Received: 19 March 2021 | Accepted: 14 September 2021

DOI: 10.1002/mbo3.1239

ORIGINAL ARTICLE

MicrobiologyOpen Open Access WILEY

Characterization of endogenous promoters of *GapC1* and *GS* for recombinant protein expression in *Phaeodactylum tricornutum*

Erdenedolgor Erdene-Ochir[1,2] | Bok-Kyu Shin[3] | Md Nazmul Huda[1,2] | Eun Ha Lee[1] | Dae-Geun Song[1] | Choonkyun Jung[4,5] | Cheol-Ho Pan[1,2,6]

[1]Natural Product Informatics Research Center, KIST Gangneung Institute of Natural Products, Gangneung, Republic of Korea

[2]Division of Bio-Medical Science and Technology, KIST School, Korea University of Science and Technology, Seoul, Republic of Korea

[3]Algaeprona Inc, Gangneung, Republic of Korea

[4]Department of International Agricultural Technology and Crop Biotechnology Institute/GreenBio Science and Technology, Seoul National University, Pyeongchang, Republic of Korea

[5]Department of Agriculture, Forestry, and Bioresources and Integrated Major in Global Smart Farm, College of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea

[6]Microalgae Ask Us Co., Ltd., Gangneung, Republic of Korea

Correspondence
Choonkyun Jung, Graduate School of International Agricultural Technology and Crop Biotechnology Institute/GreenBio Science and Technology, Seoul National University, Pyeongchang 25354, Republic of Korea.
Email: jasmin@snu.ac.kr

Cheol-Ho Pan, Natural Product Informatics Research Center, KIST Gangneung Institute of Natural Products, Gangneung 25451, Republic of Korea.
Email: panc@kist.re.kr

Funding information
Korea Institute of Science and Technology Intramural Research Grant, Grant/Award Number: 2Z06483; Seoul National University

Abstract

Although diatoms have been utilized as a cellular factory to produce biopharmaceuticals, recombinant proteins, and biofuels, only a few numbers of gene promoters are available. Therefore, the development of novel endogenous promoters is essential for the production of a range of bioactive substances. Here, we characterized the activities of endogenous promoters *glyceraldehyde-3-phosphate dehydrogenase* (*GapC1*) and *glutamine synthetase* (*GS*) of *Phaeodactylum tricornutum* using green fluorescent protein (GFP) under different culture conditions. Compared with the widely used fucoxanthin chlorophyll-binding protein A (*fcpA*) promoter, the *GS* promoter constitutively drove the expression of GFP throughout all growth phases of *P. tricornutum*, regardless of culture conditions. Additionally, the GFP level driven by the *GapC1* promoter was the highest at the log phase, similar to the *fcpA* promoter, and increased light and nitrogen-starvation conditions reduced GFP levels by inhibiting promoter activity. These results suggested that the *GS* promoter could be utilized as a strong endogenous promoter for the genetic engineering of *P. tricornutum*.

KEYWORDS
diatom, endogenous promoter, glutamine synthetase, glyceraldehyde-3-phosphate dehydrogenase, *Phaeodactylum tricornutum*

1 | INTRODUCTION

Diatoms are unicellular, eukaryotic phytoplankton that thrives since the Oligocene about 30 million years ago (Falkowski et al., 2004). Diatoms live in both marine and freshwater environments and account for about 20% of the total photosynthetic productivity (Bowler et al., 2008; Maheswari et al., 2010). They are currently considered among the most productive and flexible microalgae, with leading roles in the ocean food chain.

The entire genome of *Phaeodactylum tricornutum* is about 27.6 Mb and contains 33 chromosomes harboring 12,177 predicted genes (Rastogi et al., 2018). Previously, 130,000 expressed sequence tags (ESTs) were determined from *P. tricornutum* cells grown in 16 different conditions, including various nitrogen sources; different carbon dioxide, silicate, and iron concentrations; different morphotypes and lighting sources; and abiotic stress, including low temperature and low salinity (Maheswari et al., 2005, 2010). Molecular tools have also been developed for the genetic manipulation of *P. tricornutum* (Apt et al., 1996; De Riso et al., 2009; Karas et al., 2015; Maheswari et al., 2005; Nymark et al., 2016; Rastogi et al., 2018; Siaut et al., 2007).

