3 of the centric diatom *Thalassiosira weissflogii* [40]. Sig1 orthologs were later shown to be located on the mastigonemes of stramenopile flagella [41], suggesting their upregulation is related to the differentiation of flagellated male gametes in centric diatoms. We could not detect orthologs of Sig1–3 in the genome sequence of *S. robusta* (BLASTp, E < 0.001), in line with the lack of flagellated stages in pennate diatoms. Nevertheless, the presence of EGF-like domains in sexually induced genes in pennate and centric diatoms suggests that genes containing EGF-like domains play a role in diatom cell–cell communication or gamete fusion.

Conventional DE analysis uncovers mating-related processes differentially regulated in both mating types

We used flow cytometry to confirm a sex pheromone induced G1 phase arrest, which was proposed for *S. robusta* and *Pseudo-nitzschia multistriata* [16, 17]. Treatment with SIP significantly decreased the proportion of G2/M phase cells in MT+ after 3 h (control 12.4% versus SIP 3.5%, $p = 0.0023$) and 9 h (control 20.9% versus SIP 9.4%, $p = 0.0123$) and after 9 h in MT− (control 10.5% versus SIP 3.3%, $p < 0.0001$) (Fig. 5a, b). A temporary arrest of the cell cycle appears to be a prerequisite for a switch to meiosis in many diatoms [16, 17, 42, 43], although in the centric diatom *Skeletonema marinoi* no growth arrest was observed during sexual reproduction [44]. The cell cycle arrest is reflected in the transcriptomic data as a sequential downregulation of cell cycle genes in the conventional DE results (Fig. 4a), causing an enrichment in cell cycle related GO terms (Supplementary Fig. 2). In eukaryotes, S phase progression is controlled by E2F transcription factors forming a heterodimer with Dimerization Partner (DP) transcription factors [45]. In accordance with plants and animals [46, 47], we observe an increase in expression of E2F transcription factors in control conditions as cells go through the cell cycle. SIP treatment significantly repressed expression of two E2Fs (*Sro2696_g334910* and *Sro1798_g298290*) in both mating types and MT−, respectively (Supplementary Fig. 6A). The transcriptional repression of E2Fs is likely caused by the inability of cells to enter S phase as a result of the G1 phase arrest, although the activity of E2Fs is generally also regulated by other mechanisms such as Retinoblastoma-related (Rb) protein binding and phosphorylation which we did not assess [46, 47]. Furthermore, two DP genes were DE in MT− in response to SIP+; one (*Sro905_g218540*) was repressed, while the other (*Sro905_g218570*) was induced, suggesting they play a contrasting role in the cell cycle arrest (Supplementary Fig. 6A).

During the S and G2 phase of the mitotic cell cycle, fission of both chloroplasts of *S. robusta* results in four daughter chloroplasts by the start of the M phase [48]. FtsZ and dynamin-related protein 5B (DRP5B) are key factors in the formation of a multi-ring structure that constricts the chloroplast during fission [49, 50]. In control conditions, we observe a expression peak of two FtsZ homologs (*Sro1409_g270150* and *Sro931_g221490*) and DRP5B (*Sro814_g206290*), potentially coinciding with the timing of chloroplast division in *S. robusta* [48] (Supplementary Fig. 6B). After exposure to SIP, expression of these genes was significantly repressed in both mating types (Supplementary Fig. 6B), indicating that the cell cycle arrest results in an inhibition of chloroplast fission. Indeed, chloroplast division ceases during sexual reproduction of *S. robusta* so that each gamete will contain one chloroplast and the auxospore inherits two chloroplasts [51]. Noteworthy, while DRP5B is restricted to the chloroplast [52], stramenopiles can encode a mitochondrium-targeted FtsZ that is involved in mitochondrial division [53]. Thus, it is possible that the downregulation of FtsZ genes is linked with mitochondrial instead of chloroplastic division.

In the transcriptomic data, seven cyclins were downregulated in both mating types after treatment with SIP (Fig. 4a). Six out of seven appear in a cluster of genes, which peak late in the time series in control conditions (6 h, 9 h and 10 h; Supplementary Fig. 7), suggesting that they are mitotic cyclins involved in G2/M transition [54]. To determine which cyclin family they represent, a maximum likelihood phylogenetic tree of *S. robusta* cyclins was constructed (Supplementary Fig. 8). The repressed cyclins consist of three diatom-specific cyclins (dsCyc), three A/B type cyclins and one cyclin D (Supplementary Table 3).

Several genes involved in mitosis were downregulated in both mating types, including MAD2 (*Sro3109_g343920*),

