

# SCIENTIFIC REPORTS



OPEN

## A sex-inducing pheromone triggers cell cycle arrest and mate attraction in the diatom *Seminavis robusta*

Received: 25 September 2015

Accepted: 04 December 2015

Published: 20 January 2016

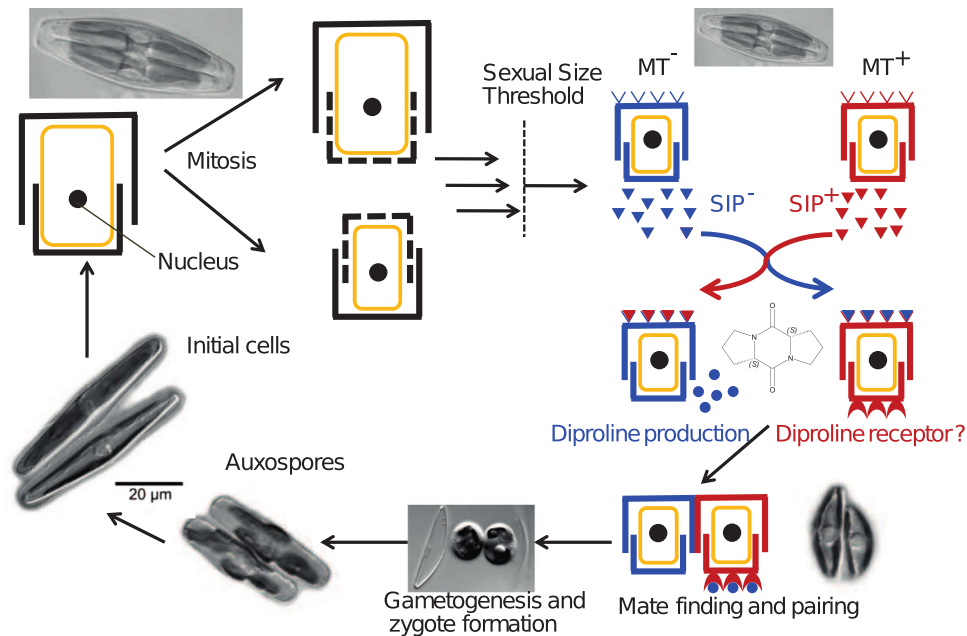
Sara Moeys<sup>1,2,3,\*</sup>, Johannes Frenkel<sup>4,\*</sup>, Christine Lembke<sup>4,\*</sup>, Jeroen T. F. Gillard<sup>1,5</sup>, Valerie Devos<sup>1,2,3</sup>, Koen Van den Berge<sup>6</sup>, Barbara Bouillon<sup>1,2,3</sup>, Marie J. J. Huysman<sup>2,3</sup>, Sam De Decker<sup>1</sup>, Julia Scharf<sup>4</sup>, Atle Bones<sup>7</sup>, Tore Brembu<sup>7</sup>, Per Winge<sup>7</sup>, Koen Sabbe<sup>1</sup>, Marnik Vuylsteke<sup>2,3</sup>, Lieven Clement<sup>6</sup>, Lieven De Veylder<sup>2,3</sup>, Georg Pohnert<sup>4</sup> & Wim Vyverman<sup>1</sup>

Although sexual reproduction is believed to play a major role in the high diversification rates and species richness of diatoms, a mechanistic understanding of diatom life cycle control is virtually lacking. Diatom sexual signalling is controlled by a complex, yet largely unknown, pheromone system. Here, a sex-inducing pheromone (SIP<sup>+</sup>) of the benthic pennate diatom *Seminavis robusta* was identified by comparative metabolomics, subsequently purified, and physicochemically characterized. Transcriptome analysis revealed that SIP<sup>+</sup> triggers the switch from mitosis-to-meiosis in the opposing mating type, coupled with the transcriptional induction of proline biosynthesis genes, and the release of the proline-derived attraction pheromone. The induction of cell cycle arrest by a pheromone, chemically distinct from the one used to attract the opposite mating type, highlights the existence of a sophisticated mechanism to increase chances of mate finding, while keeping the metabolic losses associated with the release of an attraction pheromone to a minimum.

Diatoms are a key ecological and extremely diverse group of microalgae<sup>1</sup> that contribute to about 20 percent of the global carbon fixation<sup>2</sup>. Their unique diplontic life cycle, involving a size-dependent transition from mitotically dividing cells to cells capable of sexual reproduction<sup>3,4</sup>, is a defining feature of diatom biology<sup>5</sup>. This size-dependent transition is connected to their cell wall structure, which consists of two overlapping halves or thecae. After cell division, each daughter cell inherits one theca and synthesizes a new one while still enclosed in the cell wall of the mother cell, leading to a decrease in mean cell size of the population with every cell division (Fig. 1). Restoration of cell size typically occurs by zygote (auxospore) expansion following sexual reproduction. Cells that fail to reproduce sexually will continue to divide mitotically until they reach a critical minimal size, upon which they die. Strikingly, cells are only capable of sexual reproduction once a certain cell size (the sexual size threshold or SST) is reached<sup>3</sup>.

Major clades within the diatoms have evolved a diversity of sexual strategies, including homothallic reproduction within one clone and heterothallic reproduction with each clone belonging to one of two mating types (MT<sup>+</sup> or MT<sup>-</sup>)<sup>4,5</sup>. Although it has been long suspected that diatoms employ chemical cues to facilitate mate-finding and sexual reproduction<sup>4,6-9</sup>, the first molecular structure of a diatom pheromone was only recently revealed for the model species *Seminavis robusta*<sup>10</sup>. In this species, a two-step signalling system has been identified, in which mating-compatible cells below the SST produce sex-inducing pheromones (in the following abbreviated as SIP) that trigger the production of the proline-derived diketopiperazine (in the following abbreviated as diproline)<sup>10</sup> (Fig. 1). Diproline then functions as an attraction pheromone to direct MT<sup>+</sup> cells towards MT<sup>-</sup> cells, followed by pairing of the cells and meiosis. Compared to other microalgae<sup>11-13</sup>, this pheromone is active at relatively

<sup>1</sup>Protistology and Aquatic Ecology, Department of Biology, Ghent University, 9000 Ghent, Belgium. <sup>2</sup>Department of Plant Systems Biology, VIB, B-9052 Ghent, Belgium. <sup>3</sup>Department of Plant Biotechnology and Bioinformatics, Ghent University, B-9052 Ghent, Belgium. <sup>4</sup>Institute for Inorganic and Analytical Chemistry, Bioorganic Analytics, Friedrich Schiller University, 07743 Jena, Germany. <sup>5</sup>Department of Biology, California State University, 9001 Stockdale Highway, Bakersfield, CA 93311, USA. <sup>6</sup>Department of Applied Mathematics, Computer Science and Statistics, Ghent University, 9000 Ghent, Belgium. <sup>7</sup>Department of Biology, Norwegian University of Science and Technology, 7491 Trondheim, Norway. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to L.D.V. (email: Lieven.DeVeylder@psb.vib-ugent.be) or G.P. (email: Georg.Pohnert@uni-jena.de) or W.V. (email: Wim.Vyverman@ugent.be)



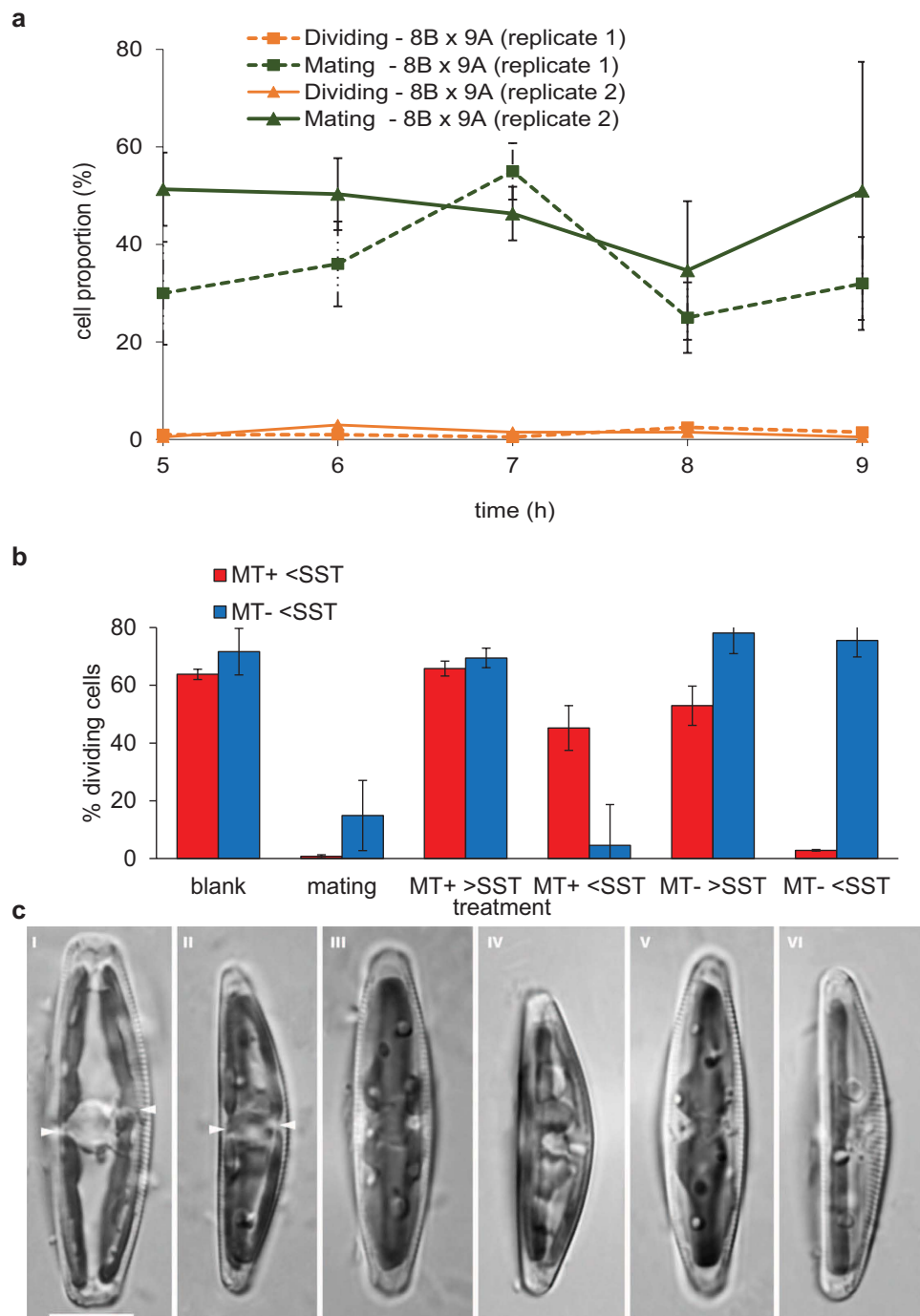
**Figure 1.** The life cycle of *Seminavis robusta*, modified after Frenkel, *et al.*<sup>5</sup> with microscopic images from Chepurinov, *et al.*<sup>24</sup> and Chepurinov, *et al.*<sup>9</sup>. During the vegetative phase, average cell size decreases. Once below a certain cell size, called the sexual size threshold (SST), diatoms become capable of sexual reproduction. When passing this SST, both  $MT^+$  and  $MT^-$  cells start the production of sex-inducing pheromones,  $SIP^+$  and  $SIP^-$  respectively. Under the influence of these pheromones,  $MT^-$  secretes the attraction pheromone diproline and  $MT^+$  probably expresses a diproline receptor. Diproline signalling then leads to mate finding and pair formation, after which gametes and zygotes are formed. The specialized zygotes, called auxospores, elongate to initial cell size.