Diatoms have been extensively studied for various biotechnological purposes and can be utilized to produce biopharmaceuticals and secondary metabolites (Hempel et al., 2011; Mathieu-Rivet et al., 2014). A constitutive promoter driving high recombinant protein yields is not only essential for developing a cost-efficient expression system but also necessary for metabolic engineering by gene regulation. Heterologous promoters originating from various species have been used to express recombinant proteins in *P. tricornutum* (Gorman et al., 1982; Harada et al., 2005; Poulsen & Kroger, 2005; Sanders et al., 1987; Tomaru et al., 2008, 2011, 2012). Additionally, endogenous promoters for inducible nitrate reductase (Chu et al., 2016; Hempel et al., 2011; Niu et al., 2012) and light-inducible fucoxanthin chlorophyll of light-harvesting antennae complexes (*fcp*) encoding fcpA–E (Apt et al., 1996; De Riso et al., 2009; Joshi-Deo et al., 2010; Siaut et al., 2007; Zaslavskaia et al., 2000) have been used in *P. tricornutum*. Furthermore, the promoters of *elongation factor 2*, *β-carbonic anhydrase 1*, *acyl-CoA*: *diacylglycerol acyltransferase 1*, and *highly abundant secreted protein 1* (*HASP1*) from *P. tricornutum* were fused with a reporter gene to evaluate reporter expression (Erdene-Ochir et al., 2019; Harada et al., 2005; Ohno et al., 2012; Shemesh et al., 2016). These studies focused on evaluating strong constitutive promoters capable of expressing large quantities of protein inside or secreted from *P. tricornutum*. The highest level of protein amount is required during the stationary phase of cell culture to maximize productivity.

Here, we searched for a novel candidate promoter of genes encoding proteins strongly expressed during the stationary phase. We identified *glyceraldehyde-3-phosphate dehydrogenase* (*GapC1*) and *glutamine synthetase* (*GS*) promoters for constitutive expression of recombinant protein in *P. tricornutum* and constructed a green fluorescent protein (GFP)-reporter system using a truncated version of their promoter regions. Following transformation of *P. tricornutum* with these constructs, we tested them for their ability to constitutively express downstream gene products under different culture conditions.

2 | MATERIALS AND METHODS

2.1 | Cell culture

P. tricornutum Bohlin UTEX 646 strain was purchased from the UTEX Culture Collection of Algae (The University of Texas, Austin, TX, USA). *P. tricornutum* was cultivated in F/2 media (Guillard et al., 1975), at 20°C with shaking at 200 rpm and with or without nitrogen under constant lighting from white fluorescent lamps (1600 or 3000 lux).

2.2 | Protein identification

Using cell culture at stationary phase, SDS-PAGE, in-gel digestion, and LC-MS/MS analysis were performed, and proteins were identified by database searches as previously described (Erdene-Ochir et al., 2016, 2019).

2.3 | *In silico* analysis of potential regulatory elements in *GapC1* and *GS* promoters

The 5' upstream regions of *GapC1* (NCBI ID: XP_002182291; Uniprot accession number: B7G5Q1) and *GS* (NCBI ID: XP_002182898; Uniprot accession number: B7G6Q6) were extracted from EnsemblProtists (Kersey et al., 2014) using the Biomart tool (Smedley et al., 2015) and analyzed for *cis*-acting elements by PlantCARE (Lescot et al., 2002). Sequence-based single-site analysis (SSA) and transcription factor-binding site (TFBS) cluster analysis (TCA) using oPOSSUM (v.3.0) (Kwon et al., 2012) were performed to identify consensus TFBSs in *GapC1* and *GS* promoters. These were also checked using the Melina II web tool (Okumura et al., 2007).

2.4 | Rapid amplification of complementary DNA ends (RACE)

Total RNA was isolated using RNAiso Plus reagent (Takara Bio, Shiga, Japan) according to the manufacturer's instructions. Total RNA (1 μg) was subjected to 5' and 3' RACE, performed as previously described (Pinto & Lindblad, 2010) with minor modifications. The primers used in RACE are listed in Table A1.

2.5 | Construction of plasmid vectors

The *CIP1* promoter (Kadono et al., 2015) and fragments of the GS (501 and 996 bp) (Erdene-Ochir et al., 2016) and *GapC1* (500 and

1086 bp) promoters were amplified by PCR from genomic DNA and cloned into the pPha-T1 vector using *Nde*I and *Eco*RI sites (Zaslavskaia et al., 2000). The primers are listed in Table A1. The GFP-encoding gene was amplified by PCR from the pEGFP-C2 vector and cloned into the pPha-T1 vector using *Eco*RI and *Bam*HI sites.

2.6 | Transformation of *P. tricornutum*

Particle bombardment-mediated transformation and PCR-based transformant selection were performed as previously described (Erdene-Ochir et al., 2019). Primers used in genomic DNA PCR are listed in Table A1.

