

SCIENTIFIC REPORTS

OPEN

Identification and characterisation of the novel endogenous promoter *HASP1* and its signal peptide from *Phaeodactylum tricornutum*

Erdenedolgor Erdene-Ochir^{1,2}, Bok-Kyu Shin³, Byeori Kwon³, Choonkyun Jung⁴ & Cheol-Ho Pan^{1,2}

Although diatoms have been extensively studied as bioreactors, only a limited number of efficient gene promoters are available. Therefore, the development of new endogenous promoters is important for the heterologous production of a variety of recombinant proteins. Herein, we identified the most abundant secreted protein in *Phaeodactylum tricornutum*, designated 'highly abundant secreted protein 1' (*HASP1*), and characterised the activities of its promoter and signal peptide using green fluorescent protein (GFP) as a reporter. The *HASP1* promoter strongly drove GFP expression during all growth phases of *P. tricornutum* in culture, in contrast to the commonly used *fcpA* promoter, which is less active during the stationary phase. The *HASP1* signal peptide was also sufficient for facilitating efficient secretion of GFP by *P. tricornutum*. Our findings suggest that both the promoter and the signal peptide of *HASP1* can be utilized as novel tools for the overexpression and secretion of recombinant proteins in *P. tricornutum*.

Diatoms are unicellular, eukaryotic phytoplankton that inhabit marine and freshwater environments and are responsible for about 20% of the primary photosynthetic productivity on earth¹. Compared with other microalgae, diatoms have a higher carbon fixation ability². *Phaeodactylum tricornutum* is a pleiomorphic coastal marine diatom that utilizes silicic acid in a facultative manner³. The complete genome of *P. tricornutum* is approximately 27.6 Mb in size and consists of 33 chromosomes containing 12,177 predicted protein-coding genes⁴. A large number of expressed sequence tags (ESTs) have been identified in *P. tricornutum* cells grown under various different conditions, including different sources of nutrients, morphotypes, lighting sources, and abiotic stresses^{5,6}. Furthermore, tools for genetic manipulation of *P. tricornutum*, such as gene overexpression, gene silencing, gene editing, and plasmid delivery methods, are well established^{7–11}.

Diatoms are being actively studied for diverse biotechnological applications and can currently be used as bioreactors for the production of biopharmaceuticals and secondary metabolites. For example, *P. tricornutum* has been genetically engineered to produce pharmaceutically important proteins such as human IgG antibodies¹². In this regard, the N-glycosylation pathways in *P. tricornutum* can be remodelled to have a human-compatible glycosylation profile¹³. The development of an optimised expression system using a suitable promoter is an essential prerequisite for heterologous gene expression. A strong promoter that drives high yields of recombinant protein is not only crucial for developing a cost-effective expression system but is also indispensable for metabolic engineering through gene regulation. The endogenous promoter from the gene encoding fucoxanthin chlorophyll a/c binding protein (FCP) has been commonly used for both genetic and metabolic engineering of *P. tricornutum* for a variety of different applications^{14–18}. However, the *fcpA* promoter is light-inducible and is not always able to drive constitutive expression of downstream transgenes throughout all growth phases of the culture^{19–21}. The nitrate reductase gene promoter has also been widely used as a strong endogenous promoter for many purposes,

¹Natural Product Informatics Research Center, KIST Gangneung Institute of Natural Products, Gangneung, 25451, Republic of Korea. ²Division of Bio-Medical Science and Technology, KIST School, Korea University of Science and Technology, Seoul, 02792, Republic of Korea. ³Algaeprona Inc, Gangneung, 25451, Republic of Korea. ⁴Graduate School of International Agricultural Technology and Crop Biotechnology Institute/GreenBio Science and Technology, Seoul National University, Pyeongchang, 25354, Republic of Korea. Correspondence and requests for materials should be addressed to C.J. (email: jasmin@snu.ac.kr) or C.-H.P. (email: panc@kist.re.kr)

such as the heterologous expression of human antibodies¹². However, the disadvantage of the nitrate reductase promoter is that it requires an exogenous nitrogen source to induce downstream gene expression. Recently, other endogenous promoters such as *V-ATPase C* and *Pt211* have been identified as strong promoters^{22,23}. On the other hand, heterologous promoters, such as the *CIP1* promoter of the putative replication-associated protein gene from the *Chaetoceros lorenzianus*-infecting DNA virus, are less active than the diatom endogenous gene promoter^{19,24}.

The secretion of recombinant proteins into the culture medium is a powerful and cost-efficient platform for producing pharmaceutically important proteins because the steps needed to purify secreted recombinant proteins are less complicated than the steps needed for the purification of intracellular proteins. Endogenous signal peptides from the ARS1 and gametolysin proteins have been used to promote secretion of recombinant proteins into the culture medium in *Chlamydomonas reinhardtii*^{25,26}. A human IgG antibody without an endoplasmic reticulum retention signal has also been shown to be efficiently secreted into the culture medium of *P. tricornutum*²⁷. Despite their great advantages, protein secretion pathways have been rarely investigated in *P. tricornutum*.

In this study, we performed a secreted proteome profile study of *P. tricornutum* using liquid chromatography tandem mass spectrometry (LC-MS/MS) to examine the secreted proteins present in the culture medium. We first identified several hundred proteins that are secreted at high levels into the extracellular space in the stationary phase and designated these as HASPs (highly abundant secreted proteins) for further analysis. Among these, we found that the HASP1 protein was the most abundant protein in the culture medium. In order to develop a novel expression system for producing secreted recombinant proteins in *P. tricornutum*, we characterised promoter and its signal peptide using green fluorescent protein (GFP) as a reporter. The *HASP1* promoter strongly drove GFP expression, especially in the stationary phase of *P. tricornutum* culture, to a much higher level than that of the *fcpA* promoter. The signal peptide of the HASP1 protein was sufficient to result in GFP secretion into the culture medium of *P. tricornutum*. Taken together, our findings provide evidence that the endogenous *HASP1* promoter is a novel strong promoter that can be used as a suitable tool for overexpressing recombinant proteins in *P. tricornutum*.