high concentrations. Here, we used physiological, metabolomic, and transcriptomic methods to provide a deep molecular insight into the two-step signalling system of *S. robusta*. Specifically, we isolated and physicochemically characterized the sex-inducing pheromone produced by  $MT^+$  cells ( $SIP^+$ ) that initiates the biosynthesis and secretion of diproline by  $MT^-$  cells. This isolated  $SIP^+$  was then used in RNA sequencing experiments to study the pheromone signalling on the transcriptional level. We also verified the existence of a chemically distinct SIP that is produced by  $MT^-$  cells ( $SIP^-$ ) and found that both chemically distinct SIPs, besides sexualizing cells, reciprocally arrest the cell cycle at the G<sub>1</sub> phase.

## Results

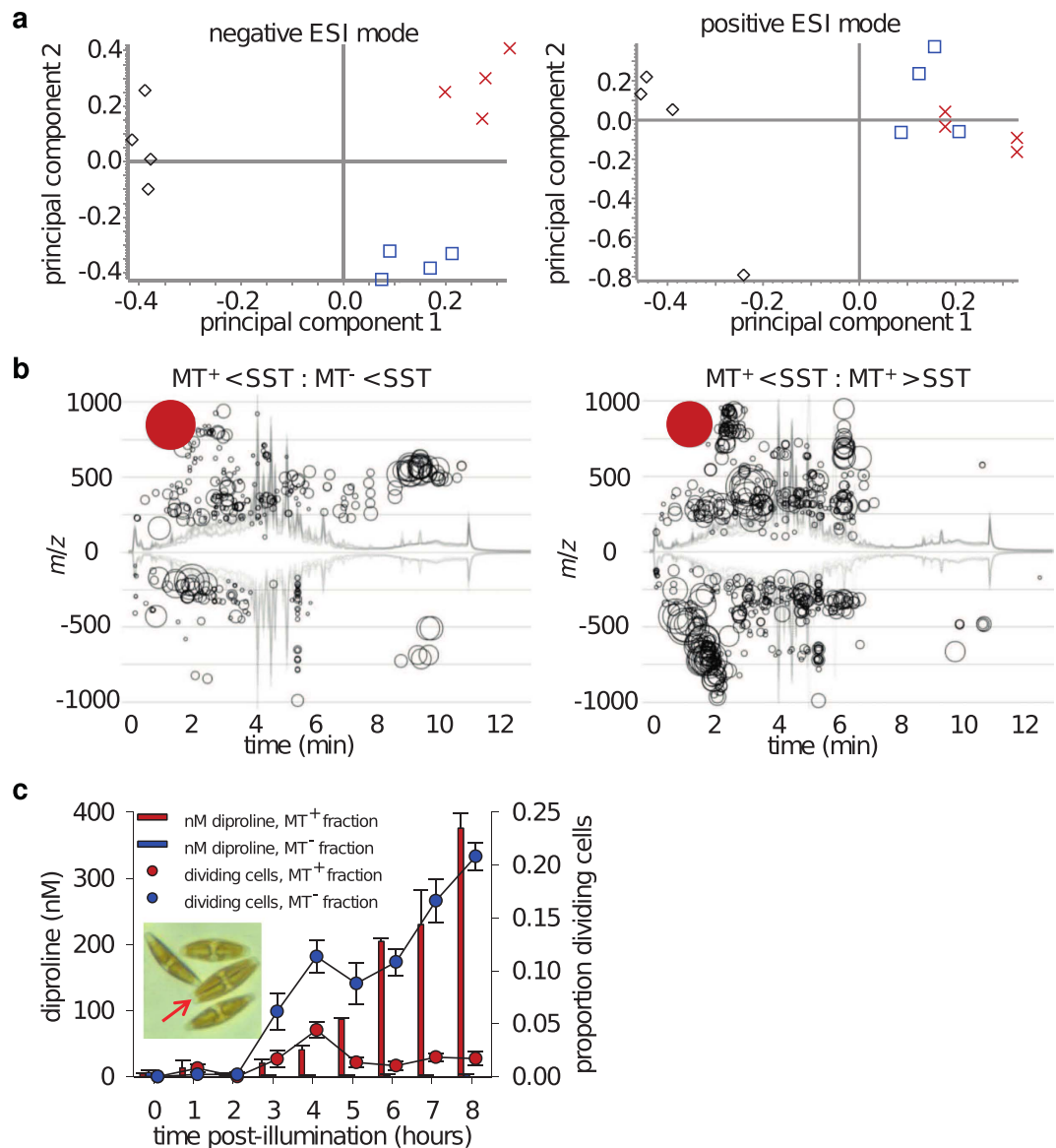
**Cells below the SST produce cytostatic pheromones.** When two strains of opposite mating types with cell sizes below the SST are mixed, mating pairs are formed. In these pairs, gamete formation is initiated (Fig. 1), implying a switch from mitotic cell cycle progression to meiosis. Surprisingly, while only about half of the cells engage in mating pair formation, virtually none of the cells in the co-culture continued mitotic cell division (Fig. 2a). Replacing the culture medium of  $MT^+$  or  $MT^-$  cells with a size below the SST by filtered culture medium of a mating culture (containing all excreted metabolites, but no cells) resulted for both in the induction of cell cycle arrest. Experiments with combinations of opposite mating types showed that two specific signals are involved, as only filtered medium from the  $MT^+$  culture induced cell cycle arrest in the  $MT^-$  culture and vice versa (Fig. 2b). Cells with a size above the SST did not respond to the same treatments, nor were they able to induce cell cycle arrest in other cultures (Supplementary Fig. S1). Microscopic observations showed that growth-arrested cells had undivided chloroplasts at the girdle (Fig. 2c and Supplementary Fig. S2). During vegetative division, chloroplasts divide and translocate from the girdle to the valves during the S/G<sub>2</sub> phase<sup>14</sup>. The location and morphology of the chloroplasts of growth-arrested cells are thus indicative for the G<sub>1</sub> phase. Together, these results suggest that cells below the SST secrete mating type-specific pheromones with the ability to arrest the cell cycle of the cells of the compatible mating type in the G<sub>1</sub> phase.

**Metabolomics-assisted identification of  $SIP^+$ .** In addition to inducing cell cycle arrest, medium of  $MT^+$  cells below the SST activates the production of the attraction pheromone diproline in  $MT^-$  cells<sup>10</sup>. To assess if the signal leading to cell cycle arrest is identical to the compound inducing diproline production, we aimed to purify the putative pheromone(s). Based on the above results, we reasoned that cell size-dependent production of a signal molecule would be reflected by changes in the exo-metabolome (i.e. all metabolites released into the medium)<sup>15</sup>. We therefore performed a comparative exo-metabolomic study of  $MT^+$  versus  $MT^-$  cells below the SST and of  $MT^+$  cells below versus  $MT^+$  cells above the SST to identify unique compounds in the pheromone-emitting cultures. For this approach, suitable extraction procedures for  $SIP^+$  were developed using