2.7 | Total RNA isolation and real-time PCR analysis

Eight or four milliliters of cell culture grown for 6 or 11 days in culture Condition 1 were centrifuged at 1200 *g* for 15 min at 4°C. Total RNA isolation and RT-PCR analysis were performed as previously described (Erdene-Ochir et al., 2019).

2.8 | GFP fluorescence measurement

Fluorescence was measured as previously described (Erdene-Ochir et al., 2016, 2019). The autofluorescence value of the *fcpApro* construct was removed from the GFP fluorescence value obtained with the *CIP1*, *GapC1*, and *GS* constructs. Using a recombinant *E. coli* GFP protein (ab119740; Abcam, Cambridge, UK), a GFP standard curve was generated. Measurements were conducted using biological triplicates.

2.9 | Western blot analysis

Cell lysis and protein quantification were performed as described previously (Erdene-Ochir et al., 2019). Total soluble protein (7 μg) was resolved on 12% Tris-glycine SDS-PAGE and transferred to PVDF membrane, which was incubated with anti-GFP goat antibody (Abcam, Cambridge, UK) and anti-goat HRP-conjugated bovine antibody (Santa Cruz Biotechnology, Dallas, TX, USA). Western blot signals were detected using SuperSignal West Femto substrate (Thermo Fisher Scientific, Waltham, MA, USA).

2.10 | Subcellular localization of GFP

GFP images at mid-log and stationary phases were obtained using a Leica confocal microscope (Leica Biosystems, Wetzlar, Germany) (Erdene-Ochir et al., 2019; Tanaka et al., 2005).

2.11 | Statistical analysis

Data are expressed as mean ± SD. Statistical analysis was conducted using Student's *t*-test and one-way analysis of variance, followed by Duncan test for multiple comparisons. A $p < 0.05$ was regarded as statistically significant.

3 | RESULTS

3.1 | Proteomics-based identification of the most abundant proteins at the stationary phase

A previous study used LC-MS/MS analysis to identify a total of 1,836 proteins abundant during the stationary phase (Erdene-Ochir et al., 2016). The most abundant of these proteins was fcp binding protein E (FcpE), identified by database searching with a 23% sequence coverage (Apt et al., 1996). The second most abundant protein (PHATRDRAFT_22357) was annotated via homology as GS, involved in nitrogen assimilation (Erdene-Ochir et al., 2016). The third most abundant protein (PHATRDRAFT_22122) was GapC1 (Erdene-Ochir et al., 2016); therefore, we selected GS and GapC1 for further analysis. LC-MS/MS sequence coverage and the spectra for GapC1 are shown in Figure A1.

3.2 | *In silico* analysis of potential promoters

Using PlantCARE, the *GapC1* and *GS* promoter regions were analyzed for cis-acting regulatory elements (Lescot et al., 2002). All the light-responsive elements shown in Figures A2 and A3 were identified by PlantCARE. We evaluated the SSA and TCA using oPOSSUM (Kwon et al., 2012) and default parameters, with a threshold of >95% and *Z*-score >4 used as a threshold for SSA (Figures A2 and A3). The identified transcription factors were cross-checked against previous results (Rayko et al., 2010). Consensus sequences in the *GapC1* and *GS* promoters analyzed using Melina II (Okumura et al., 2007) identified two conserved motifs (CACACACA and GACACACG).

3.3 | RACE

GapC1 and *GS* are located on chromosomes 15 and 17 of *P. tricornutum*, respectively (Fabris et al., 2012). The transcription start site (TSS) for *GS* was identified by 5' RACE along with an initiator-like sequence (Kadono et al., 2015) in the *GS* promoter (Figure A3). Additionally, we identified the untranslated 5' and 3' regions (UTRs) of *GS* as 214 bp and 144 bp from the start and stop codons, respectively (Figure A4a). The 3' UTR of *GapC1* was 279 bp from the stop codon; however, we were unable to determine the *GapC1* TSS, although the predicted TSS is 61 bp (Grillo et al., 2010).

3.4 | Isolation of endogenous promoters of *GapC1* and *GS*

The pPha-T1 vector containing a *fcpA* promoter-driven zeocin-resistance gene (Zaslavskaia et al., 2000) was used for all plasmid constructions as a backbone (Figure A4b). The promoters widely used for the genetic manipulations of *P. tricornutum* are ~500 bp (Apt et al., 1996; Kadono et al., 2015). Using the predicted 61-bp long 5' UTR for *GapC1*, we cloned 500- and 1086-bp 5' UTRs as potential *GapC1* promoter regions (Grillo et al., 2010). *GS* 5' UTRs of 501 bp and 996 bp were extracted using the Biomart tool from Ensembl Protists (Kinsella et al., 2011). The *fcpA* (442 bp) and *CIP1* (502 bp) promoters were used to drive reporter-protein expression as endogenous and heterologous constitutive promoters, respectively (Apt et al., 1996; Kadono et al., 2015). The *fcpA* promoter activity was the highest at the log phase, whereas the *CIP1* promoter activity was the highest at the stationary phase. As a mock construct, the pPha-T1 vector carrying the *fcpA* promoter but without the *gfp* gene was used (Figure A4b).