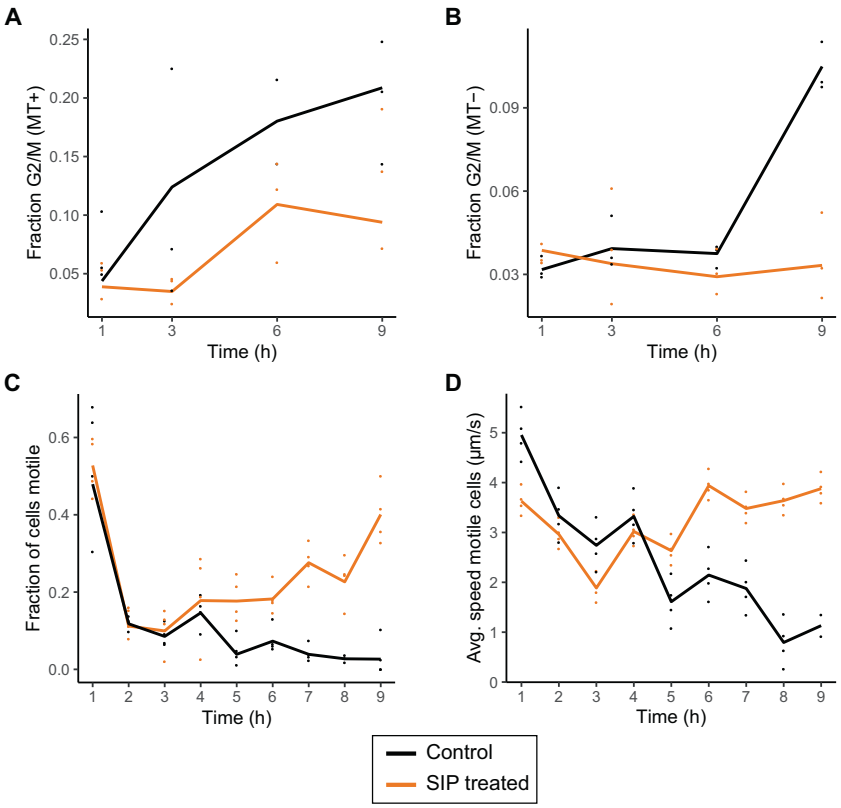

Fig. 5 Physiological responses to SIP. a, b The fraction of cells in the G2 or M phase during the RNA-seq experiment for MT+ (**a**), and for MT− (**b**) as determined by flow cytometry ($n = 3$). For both MT+ and MT−, SIP induces a G1 phase cell cycle arrest, apparent by a significant lower fraction of cells in G2 and M phase after 3 h and 9 h (MT+) and 9 h (MT−). **c, d** Motility of MT+ cells over time after treatment with a 1/10 dilution of SIP filtrate ($n = 4$). Both the proportion of motile cells in the culture was determined (**c**) as well as the gliding speed of cells that are motile (**d**). Note that an initial high proportion of motile cells are observed after 1 h in all treatments, likely the result of a phototactic response following illumination after a prolonged period of darkness. For all panels, points show individual data points while solid lines show the average per time point. Untreated (control) cultures are indicated in black, while SIP treated cultures are represented in orange.

MAD3 (*Sro690_g187690*) and CDC20 (*Sro589_g171680*), which form the Mitotic Checkpoint Complex. Furthermore, we observe downregulation of five genes coding for subunits of the condensin complex, which play a central role in chromosome organization during mitosis and meiosis [55] (Fig. 4a).

Finally, at the end of each mitotic cell cycle, diatom cells produce a new valve in a silica deposition vesicle (SDV) prior to cytokinesis. As it was previously shown [16], treating cultures with SIP of the compatible mating type reduces the fraction of cytokinetic cells after 14 h (Supplementary Fig. 9A, $p < 0.0001$ for both MT+ and MT−). Accordingly, we observed downregulation of genes known to be important for silica cell wall formation after treatment with SIP (Fig. 4a), including two out of five *S. robusta* silicic acid transporter homologs, involved in the uptake of silicic acid from the environment [56]. In addition, we find two (out of a total of three) silicanin1-like genes as well as four (out of thirteen) frustulins, respectively proteins embedded in the SDV membrane and proteins which are found in the organic casing surrounding the cell wall [57, 58]. Moreover, four genes making up various subunits of V-type ATPase complexes were significantly downregulated in both species upon treatment with SIP. These proton pumps play a role in biomineralization of silica by acidifying the SDV [59].

While mitotic cell cycle genes were generally downregulated by SIP treatment, we observed an increase in meiotic gene expression, accompanied by an enrichment in the "meiotic chromosome condensation" GO term in both mating types (Supplementary Fig. 2). Surprisingly, one cyclin which is an SRB (*Sro299_g111470*, Fig. 4a) is upregulated rather than downregulated in response to the pheromone. Its expression pattern suggests it plays a role either in SIP induced physiological responses such as mate finding or in preparing the cell for meiotic cell cycle progression. The gene has presumably evolved independently from the sexually induced cyclins of other major eukaryotic clades, as phylogenetic analysis places the gene among diatom specific cyclins (dsCyc) (Supplementary Fig. 8). The

expansion of the cyclin family in diatoms compared to other members of the SAR clade [60] might have been instrumental to allow the diversification of this sexually induced cyclin [61–64].

To investigate the dynamics of other meiotic genes throughout the experiment, we explored the expression of a set of 42 known meiotic and bifunctional mitotic/meiotic diatom genes [5], some of which were induced in response to sex pheromones in pennates [16, 17] and during sexual reproduction in the centric diatom *S. marinoi* [44]. Here, we found ten meiotic markers significantly responding to SIP in at least one time point in both mating types (Fig. 4a). These include genes encoding DNA replication licensing factors MCM4, MCM6, MCM7, which are involved in the initiation of replication during the mitotic and meiotic S phase [5]. Two homologues of SPO11 (SPO11–2 and SPO11–3) are upregulated in both mating types. SPO11 plays a role in the formation of double strand DNA breaks during homologous recombination. In diatoms and plants, the SPO11–2 homologue was shown to be meiotic while SPO11–3 is involved in vegetative growth [5, 17, 65]. However, the observed upregulation of SPO11–3 in response to SIP in *S. robusta* and during sexual reproduction in the centric *S. marinoi* [44] suggest that in addition to SPO11–2, SPO11–3 might also be involved in meiotic homologous recombination in diatoms. Three genes involved in DNA repair after the induction of double stranded breaks by SPO11 were upregulated in both mating types: MRE11, RAD50, and RAD51D [5]. Finally, we observed the upregulation of two genes not yet described in the meiotic toolkit of diatoms: Ataxia Telangiectasia Mutated (ATM, *Sro154_g070060*) which codes for a protein that controls double-strand break formation by SPO11 during meiosis [66] and Replication Protein A (RPA, *Sro944_g222890*), involved in the binding of ssDNA during replication and homologous recombination [67] (Fig. 4a). Interestingly, although meiotic genes are upregulated in response to SIP, the G1 phase arrest inhibits progression through the meiotic cell cycle and we did not observe gametogenesis in MT+ cells attracted to diproline loaded beads. Accordingly, chloroplast rearrangements and indications of the meiotic prophase were only observed microscopically after compatible cells form a mating pair [51]. We therefore hypothesize that a separate, local signal during cell pairing is required to break the G1 phase arrest and induce gametogenesis.

In response to SIP, 17 genes with a guanylate cyclase domain (GC) were significantly DE in both mating types, 8 of which form a bifunctional guanylate cyclase/phosphodiesterase (GC/PDE) fusion enzyme. Interestingly, the GC and PDE domains in these genes have contrasting functions, respectively synthesizing and breaking down the secondary metabolite cGMP. Although the genes with only a GC domain do not show a general direction of regulation, all GC/PDE genes were upregulated (Supplementary Fig. 10), including *Sro991_g228730*, which was previously described to be responding to SIP in *S. robusta* [16]. All significant GC/PDEs show the topology described by Moeys et al. (2016) [16] a small N-terminal intracellular domain, an extracellular stretch and a long C-terminal intracellular part containing the GC and PDE domains. Thus, cGMP signaling appears to be a conserved response to sex pheromones in pennate diatoms, as guanylate cyclases are also upregulated in *P. multistriata* [17].