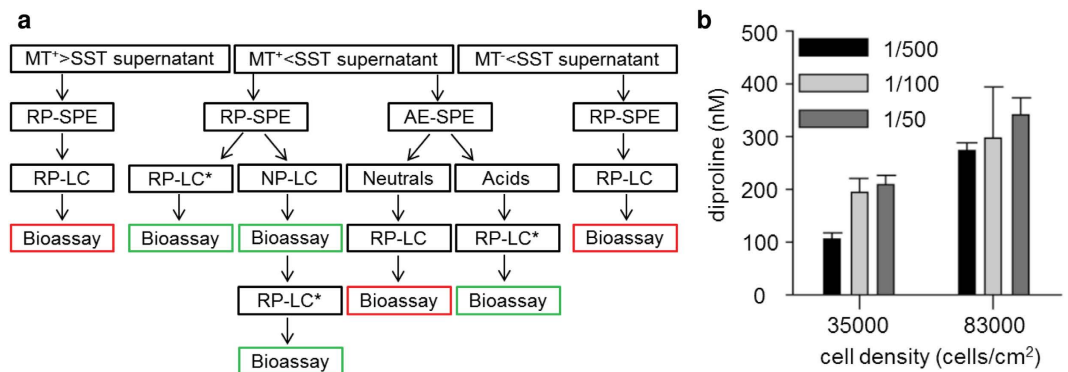
**Figure 2. Cells below the SST produce a cytostatic pheromone.** (a) The proportion of dividing and mating cells was determined by counting the different cell types under an inverted microscope in two mixed cultures of MT<sup>+</sup> (8B) × MT<sup>-</sup> (9A) strains after dark-synchronization. A minimum of 200 instances was evaluated for every time point. Error bars are ± s.e. of 3 replicate estimations. (b) Means ± s.e. (n = 3) of proportions of mitotically dividing MT<sup>+</sup> (red) or MT<sup>-</sup> (blue) cells below the SST, 9 h after illumination, depending on the administration of fresh medium (blank), medium from a mating culture or medium from the opposite mating partner (above or below the SST). (c) Light microscopy images of MT<sup>-</sup> (85A) cells treated for 8 h with control medium (i–ii), with filtered medium of MT<sup>+</sup> (85B) cells below the SST (iii–iv), or with purified SIP<sup>+</sup> (v–vi). Cells are shown in the girdle (i, iii, v) or valve (ii, iv, vi) view. Arrowheads in i and ii point to constricting plastids. Scale bar = 10 μm.

bioassays to confirm that the extract could induce diproline production in MT<sup>-</sup> and therefore contained the pheromone. Positive bioassay results were observed when the medium was extracted using reversed-phase solid-phase extraction (RP-SPE) cartridges. Similar to the filtered culture medium of MT<sup>+</sup> cells below the SST, these partially



**Figure 3. Metabolomics-Assisted Identification of SIP<sup>+</sup>.** (a) Principal component analysis of Liquid-chromatography (LC)-mass spectrometry data in positive and negative ESI mode from extracts ( $n = 4$ ) of MT<sup>+</sup> cells above the SST (black), MT<sup>+</sup> cells below the SST (red) and MT<sup>-</sup> cells below the SST (blue). (b) Screening of up- (positive Y-axis) and down- (negative Y-axis) regulated metabolites in MT<sup>+</sup> cells below the SST compared with MT<sup>-</sup> cells below the SST (left) and MT<sup>+</sup> cells above the SST (right) in the chromatograms acquired in negative ESI mode ( $n = 4$ ). The radius of the circle shows the relative degree of regulation of the signal with the indicated mass at the respective retention time in the chromatogram. The red circle corresponds to 600x (left) or 320x (right) upregulation in comparison of the treatments and indicates a signal with  $m/z = 842$ . (c) Means  $\pm$  s.e. ( $n = 3$ ) of diproline concentration in the medium and dividing MT<sup>-</sup> cells below the SST treated with purified fractions of MT<sup>+</sup> below the SST containing SIP<sup>+</sup> (red) and fractions of equally purified fractions of MT<sup>-</sup> below the SST (blue), acting as control. The red arrow shows a dividing cell.

purified extracts induced diproline production in MT<sup>-</sup>, indicating that SIP<sup>+</sup> is extractable with the applied technique. Subsequently, the extracts were submitted to reversed-phase ultra-performance liquid chromatography/mass spectrometry (RP-UPLC/MS) and measured in positive and negative electrospray-ionization (ESI) mode. The resulting chromatograms were highly complex, covering diatom metabolites and medium ingredients, as well as contaminants from solvents and plasticizers. In order to identify metabolites specific for MT<sup>+</sup> cells below the SST, data were resolved with principal component analysis. The replicates of the exo-metabolomes measured in negative mode grouped according to their respective cell size and mating type; data recorded in positive ionization mode revealed fewer differences (Fig. 3a). Chromatograms measured in negative mode were screened for metabolites that varied in their concentration using XCMS online, a platform for processing untargeted metabolomics data<sup>16</sup>. A metabolite having a pseudomolecular ion of  $[M-H]^- = 842$  was identified with the most pronounced upregulation in MT<sup>+</sup> cells below the SST compared with MT<sup>-</sup> cells below the SST (Fig. 3b). It also



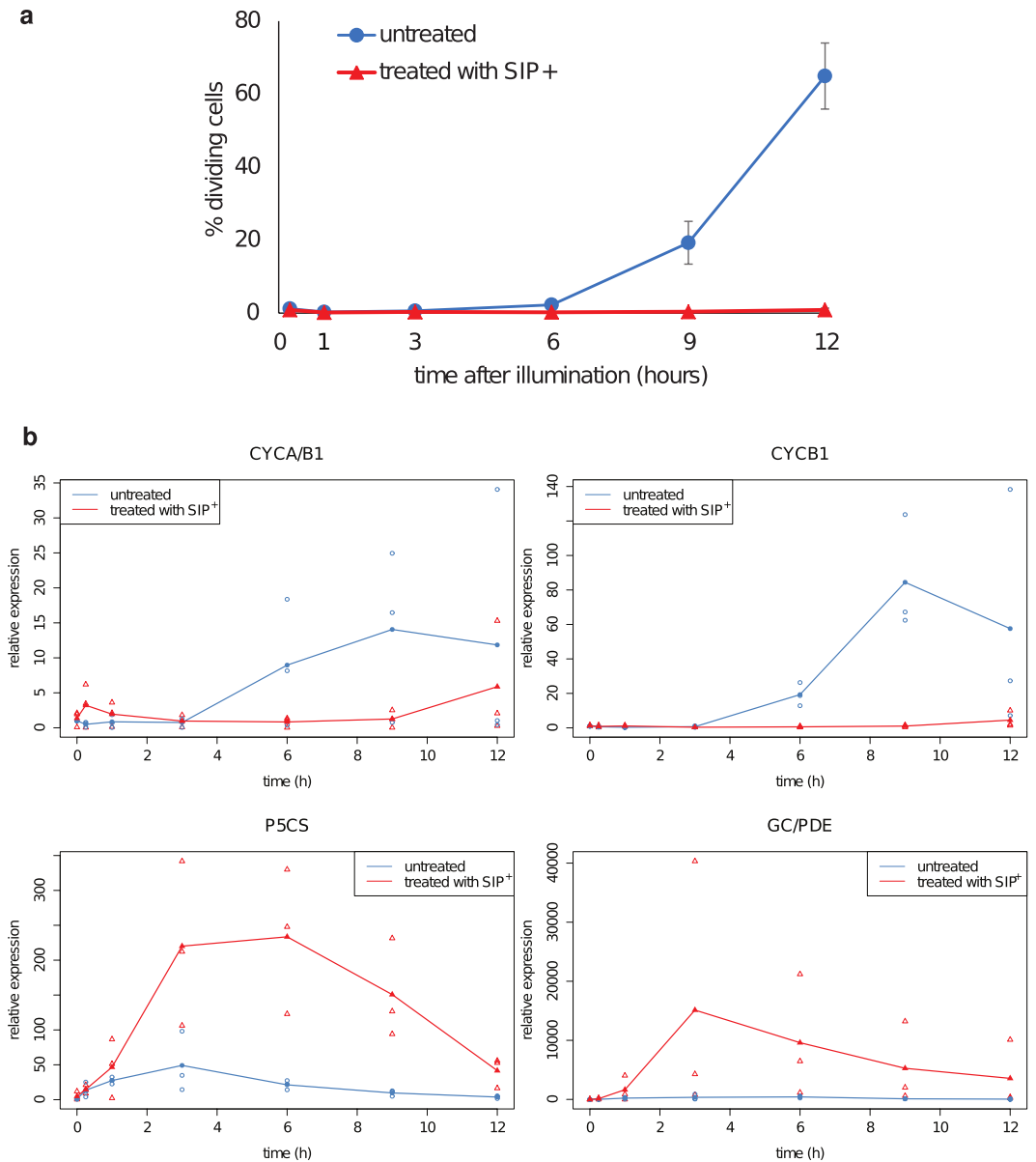
**Figure 4. Further purification of SIP<sup>+</sup>.** (a) Orthogonal purification of this metabolite through a combination of reversed-phase solid-phase extraction (RP-SPE), anion exchange (AE)-SPE, RP-liquid chromatography (LC) and normal-phase (NP)-LC. The algae supernatant, SPE-extract or LC-fraction were applied in diproline-induction assays. A positive result was only observed when the signal with  $m/z = 842$  was detected (green), whereas there was no bioactivity when the metabolite could not be found (red). \*RP-LC was accomplished with two different gradients. (b) Means  $\pm$  s.e. ( $n = 3$ ) of diproline concentration in the medium of MT<sup>-</sup> cells below the SST after treatment with different amounts of SIP<sup>+</sup>.

displayed the second highest upregulation in MT<sup>+</sup> cells below the SST compared with MT<sup>+</sup> cells above the SST (Fig. 3b). This metabolite showed a rather high polarity eluting from a C18 RP-UPLC column at a solvent composition of 15% acetonitrile in water. A UPLC fraction containing this purified metabolite indeed induced cell cycle arrest and the production of the attraction pheromone diproline in MT<sup>-</sup> (Fig. 3c), thereby supporting the compound as a candidate for SIP<sup>+</sup>.