3.5 | *P. tricornutum* transformation and transformant selection

All constructs were transformed into stationary phase cells. After 4 weeks, we observed 459 zeocin-resistant colonies following selection of transformants on f/2 agar including 100 μg/ml zeocin. Transformation of the *GS-501pro:GFP* construct resulted in 175 zeocin-resistant colonies, whereas other constructs showed relatively low numbers of resistant colonies. All zeocin-resistant colonies were moved to liquid f/2 medium including 100 μg/ml zeocin, followed by selection by PCR analysis (Figure A4c); 72% of the zeocin-resistant colonies contained the appropriate promoter and *gfp*. These colonies were then selected by GFP fluorescence, with 42% of the colonies expressing the GFP reporter. Based on these findings, we selected three colonies for each construct for further analysis.

3.6 | Assessment of culture conditions

Multiple factors, including temperature, lighting intensity, nutrition source, and aeration, influence cell growth. To determine the most favorable conditions for promoter function, the selected colonies were cultivated under different culture conditions. First, the selected colonies were grown in f/2 liquid medium including 50% artificial seawater, 100 μg/ml zeocin, and mixed antibiotics at 20°C and 200 rpm under continuous aeration and constant lighting (1600 lux), until the stationary phase (Condition 1). The cells were seeded at 10^5 cells/ml on day 0 and cultivated to ~10^7 cells/ml on day 10, with cell density and GFP expression checked daily. The cell-growth curve revealed days 6 and 11 as mid-log and stationary phases, respectively (Figure 1a). The presence of the transgene did not affect the growth rate of *P. tricornutum* cells in all cases. Cell autofluorescence driven by the blank construct *fcpApro* was subtracted from GFP fluorescence in the target cells to assess promoter-specific fluorescence intensity. The *fcpApro*: *GFP*, *GapC1-500pro*:*GFP*, and *GapC1-1086pro*:*GFP* constructs showed peak GFP-expression levels at the log phase; thereafter, it decreased until the stationary phase (day 11) (Figure 1c). The reporter-protein levels relative to *GapC1-500pro*:*GFP* and *GapC1-1086pro*:*GFP* expression were similar to that of *fcpApro*:*GFP* (Figure 1b, c). Interestingly, *GS-501pro*:*GFP* and *GS-996pro*:*GFP* promoter-driven constructs indicated constitutive GFP expression in proportion to cell number from the early log to the stationary phase, with GFP expression in *GS-501pro*:*GFP* and *GS-996pro*:*GFP* constructs >fourfold and >sixfold higher, respectively, than that of *fcpApro*:*GFP* construct during the stationary phase (day 11) (Figure 1b, d). These results obtained from immunoblotting and fluorescence measurements were consistent with the levels of GFP