Finally, orthologs of three uncharacterized genes induced during sexual reproduction in the centric diatom *S. marinoi* and the pennate diatom *P. multistriata* [44], were upregulated in response to SIP (Supplementary Fig. 6C). *Sro587_g171310* (orthologous to *STRINITY_DN12692_c0_g1_i1* and *Pmu0010180*) was significantly upregulated in both mating types. Protein domain analysis revealed the presence of a Homologous-pairing protein 2 (Hop2) domain in its *S. robusta* (IPR010776), *P. multistriata* (IPR010776) and *S. marinoi* (IPR040461) sequence. Although Hop2's role in homologous recombination would be in accordance with the observed expression, more work is needed to confirm this gene as a Hop2 homologue, as Hop2 is presumed to be absent in diatoms [5, 44]. Another conserved sexual gene (*Sro131_g062230*, orthologous to *Pc15065_g1_i1* and *Pmu0009930*), is upregulated in response to SIP only after 10 h of SIP treatment in MT−. Finally, *Sro637_g179400* (ortholog of *MTRINITY_DN9343_c0_g1_i2* and *Pmu0061540*) was upregulated after 3 h and 10 h in MT−. Interestingly, domain predictions show that the protein consists of an unknown N-terminal domain, followed by one transmembrane helix and a C-terminal beta-propeller domain in all three species (IPR013519, IPR011043 and IPR015943 in *S. robusta, P. multistriata* and *S. marinoi* respectively). This topology suggests a function as a receptor or in adhesion [68]. Since these three genes are upregulated in two pennate and one centric diatom species, they are interesting candidate marker genes for sexual reproduction in diatoms. Notably, we discovered five additional conserved sexual genes in the *S. robusta* genome which are not differentially expressed to SIP. If their function is conserved, we expect these genes to be upregulated during zygote or auxospore formation rather than during pheromone signaling, since the physiology of mate finding strongly differs between species.

Responses to SIP specific for MT+

We identified 12 genes displaying a MT+ specific response to SIP (SRPs, Fig. 3a, Supplementary Table 2), among them a Myb-like transcription factor (*Sro94_g048900*), which is likely regulating downstream events associated with the perception of SIP− by MT+. One such downstream response

unique for MT+ is responsiveness to the attraction pheromone diproline [16], which we confirmed using a bead attraction assay (Supplementary Fig. 9B). This might be caused by SIP− induced expression of a diproline receptor in MT+. While biosynthetic pathways of 2,5-diketopiperazine cyclodipeptides such as diproline have been elucidated, cyclodipeptide receptors are not well characterized [69]. Among the SRPs, we have identified two functionally unidentified genes which satisfy the requirements for a potential diproline receptor: *Sro1719_g293450* and *Sro1_g000260*. Both encode transmembrane proteins and show high expression uniquely in response to SIP− in MT+ (Fig. 3b). Interestingly, the former gene is predicted to contain 7–8 transmembrane helices, reminiscent of G-protein coupled receptors (GPCR) which often associate with cyclodipeptides in humans [69] and, moreover, a GPCR was upregulated in response to sex pheromones in *P. multistriata* [17]. Sexual *S. robusta* MT− cells were shown to predominantly secrete the cyclo(L-Pro-L-Pro) enantiomer [14]. However, while pheromone receptors are usually stereosensitive, synthetic cyclo (D-Pro-D-Pro) is also biologically active, suggesting a highly unusual stereo-insensitive receptor or racemization of the pheromone [14, 70]. Further, the SRPs include 6 genes lacking functional annotation, among which is SRP12 (*Sro2882_g339270*) that belongs to a gene family containing five *S. robusta* genes of which three are located adjacently on the same genomic contig (*Sro2882_g339270, Sro2882_g339280* and *Sro2882_g339290)*. Four out of five show a MT + specific upregulation to SIP− with very high expression ranging up to 1.5% of the total transcriptome library (Fig. 3c). Furthermore, all five genes exhibit a high baseline MT+ expression and negligible expression in control MT− conditions (Fig. 3c), suggesting they play a role in mating type differentiation, comparable to MRP genes in the pennate *P. multistriata* [71].

When exploring ROS signaling related genes, the conventional DE analysis uncovered an NADPH oxidase (NOX, Sro1572_*g283390*) with a pronounced MT+ specific response to SIP (Fig. 4b), catalyzing extracellular production of superoxide anions by moving an electron from NADPH through the plasma membrane to molecular oxygen. The observed NOX contains six transmembrane domains [72] and an EF-hand domain suggesting a potential link with calcium signaling. As superoxide is cell impermeable, extracellular superoxide is most likely dismutated to H_2O_2, which can enter neighboring cells for ROS signaling, e.g., through aquaporin channels [73, 74]. NOX activity during sexual reproduction is a common theme in eukaryotes: it is required for gametogenesis and fertilization in plants [75] and is required for the formation of sexual fruiting bodies in the fungus *Aspergillus nidulans* [32]. Furthermore, gametophytes of the kelp *S. latissima* show female specific expression of NOX, suggesting mating type specific expression of NOX during sexual reproduction may be conserved among stramenopiles [26]. Notably, enzymes from the NADPH producing PPP were upregulated in both mating types, potentially supplying NADPH to support NOX-mediated superoxide production.

To assess whether SIP− affects the motility of MT+, the displacement of control and SIP− filtrate treated cells was tracked for 30 sec over the course of 9 h (Supplementary Fig. 11). We observed mate-searching behavior starting 6 h–7 h after treatment with SIP−. Not only did the fraction of motile cells (i.e., cells showing any displacement) in the culture increase (Omnibus test $p < 0.001$, Fig. 5c, Supplementary Fig. 12A), the speed of the subset of motile cells also increased: after 9 h, the average speed of motile SIP− treated cells was 3.8 µm/s compared to 1.1 µm/s for untreated cells (Omnibus test $p < 0.0001$, Fig. 5d, Supplementary Fig. 12B). Thus, exposure to SIP does not only prime MT+ cells to become responsive to the attraction pheromone diproline, the increased motility of gametangia further increases the probability to encounter an immotile, diproline producing MT− cell. Furthermore, the pre-activation of cellular motility machinery by SIP− might explain the almost immediate attraction to a new diproline source [76]. The timing of the mating behavior in *S. robusta* appears to be remarkably synchronized between mating types: significant increases in motility occurred simultaneously with the first noticeable amount of diproline produced by MT− [14] and the onset of responsiveness of MT+ to diproline (Supplementary Fig. 9B). As cell motility in raphid diatoms is achieved through the secretion of adhesive molecules that attach the cell to the substratum and provide traction for their gliding movement [77], we performed BLAST searches to identify adhesive proteins containing a GDPH-domain, named after a conserved Gly-Asp-Pro-His amino acid motif [78]. Among 87 GDPH-containing proteins identified, four are upregulated in MT+ in response to SIP− treatment in the conventional DE analysis (Supplementary Fig. 6D). However, three out of four GDPH-domain-containing genes are also significantly upregulated in immotile SIP+ treated MT− cells and one is upregulated uniquely in MT−. Therefore, as previously suggested, the GDPH-domain containing proteins may play additional roles related to extracellular adhesion in addition to cell motility, such as mucilage pad, stalk and chain formation as well as cell–cell adhesion during mating [78].