Results and Discussion

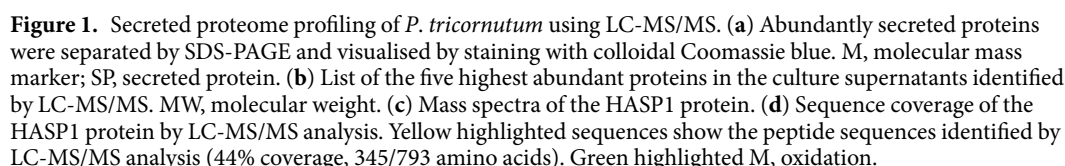
Identification of highly abundant secreted proteins in stationary phase cultures of *P. tricornutum*.

The abundantly secreted proteins in the stationary phase culture medium of *P. tricornutum* were resolved by SDS-PAGE and visualised by staining with colloidal Coomassie blue (Fig. 1a). LC-MS/MS analysis of the most abundant secreted proteins identified a total of 468 proteins. The five highest secreted proteins, whose relative abundance in the culture medium was compared based on their spectral counts in LC-MS/MS analysis, are shown in Fig. 1b. Among these, we found that the most abundantly secreted protein was PHATRDRAFT_47612 (NCBI accession number: XM_002181840.1), identified through database searching with a 44% sequence coverage; we have designated this protein as HASP1 (highly abundant secreted protein 1; consists of 793 amino acids; Fig. 1c,d). The HASP1 protein contains a phytase-like domain in the C-terminal region and therefore belongs to the phytase superfamily. The fourth most abundant protein (PHATRDRAFT_54681) was annotated via homology as the enzyme endo-1,3-beta-glucanase which is involved in carbohydrate metabolism²⁸. In addition, the fifth most abundant protein (PHATRDRAFT_49678) encodes an alkaline phosphatase, whose activity has been experimentally proven²⁹. The protein levels of HASP1 were drastically higher than those of any of the other protein in the culture medium. Therefore, we selected the HASP1 protein for further analysis and isolated its potential promoter region and putative signal peptide (Fig. 2a,b) to test their effects on the production and secretion of recombinant proteins in *P. tricornutum*. The *fcpA* promoter, a widely used promoter in *P. tricornutum*, was used as a control³⁰.

Isolation of the potential promoter and the putative signal peptide of the *HASP1* gene.

The *HASP1* gene is located on chromosome 14 of the *P. tricornutum* genome³¹. The intergenic region between *HASP1* and the upstream *Phatdraft_52260* gene was 3,534 bp long, extending from after the stop codon of *Phatdraft_52260* to before the start codon of *HASP1* (Fig. 2a). The *P. tricornutum* genome has a relatively higher gene density than other higher eukaryotic genomes³². Approximately, 500-bp upstream regions of endogenous genes are sufficient for full promoter activity by strongly driving transgene expression^{7,12,15,19,22,33}. Therefore, we considered the 499-bp upstream region before the start codon of the *HASP1* gene as the potential promoter. The *HASP1* potential promoter contains a 5'-untranslated region, an initiator-like sequence, and potential *cis*-acting regulatory elements that are recognised by transcription factors such as Myb and bZIP (Supplementary Fig. S1)^{24,29,32,34–39}. Using SignalP software, the 18-amino acid long signal peptide (SP) responsible for HASP1 secretion was found in the N-terminus³³ (Fig. 2b). The nucleotide sequences of the 499-bp long *HASP1* potential promoter and the putative signal peptide were obtained using the Biomart tool in EnsemblProtists⁴⁰. The potential promoter fragment of *HASP1* and the *fcpA* promoter were amplified by PCR and then cloned into a modified pPha-T1 vector to drive GFP expression (*HASP1pro:GFP* and *fcpApro:GFP*). The putative signal peptide of the HASP1 protein was fused in frame to the N-terminus of GFP to allow for secretion of the GFP protein into the extracellular space (*HASP1pro:SP-GFP* and *fcpApro:SP-GFP*). The GFP levels in the supernatant of *P. tricornutum* cultures, either with or without the putative signal peptide, were compared to assess the efficiency of GFP secretion into the culture medium. It should be noted that in order to enhance GFP expression, a minimal Kozak sequence (ACC) was placed directly before the putative signal peptide derived from the HASP1 protein. In contrast, this Kozak sequence was not inserted into the vectors lacking the putative HASP1 signal peptide. A promoter-less vector harbouring only the *GFP* gene was used as a negative control (Supplementary Fig. S2).

Transformation of *P. tricornutum* and selection of transformants. *P. tricornutum* cells grown to the stationary phase were transformed with the five different constructs (*fcpApro:SP-GFP*, *HASP1pro:SP-GFP*, *fcpApro:GFP*, *HASP1pro:GFP*, and *promoter-less GFP*) by particle bombardment. Putative transformants were selected on f/2 agar medium containing zeocin for at least 3 to 4 weeks. Following this, the zeocin-resistant colonies were



Quantification of the *HASP1* promoter activity driving *GFP* expression. Transgenic *P. tricornutum* cells were seeded at a density of 5×10^5 cells/mL and grown for 22 days with a periodic assessment of cell number to generate growth curves (Fig. 3a). From the *P. tricornutum* culture growth curves of all transgenic lines, it took approximately 4 and 8 days to reach the log and stationary phases, respectively. For all transgenic lines, the cell densities exceeded 1×10^7 cells/mL at the stationary phase. It should be noted that the transgenic lines harbouring the *HASP1* promoter showed slightly slower growth than that of the other transgenic lines. Nonetheless, the data indicated that in all cases, transgene presence did not have any harmful effects on the growth rate of *P. tricornutum* cells. We then extracted total RNA from each transgenic cell line at both the log (day 4) and stationary phases (day 8) to compare the activity of the *HASP1* promoter with that of the *fcpA* promoter by measuring the levels of *GFP*-containing transcripts by real-time RT-PCR. The levels of *GFP*-containing transcripts in the transgenic line expressing *HASP1pro:GFP* were 3 and 44-fold higher than those expressing *fcpApro:GFP* on day 4 and 8, respectively. When compared with the promoter-less *GFP* control, the levels of *GFP* expression driven by the *HASP1* promoter increased by 35 and 764-fold on day 4 and 8, respectively (Fig. 3b). It is well known that the *fcpA* promoter is primarily active during the log phase¹⁹. Notably, the *HASP1* promoter activity was higher than that of the *fcpA* promoter during the same phase. In addition, with respect to the level of *GFP* transcripts, *HASP1* promoter activity was maintained at an even higher level in the stationary phase than in the log phase. These results are consistent with expression levels of endogenous *fcpA* and *HASP1* genes. In 8-day grown wild-type *P. tricornutum* cells, the level of endogenous *HASP1* mRNA was approximately 40-fold higher than that of endogenous *fcpA* mRNA. (Supplementary Fig. S3).