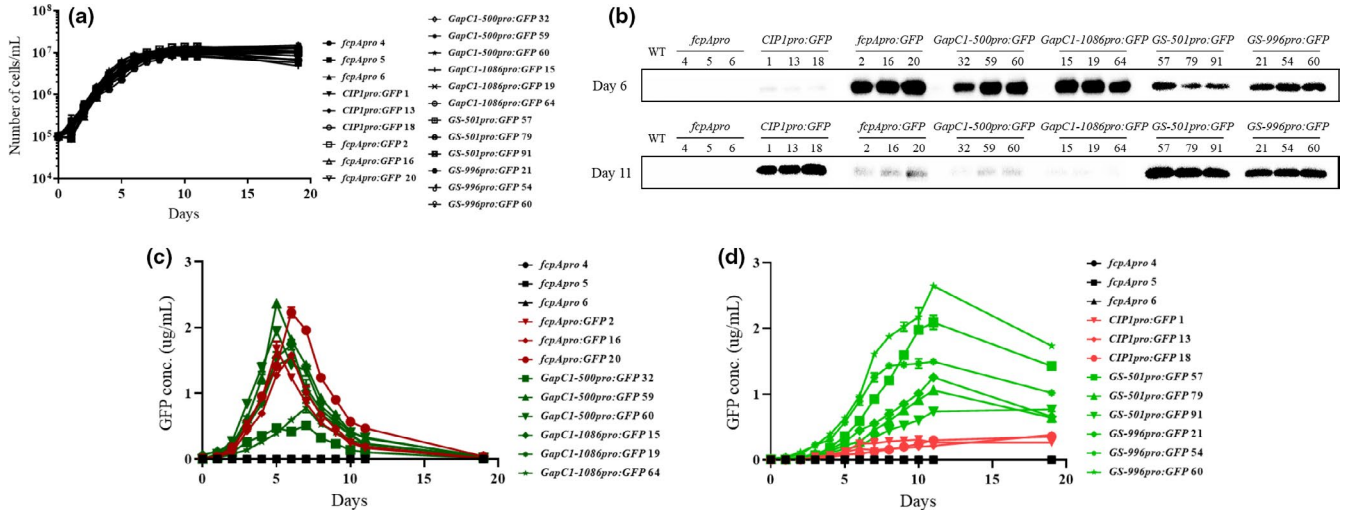


FIGURE 1 Growth curves of all transgenic lines and GFP expression level in Condition 1. (a) Growth curves of *P. tricornutum* cultures. All transgenic lines were cultivated for 19 days. (b) GFP protein levels in cell lysates on days 6 and 11 and determined by immunoblot. Levels of GFP fluorescence in cell lysates of (c) *GapC1pro*:*GFP* and (d) *GSpro*:*GFP* transgenic lines were measured by a fluorometer

mRNAs at the mid-log (day 6) and stationary (day 11) phases of cultivation under Condition 1, showing that GFP expression was driven by *GapC1* and *GS* promoters (Figure A5). However, we did not observe the expected result from the *CIP1pro*:*GFP* construct (Kadono et al., 2015) under these culture conditions, and GFP expression by *CIP1pro*: *GFP* was lower than that by other promoters (Figure 1d). Western blot results agreed with all observed patterns of GFP fluorescence (Figure 1b).

Light intensity is a key factor for microalgal growth, as they are eukaryotic phytoplankton capable of fixing carbon and nitrogen while producing oxygen through photosynthesis (Saade & Bowler, 2009). Therefore, we changed the lighting intensity to 3000 lux and incubated cells at 20°C with continuous aeration and constant lighting until the stationary phase (Condition 2). Cells were seeded at 10^6 cells/mL on day 0 and cultivated to ~10^7 cells/ml on day 8, with cell-growth curves showing that days 4 and 8 represented the mid-log phase and stationary phases, respectively (Figure 2a). GFP expression, driven by *fcpApro*:*GFP*, was twofold lower than that by the same promoter under 1600 lux at log phase, whereas GFP expression by *CIP1pro*:*GFP* gradually increased from the early log to the stationary phase and was higher than that by the same promoter in Condition 1 (Figure 2b). Interestingly, GFP expression driven by *GS-501pro*:*GFP* and *GS-996pro*:*GFP* increased from the early log to the stationary phase, which was not tested in the previous study (Erdene-Ochir et al., 2016). Additionally, GFP expression in the *GapC1pro*:*GFP* construct was twofold higher than that of *fcpApro*:*GFP* at the log phase but twofold lower than that by the same promoter in Condition 1 (Figure 2b). Western blot results agreed with all observed patterns of GFP fluorescence (Figure 2c).

Most industrial applications of *P. tricornutum* are related to the development of oil-producing cell lines under starvation conditions, such as nitrogen-free medium. Considering this, a promoter capable of driving strong constitutive expression of a protein of interest under nitrogen-free conditions will be important for engineering cells for oil production. Therefore, cells were seeded at 10^6 cells/ml in nitrogen-free f/2 medium and cultivated at 20°C with continuous aeration and constant lighting at 3000 lux (Condition 3). During 8-day cultivation, cells showed decreased growth relative to that under previous culture conditions (Figure A6a), and GFP levels in the *fcpA pro*:*GFP*, *GapC1-500 pro*:*GFP*, and *GapC1-1086 pro*:*GFP* constructs were <20 ng/ml (Figure A6b), with the *CIP1pro*:*GFP* construct showing higher GFP expression than *fcpApro*:*GFP*, *GapC1-500 pro*:*GFP*, and *GapC1-1086 pro*:*GFP*. Although GFP expression by *GS-501pro*:*GFP* and *GS-996pro*:*GFP* was lower than that by the same promoter under Conditions 1 and 2, GFP levels were higher than other constructs, which was not tested in the previous study (Erdene-Ochir et al., 2016).