Responses to SIP specific for MT−

A relatively high number (70) of SRMs, i.e., genes with a MT− specific response to SIP, were discovered (Fig. 3a) of which 16 lack any functional annotation (Supplementary Table 2). Molecular functions of SRMs are diverse, including receptors, membrane channels, guanylate cyclases and other

signaling enzymes. Among the SRMs are two E3 ubiquitin ligase genes (*Sro1305_g261220* and *Sro25_g017160*) representing the Ring-Between-Ring (RBR) and U-box family, respectively. Since ubiquitin ligases are important players in signaling by targeting downstream proteins for ubiquitination [79], these genes are potential key regulators of MT− specific responses such as the production of diproline. Ubiquitin ligases also play a role in meiosis across eukaryotes by targeting proteins for proteasomal degradation [80]. However, a function in meiosis for the identified ubiquitin ligases is hard to reconcile with their mating type specific response, as meiosis occurs in both partners. SRMs further include a gene with a DOMON domain, which is a presumed heme or sugar sensor domain (*Sro7_g006210*) [81] (Fig. 3b). Interestingly, two SRM genes belong to an unknown and *S. robusta* specific gene family containing a zinc finger domain (Supplementary Fig. 6E).

Our conventional DE analysis also shows that glutamate-to-proline conversion enzymes Δ1-pyrroline-5-carboxylate synthetase (*Sro2012_g310890*, P5CS), and Δ1-pyrroline-5-carboxylate reductase (*Sro216_g089310*, P5CR) were significantly upregulated after SIP treatment in MT− after 3 h and 10 h, while in MT+ both genes were not significantly responding (Fig. 4b). The unique increase of proline biosynthesis in MT− is further supported by the enrichment of the "proline biosynthetic process" GO term in the 3 h time point only in MT− (Supplementary Fig. 2). Their MT− specific response supports the hypothesis that upregulation of P5CS and P5CR increases the cellular proline pool as a precursor for diproline biosynthesis [16]. Furthermore, a proline-tRNA ligase (*Sro27_g018090*) exhibits a strong significant upregulation uniquely in MT−, with expression levels exceeding 4000 CPM (Fig. 4b). This enzyme attaches proline to transfer RNA (tRNA), which serves as a substrate for ribosomal protein synthesis, explaining its consistent expression in control conditions in this dataset (Fig. 4b). The cyclodipeptide ring of diketopiperazines such as diproline is typically synthesized by either nonribosomal peptide synthetases or cyclodipeptide synthases (CDPS) [82]. Interestingly, CDPS require aminoacyl-tRNA as a substrate for the reaction instead of a free amino acid [83]. Thus, the observed MT− specific upregulation of a proline-tRNA ligase likely caters to the increased need for proline-tRNA^{pro} driving CDPS dependent diproline biosynthesis. BLAST searches in the *S. robusta* genome revealed several candidate CDPS genes, with rather low levels of conservation (Supplementary Table 4), of which none show significant regulation at any time points in any mating type. Thus, either an unidentified, transcriptionally controlled CDPS exists or one of the identified CDPS is non-transcriptionally regulated to be active only in SIP+ treated MT− cells with a cell size below the SST.

Conclusion

Our study shows that, in pennate diatoms, the perception of extracellular chemical cues triggers behavioral changes and alters gene expression of gametangial cells, preparing them for cell pairing and gamete formation (Fig. 6). The response of a number of genes to SIPs was shared by both mating types, including bifunctional GC/PDE genes, an EGF-like transmembrane gene similar to sperm-egg recognition factors in animals [37–39] and several potential adhesive genes unique to diatoms [78]. Furthermore, we confirmed the induction of a G1 arrest in both mating types, reflected in a downregulation of essential genes involved in S phase progression, organelle division, mitosis and cell wall formation. To some extent, parallels can be drawn with yeasts, where compatible sex pheromones also induce a G1 phase arrest, resulting in the downregulation of key cell cycle genes [84–86]. In addition, we observed that in the diploid *S. robusta* gametangial cells of both mating types, the expression of meiotic genes was triggered by SIP, including the first sexually induced cyclin characterized in diatoms. Sex pheromone exposure in *S. robusta* thus activates both mate-finding and meiotic transcriptional programs, a feature also observed in the yeasts *Schizosaccharomyces pombe* and *Candida lusitaniae*, while mating and meiotic programs are strictly separated in the model yeast *Saccharomyces cerevisiae* [87].

Treatment with SIP also elicited mating type specific behavior and gene expression. The MT− specific response of proline biosynthesis and proline-tRNA ligase genes suggests that the synthesis of the attraction pheromone diproline is cyclodipeptide synthase (CDPS) dependent, while two SRPs encode potential diproline receptors. To our knowledge, evidence for such a complex multi-step pheromone signaling system is missing for other micro-algae. Sex pheromones are known to also induce an asymmetric expression of pheromone and receptor genes in the charophycean green alga *Closterium*, but the presence of an additional attraction pheromone remains to be confirmed [88–90]. The MT+ specific response to SIP shows some similarities with land plants, where superoxide-producing NOX also plays an important role in sexual reproduction [91, 92]. Similar to other stramenopiles, where a small set of strongly sex-biased genes underlies mating type differentiation [71, 93], we identified six genes expressed uniquely in MT+ in control conditions that are upregulated following exposure to SIP, including five members of an uncharacterized gene family and a Myb transcription factor. In plants too, male/female determining Myb transcription factors with gender specific expression are found [94–96], while they are also implicated in the sexual phase of fungi and ciliates [17, 95, 97, 98]. Finally, we report several unknown SIP responsive genes, which may include novel