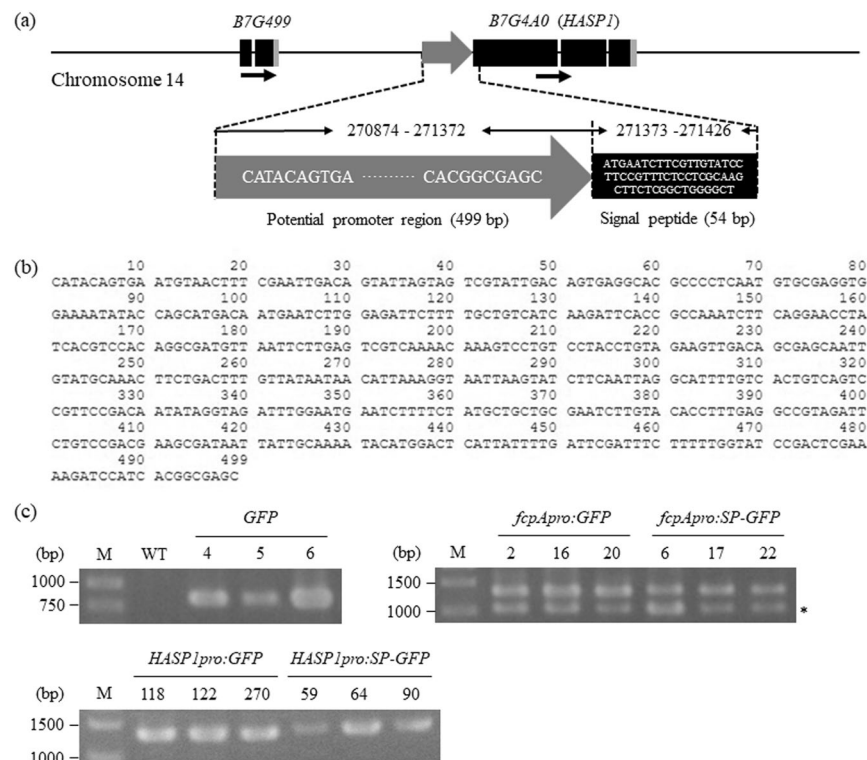


Figure 2. Isolation of the potential promoter and putative signal peptide of the *HASP1* gene and selection of transformants by PCR analysis with genomic DNA. (a) Schematic representation of the *HASP1* gene location on chromosome 14. The grey arrows indicate the potential promoter of the *HASP1* gene. Black arrows indicate the gene orientation. (b) Nucleotide sequence of the potential promoter region of the *HASP1* gene. (c) PCR amplification of transgenes from the genomic DNA of transformants. The numbers indicate independent transgenic lines for each construct. Full-length agarose gels are presented in Supplementary Fig. S4. Asterisk shows the nonspecific PCR products. M, molecular size marker.

We have therefore demonstrated that—compared with a previously reported promoter—the *HASP1* promoter is more active in *P. tricornutum*^{7,11,19,24,41–45}. Moreover, the transcriptional level of *SP-GFP* in the *HASP1pro:SP-GFP* line was 3 and 147-fold higher than that of the *fcpApro:SP-GFP* line on day 4 and 8, respectively. *SP-GFP* transcript levels driven by the *HASP1* promoter were 19 and 395-fold higher than those of the promoter-less *GFP* control on day 4 and 8, respectively (Fig. 3b). Although the mRNA levels of *SP-GFP* driven by the *HASP1* promoter were slightly lower than those of *GFP* alone driven by the *HASP1* promoter, *SP-GFP* levels were still higher than those of the *fcpA* promoter in both the log and stationary phases, suggesting that the presence of the SP sequence somehow affects the steady-state level of *GFP* mRNA. These results suggest that the *HASP1* promoter is a suitable tool for overexpressing recombinant proteins in *P. tricornutum*.

Relative fluorescence and GFP protein levels in the cell lysates and culture supernatants. To assess promoter activities and GFP secretion at the protein level, protein samples were prepared from both the *P. tricornutum* cells and culture supernatants. Subsequently, GFP fluorescence was measured using a fluorometer and GFP protein levels were examined by western blotting (Fig. 4).

The fluorescent signal of intracellular GFP in the *fcpApro:GFP* transgenic line gradually increased from the lag to the log phase (day 4) but decreased from the log to the stationary phase (day 8). The intracellular GFP concentration calculated from the fluorescent signal reached a maximal level of 1 µg/mL at the log phase (Fig. 4a). This result supports previous studies that showed that the *fcpA* promoter is maximally active in the log phase and less active in the stationary phase^{19–21}. In contrast, the fluorescent signal of intracellular GFP in the *HASP1pro:GFP* transgenic line rapidly increased from the log to the early stationary phase (day 8) which was maintained at a high level throughout the stationary phase until day 22 (Fig. 4a). The intracellular GFP concentration calculated from the fluorescent signal reached a level of 6 µg/mL at the stationary phase. Compared with the *fcpA* promoter, GFP protein levels driven by the *HASP1* promoter were slightly lower in the log phase, but rapidly increased by 6-fold during the early stationary phase, reaching an approximate 300-fold increase by the late stationary phase (Fig. 4a). In contrast, we could not detect GFP fluorescence in the culture supernatants of the transgenic lines expressing *HASP1pro:GFP* or *fcpApro:GFP* due to their lack of a putative signal peptide (Fig. 4b).