**Further purification and partial characterization confirmed the identity of SIP<sup>+</sup>.** In order to rule out the possibility that co-eluting metabolites in the fraction containing the SIP<sup>+</sup> candidate cause the attraction pheromone induction in MT<sup>-</sup> cells, the metabolite was further purified, guided by a targeted UPLC-MS analysis (Fig. 4a). First, to achieve a higher selectivity during the extraction process, strong anion-exchange SPE (AE-SPE) was used instead of RP-SPE to enrich acidic metabolites from the culture medium. Indeed, the extract, and not the flow through, induced diproline production in MT<sup>-</sup> cultures. Accordingly, only the AE-SPE extract contained the SIP<sup>+</sup> candidate with a pseudomolecular mass of  $[M-H]^- = 842$ . To exclude potential artefacts during purification, a flatter gradient was developed for elution; again, only the fraction containing the SIP<sup>+</sup> candidate was bioactive (Fig. 4a), confirming that the metabolite with  $[M-H]^- = 842$  is the diproline-inducing SIP<sup>+</sup>. For further characterization of the metabolites, high-resolution (HR) ESI-MS/MS experiments were conducted. Analysis of the isotope pattern corroborated the presence of at least one sulphur atom in the molecular formula and fragmentation supported the presence of a sulphate functionality in the metabolite (Supplementary Fig. S3). Due to the very low amounts of released SIP<sup>+</sup>, further attempts for structure elucidation after scale-up proved to be unsuccessful. Whereas purification of one litre of MT<sup>-</sup> culture was sufficient to extract 100  $\mu$ g diproline, not even 20 litres of MT<sup>+</sup> medium could yield any weighable amounts of SIP<sup>+</sup>. Considering the low amount of SIP<sup>+</sup> in the medium, one must conclude that its activity is remarkably high. The medium of a culture of 35,000 cells/cm<sup>2</sup> could be diluted 500-fold before a dose-dependent reduction in diproline production in induced MT<sup>+</sup> cells could be observed (Fig. 4b).

**SIP<sup>+</sup> is responsible for the mitosis-to-meiosis switch in MT<sup>-</sup> cells.** The cell cycle arrest induced by SIP<sup>+</sup> was verified by expression profiling of cell cycle marker genes. Cultures of MT<sup>-</sup> cells below the SST were grown for three days in a 12-h:12-h light:dark regime, after which the cell cycle was synchronized by extending the dark period for twelve additional hours. After this prolonged dark period, the majority of the cells were arrested in the G<sub>1</sub> phase of the cell cycle<sup>14</sup>. Half of the cultures were treated with SIP<sup>+</sup> before illumination. The expression levels of two mitosis markers were measured by real-time quantitative PCR (RT-qPCR) to investigate the regulation of cell cycle arrest. The *CYCB1* and *CYCA/B1* markers were selected because they are highly expressed during mitosis in the diatom *Phaeodactylum tricornutum*<sup>17</sup>. *CYCB1* and *CYCA/B1* were indeed expressed during mitosis in untreated *S. robusta* cells. In contrast, their expression was repressed when the cells were treated with SIP<sup>+</sup>, which corroborates the observed cell cycle arrest (Fig. 5a,b).

In order to obtain a more global view on the transcriptional changes induced by SIP<sup>+</sup>, an RNA-seq experiment was conducted. Therefore, purified SIP<sup>+</sup> from the RP-UPLC was added to MT<sup>-</sup> cultures, 15 min before light onset after a prolonged dark period. MT<sup>-</sup> cultures without extract addition constituted the control conditions. Samples were taken after 15 min, 1 h and 3 h of illumination. An additional, non-treated MT<sup>-</sup> culture was harvested before illumination. RNA extracted from these samples was sequenced and resulting reads were assembled in a transcriptome containing 50,231 transcripts (Supplementary Table S1). Comparing gene expression data of treated and untreated samples for each time point showed 3,461 unique transcripts being differentially expressed in at least one time point (Supplementary Table S2). The majority of these differentially expressed transcripts was observed at 3 h (2,697 transcripts), while 192 and 544 transcripts were differentially expressed at 1 h and 15 min, respectively (Fig. 6a). Furthermore, a significant disparity in fold change (FC) was observed for 289, 1,308

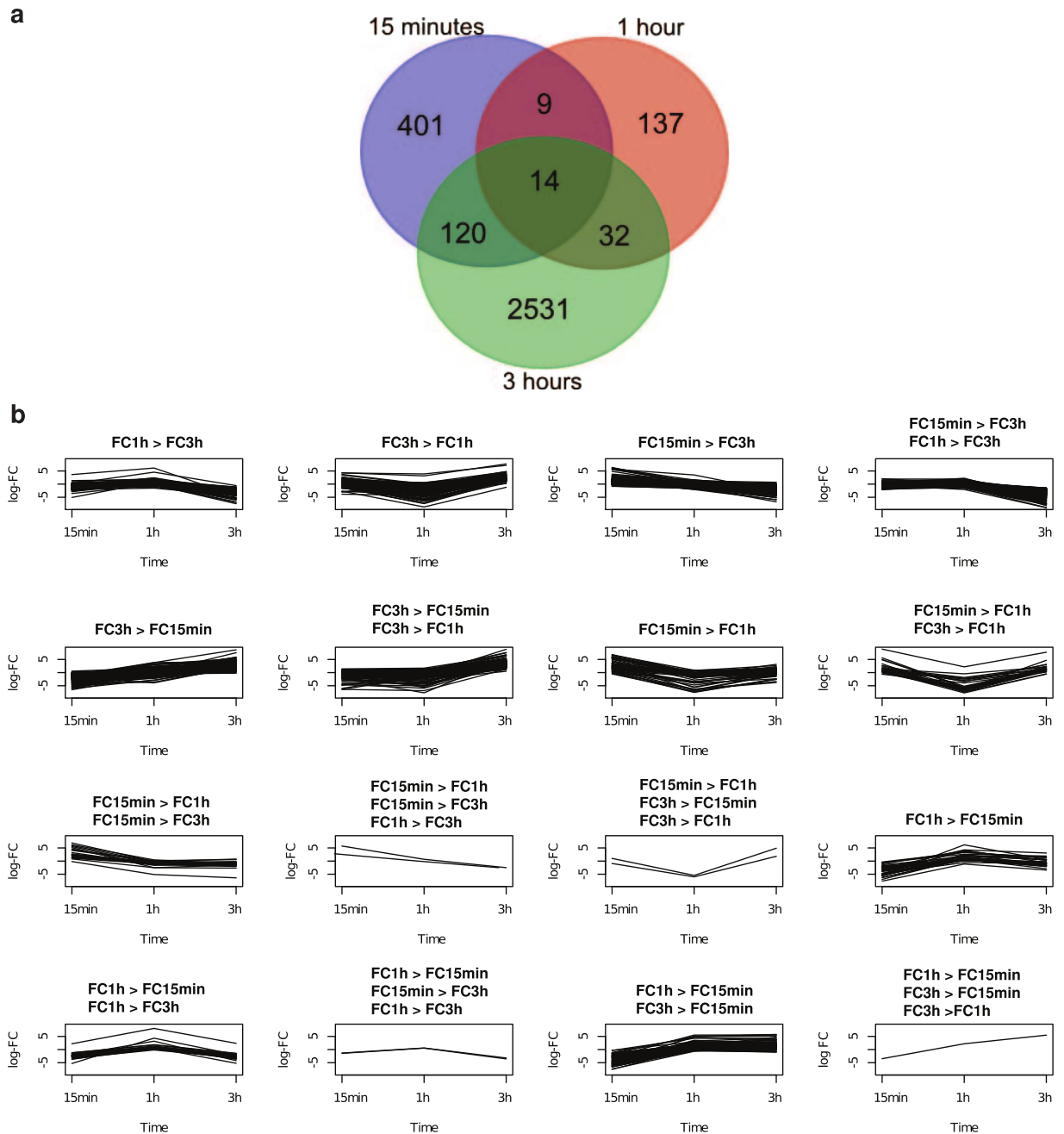


**Figure 5. Transcriptional effects of SIP<sup>+</sup> on MT<sup>-</sup> cells below the SST determined by qRT-PCR.**

(a) Mean  $\pm$  s.e. (n = 3) of the percentage of dividing MT<sup>-</sup> cells after treatment with SIP<sup>+</sup> (red) or untreated (blue). (b) Mean (n = 3) relative expression of *CYCA/B1*, *CYCB1*, *P5CS* and *GC/PDE* in SIP<sup>+</sup>-treated (red) and untreated (blue) MT<sup>-</sup> cultures. Biological replicates (n = 3) are shown as separate data points. Relative expression values are normalized to two constitutively expressed genes.

and 1,263 transcripts in the 1 h versus 15 min, 3 h versus 15 min and 3 h versus 1 h comparisons, respectively. These results were used to classify the differentially expressed transcripts according to the significant FC patterns (Fig. 6b).

Amongst the differentially expressed genes after treatment with SIP<sup>+</sup>, four meiosis-related genes were identified, all being significantly higher expressed in treated cultures compared with untreated ones after 3 h of illumination (Supplementary Fig. S4). Although some show already small fold changes at 15 min and 1 h, the fold change is most outspoken at 3 h (ranging from 4.7 to 13.4). MRE11 and RAD50 are part of the Mre11-Rad50-Nbs1 complex that is involved in repairing DNA double-strand breaks and homologous recombination during meiosis<sup>18,19</sup>. MSH4 is a mismatch repair protein that functions specifically in meiosis<sup>20</sup>. MCM8 is involved in meiotic recombination<sup>21,22</sup>. Although not much is known about the exact role of these four proteins in diatoms, there are indications that they are indeed involved in meiosis<sup>23</sup>. Along with the observed cell cycle arrest, these data indicate that SIP<sup>+</sup> sensing by MT<sup>-</sup> cells below the SST results in a switch from mitosis-related genes to meiosis-related genes, although meiosis itself will only start when mating pairs have been formed and gametogenesis can start<sup>24</sup>.