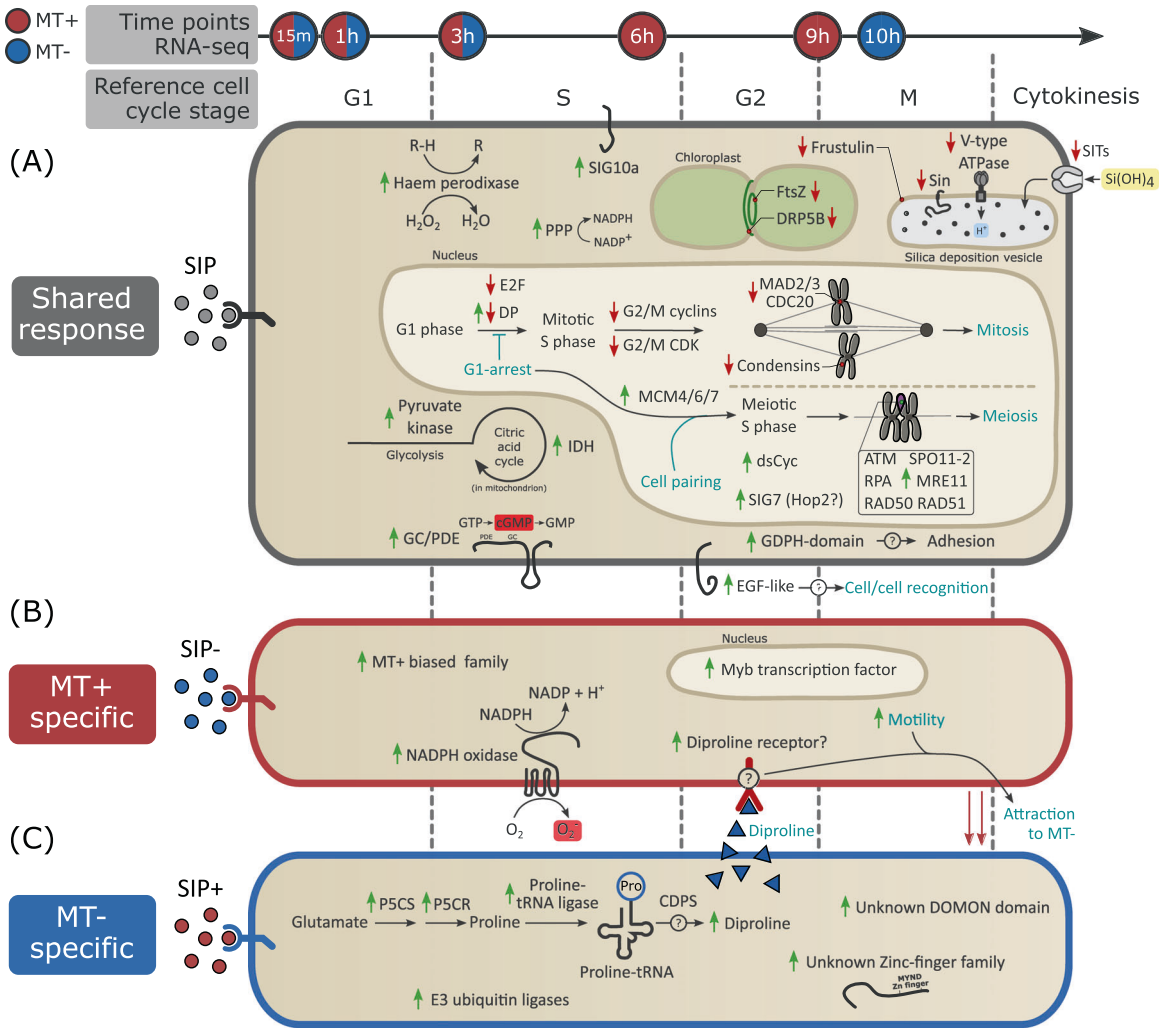

Fig. 6 Overview of molecular and physiological processes in response to SIP in both mating types of *Seminavis robusta*. The arrow on top displays the harvest time of RNA-seq samples. Blue filled time points represent samples taken from MT−, red filled time points are samples from MT+ and red/blue filling indicates sampling for both mating types. The approximate timing of different cell cycle phases is shown below the arrow. **a** Cellular processes taking place in both mating types in response to SIP, (**b, c**) cellular responses to SIP unique for MT+ and MT−, respectively. Genes that are significantly up- or downregulated in response to SIP are depicted with a green or red arrow, respectively. Physiological events are indicated in cyan. PPP = pentose phosphate pathway, IDH = isocitrate dehydrogenase, P5CS = Δ1-pyrroline-5-carboxylate synthetase, P5CR = Δ1-pyrroline-5-carboxylate reductase and CDPS = cyclodipeptide synthase. "SIG" follows the nomenclature for conserved sexual genes used in Ferrante et al. (2019) [44].

cell–cell communication and gamete fusion genes. Among them, genes conserved among pennate diatoms should be prime targets for further research. These potential new marker genes provide a much-needed tool for in situ monitoring of the phenology of sexual reproduction in the natural diatom population.