To investigate the role of the putative signal peptide of *HASP1* in GFP secretion, we measured the fluorescent signals of intracellular and secreted GFP from the same culture of *P. tricornutum* over time. The fluorescent signal from intracellular GFP in the *HASP1pro:SP-GFP* transgenic line was similar to that of the *fcpApro:SP-GFP* transgenic line during the log phase but was 3-fold higher during the stationary phase (Fig. 4a). Secreted GFP levels in

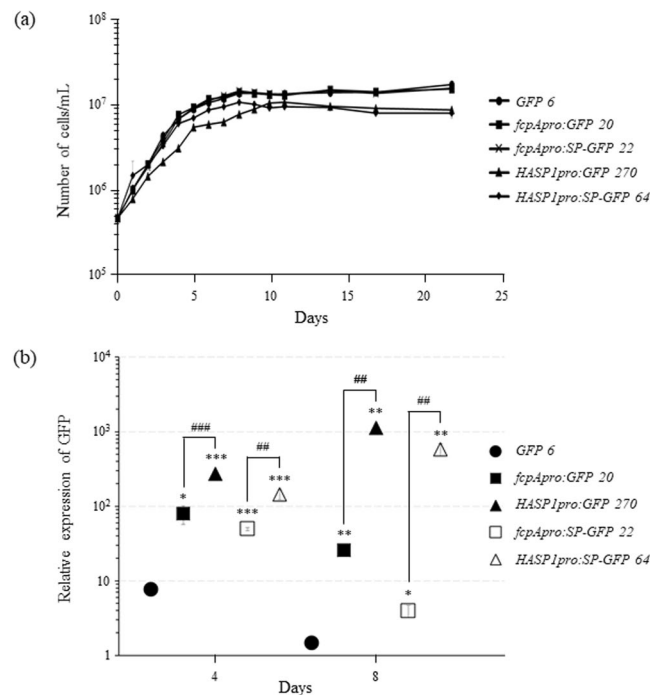


Figure 3. Growth curves of transgenic *P. tricornutum* and levels of GFP transcript in the selected transgenic lines. **(a)** Cell growth curves of *P. tricornutum* cultures. The selected transgenic lines were grown for 22 days. **(b)** Relative levels of GFP mRNA in the selected transgenic lines. GFP expression levels were normalised to *TBP* (TATA-box binding protein) expression. Data are expressed as the mean $\pm$ SD of three replicates. The asterisk (*) and hash (#) indicate statistically significant differences vs. the transformant groups of the negative (promoter-less GFP) and positive controls, respectively. ns, not significant ($p > 0.05$); * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (ANOVA).

the *fcpApro:SP-GFP* transgenic line increased slightly to 0.01 μ g/mL at the stationary phase (Fig. 4b). In contrast, secreted GFP levels in the *HASPIpro:SP-GFP* transgenic line steadily increased from the log to the stationary phases; extracellular GFP concentrations reached 0.01, 0.1, and 0.3 μ g/mL at the log, early stationary, and late stationary phases, respectively. These values correspond to a 3, 8, and 19-fold increase compared with that of the *fcpApro:SP-GFP* transgenic line (Fig. 4b). Moreover, secreted GFP levels under the control of the *HASPI* promoter were comparable to those secreted with the gametolysin signal sequence reported previously in *C. reinhardtii*²⁵.

To confirm the GFP fluorescence results obtained by the fluorometer, we also performed western blot analysis of cell lysates and culture supernatants over the same time course. With the exception of day 1, protein levels of both intracellular and secreted GFP in the cell lysates and culture supernatants were consistent with the fluorescent signals seen for intracellular and secreted GFP. On day 1, the protein levels of both intracellular GFP and SP-GFP were abnormally high. It should be noted that GFP fluorescence measurements were made with protein samples prepared from the same culture volumes (200 μ L), whereas for western blotting, the protein levels of GFP and SP-GFP were determined using the same amount of total soluble protein (10 μ g). Intracellular GFP was detected at high levels in all transgenic lines expressing *fcpApro:GFP*, *fcpApro:SP-GFP*, *HASPIpro:GFP*, and *HASPIpro:SP-GFP* except for the promoter-less GFP control (Fig. 4c; uncropped Western blot images used to generate Fig. 4c are shown in Supplementary Fig. S4). Protein levels of intracellular GFP driven by the *fcpA* promoter were the highest at the log phase, whereas those by the *HASPI* promoter were the highest at the late stationary phase. In contrast to the intracellular GFP levels, extracellular GFP secreted from the *P. tricornutum* cells could only be detected in the *HASPIpro:SP-GFP* transgenic line (Fig. 4d; uncropped Western blot images used to generate Fig. 4d are shown in Supplementary Fig. S4). This result is inconsistent with the GFP fluorescence measurements, likely due to differences in detection limits between a fluorometer and western blotting.

Overall, these data demonstrate that the *HASPI* promoter is much stronger than the *fcpA* promoter, especially during the stationary phase, and that the signal peptide derived from the *HASPI* protein is sufficient to facilitate efficient GFP secretion. There have been several studies that tested endogenous and exogenous promoter ability in driving the expression of heterologous genes in *P. tricornutum*^{7,15,19,41,44–46}. Compared to previously reported promoters, our data provide several lines of evidence that both the promoter and signal peptide of the *HASPI* gene can be powerful tools for the efficient expression and secretion of heterologous recombinant proteins throughout all growth phases of a *P. tricornutum* culture.

Subcellular localisation of GFP with or without the signal peptide of the *HASPI* protein. To observe the subcellular localisation of GFP with or without the *HASPI* signal peptide, GFP fluorescence (green) and chlorophyll autofluorescence (red) in all transgenic lines were visualised by confocal laser scanning microscopy. The transgenic line carrying the promoter-less GFP was analysed in parallel as a control (Fig. 5a). GFP