3.7 | GFP localization

Images of GFP localization at the mid-log (day 6) and stationary (day 11) phases of cultivation under Condition 1 (Figures 3 and A7) showed that GFP signals in *fcpApro*:*GFP*, *CIP1pro*:*GFP*, *GapC1-500pro*:*GFP*, *GapC1-1086pro*:*GFP*, *GS-501pro*:*GFP*, and *GS-996pro*:*GFP* constructs accumulated in the cytoplasm and were directly proportional to the strength of the promoters at each growth phase (Figure 3a, b). GFP fluorescence in the *GapC1pro*:*GFP* transgenic line was the highest at the log phase but almost disappeared at the stationary phase. In contrast, GFP fluorescence in the *GSpro*:*GFP* transgenic line largely increased from the log to the stationary phase. These results agreed

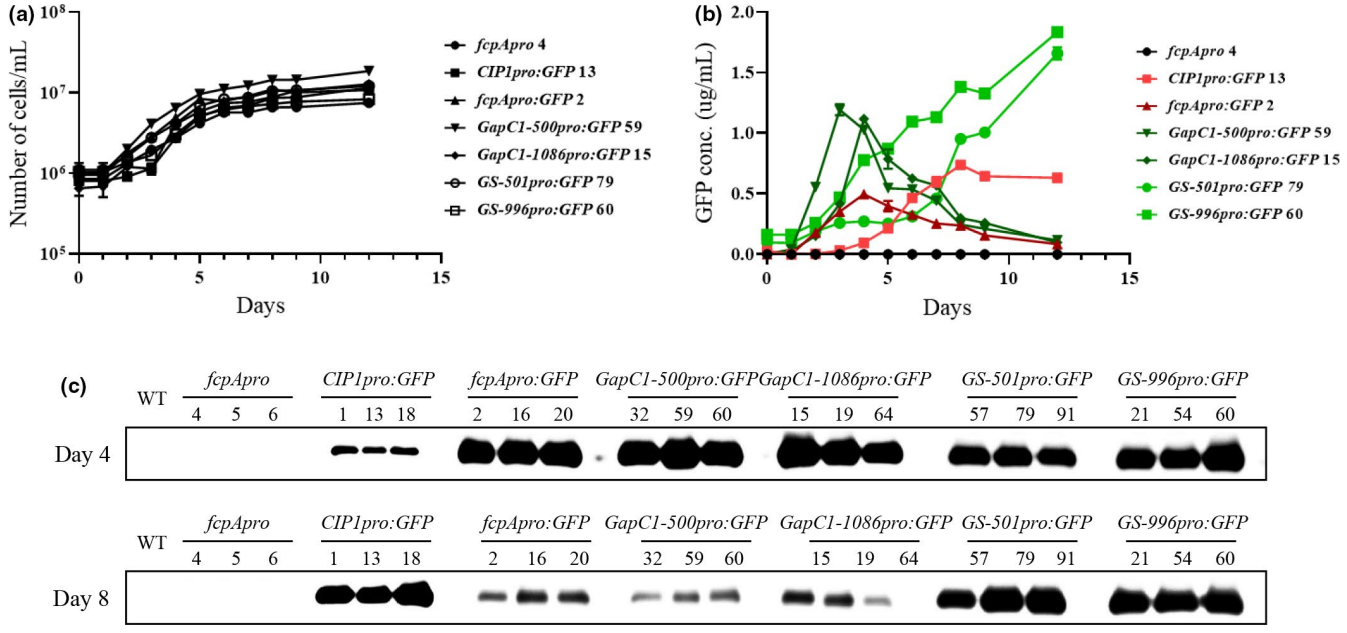


FIGURE 2 Growth curves of all transgenic lines and GFP expression levels in Condition 2. (a) Growth curves of *P. tricornutum* cultures. All transgenic lines were cultivated for 12 days. (b) Levels of GFP fluorescence in cell lysates of *GapC1pro*:*GFP* and *GSpro*:*GFP* transgenic lines were measured by a fluorometer. (c) GFP protein levels in cell lysates on days 6 and 11 were determined by immunoblot