Acknowledgements We would like to thank Emmelien Vancaester for valuable contributions regarding gene family prediction. We are grateful to dr Jeroen Gillard for providing some of the microscopic pictures of *S. robusta* sexual reproduction stages used in Fig. 1. GB was supported by a Research Foundation Flanders (FWO) Aspirant grant (3F001916). KVdB is a postdoctoral fellow of the Belgian American Educational Foundation (BAEF), and was supported by FWO grants 1S41818N and 1246220N. SDD was supported by FWO grant No. G0D6114N and EB was supported by the research council of Ghent University (BOF/GOA No. 01G01715). NP was supported by the Deutsche Forschungsgemeinschaft (PO 2256/1–1). PB was supported by the Erwin Schrödinger fellowship from Austrian Science Fund (FWF) (J3692-B22) and FWO project G0D6114N. JD was supported by European Research Council grant ERC-ADG-670370. The research leading to the results presented in this publication was carried out with infrastructure funded by EMBRC Belgium—FWO project GOH3817N and funding from the research council of Ghent University (BOF/GOA No. 01G01715).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Goodenough U, Heitman J. Origins of eukaryotic sexual reproduction. Cold Spring Harb Perspect Biol. 2014;6:a016154.
2. Field CB. Primary production of the biosphere: integrating terrestrial and oceanic components. Science. 1998;291:237–40.
3. Sims PA, Mann DG, Medlin LK. Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia. 2006;45:361–402.
4. Chepurnov VA, Mann DG, Sabbe K, Vyverman W. Experimental studies on sexual reproduction in diatoms. Int Rev Cytol. 2004;237:91–154.
5. Patil S, Moeys S, von Dassow P, Huysman MJJ, Mapleson D, De Veylder L, et al. Identification of the meiotic toolkit in diatoms and exploration of meiosis-specific SPO11 and RAD51 homologs in the sexual species *Pseudo-nitzschia multistriata* and *Seminavis robusta*. BMC Genom. 2015;16:930.
6. Rastogi A, Vieira FRJ, Deton-Cabanillas A-F, Veluchamy A, Cantrel C, Wang G, et al. A genomics approach reveals the global genetic polymorphism, structure, and functional diversity of ten accessions of the marine model diatom *Phaeodactylum tricornutum*. ISME J. 2020;14:347–63.
7. Koester JA, Berthiaume CT, Hiranuma N, Parker MS, Iverson V, Morales R, et al. Sexual ancestors generated an obligate asexual and globally dispersed clone within the model diatom species *Thalassiosira pseudonana*. Sci Rep. 2018;8:10492.
8. Crawford R. The role of sex in the sedimentation of a marine diatom bloom. Limnol Oceanogr. 1995;40:200–4.
9. D'Alelio D, d'Alcalà MR, Dubroca L, Sarn D, Zingone A, Montresor M. The time for sex: a biennial life cycle in a marine planktonic diatom. Limnol Oceanogr. 2010;55:106–14.
10. Holtermann KE, Bates SS, Trainer VL, Odell A, Virginia Armbrust E. Mass sexual reproduction in the toxigenic diatoms *Pseudo-nitzschia australis* and *P. pungens* (Bacillariophyceae) on the Washington coast, USA. J Phycol. 2010;46:41–52.
11. Assmy P, Henjes J, Smetacek V, Montresor M. Auxospore formation by the silica-sinking, oceanic diatom *Fragilariopsis kerguelensis* (Bacillariophyceae). J Phycol. 2006;42:1002–6.
12. Frenkel J, Vyverman W, Pohnert G. Pheromone signaling during sexual reproduction in algae. Plant J. 2014;79:632–44.
13. Font-Muñoz JS, Jeanneret R, Arrieta J, Anglès S, Jordi A, Tuval I, et al. Collective sinking promotes selective cell pairing in planktonic pennate diatoms. Proc Natl Acad Sci USA. 2019;116:15997–6002.
14. Gillard J, Frenkel J, Devos V, Sabbe K, Paul C, Rempt M, et al. Metabolomics enables the structure elucidation of a diatom sex pheromone. Angew Chem Int Ed Engl. 2013;52:854–7.
15. Sato S, Beakes G, Idei M, Nagumo T, Mann DG. Novel sex cells and evidence for sex pheromones in diatoms. PLoS ONE. 2011;6:e26923.
16. Moeys S, Frenkel J, Lembke C, Gillard JTF, Devos V, Van den Berge K, et al. A sex-inducing pheromone triggers cell cycle arrest and mate attraction in the diatom *Seminavis robusta*. Sci Rep. 2016;6:19252.
17. Basu S, Patil S, Mapleson D, Russo MT, Vitale L, Fevola C, et al. Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. N Phytol. 2017;215:140–56.
18. Cirri E, De Decker S, Bilcke G, Werner M, Osuna-Cruz CM, De Veylder L, et al. Associated bacteria affect sexual reproduction by altering gene expression and metabolic processes in a biofilm inhabiting diatom. Front Microbiol. 2019;10:1790.
19. Osuna-Cruz CM, Bilcke G, Vancaester E, De Decker S, Bones AM, Winge P, et al. The *Seminavis robusta* genome provides insights into the evolutionary adaptations of benthic diatoms. Nat Commun. 2020;11:3320.
20. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14:417–9.
21. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
22. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
23. Vandepoele K, Van Bel M, Richard G, Van Landeghem S, Verhelst B, Moreau H, et al. pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. Environ Microbiol. 2013;15:2147–53.
24. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol. 2017;34:2115–22.
25. Enright AJ. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002;30:1575–84.
26. Pearson GA, Martins N, Madeira P, Serrão EA, Bartsch I. Sex-dependent and -independent transcriptional changes during haploid phase gametogenesis in the sugar kelp *Saccharina latissima*. PLoS ONE. 2019;14:e0219723.
27. Garcia CHS, Depoix D, Queiroz RML, Souza JMF, Fontes W, de Sousa MV, et al. Dynamic molecular events associated to *Plasmodium berghei* gametogenesis through proteomic approach. J Proteom. 2018;180:88–98.
28. Matthews GM, Butler RN. Cellular mucosal defense during *Helicobacter pylori* infection: a review of the role of glutathione and the oxidative pentose pathway. Helicobacter. 2005;10:298–306.
29. Osada K, Maeda Y, Yoshino T, Nojima D, Bowler C, Tanaka T. Enhanced NADPH production in the pentose phosphate pathway accelerates lipid accumulation in the oleaginous diatom *Fistulifera solaris*. Algal Res. 2017;23:126–34.
30. Shaw WR, Teodori E, Mitchell SN, Baldini F, Gabrieli P, Rogers DW, et al. Mating activates the heme peroxidase HPX15 in the sperm storage organ to ensure fertility in *Anopheles gambiae*. Proc Natl Acad Sci USA. 2014;111:5854–9.
31. Scherer M, Wei H, Liese R, Fischer R. *Aspergillus nidulans* catalase-peroxidase gene (cpeA) is transcriptionally induced during sexual development through the transcription factor StuA. Eukaryot Cell. 2002;1:725–35.
32. Lara-Ortíz T, Riveros-Rosas H, Aguirre J. Reactive oxygen species generated by microbial NADPH oxidase NoxA regulate sexual development in *Aspergillus nidulans*. Mol Microbiol. 2003;50:1241–55.