**Figure 6. RNA-seq experiment of  $SIP^+$ -treated and untreated  $MT^-$  cultures.** (a) Venn diagram showing the number genes that are differentially expressed between  $SIP^+$ -treated and untreated cultures at 15 min, 1 h and 3 h after illumination. (b) Transcripts for which the fold change (treated versus untreated cells) changed over time, divided in classes based on the direction of the difference in logFC between the different time points. FC15 min = log fold change at 15 min, FC1 h = log fold change at 1 h, FC3 h = log fold change at 3 h.

### The glutamate-to-proline pathway is transcriptionally upregulated in response to $SIP^+$ treatment.

Four hours after illumination and conditioning with  $SIP^+$  extracts, diproline was detected in the medium of the dark-synchronized  $MT^-$  cultures (Fig. 3c). Thus, we expected the genes responsible for diproline biosynthesis to be upregulated in the transcriptome after administration of  $SIP^+$ . Correspondingly,  $\Delta 1$ -pyrroline-5-carboxylate synthetase (*P5CS*; asmb1\_5165 Table S2) was time-dependently upregulated in treated cultures (Fig. 6b, subset FC3h > FC15min, FC3h > FC1h), with a strong increase in expression at 3 h. This enzyme has both  $\gamma$ -glutamyl kinase and glutamic- $\gamma$ -semialdehyde dehydrogenase activity and mediates the first two steps in the conversion of glutamate to proline<sup>25</sup>. Consequently, this enzyme could catalyse the first steps in diproline biosynthesis. The second enzyme of the pathway,  $\Delta 1$ -pyrroline-5-carboxylate reductase (*P5C*; asmb1\_35385 Table S2), was also significantly higher expressed in treated cultures compared with untreated ones 3 h after illumination. Because diproline is detected in treated  $MT^-$  cultures from 4 up to 8 h after

illumination (Fig. 3c), the expression of *P5CS* was monitored with a higher resolution over 12 h by RT-qPCR, showing that *P5CS* expression is already slightly induced 1 h after illumination and peaks at 3–6 h in the treated cultures (Fig. 5b). This time profile corresponded with the timing of mate pairing in *S. robusta* (5–10 h after illumination)<sup>10</sup>.

**The expression of a bifunctional guanylyl cyclase/phosphodiesterase increases upon treatment with SIP<sup>+</sup>.** The only transcript for which the FC in expression in treated cultures at each time point was statistically higher than at the previous time point (class FC3h > FC1h, FC3h > FC15min, FC1h > FC15min) (Fig. 6b), encodes a putative bifunctional protein containing an adenylyl/guanylyl cyclase domain and a cyclic nucleotide phosphodiesterase (PDE) domain (asmb1\_30773 Table S2). These domains are responsible for cyclic nucleotide (cAMP or cGMP) synthesis and breakdown, respectively, pointing to the potential involvement of one of these secondary messengers in pheromone signalling in *S. robusta*. The catalytic core of the cyclase domain was aligned to that of several adenylyl and guanylyl cyclases that were previously used to model the catalytic mechanism of nucleotide cyclases (NCs)<sup>26</sup>. Of the guanine binding residues, all but two are conserved and except for the pyrophosphate-binding site, all catalytic sites are conserved (Supplementary Fig. S5). Richter *et al.*<sup>27</sup> identified eight amino acids that are conserved in cAMP-specific PDEs but not in cGMP-specific PDEs. These amino acids are not conserved in the PDE domain of this *S. robusta* protein (Supplementary Fig. S5), indicating that it probably has cGMP-hydrolyzing capacity.

Guanylyl cyclases can be either transmembrane or soluble proteins. Prediction of the transmembrane topology of the bifunctional guanylyl cyclase (GC)/PDE<sup>28</sup> indicated two predicted transmembrane domains N-terminal from the two catalytic domains, suggesting that this probably is a transmembrane guanylyl cyclase. The transcription of the *GC/PDE* gene was followed during a longer time course by RT-qPCR, which indicated that its expression peaks at three hours after supplementing with SIP<sup>+</sup> (Fig. 5b).

## Discussion

Using a comparative metabolomics approach, a pheromone, called SIP<sup>+</sup>, that is exclusively produced by MT<sup>+</sup> cells below the SST was purified. SIP<sup>+</sup> triggers both cell cycle arrest and diproline production in MT<sup>-</sup> cells. Despite the minute amount of produced pheromone, metabolomics-assisted identification of a SIP<sup>+</sup> candidate, in combination with orthogonal separation protocols could unambiguously identify a metabolite with a pseudomolecular ion of [M-H]<sup>-</sup> = 842 in negative ESI-MS. HR-MS and fragmentation analysis identified this compound as a sulfated metabolite. Interestingly, *S. robusta* uses the cyclic dipeptide diproline as an attraction pheromone, but the mass spectral data suggest that SIP<sup>+</sup> is structurally unrelated. SIP<sup>+</sup> is more polar, potentially sulphated and of significantly higher molecular weight compared to diproline. In general, the pheromone chemistry of (micro) algae is highly diverse and covers a vast range of polarity and molecular weight<sup>5</sup>. Our identification of SIP<sup>+</sup> indicates that diatoms invoke most likely multiple metabolic pathways in producing chemically distinct metabolites as pheromones, even within a single species. Strikingly, SIP<sup>+</sup> appears to possess extreme high activity. Likewise, some known pheromones in aquatic systems are active at very low concentrations. Lurlicin acid, which is produced by MT<sup>-</sup> cells of *Chlamydomonas allensworthii* to attract MT<sup>+</sup> cells, is active at picomolar concentrations<sup>11</sup>, whereas in *Volvox* species pheromones are known to induce gametogenesis at even lower concentrations (from 10<sup>-12</sup> to even 10<sup>-16</sup>)<sup>12,29</sup>.

As in yeast, mammals and plants, commitment to meiosis in *S. robusta* is made during the G<sub>1</sub> phase, before the onset of the S phase<sup>30</sup>, as evident from the chloroplast behaviour<sup>14,24</sup>. Similarly, in the centric diatom *Thalassiosira weissflogii*, cells can only be induced to undergo spermatogenesis when they are in the early G<sub>1</sub> phase<sup>31</sup>. In accordance with these observations, we showed that SIP<sup>+</sup> induces cell cycle arrest in the G<sub>1</sub> phase in *S. robusta*. Likewise, MT<sup>-</sup> cells induce cell cycle arrest in MT<sup>+</sup> cells via the production of a so far unidentified SIP<sup>-</sup> pheromone. This reciprocal pheromone signalling is reminiscent of the system known in yeast where pheromones induce cell cycle arrest in G<sub>1</sub>-phase preceding mating<sup>32</sup>. Whereas in yeast, cell cycle arrest precedes directional cell growth leading to fusion, cell cycle arrest in *S. robusta* accompanies the secretion of the attraction pheromone diproline. This two-step signalling thus effectively increases chances of finding a mating partner within the right time frame and physiological conditions, which appears particularly relevant in diffusion-controlled environments, such as biofilms. It allows signalling compounds to operate on a local scale and develop concentration gradients on spatial scales relevant to those covered by cells gliding through a structurally complex matrix.

Transcriptome analysis revealed that SIP<sup>+</sup> induced a significant modulation of gene expression patterns. About 6% of the transcripts were differentially expressed between controls and SIP<sup>+</sup>-treated cells, providing a first insight into the molecular pathways involved. Notably, the genes  $\Delta$ 1-pyrroline-5-carboxylate synthetase and  $\Delta$ 1-pyrroline-5-carboxylate reductase were upregulated in response to SIP<sup>+</sup>. These genes encode two enzymes from the pathway responsible for the conversion of glutamate to proline, suggesting that the attraction pheromone diproline is synthesized from an increasing intracellular pool of proline. However, candidate enzyme(s) implicated in the last steps (proline to the cyclic dipeptide (CDP) diproline) could not yet be identified. CDPs are widespread in nature and are predominantly synthesized by microorganisms. They are thought to be involved in cell-to-cell communication, for instance in bacterial quorum sensing or in bacteria-plant interactions<sup>33,34</sup>. Two distinct protein families are known to be involved in CDP biosynthesis, namely non-ribosomal peptide synthetases (NRPSs) and CDP synthetases (CDPSs) (31). While CDPSs are highly specific for substrates with only the canonical amino acids, NRPSs can use a wider range of substrates. Several putative NRPS homologues were expressed in the *S. robusta* transcriptome, but none of them was upregulated upon treatment with SIP<sup>+</sup>. Nevertheless, it is possible that these genes are constitutively expressed and that their activity is regulated at the posttranscriptional level or by substrate availability, which would be in accordance with the results described above. Further, although no CDPS homologue could be identified using PSI-blast, it cannot be excluded that



an enzyme with a similar function is present, because sequence similarity between CDPS orthologues is only 19–27%<sup>35</sup>.