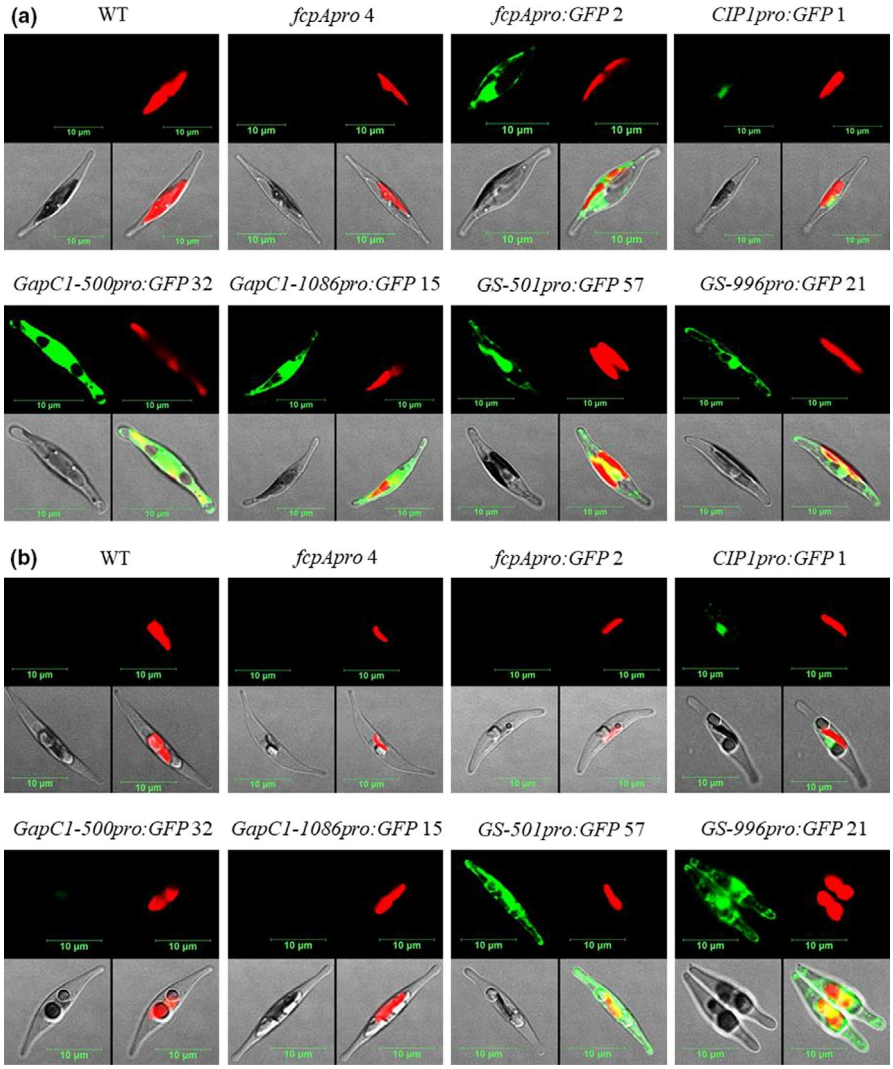


FIGURE 3 Subcellular localization of GFP in transgenic *P. tricornutum*. GFP fluorescence and chlorophyll fluorescence in transgenic lines at (a) mid-log (day 6) and (b) stationary (day 11) phases and visualized by confocal microscopy. Numbers on the images show independent transgenic lines for each construct. Scale bars = 10 µm

with all observed patterns of GFP fluorescence and Western blot results.

4 | DISCUSSION

A strong constitutive promoter able to drive the expression of large quantities of protein in a host organism is one of the most significant genetic engineering tools for foreign protein expression and metabolic engineering. To maximize the productivity of protein of interest, the present study focused on identifying novel candidate promoters driving strong protein expression during the stationary phase of *P. tricornutum*. Thus, we identified GapC1 and GS among 1836 proteins (Figure A1) (Erdene-Ochir et al., 2016) and cloned their promoters into transformation vectors to evaluate their efficacy for overexpression of target proteins in the *P. tricornutum* host, with the previously reported *fcpA* and *CIP1* promoters used as positive controls.

The *fcpApro*:*GFP* construct showed increased GFP expression from the lag to log phase, followed by decreased expression from the log to stationary phase, with similar levels to those of the *CIP1pro*:*GFP* construct during the early stationary phase (Figure 1b, c). In a previous study, *CIP1* promoter resulted in threefold higher levels of reporter-protein expression relative to that driven by the *fcpA* promoter during the stationary phase (Kadono et al., 2015). This observed difference in the *CIP1* promoter activity could be due to the different experimental conditions, especially light intensity; therefore, we increased the light intensity to 3000 lux during cultivation, which resulted in the reported 3:1 *CIP1*:*fcpA* ratio of GFP expression, suggesting that the *CIP1* promoter could be a light-responsive (Figure 2b, c). Additionally, the *GS* promoter regions (501 and 996 bp) were able to drive downstream gene expression, resulting in up to fourfold higher reporter-protein expression relative to that of the *fcpA* promoter during the stationary phase and under different growth conditions (Figures 1, 2, and A6). Consequently, the *GS* promoter drove strong constitutive expression of the reporter protein, irrespective of the cell-growth phase. Moreover, these levels were also higher than the GFP expression driven by the *CIP1* promoter under optimal conditions (Figures 1, 2, and A6). Although GFP-expression levels driven by *GS-501pro*:*GFP* and *GS-996pro*:*GFP* constructs differed according to culture condition, the expression patterns were similar. Furthermore, GFP levels and patterns driven