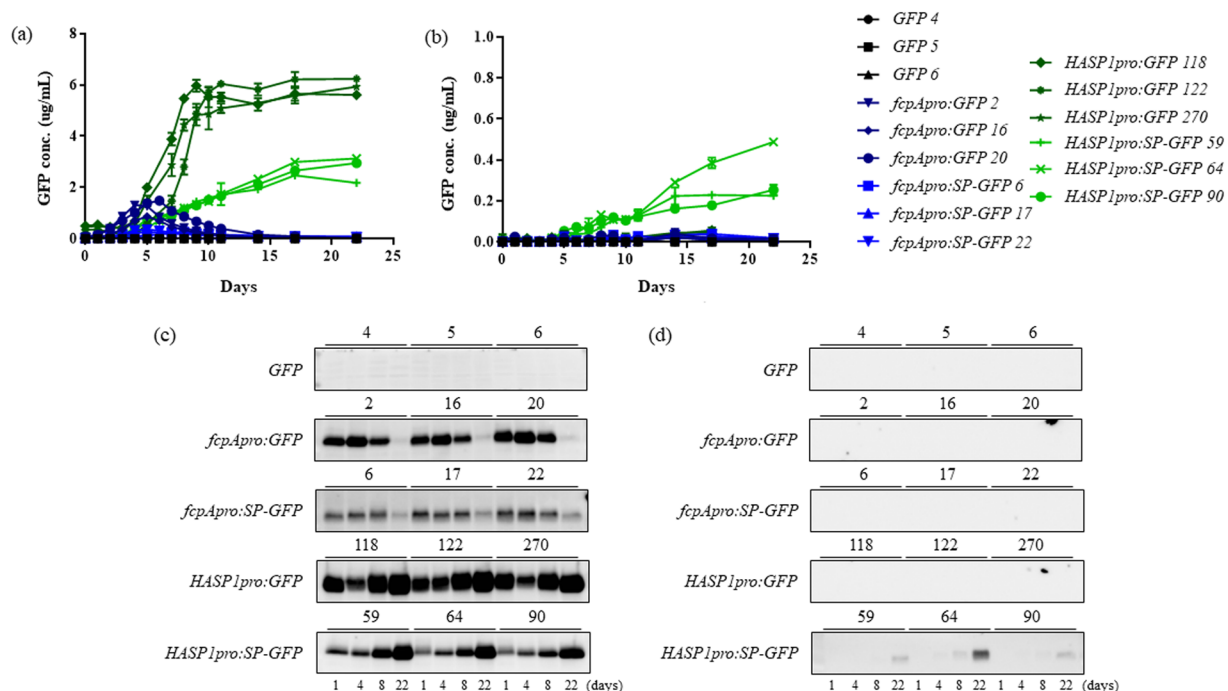


Figure 4. Relative protein expression levels of GFP in the cell lysates and culture supernatants. Levels of GFP fluorescence in (a) cell lysates and (b) culture supernatants were measured by a fluorometer. The protein levels of GFP in (c) cell lysates and (d) culture supernatants were determined by immunoblotting. Full-length Western blots are presented in Supplementary Fig. S4.

fluorescence in the *fcpApro:GFP* transgenic line was diffusely localised throughout the cytoplasm at both the log and early stationary phase but suddenly decreased at the late stationary phase (Fig. 5b). In contrast, GFP fluorescence in the *HASP1pro:GFP* transgenic line strongly increased throughout the entire cytoplasm from the lag to the early stationary phase and was maintained at high levels during the stationary phase (Fig. 5c). Interestingly, the signal peptide of HASP1 protein caused considerable changes in the subcellular localisation of GFP during all growth phases. GFP fluorescence in the *fcpApro:SP-GFP* lines during all growth phases was weakly localised near the chloroplast compartment, such as in the chloroplast endoplasmic reticulum^{41,47,48} and a blob-like structure found between the two plastid lobes^{47,49,50} (Fig. 5d). Similarly, GFP fluorescence in the *HASP1pro:SP-GFP* lines was weakly localised near the chloroplast compartment in both the lag and log phase. However, GFP fluorescence in the *HASP1pro:SP-GFP* lines rapidly increased from the log to the early stationary phase (day 8), and was maintained at a high level throughout the stationary phase until day 22 (Fig. 5e). This finding suggests that the HASP1 signal peptide leads to GFP entry into the secretory pathway. Thus far, the pathways utilized for protein secretion in *P. tricornutum* remain mostly unknown. Therefore, elucidating the pathways by which proteins are secreted in *P. tricornutum* would be an interesting research direction.

Conclusions

The HASP1 protein is the most abundant protein secreted into the culture medium of *P. tricornutum*. Compared with the *fcpA* promoter, the *HASP1* promoter strongly drives GFP expression during all growth phases, especially during the stationary phase. The HASP1 signal peptide also efficiently facilitates GFP secretion into the culture medium. The endogenous *HASP1* promoter as a novel strong promoter can therefore be an alternative to the commonly used *fcpA* promoter. Overall, the combination of the *HASP1* gene promoter and signal peptide demonstrated great potential for improving the production of secreted recombinant proteins by *P. tricornutum*.

Materials and Methods

Cell culture and cell density measurement. The *P. tricornutum* Bohlin UTEX 646 strain was obtained from the UTEX Culture Collection of Algae (The University of Texas at Austin, TX, USA). The strain was grown in 200 mL of f/2 medium⁵¹. Fifty percent artificial seawater, prepared with Royal Nature Advanced Pro Formula Salt (Royal Nature, Nesher, Israel) and f/2 medium, was autoclaved before addition into the final medium. The vitamin and metal solutions were added after autoclaving. A mixture of antibiotics containing 50 µg/mL ampicillin, 10 µg/mL kanamycin, and 50 µg/mL streptomycin was added to the liquid medium and agar plates to suppress bacterial contamination. Cultures were incubated at 20 °C with shaking at 200 rpm, constant aeration, and continuous lighting from a cool white fluorescent lamp (1600 lux). Cell density measurements at each sampling point were made by measuring the optical density of 200 µL cell culture at 530, 670, and 730 nm (Multi-Detection Microplate Reader Synergy HT; Biotek, Winooski, VT, USA).