33. Fawal N, Li Q, Savelli B, Brette M, Passaia G, Fabre M, et al. PeroxiBase: a database for large-scale evolutionary analysis of peroxidases. Nucleic Acids Res. 2013;41:441–4.
34. Nedelcu AM, Marcu O, Michod RE. Sex as a response to oxidative stress: a twofold increase in cellular reactive oxygen species activates sex genes. Proc Biol Sci. 2004;271:1591–6.
35. Appella E, Weber IT, Blasi F. Structure and function of epidermal growth factor-like regions in proteins. FEBS Lett. 1988;231:1–4.
36. Baquero E, Fedry J, Legrand P, Krey T, Rey FA. Species-specific functional regions of the green alga gamete fusion protein HAP2 revealed by structural studies. Structure. 2019;27:113–24.e4.
37. Ensslin MA, Shur BD. Identification of mouse sperm SED1, a bimotif EGF repeat and discoidin-domain protein involved in sperm-egg binding. Cell. 2003;114:405–17.
38. Herlyn H, Zischler H. The molecular evolution of sperm zonadhesin. Int J Dev Biol. 2008;52:781–90.
39. Singson A, Mercer KB, L'Hernault SW. The *C. elegans* spe-9 gene encodes a sperm transmembrane protein that contains EGF-like repeats and is required for fertilization. Cell. 1998;93:71–9.
40. Virginia Armbrust E. Identification of a new gene family expressed during the onset of sexual reproduction in the centric diatom *Thalassiosira weissflogii*. Appl Environ Microbiol. 1999;65:3121–8.
41. Honda D, Shono T, Kimura K, Fujita S, Iseki M, Makino Y, et al. Homologs of the sexually induced gene 1 (sig1) product constitute the stramenopile mastigonemes. Protist. 2007;158:77–88.
42. Armbrust EV, Chisholm SW, Olson RJ. Role of light and the cell cycle on the induction of spermatogenesis in a centric diatom. J Phycol. 1990;26:470–8.
43. Davidovich NA. Transition to sexual reproduction and control of initial cell size in *Nitzschia lanceolata*. Diatom Res. 1998;13:29–38.
44. Ferrante MI, Entrambasaguas L, Johansson M, Töpel M, Kremp A, et al. Exploring molecular signs of sex in the marine diatom *Skeletonema marinoi*. Genes. 2019;10:494.
45. Ramirez-Parra E, del Pozo JC, Desvoyes B, de la Paz Sanchez M, Gutierrez C. E2F–DP Transcription factors. cell cycle control and plant development. Annu Plant Rev. 2019;32:138–63.
46. Dyson N. The regulation of E2F by pRB-family proteins. Genes Dev. 1998;12:2245–62.
47. Mariconti L, Pellegrini B, Cantoni R, Stevens R, Bergounioux C, Cella R, et al. The E2F family of transcription factors from *Arabidopsis thaliana*. Novel and conserved components of the retinoblastoma/E2F pathway in plants. J Biol Chem. 2002;277:9911–9.
48. Gillard J, Devos V, Huysman MJJ, De Veylder L, D'Hondt S, Martens C, et al. Physiological and transcriptomic evidence for a close coupling between chloroplast ontogeny and cell cycle progression in the pennate diatom *Seminavis robusta*. Plant Physiol. 2008;148:1394–411.
49. Yoshida Y. Insights into the mechanisms of chloroplast division. Int J Mol Sci. 2018;19:733.
50. Miyagishima S-Y, Kuwayama H, Urushihara H, Nakanishi H. Evolutionary linkage between eukaryotic cytokinesis and chloroplast division by dynamin proteins. Proc Natl Acad Sci USA. 2008;105:15202–7.
51. Chepurnov VA, Mann DG, Vyverman W, Sabbe K, Danielidis DB. Sexual reproduction, mating system, and protoplast dynamics of *Seminavis* (Bacillariophyceae). J Phycol. 2002;38:1004–19.
52. Miyagishima S-Y. Mechanism of plastid division: from a bacterium to an organelle. Plant Physiol. 2011;155:1533–44.
53. Leger MM, Petrů M, Žárský V, Eme L, Vlček Č, Harding T, et al. An ancestral bacterial division system is widespread in eukaryotic mitochondria. Proc Natl Acad Sci USA. 2015;112:10239–46.
54. Huysman MJJ, Vyverman W, De Veylder L. Molecular regulation of the diatom cell cycle. J Exp Bot. 2014;65:2573–84.
55. Hirano T. Condensins: universal organizers of chromosomes with diverse functions. Genes Dev. 2012;26:1659–78.
56. Hildebrand M, Volcani BE, Gassmann W, Schroeder JI. A gene family of silicon transporters. Nature. 1997;385:688–9.
57. 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;15:65.
58. van de Poll WH, Vrieling EG, Gieskes WWC. Location and expression of frustulins in the pennate diatom *Cylindrotheca fusiformis*, *Navicula pelliculosa* and *Navicula salinarum* (Bacillariophyceae). J Phycol. 1999;35:1044–53.
59. Yee DP, Hildebrand M, Tresguerres M. Dynamic subcellular translocation of V-type H -ATPase is essential for biomineralization of the diatom silica cell wall. N Phytologist. 2019;225:2411–22.
60. Huysman MJJ, Martens C, Vandepoele K, Gillard J, Rayko E, Heijde M, et al. Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. Genome Biol. 2010;11:R17.
61. Wolgemuth DJ, Roberts SS. Regulating mitosis and meiosis in the male germ line: critical functions for cyclins. Philos Trans R Soc B: Biol Sci. 2010;365:1653–62.
62. Carlile TM, Amon A. Meiosis I is established through division-specific translational control of a cyclin. Cell. 2008;133:280–91.
63. Yokoo R, Zawadzki KA, Nabeshima K, Drake M, Arur S, Villeneuve AM. COSA-1 reveals robust homeostasis and separable licensing and reinforcement steps governing meiotic crossovers. Cell. 2012;149:75–87.
64. Bulankova P, Akimcheva S, Fellner N, Riha K. Identification of *Arabidopsis* meiotic cyclins reveals functional diversification among plant cyclin genes. PLoS Genet. 2013;9:e1003508.
65. Stacey NJ, Kuromori T, Azumi Y, Roberts G, Breuer C, Wada T, et al. *Arabidopsis* SPO11-2 functions with SPO11-1 in meiotic recombination. Plant J. 2006;48:206–16.
66. Lange J, Pan J, Cole F, Thelen MP, Jasin M, Keeney S. ATM controls meiotic double-strand-break formation. Nature. 2011;479:237–40.
67. Zou Y, Liu Y, Wu X, Shell SM. Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. J Cell Physiol. 2006;208:267–73.
68. Campbell ID, Humphries MJ. Integrin structure, activation, and interactions. Cold Spring Harb Perspect Biol. 2011;3:a004994.
69. Martins MB, Carvalho I. Diketopiperazines: biological activity and synthesis. Tetrahedron. 2007;63:9923–32.
70. Frenkel J, Wess C, Vyverman W, Pohnert G. Chiral separation of a diketopiperazine pheromone from marine diatoms using supercritical fluid chromatography. J Chromatogr B Anal Technol Biomed Life Sci. 2014;951–2:58–61.
71. Russo MT, Vitale L, Entrambasaguas L, Anestis K, Fattorini N, Romano F, et al. MRP3 is a sex determining gene in the diatom *Pseudo-nitzschia multistriata*. Nat Commun. 2018;9:5050.
72. Anderson A, Bothwell JH, Laohavisit A, Smith AG, Davies JM. NOX or not? Evidence for algal NADPH oxidases. Trends Plant Sci. 2011;16:579–81.
73. Fisher AB. Redox signaling across cell membranes. Antioxid Redox Signal. 2009;11:1349–56.
74. Bienert GP, Chaumont F. Aquaporin-facilitated transmembrane diffusion of hydrogen peroxide. Biochim Biophys Acta. 2014;1840:1596–604.
75. Jiménez-Quesada MJ, Traverso JÁ, Alché J, de D. NADPH oxidase-dependent superoxide production in plant reproductive tissues. Front Plant Sci. 2016;7:359.
76. Bondoc KGV, Lembke C, Vyverman W, Pohnert G. Searching for a mate: pheromone-directed movement of the benthic diatom *Seminavis robusta*. Micro Ecol. 2016;72:287–94.