Interestingly, a bifunctional guanylyl cyclase/phosphodiesterase (GC/PDE) was found for which the expression difference between treated and untreated cells strongly increased over time. cGMP is a secondary messenger synthesized from GTP by GCs and is broken down to GMP by PDEs. GCs can be soluble or membrane-bound. In animals, soluble GCs are activated by nitric oxide, while membrane-bound GCs are activated by ligands that bind to their extracellular domain<sup>36</sup>. The SIP<sup>+</sup>-responsive GC/PDE is predicted to have two transmembrane domains, and both the putative GC and PDE catalytic domains are positioned at the inside of the cell. We hypothesize that this enzyme is activated by a ligand (most likely SIP<sup>+</sup>) that binds to its extracellular domain. Mammalian membrane-bound GCs have one transmembrane domain with the catalytic domain on the inside of the cell and a ligand-binding domain on the outside, but other conformations have been found in other eukaryotes<sup>37</sup>. The main targets of cGMP are cGMP-dependent protein kinase, cGMP-gated ion channels, and proteins containing a GAF-domain. cGMP is known to regulate many processes in mammals, including smooth muscle relaxation and phototransduction<sup>38,39</sup>. cGMP signalling is often involved in motility, for example in sperm chemotaxis in the sea urchin *Arbacia punctulata* or myosin II regulation in *Dictyostelium*<sup>40,41</sup>. Up till now, the function of cGMP signalling in diatoms remains unclear. The fact that the GC and PDE domains are part of a fusion protein could indicate a tight regulation of cGMP levels. To our knowledge, this combination of domains is only found in diatoms. We hypothesize that SIP<sup>+</sup> is sensed by either the extracellular domain of GC/PDE or by an activator of this protein, leading to cGMP synthesis, which can then activate the downstream signalling cascade, probably resulting in cell cycle arrest and diproline production.

In conclusion, the data presented here combined with the hypothesized existence of a yet unidentified SIP<sup>-</sup> that is produced by MT<sup>-</sup> cells and putatively induces or activates the diproline receptor in MT<sup>+</sup>, indicate that *S. robusta* has developed a rather complex reciprocal signalling system that controls the initial steps of sexual reproduction (Fig. 1). The control of diproline production and perception by two sex-inducing pheromones seems to be a high investment compared with a continuous secretion of diproline by *S. robusta* once cell size drops below the SST. However, SIP<sup>+</sup> and putatively SIP<sup>-</sup> are secreted in much lower concentrations than diproline. These pheromones only transmit the information of the presence of the algae and not their location. Only when the opposite sex perceives the presence of its mating partner, it starts the potentially more expensive production of the N-rich diproline in elevated concentrations as well as its receptor. Consequently, metabolic losses of resources are minimized since their release into the environment is limited to a very short time frame.

## Methods

**Strains and culture conditions.** *Seminavis robusta* strains 8B and 85A (MT<sup>+</sup>) and 9A and 85B (MT<sup>-</sup>) were obtained from the diatom culture collection of the Belgian Coordinated Collection of Micro-organisms (BCCM/DCG, <http://bccm.belspo.be>, accession numbers DCG 0097, DCG 0105, DCG 0096 and DCG 0107). Cultures were either grown in F/2 medium made with autoclaved, filtered natural sea water collected from the North Sea and Guillard's F/2 solution (Sigma-Aldrich)<sup>42</sup> or in artificial sea water<sup>43</sup>.

**Metabolomic screening by RP-UPLC/MS analysis.** Three strains of *S. robusta* (cell size: MT<sup>+</sup> cells below the SST = 22.4 μm, MT<sup>-</sup> cells below the SST = 23.5 μm, MT<sup>+</sup> cells above the SST = 51.9 μm; four replicates each) were grown in 200 ml medium until the culture reached a density of around 100,000 cells/cm<sup>2</sup> in 185-cm<sup>2</sup> flasks. The cultures were filtered five hours after illumination and the filtrate was extracted on 200 mg HLB-SPE cartridges (Oasis<sup>®</sup>, Waters, Eschborn, Germany), following the manufacturer's instructions. The eluate was evaporated to dryness under a stream of nitrogen and dissolved in 100 μl bidistilled water. After centrifugation, 20 μl of the supernatant was analysed with an UPLC system coupled to a time-of-flight mass spectrometer with electrospray ionization (ESI-MS) (Acquity UPLC<sup>®</sup>/Q-ToF micro Waters, Manchester, UK) with an Acquity UPLC BEH C<sub>18</sub> column (2.1 × 50 mm, 1.7 μm, Waters). The composition of the mobile phase was set to 100% A (0.1% formic acid and 2% acetonitrile in water) and ramped to 100% B (0.1% formic acid in acetonitrile) in a linear gradient within 10 min with a flow of 0.6 ml/min. The solvent composition was held at 100% B for 1 min, returned to 100% A in 0.1 min and was kept at 100% A for 2.4 min. The scanned mass range was between 50 and 1000 *m/z* with a scan rate of 0.9 scans sec<sup>-1</sup> and an inter-scan delay of 0.1 sec. The sample cone was set to 25 V, the cone gas flow was held at 20 l N<sub>2</sub> h<sup>-1</sup> and the desolvation gas flow at 700 l N<sub>2</sub> h<sup>-1</sup>. Principal component analysis was done using an established method by means of the Software MarkerLynx (Waters, Milford, USA)<sup>44</sup>. The screening for up-regulated metabolites was performed with XCMS Online (Scripps Center for Metabolomics, La Jolla, USA)<sup>16</sup>. SIP<sup>+</sup> was fractionated at 1.4 min (15% B) and added to a dark-synchronized<sup>14</sup> MT<sup>-</sup> culture (cell size: 23 μm, three replicates per fraction), which was grown in 20 ml medium in 25-cm<sup>2</sup> flasks and had a cell density of around 50,000 cells/cm<sup>2</sup>. The diproline concentration in the medium was checked eight hours after illumination following an established method by extracting the medium with HLB-SPE cartridges (Oasis<sup>®</sup>, Waters, Eschborn, Germany) and investigating the extract using gas chromatography coupled to mass spectrometry on an ISQ Trace GC-Ultra GC-MS system (Thermo Fisher, Dreieich, Germany)<sup>10</sup>.

**Purification of SIP<sup>+</sup>.** Besides RP-SPE with HLB-SPE cartridges, alternative extraction of SIP<sup>+</sup> from the culture medium was done with AE-SPE, using mixed mode MAX SPE-cartridges (Oasis<sup>®</sup>, Waters, Eschborn, Germany) following the manufacturer instructions. For optimization of the separation of SIP<sup>+</sup> in RP-UPLC/MS, all factors in the UPLC-MS analysis were kept unchanged compared to the above protocol, except the solvent gradient, which was ramped from 0% to 10% B in 8 min, and fractions containing SIP<sup>+</sup> were gathered at 4.6 min (6% B). For NP-HPLC/MS, the analytical column was changed to a SeQuant<sup>®</sup> ZIC<sup>®</sup>-HILIC HPLC column (150 × 2.1 mm, 5 μm, Merck). The composition of the mobile phase was kept at 100% B (90% acetonitrile with

5 mM NH<sub>4</sub>Ac) for 1 min and ramped to 80% A (0.1% formic acid and 2% acetonitrile in water) within 5.5 min, returned to 100% B within 0.6 min and was held for 2.9 min. The pheromone was fractionated at 3.9 min (42% A). All gathered fractions were tested for their diproline-inducing capability in MT<sup>-</sup> as mentioned above.

**Mass spectrometric characterization of SIP<sup>+</sup>.** For structural characterization of SIP<sup>+</sup>, 10 µl of MT<sup>+</sup> medium extract was analysed by HPLC (Dionex UltiMate<sup>®</sup> 3000, Thermo Fisher Scientific, Dreieich, Germany), coupled to an ESI-Orbitrap MS (Q-Exactive Plus, Thermo Fisher Scientific, Dreieich, Germany).

Liquid chromatography was done with a Kinetex<sup>®</sup> C18 column (2.1 × 50 mm, 1.7 µm, Phenomenex, Torrance, CA, USA). The composition of the mobile phase was set to 100% A (0.1% formic acid and 2% acetonitrile in water) for 0.2 min and ramped to 100% B (0.1% formic acid in acetonitrile) in a linear gradient within 7.8 min. The solvent composition was held at 100% B for 1 min, returned to 100% A in 0.1 min and held at 100% A for 0.9 min. The flow rate was adjusted to 0.4 ml min<sup>-1</sup>.

Ionization was performed with a spray voltage of 3.3 kV and a capillary temperature of 360 °C. The sheath gas flow rate was kept at 60 arbitrary units. Nitrogen was used as desolvation gas. The scanned mass range was between  $m/z$  100 and 1,500, at  $m/\Delta m$  280,000 resolving power for high-resolution mass spectrometry and at  $m/\Delta m$  17,500 resolving power for MS<sup>2</sup> in negative ion mode. MS<sup>2</sup> was done by fragmenting the precursor pseudomolecular ion  $[M-H]^- = 842.20 m/z$  with a collision energy of 20 eV.

**Cell cycle arrest.** Triplicate MT<sup>+</sup> and MT<sup>-</sup> cultures with cell sizes above and below the SST were grown separately in 12-well tissue culture plates and G1-phase synchronized using the dark-arrest method described by Gillard *et al.*<sup>14</sup>. Then, their supernatant medium was replaced with filtered medium from either MT<sup>+</sup> cultures (with a size above or below the SST), MT<sup>-</sup> cultures (with a size above or below the SST), a mixed mating culture (MT<sup>+</sup> × MT<sup>-</sup>) or fresh culture medium. The cultures were then kept in the dark for two more hours and cell cycle progression was assessed at 9 h after illumination by counting the number of dividing cells (minimum of 200 instances were evaluated).

For testing the cell cycle arresting capability of SIP<sup>+</sup>, MT<sup>+</sup> and MT<sup>-</sup> cells with a size below the SST were cultured in 200 ml medium in 185-cm<sup>2</sup> tissue culture flasks and the medium was extracted with HLB-SPE as described above. The extracts were submitted in sequence to RP-UPLC/MS, fractionated at 4.6 min (6% B) and evaporated under a stream of nitrogen. After re-dissolving in one ml of water, both extracts were equally distributed to each 27 dark-synchronized MT<sup>-</sup> cultures grown in 20 ml medium in 25-cm<sup>2</sup> flasks with a cell density of around 15,000 cells/cm<sup>2</sup>. After re-illumination, three replicates were processed per hour by counting the amount of dividing cells and measuring the extracellular concentration of diproline as described above.