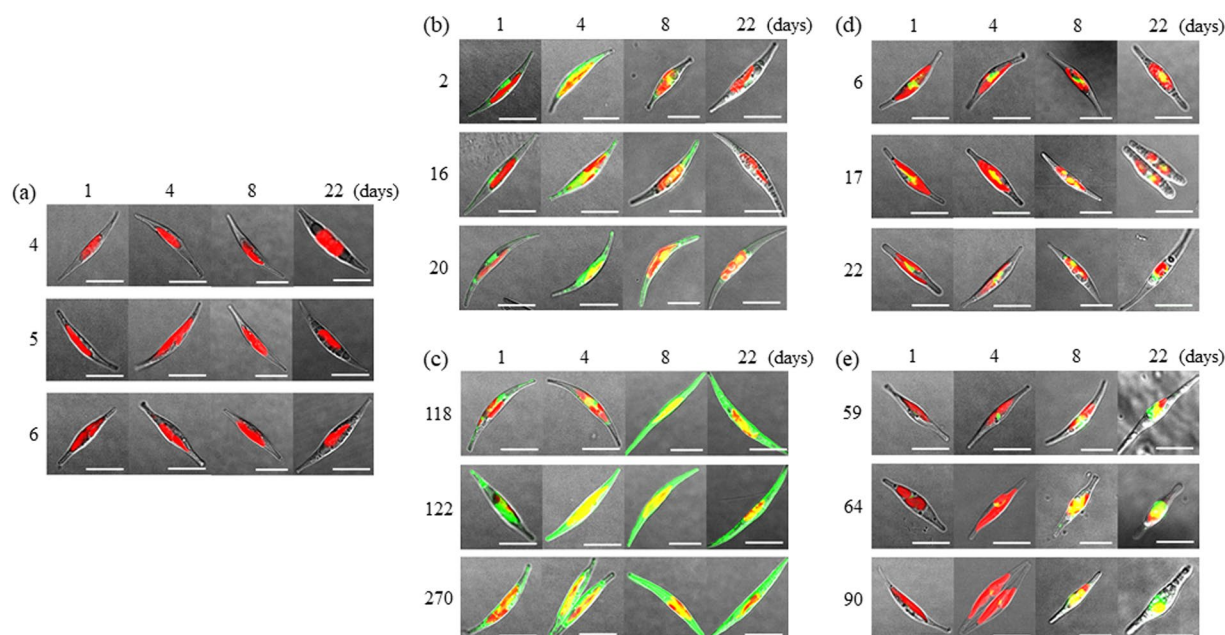


Figure 5. Effect of the HASP1 signal peptide on subcellular localisation of GFP. GFP fluorescence and chlorophyll autofluorescence in transgenic lines expressing the (a) promoter-less GFP, (b) *fcpApro::GFP*, (c) *HASP1pro::GFP*, (d) *fcpApro::SP-GFP*, and (e) *HASP1pro::SP-GFP* were visualised by confocal laser scanning microscopy. The numbers on the left side of the images indicate three independent transgenic lines for each construct. Scale bars = 10 μ m.

Preparation of secreted proteins from the *P. tricornutum* culture in the stationary phase. To profile the proteins secreted by *P. tricornutum*, cultures were grown to a density of 1×10^7 cells/mL as described above, and the culture was centrifuged at 3500 rpm for 15 min at 4 °C to pellet the cells. Supernatants were filtered through a 0.2- μ m vacuum filter system (Corning Inc., Corning, NY, USA) and concentrated using centrifugal filter devices (Vivaspin 20 and 500; Satorius, Goettingen, Germany) with 5 and 10 kDa cutoff membranes according to manufacturer's instructions. Protein concentrations were determined by the Bradford assay (Bio-Rad Laboratories, Hercules, CA, USA). Samples were prepared from three independent experiments.

SDS-PAGE and in-gel digestion. Protein samples (30 μ g) were resolved by electrophoresis on 12% SDS-PAGE gels and visualised by Coomassie blue staining. Abundant proteins were excised from the gels and subjected to in-gel digestion⁵².

LC-MS/MS analysis and protein identification through a database search. LC-MS/MS analysis and protein identification were performed as previously described¹⁹.

Construction of plasmid vectors. The pPha-T1 expression vector was used as a backbone for all constructs³⁰. The *fcpA* promoter fragment harbouring NdeI/EcoRI restriction sites³⁰ and the potential promoter fragment of *HASP1* (Fig. 2) were amplified by PCR from the pPha-T1 vector and *P. tricornutum* genomic DNA, respectively. The GFP gene was amplified from a pEGFP-C2 vector with three different restriction enzyme combinations (NheI/BamHI, EcoRI/BamHI, and NdeI/BamHI), which were later used to construct the five different expression vectors. Next, the sequence encoding the putative signal peptide of the HASP1 protein (Fig. 2) was cloned into the constructs using a long primer harbouring EcoRI/NheI restriction sites, the Kozak sequence, the putative signal peptide sequence, and the 5' end of GFP. The promoter-less pPha-T1 vector, harbouring only GFP, was used as a negative control. All ligation reactions were performed using a DNA ligation kit (6023; Takara Bio, Kusatsu, Japan). All constructs were verified by sequencing. All primers used in vector construction are listed in Supplementary Table S1.

Transformation of *P. tricornutum* by particle bombardment. The activation of 0.6- μ m microcarriers/nanogold particles (9204298; Bio-Rad Laboratories) and DNA coating of microcarriers were performed according to manufacturer's instructions for the Biolistic® PDS-1000/He Particle Delivery System (165–2257; Bio-Rad Laboratories). *P. tricornutum* cells cultivated in 2 L of f/2 medium for 7 days at 20 °C under continuous white fluorescent light ($40 \mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$) were centrifuged at 3500 rpm for 5 min at 15 °C and then plated on the central third of the f/2 agar plates, which also contained mixed antibiotics. Transformation of the *P. tricornutum* cells was conducted as previously described¹⁹.

PCR analysis of putative transformants. Harvested *P. tricornutum* cells from 20 µL cell culture were lysed in 10 µL of 5 mM Tris-HCl (pH 7.5) buffer containing 0.8% Tween-20⁵³. The lysates were centrifuged at 3500 rpm for 5 min at 20 °C and the supernatants were discarded. The pellet was used for PCR analysis of genomic DNA with a Maxime PCR Premix (i-StarTaq; 25167; iNtRON Biotechnology, Seongnam, South Korea). Primers used for PCR analysis are listed in Supplementary Table S1.

Total RNA preparation and real-time RT-PCR analysis. Five millilitres of cell culture grown for 4 or 8 days was centrifuged at 3500 rpm for 15 min at 4 °C. The pellet was washed twice with 1 mL ice-cold Dulbecco's phosphate buffered saline (DPBS; Corning™ 21-031-CVR; Thermo Fisher Scientific, Waltham, MA, USA) via centrifugation at 6000 rpm for 5 min at 4 °C. After the last centrifugation step, the pellet was suspended in 1 mL ice-cold DPBS and ground in liquid nitrogen. Total RNA was extracted according to manufacturer's instructions for the RNeasy Plus Total RNA Extraction Reagent (9109; Takara Bio). First-strand cDNA was then synthesised using a QuantiTect® Reverse Transcription Kit (205311; Qiagen, Hilden, Germany) according to manufacturer's instructions. Real-time RT-PCR analysis was performed as described in the LightCycler® 480 SYBR® Green I Master (04707516001; Roche Diagnostics, Risch-Rotkreuz, Switzerland) protocol. Primers used for real-time RT-PCR analysis are listed in Supplementary Table S1.