77. Poulsen NC, Spector I, Spurck TP, Schultz TF, Wetherbee R. Diatom gliding is the result of an actin-myosin motility system. Cell Motil Cytoskeleton. 1999;44:23–33.
78. Lachnit M, Buhmann MT, Klemm J, Kröger N, Poulsen N. Identification of proteins in the adhesive trails of the diatom *Amphora coffeaeformis*. Philos Trans R Soc Lond B Biol Sci. 2019;374:20190196.
79. Deshaies RJ, Joazeiro CAP. RING domain E3 ubiquitin ligases. Annu Rev Biochem. 2009;78:399–434.
80. Bolaños-Villegas P, Xu W, Martínez-García M, Pradillo M, Wang Y. Insights into the role of ubiquitination in meiosis: fertility, adaptation and plant breeding. Arabidopsis Book. 2018;16: e0187.
81. Iyer LM, Anantharaman V, Aravind L. The DOMON domains are involved in heme and sugar recognition. Bioinformatics. 2007;23:2660–4.
82. Gu B, He S, Yan X, Zhang L. Tentative biosynthetic pathways of some microbial diketopiperazines. Appl Microbiol Biotechnol. 2013;97:8439–53.
83. Gondry M, Sauguet L, Belin P, Thai R, Amouroux R, Tellier C, et al. Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. Nat Chem Biol. 2009;5:414–20.
84. Wittenberg C, Sugimoto K, Reed SI. G1-specific cyclins of *S. cerevisiae*: cell cycle periodicity, regulation by mating pheromone, and association with the p34CDC28 protein kinase. Cell. 1990;62:225–37.
85. Merlini L, Dudin O, Martin SG. Mate and fuse: how yeast cells do it. Open Biol. 2013;3:130008.
86. Côte P, Whiteway M. The role of *Candida albicans* FAR1 in regulation of pheromone-mediated mating, gene expression and cell cycle arrest. Mol Microbiol. 2008;68:392–404.
87. Sherwood RK, Scaduto CM, Torres SE, Bennett RJ. Convergent evolution of a fused sexual cycle promotes the haploid lifestyle. Nature. 2014;506:387–90.
88. Sekimoto H, Tanabe Y, Tsuchikane Y, Shirosaki H, Fukuda H, Demura T, et al. Gene expression profiling using cDNA microarray analysis of the sexual reproduction stage of the unicellular charophycean alga *Closterium peracerosum-strigosum-littorale* complex. Plant Physiol. 2006;141:271–9.
89. Kanda N, Ichikawa M, Ono A, Toyoda A, Fujiyama A, Abe J, et al. CRISPR/Cas9-based knockouts reveal that CpRLP1 is a negative regulator of the sex pheromone PR-IP in the *Closterium peracerosum-strigosum-littorale* complex. Sci Rep. 2017;7:17873.
90. Hirano N, Marukawa Y, Abe J, Hashiba S, Ichikawa M, Tanabe Y, et al. A receptor-like kinase, related to cell wall sensor of higher plants, is required for sexual reproduction in the unicellular charophycean alga, *Closterium peracerosum-strigosum-littorale* Complex. Plant Cell Physiol. 2015;56:1456–62.
91. Duan Q, Kita D, Johnson EA, Aggarwal M, Gates L, Wu H-M, et al. Reactive oxygen species mediate pollen tube rupture to release sperm for fertilization in *Arabidopsis*. Nat Commun. 2014;5:3129.
92. Kurusu T, Kuchitsu K. Autophagy, programmed cell death and reactive oxygen species in sexual reproduction in plants. J Plant Res. 2017;130:491–9.
93. Lipinska A, Cormier A, Luthringer R, Peters AF, Corre E, Gachon CMM, et al. Sexual dimorphism and the evolution of sex-biased gene expression in the brown alga *Ectocarpus*. Mol Biol Evol. 2015;32:1581–97.
94. Hisanaga T, Okahashi K, Yamaoka S, Kajiwara T, Nishihama R, Shimamura M, et al. A cis -acting bidirectional transcription switch controls sexual dimorphism in the liverwort. EMBO J. 2019;38:e100240.
95. Makkena S, Lee E, Sack FD, Lamb RS. The R2R3 MYB transcription factors FOUR LIPS and MYB88 regulate female reproductive development. J Exp Bot. 2012;63:5545–58.
96. Borg M, Brownfield L, Khatab H, Sidorova A, Lingaya M, Twell D. The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in *Arabidopsis*. Plant Cell. 2011;23:534–49.
97. Kim Y, Kim H, Son H, Choi GJ, Kim J-C, Lee Y-W. MYT3, a Myb-like transcription factor, affects fungal development and pathogenicity of *Fusarium graminearum*. PLoS ONE. 2014;9: e94359.
98. Miao W, Xiong J, Bowen J, Wang W, Liu Y, Braguinets O, et al. Microarray analyses of gene expression during the *Tetrahymena thermophila* life cycle. PLoS ONE. 2009;4:e4429.