**RNA-seq on SIP<sup>+</sup>-treated MT<sup>-</sup> cultures.** Cultures of strain 85B (MT<sup>-</sup>) with an average cell size below the SST were grown at 18 °C in a 12-h:12-h light:dark regime with cool white fluorescent lamps at approximately 80 µmol photons m<sup>-2</sup> s<sup>-1</sup>. Before sampling, the dark period was extended with 12 h to synchronize the cells in the G<sub>1</sub> phase. A first control sample was taken in darkness (T<sub>0</sub>). Then, SIP<sup>+</sup> was added to half of the cultures in darkness. After 15 min, the light was turned on and 15 min later both treated and untreated cultures were harvested by filtration on Versapor filters (3-µm pore size, 25-mm diameter, PALL). The filters were rinsed with 1 ml PBS and put in liquid N<sub>2</sub>. This was repeated 1 h and 3 h after illumination. Sampling was carried out in triplicate. The protocol used for RNA extraction was based on the method developed by Apt, Clendennen, Powers and Grossman<sup>45</sup> with some modifications. The frozen filters were put in extraction buffer (100 mM Tris-HCl pH 7.5, 2% CTAB, 1.5 M NaCl, 50 mM EDTA, 10% β-mercaptoethanol) and sharp carbid beads were added. The samples were shaken in a bead mill (Retsch) at room temperature for 30 min. 100 µl of 10X Chelex-100 was added and the sample was incubated for 15 min at 56 °C, then one volume of chloroform:isoamyl alcohol (24:1, v/v) was added and the samples were shaken for 25 min at low speed. After centrifugation, the upper phase was transferred to a new tube and mixed with 0.3 volumes of absolute ethanol to precipitate the polysaccharides, and extracted with 1 volume of chloroform. After centrifugation, the upper phase was transferred to a fresh tube and RNA was precipitated overnight at -20 °C by addition of 0.25 volumes of 12 M LiCl and β-mercaptoethanol to 1% final concentration. After centrifugation, the pellet was washed with 70% ethanol, air-dried and suspended in RNase-free water for DNase treatment by an RNase-free DNase I (Turbo DNase, Ambion) according to the manufacturer's instructions. An extraction was then carried out by adding phenol-chloroform (1:1, V/V). After centrifugation, the upper phase was transferred to a fresh tube, and extracted with 1 volume of chloroform:isoamyl alcohol (24:1, V/V) and centrifuged again. The upper phase was precipitated with 0.3 M NaAc pH 5.5 and 75% ice cold ethanol by incubating overnight at 20 °C. After centrifugation, the pellet was washed with 70% ethanol and air-dried. The pellet was resuspended in an appropriate volume of RNase-free water. RNA quality was evaluated by spectrophotometry (Nanodrop) and Bioanalyzer (Agilent Technologies).

Sequencing libraries were prepared using Illumina TruSeq Stranded mRNA kit. All libraries were analysed in a 2 × 150 bp run on 2 lanes of a flowcell of the Illumina HiSeq 2500 at the VIB Nucleomics Core (www.nucleomics.be). Every library was sequenced twice, as all libraries were pooled together in every lane. After adaptor trimming with CLC Assembly Cell 4.2.0 (CLC Inc, Aarhus, Denmark), all paired reads were mapped to an in-house draft genome of *S. robusta* using GSNAP (GMAP-GSNAP version 2013-06-27)<sup>46</sup> with default settings. Based on this mapping, a genome-guided trinity approach was used to assemble the transcriptome, using Trinity r20131110<sup>47</sup> and PASA r20130907<sup>48</sup>. All reads were mapped to this transcriptome using GSNAP<sup>46</sup> after which all unmapped reads were extracted and de novo assembled using Trinity<sup>47</sup> with default settings. This *de novo* assembly was joined together with the genome-guided transcriptome assembly. CD\_HIT 4.6.1 (cd-hit-est)<sup>49,50</sup> was used to optimize the assembly with the following settings: -c 0.9 -n 8 -t 1. Finally, all transcripts shorter than 500 bp were discarded. For functional annotation, the final transcriptome was loaded into TRAPID (reference database: Plaza 2.5)<sup>51,52</sup>. For every library, the reads were mapped to the final transcriptome using GSNAP<sup>46</sup> with default settings

and mapped reads per transcript per library were counted. Transcripts with low overall counts (threshold of at least 1 cpm in at least six samples) have been removed from the analysis because they have little power for detecting differential expression. Hence, filtering these transcripts leads to a negligible information loss and results in a gain of statistical power<sup>53</sup>. Upon filtering, the sequencing libraries were normalized using TMM normalization<sup>54</sup>. Transcript-wise RNA-seq counts were analysed using a negative binomial model with a factor with 7 levels, one for each treatment (Control, SIP<sup>+</sup>) × time (T0, T1, T2, T3) combination (C0, C1, C2, C3, SIP<sup>+</sup> 1, SIP<sup>+</sup> 2, SIP<sup>+</sup> 3). The control taken in darkness (C0) was used as the reference level. Technical replicates have been pooled to avoid underestimation of the dispersion parameter. Both differential expression (DE) between SIP<sup>+</sup>-treated samples and untreated samples within each time point as well as DE changes over time have been assessed on a log scale using appropriate contrasts (DE1 = SIP<sup>+</sup> 1 – C1, DE2 = SIP<sup>+</sup> 2 – C2, DE3 = SIP<sup>+</sup> 3 – C3, DE3–DE1, DE2–DE1, DE3–DE2). A hierarchical two-stage approach has been used for assessing statistical significance, which controls the overall false discovery rate (FDR) at the transcript level<sup>55</sup>. In the first stage, the overall null hypothesis that no differential expression occurred at none of the time points (DE1 = DE2 = DE3 = 0) has been assessed. Next, the six contrasts of interest have been tested for all genes that passed the screening stage. An advanced, within-gene, multiple testing correction that accounts for the relatedness of different hypotheses (contrasts) has been adopted in the second stage<sup>56</sup>. Hence, the overall FDR has efficiently been controlled at a 5% level, while maintaining significant power for testing the hypotheses of interest. All analyses were conducted with the R/BioConductor package edgeR (version 3.8.5)<sup>57</sup> in R version 3.1.2 and likelihood ratio tests were used for assessing all hypotheses of interest<sup>58</sup>. RNA-Seq data have been submitted to ArrayExpress (www.ebi.ac.uk/arrayexpress), with accession number E-MTAB-4056.

**qRT-PCR on SIP<sup>+</sup>-treated MT<sup>-</sup> cultures.** Samples were taken as described for the RNA-seq on SIP<sup>+</sup>-treated MT<sup>-</sup> cultures just before illumination and at 15 min, 1 h, 3 h, 6 h, 9 h and 12 h after illumination. Progression through the cell cycle was monitored by assessing the percentage of dividing cells using an inverted microscope. The total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen). About 200 ml of culture ( $\pm 5 \times 10^6$  cells) was scraped from the bottom of a tissue culture flask (Cellstar, 175 cm<sup>2</sup> growth surface, Greiner Bio-One) and filtered on a Versapor filter (3- $\mu$ m pore size, 25-mm diameter, PALL). Filters were frozen in liquid nitrogen. RLT buffer (1 ml),  $\beta$ -mercaptoethanol (10  $\mu$ l) and silicon carbide beads (1.0 mm, BioSpec) were added to the filter. Cells were disrupted by beating on a bead mill (Retsch, 3 × 1 min). The remaining steps for the RNA extraction were done according to the manufacturer's instructions. cDNA was prepared using the iScript cDNA Synthesis Kit (Bio-Rad) as recommended by the manufacturer.

Fifteen ng of cDNA was used as a template for each qRT-PCR reaction. Samples were amplified in triplicate on the Lightcycler 480 platform with the Lightcycler 480 SYBR Green I Master mix (Roche) in the presence of 0.5  $\mu$ M gene-specific primers (primer sequences in Supplementary Table S3). The cycling conditions were: 10 min of polymerase activation at 95 °C and 45 cycles at 95 °C for 10 s, 58 °C for 15 s, 72 °C for 15 s. Afterwards, amplicon dissociation curves were recorded by heating from 65 °C to 97 °C with a ramp rate of 2.5 °C. Data were analysed in the qbase + software package (Biogazelle) with the  $\Delta$ Ct relative quantification method with two stably expressed normalization genes, previously selected based on cDNA-AFLP data (Devos V., unpublished data).