GFP fluorescence measurement. Protein samples were prepared separately from equal volumes (200 µL) of *P. tricornutum* cell culture and culture supernatants for all time points. Fluorescence measurements were performed as previously described¹⁹. The autofluorescence value of the promoter-less GFP construct was subtracted from the GFP fluorescence value obtained with the *fcpA* and *HASPI* promoters. A GFP standard curve was generated using a recombinant GFP protein (ab119740; Abcam, Cambridge, UK). Measurements were performed with three biological replicates.

Western blot analysis. Cell lysis was performed as described previously¹⁹. *P. tricornutum* culture supernatant was concentrated as described above. Total soluble protein concentrations were determined using the bicinchoninic acid (BCA) protein assay kit (23225; Thermo Fisher Scientific) according to manufacturer's instructions. The same amount of total soluble protein (10 µg) in each sample was separated by electrophoresis on 12% homemade Tris-glycine SDS-polyacrylamide gels and then transferred to polyvinylidene difluoride (PVDF) membranes. The membranes were incubated with an anti-GFP goat polyclonal antibody (ab6673; Abcam) and an anti-goat IgG-horseradish peroxidase conjugated bovine polyclonal antibody (sc-2378; Santa Cruz Biotechnology, Dallas, TX, USA). Immunoblot signals were detected using SuperSignal® West Femto Maximum Sensitivity Substrate (34095; Thermo Fisher Scientific).

Subcellular localisation of GFP. Live cell images were captured using a Leica TCS-SP5 confocal laser scanning microscope (Leica, Wetzlar, Germany). GFP was excited with a 488 nm laser and its fluorescence emission was detected between 509 and 560 nm. Chlorophyll autofluorescence was observed using excitation at 488 nm and emission at 700–750 nm⁵⁴.

Statistical analysis. Data are expressed as the mean ± SD. Statistical analysis was performed using one-way analysis of variance (ANOVA) followed by Sidak's multiple comparisons test (GraphPad Prism v7.04). *p* values < 0.05 were considered statistically significant.

Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

- Mock, T. & Medlin, L. K. Genomics and genetics of diatoms in *Advances in botanical research* (Ed. Piganeau, G.) 245–284 (Academic Press, 2012).
- Thomas, W. H., Dodson, A. N. & Reid, F. M. H. Diatom productivity compared to other algae in natural marine phytoplankton assemblages. *J. Phycol.* **14**, 250–253 (1978).
- Martino, A., De Meichenin, A., Shi, J., Pan, K. & Bowler, C. Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (Bacillariophyceae) accessions. *J. Phycol.* **43**, 992–1009 (2007).
- Rastogi, A. *et al.* Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci. Rep.* **8**, 4834 (2018).
- Maheswari, U. *et al.* Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biol.* **11**, R85 (2010).
- Maheswari, U. *et al.* The diatom EST database. *Nucleic Acids Res.* **33**, D344–D347 (2005).
- Apt, K. E., Kroth-Pancic, P. G. & Grossman, A. R. Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol. Gen. Genet.* **252**, 572–579 (1996).
- De Riso, V. *et al.* Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* **37**, e96 (2009).
- Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M. & Winge, P. A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* **6**, 24951 (2016).
- Karas, B. J. *et al.* Designer diatom episomes delivered by bacterial conjugation. *Nat. Comm.* **6**, 6925 (2015).
- Siaut, M. *et al.* Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* **406**, 23–35 (2007).
- Hempel, F., Lau, J., Klingl, A. & Maier, U. G. Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PLoS One* **6**, e28424 (2011).
- Mathieu-Rivet, E. *et al.* Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Front. Plant Sci.* **5**, 1–13 (2014).
- Joshi-Deo, J. *et al.* Characterization of a trimeric light-harvesting complex in the diatom *Phaeodactylum tricornutum* built of FcpA and FcpE proteins. *J. Exp. Bot.* **61**, 3079–3087 (2010).

15. Kadono, T. *et al.* Effect of an introduced phytoene synthase gene expression on carotenoid biosynthesis in the marine diatom *Phaeodactylum tricornutum*. *Mar. Drugs* **13**, 5334–5357 (2015a).
16. Niu, Y.-F. *et al.* Improvement of neutral lipid and polyunsaturated fatty acid biosynthesis by overexpressing a type 2 diacylglycerol acyltransferase in marine diatom *Phaeodactylum tricornutum*. *Mar. Drugs* **11**, 4558–4569 (2013).
17. Radakovits, R., Eduafo, P. M. & Posewitz, M. C. Genetic engineering of fatty acid chain length in *Phaeodactylum tricornutum*. *Metab. Eng.* **13**, 89–95 (2011).
18. Xue, J. *et al.* Genetic improvement of the microalga *Phaeodactylum tricornutum* for boosting neutral lipid accumulation. *Metab. Eng.* **27**, 1–9 (2015).
19. Erdene-Ochir, E. *et al.* Cloning of a novel endogenous promoter for foreign gene expression in *Phaeodactylum tricornutum*. *Appl. Biol. Chem.* **59**, 861–867 (2016).
20. Nymark, M. *et al.* An integrated analysis of molecular acclimation to high light in the marine diatom *Phaeodactylum tricornutum*. *PLoS One* **4**, e7743 (2009).
21. Nymark, M. *et al.* Molecular and photosynthetic responses to prolonged darkness and subsequent acclimation to re-illumination in the diatom *Phaeodactylum tricornutum*. *PLoS One* **8**, e58722 (2013).
22. Watanabe, Y. *et al.* Development of endogenous promoters that drive high-level expression of introduced genes in the model diatom *Phaeodactylum tricornutum*. *Mar. Genomics* **42**, 41–48 (2018).
23. Zou, L. G. *et al.* High-efficiency promoter-driven coordinated regulation of multiple metabolic nodes elevates lipid accumulation in the model microalga *Phaeodactylum tricornutum*. *Microb. Cell Fact.* **17**, 54 (2018).
24. Kadono, T. *et al.* Characterization of marine diatom-infecting virus promoters in the model diatom *Phaeodactylum tricornutum*. *Sci. Rep.* **5**, 18708 (2015b).
25. Ramos-Martinez, E. M., Fimognari, L. & Sakuragi, Y. High-yield secretion of recombinant proteins from the microalga *Chlamydomonas reinhardtii*. *Plant Biotechnol. J.* **15**, 1214–1224 (2017).
26. Rasala, B. A. *et al.* Robust expression and secretion of xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS One* **7**, e43349 (2012).
27. Hempel, F. & Maier, U. G. An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microb. Cell Fact.* **11**, 126 (2012).
28. Kroth, P. G. *et al.* A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One* **3**, e1426 (2008).
29. Lin, H.-Y. *et al.* Alkaline phosphatase promoter as an efficient driving element for exogenic recombinant in the marine diatom *Phaeodactylum tricornutum*. *Algal Res.* **23**, 58–65 (2017).
30. Zaslavskaya, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R. & Apt, K. E. Transformation of the diatom *Phaeodactylum tricornutum* with a variety of selectable marker and reporter genes. *J. Phycol.* **386**, 379–386 (2000).
31. Fabris, M. *et al.* The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *Plant J.* **70**, 1004–1014 (2012).
32. Matthijs, M., Fabris, M., Broos, S., Vyverman, W. & Goossens, A. Profiling of the early nitrogen stress response in the diatom *Phaeodactylum tricornutum* reveals a novel family of RING-domain transcription factors. *Plant Physiol.* **170**, 489–498 (2016).
33. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
34. Grillo, G. *et al.* UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.* **38**, D75–D80 (2009).
35. Ambawat, S., Sharma, P., Yadav, N. R. & Yadav, R. C. MYB transcription factor genes as regulators for plant responses: an overview. *Physiol. Mol. Biol. Plants* **19**, 307–321 (2013).
36. Chow, C.-N. *et al.* PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.* **44**, D1154–D1160 (2016).
37. Jakoby, M. *et al.* bZIP transcription factors in Arabidopsis. *Trends Plant Sci.* **7**, 106–111 (2002).
38. Lescot, M. *et al.* PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **30**, 325–327 (2002).
39. Rayko, E., Maumus, F., Maheswari, U., Jabbari, K. & Bowler, C. Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol.* **188**, 52–66 (2010).
40. Kinsella, R. J. *et al.* Ensembl BioMart: A hub for data retrieval across taxonomic space. *Database* **2011**, 1–9 (2011).
41. Chu, L. *et al.* Shuttling of (deoxy-) purine nucleotides between compartments of the diatom *Phaeodactylum tricornutum*. *New Phytol.* **213**, 193–205 (2017).
42. Harada, H., Nakatsuma, D., Ishida, M. & Matsuda, Y. Regulation of the expression of intracellular β -carbonic anhydrase in response to CO₂ and light in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiol.* **139**, 1041–1050 (2005).
43. Niu, Y.-F. *et al.* Transformation of diatom *Phaeodactylum tricornutum* by electroporation and establishment of inducible selection marker. *Biotechniques* **52**, 1–3 (2012).
44. Seo, S., Jeon, H., Hwang, S., Jin, E. & Chang, K. S. Development of a new constitutive expression system for the transformation of the diatom *Phaeodactylum tricornutum*. *Algal Res.* **11**, 50–54 (2015).
45. Ying-Fang, N. *et al.* Transformation of diatom *Phaeodactylum tricornutum* by electroporation and establishment of inducible selection marker. *Biotechniques* **52**, 1–3 (2012).
46. Sakaue, K., Harada, H. & Matsuda, Y. Development of gene expression system in a marine diatom using viral promoters of a wide variety of origin. *Physiol. Plant.* **133**, 59–67 (2008).
47. Apt, K. E. *et al.* In vivo characterization of diatom multipartite plastid targeting signals. *J. Cell Sci.* **115**, 4061–4069 (2002).
48. Liu, X. *et al.* Addressing various compartments of the diatom model organism *Phaeodactylum tricornutum* via sub-cellular marker proteins. *Algal Res.* **20**, 249–257 (2016).
49. Kilian, O. & Kroth, P. G. Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *Plant J.* **41**, 175–183 (2004).
50. Moog, D., Stork, S., Zauner, S. & Maier, U.-G. In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. *Genome Biol. Evol.* **3**, 375–382 (2011).
51. Guillard R. L. G. Culture of phytoplankton for feeding marine invertebrates in *Culture of marine invertebrate animals* (eds. Chanley, M.H. & Smith, W.L.) 29–60 (Springer, 1975).
52. Shevchenko, A., Tomas, H., Havliš, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2007).
53. Nguyen, T. N. M., Berzano, M., Gualerzi, C. O. & Spurio, R. Development of molecular tools for the detection of freshwater diatoms. *J. Microbiol. Methods* **84**, 33–40 (2011).
54. Tanaka, Y., Nakatsuma, D., Harada, H., Ishida, M. & Matsuda, Y. Localization of soluble β -carbonic anhydrase in the marine diatom *Phaeodactylum tricornutum*. Sorting to the chloroplast and cluster formation on the girdle lamellae. *Plant Physiol.* **138**, 207–217 (2005).

Acknowledgements

We would like to thank Nuri Oh (Seoul National University) for assistance with the statistical analysis. This study is part of a project titled 'Development of novel promoters from microalgae (*Phaeodactylum tricornutum*) and production of functional peptides (hFGF1, hFGF2 and hIGF1) as cosmetic raw materials', funded by the Ministry of Oceans and Fisheries, Korea (Grant No. 20180298). C.J. was supported in part by a Research Resettlement Fund for the new faculty of Seoul National University.

Author Contributions

E.E.O., B.K.S., and C.H.P. conceived the experiments. E.E.O., B.K.S. and B.K. conducted the experiments. E.E.O., B.K.S., C.J. and C.H.P. analysed the results. E.E.O., C.J. and C.H.P. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-45786-9>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2019