## References

- Mann, D. G. & Droop, S. J. M. Biodiversity, biogeography and conservation of diatoms. *Hydrobiologia* **336**, 19–32 (1996).
- Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
- Drebes, G. Sexuality. In *The Biology of Diatoms (Botanical Monographs Vol. 13)* (ed Werner, D.) 250–283 (University of California Press, Berkeley and Los Angeles, 1977).
- Chepurnov, V. A., Mann, D. G., Sabbe, K. & Vyverman, W. Experimental studies on sexual reproduction in diatoms. *Int. Rev. Cytol.* **237**, 91–154 (2004).
- Frenkel, J., Vyverman, W. & Pohnert, G. Pheromone signaling during sexual reproduction in algae. *Plant J.* **79**, 632–644 (2014).
- Karlson, P. & Lüscher, M. 'Pheromones': a new term for a class of biologically active substances. *Nature* **183**, 55–56 (1959).
- Sato, S., Beakes, G., Idei, M., Nagumo, T. & Mann, D. G. Novel sex cells and evidence for sex pheromones in diatoms. *PLoS ONE* **6**, e26923 (2011).
- Mann, D. G. & Chepurnov, V. A. Auxosporulation, mating system, and reproductive isolation in *Neidium* (Bacillariophyta). *Phycologia* **44**, 335–350 (2005).
- Chepurnov, V. A. *et al.* In search of new tractable diatoms for experimental biology. *BioEssays* **30**, 692–702 (2008).
- Gillard, J. *et al.* Metabolomics enables the structure elucidation of a diatom sex pheromone. *Angew. Chem. Int. Ed.* **52**, 854–857 (2013).
- Starr, R. C., Marner, F. J. & Jaenicke, L. Chemoattraction of male gametes by a pheromone produced by female gametes of *Chlamydomonas*. *Proc. Natl. Acad. Sci. USA* **92**, 641–645 (1995).
- Sumper, M., Berg, E., Wenzl, S. & Godl, K. How a sex-pheromone might act at a concentration below 10<sup>-16</sup> M. *EMBO J.* **12**, 831–836 (1993).
- Pohnert, G. & Boland, W. The oxylipin chemistry of attraction and defense in brown algae and diatoms. *Nat. Prod. Rep.* **19**, 108–122 (2002).
- Gillard, J. *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.* **148**, 1394–1411 (2008).
- Kuhlisch, C. & Pohnert, G. Metabolomics in chemical ecology. *Nat. Prod. Rep.* **32**, 937–955 (2015).
- Tautenhahn, R., Patti, G. J., Rinehart, D. & Siuzdak, G. XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal. Chem.* **84**, 5035–5039 (2012).
- Huysman, M. J. J. *et al.* Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. *Genome Biol.* **11**, R17 (2010).
- Ajimura, M., Leem, S.-H. & Ogawa, H. Identification of new genes required for meiotic recombination in *Saccharomyces cerevisiae*. *Genetics* **133**, 51–66 (1993).
- Symington, L. S. Role of *RAD52* epistasis group genes in homologous recombination and double-strand break repair. *Microbiol. Mol. Biol. Rev.* **66**, 630–670 (2002).

20. Kolas, N. K. & Cohen, P. E. Novel and diverse functions of the DNA mismatch repair family in mammalian meiosis and recombination. *Cytogenet. Genome Res.* **107**, 216–231 (2004).
21. Blanton, H. L. *et al.* REC, *Drosophila* MCM8, drives formation of meiotic crossovers. *PLoS Genet.* **1**, e40 (2005).
22. Crismani, W. *et al.* MCM8 is required for a pathway of meiotic double-strand break repair independent of DMC1 in *Arabidopsis thaliana*. *PLoS Genet.* **9**, e1003165 (2013).
23. Patil, S. *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 Genomics* **16**, 930 (2015).
24. Chepurinov, V. A., Mann, D. G., Vyverman, W., Sabbe, K. & Danielidis, D. B. Sexual reproduction, mating system, and protoplast dynamics of *Seminavis* (Bacillariophyceae). *J. Phycol.* **38**, 1004–1019 (2002).
25. Hu, C. A., Delauney, A. J. & Verma, D. P. S. A bifunctional enzyme ( $\Delta 1$ -pyrroline-5-carboxylate synthetase) catalyzes the first two steps in proline biosynthesis in plants. *Proc. Natl. Acad. Sci. USA* **89**, 9354–9358 (1992).
26. Liu, Y., Ruoho, A. E., Rao, V. D. & Hurley, J. H. Catalytic mechanism of the adenylyl and guanylyl cyclases: modeling and mutational analysis. *Proc. Natl. Acad. Sci. USA* **94**, 13414–13419 (1997).
27. Richter, W., Unciuleac, L., Hermsdorf, T., Kronbach, T. & Dettmer, D. Identification of inhibitor binding sites of the cAMP-specific phosphodiesterase 4. *Cell. Signal.* **13**, 287–297 (2001).
28. Käll, L., Krogh, A. & Sonnhammer, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35**, W429–W432 (2007).
29. Al-Hasani, H. & Jaenicke, L. Characterization of the sex-inducer glycoprotein of *Volvox carterii* f. *weismannia*. *Sex. Plant Reprod.* **5**, 8–12 (1992).
30. Pawlowski, W. P., Sheehan, M. J. & Ronceret, A. In the beginning: the initiation of meiosis. *BioEssays* **29**, 511–514 (2007).
31. Armbrust, E. V., Chisholm, S. W. & Olson, R. J. Role of light and the cell cycle on the induction of spermatogenesis in a centric diatom. *J. Phycol.* **26**, 470–478 (1990).
32. Bardwell, L. A walk-through of the yeast mating pheromone response pathway. *Peptides* **25**, 1465–1476 (2004).
33. Degrassi, G. *et al.* Plant growth-promoting *Pseudomonas putida* WCS358 produces and secretes four cyclic dipeptides: Cross-talk with quorum sensing bacterial sensors. *Curr. Microbiol.* **45**, 250–254 (2002).
34. Ortiz-Castro, R. *et al.* Transkingdom signaling based on bacterial cyclodipeptides with auxin activity in plants. *Proc. Natl. Acad. Sci. USA* **108**, 7253–7258 (2011).
35. Belin, P. *et al.* The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. *Nat. Prod. Rep.* **29**, 961–979 (2012).
36. Potter, L. R. Guanylyl cyclase structure, function and regulation. *Cell. Signal.* **23**, 1921–1926 (2011).
37. Schaap, P. Guanylyl cyclases across the tree of life. *Front. Biosci.* **10**, 1485–1498 (2005).
38. Potter, L. R., Abbey-Hosch, S. & Dickey, D. M. Natriuretic peptides, their receptors, and cyclic guanosine monophosphate-dependent signaling functions. *Endocr. Rev.* **27**, 47–72 (2006).
39. Koch, K.-W. & Dell'Orco, D. A calcium-relay mechanism in vertebrate phototransduction. *ACS Chem. Neurosci.* **4**, 909–917 (2013).
40. Kaupp, U. B., Hildebrand, E. & Weyand, I. Sperm chemotaxis in marine invertebrates—molecules and mechanisms. *J. Cell. Physiol.* **208**, 487–494 (2006).
41. Artemenko, Y., Lampert, T. J. & Devreotes, P. N. Moving towards a paradigm: common mechanisms of chemotactic signaling in *Dictyostelium* and mammalian leukocytes. *Cell. Mol. Life Sci.* **71**, 3711–3747 (2014).
42. Guillard, R. R. L. Culture of phytoplankton for feeding marine invertebrates. In *Culture of Marine Invertebrate Animals* (eds Smith, W. L. & Canley, M. H.) 29–60 (Plenum Press, New York, 1975).
43. Maier, I. & Calenberg, M. Effect of extracellular  $Ca^{2+}$  and  $Ca^{2+}$ -antagonists on the movement and chemoorientation of male gametes of *Ectocarpus siliculosus* (phaeophyceae). *Bot. Acta* **107**, 451–460 (1994).
44. Barofsky, A., Vidoudez, C. & Pohnert, G. Metabolic profiling reveals growth stage variability in diatom exudates. *Limnol. Oceanogr. Methods* **7**, 382–390 (2009).
45. Apt, K. E., Clendennen, S. K., Powers, D. A. & Grossman, A. R. The gene family encoding the fucoxanthin chlorophyll proteins from the brown alga *Macrocystis pyrifera*. *Mol. Gen. Genet.* **246**, 455–464 (1995).
46. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
47. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
48. Haas, B. J. *et al.* Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
49. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
50. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
51. Van Bel, M. *et al.* TRAPID: an efficient online tool for the functional and comparative analysis of *de novo* RNA-Seq transcriptomes. *Genome Biol.* **14**, R134 (2013).
52. Proost, S. *et al.* PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* **21**, 3718–3731 (2009).
53. Bourgon, R., Gentleman, R. & Huber, W. Independent filtering increases detection power for high-throughput experiments. *Proc. Natl. Acad. Sci. USA* **107**, 9546–9551 (2010).
54. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
55. Heller, R., Manduchi, E., Grant, G. R. & Ewens, W. J. A flexible two-stage procedure for identifying gene sets that are differentially expressed. *Bioinformatics* **25**, 1019–1025 (2009).
56. Shaffer, J. P. Modified sequentially rejective multiple test procedures. *J. Am. Stat. Assoc.* **81**, 826–831 (1986).
57. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
58. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).

## Acknowledgements

The authors wish to thank Annick Bleys for help in preparing the manuscript and Eva Pfannerstill for experimental assistance. This work was supported by grants of the Research Foundation Flanders (G.0288.13 and G.0E07.14N) and the research council of Ghent University (BOF/GOA 01G01715). M.J.J.H. is a Postdoctoral Fellow of the Research Foundation Flanders. J.G. was supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders and by an EMBO short-term fellowship (ASTF 410.00-2008). J.F.,

C.L and G.P. acknowledge the Jena School for Microbial Communication (JSMC), the International Max Planck Research School (IMPRS) and the International Leibniz Research School (ILRS) for funding.

### Author Contributions

S.M., J.F., C.L., J.T.F.G., V.D., K.V.d.B., B.B., M.J.J.H., J.S., S.D.D. T.B. and P.W. conducted the experiments and analysed results, A.B., K.S., M.V., L.C., L.D.V., G.P. and W.V. conceived the experiment(s) and analysed the results. S.M., J.F., C.L., L.D.V., G.P. and W.V. wrote the manuscript. All authors reviewed the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Moeys, S. *et al.* A sex-inducing pheromone triggers cell cycle arrest and mate attraction in the diatom *Seminavis robusta*. *Sci. Rep.* **6**, 19252; doi: 10.1038/srep19252 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>