by the *GapC1-500pro*:*GFP* and *GapC1-1086pro*:*GFP* constructs were similar to that of *fcpApro*:*GFP* (Figures 1, 2, and A6). Because the *fcpA* promoter is widely used for the genetic engineering of *P. tricornutum*, these results suggest that both the *GapC1* and the *GS* promoters can be used to genetically engineer this strain. Further study is needed to elucidate the functions of the *GapC1* and *GS* promoters for expressing specific targets, including antibodies and recombinant proteins, as well as the use of the *GS* promoter for metabolic engineering of *P. tricornutum* to promote increased oil production.

ACKNOWLEDGEMENTS

This research was funded by the Korea Institute of Science and Technology Intramural Research Grant (2Z06483). C.J. was supported by Creative-Pioneering Researchers Program through Seoul National University (SNU).

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Erdenedolgor Erdene-Ochir: Data curation (equal); Investigation (lead); Methodology (equal); Resources (lead); Visualization (equal); Writing-original draft (equal). **Bok-Kyu Shin:** Funding acquisition (supporting); Investigation (supporting); Resources (supporting). **Md Nazmul Huda:** Investigation (supporting); Resources (supporting). **Eun Ha Lee:** Investigation (supporting); Resources (supporting). **Dae-Geun Song:** Investigation (supporting); Resources (supporting). **Choonkyun Jung:** Data curation (equal); Supervision (equal); Visualization (equal); Writing-original draft (equal). **Cheol-Ho Pan:** Conceptualization (lead); Funding acquisition (lead); Methodology (equal); Project administration (lead); Supervision (lead); Writing-original draft (equal).

ETHICS STATEMENT

None required.

DATA AVAILABILITY STATEMENT

All data are provided in full in this paper.

ORCID

Choonkyun Jung https://orcid.org/0000-0001-5643-309X

REFERENCES

Apt, K. E., Kroth-Pancic, P. G., & Grossman, A. R. (1996). Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Molecular and General Genetics*, *252*, 572–579.

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., … Grigoriev, I. V. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature*, *456*, 239–244.

Chu, L., Ewe, D., Río Bártulos, C., Kroth, P. G., & Gruber, A. (2016). Rapid induction of GFP expression by the nitrate reductase promoter in the diatom *Phaeodactylum tricornutum*. *PeerJ*, *4*, e2344.

De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C., & Falciatore, A. (2009). Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Research*, *37*, e96.

Erdene-Ochir, E., Shin, B.-K., Huda, M. N., Kim, D. H., Lee, E. H., Song, D.-G., Kim, Y.-M., Kim, S. M., & Pan, C.-H. (2016). Cloning of a novel endogenous promoter for foreign gene expression in *Phaeodactylum tricornutum*. *Applied Biological Chemistry*, *59*, 861–867.

Erdene-Ochir, E., Shin, B. K., Kwon, B., Jung, C., & Pan, C.-H. (2019). Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from *Phaeodactylum tricornutum*. *Scientific Reports*, *9*, 9941.

Fabris, M., Matthijs, M., Rombauts, S., Vyverman, W., Goossens, A., & Baart, G. J. E. (2012). The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *The Plant Journal*, *70*, 1004–1014.

Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O., & Taylor, F. J. R. (2004). The evolution of modern eukaryotic phytoplankton. *Scimago Magazine*, *305*, 354–360.

Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I., & Howard, B. H. (1982). The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. *Biochemistry*, *79*, 6777–6781.

Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S., Banfi, S., Gennarino, V. A., Horner, D. S., Pavesi, G., Picardi, E., & Pesole, G. (2010). UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Research*, *38*, D75–D80.

Guillard, R. R. L., Smith, W. L., & Chanley, M. H. (1975). *Culture of Phytoplankton for Feeding Marine Invertebrates* (1st ed., pp. 29–60). Springer.

Harada, H., Nakatsuma, D., Ishida, M., & Matsuda, Y. (2005). Regulation of the expression of intracellular beta-carbonic anhydrase in response to CO_2 and light in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiology*, *139*, 1041–1050.

Hempel, F., Lau, J., Klingl, A., & Maier, U. G. (2011). Algae as protein factories: Expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PLoS One*, *6*, e28424.

Joshi-Deo, J., Schmidt, M., Gruber, A., Weisheit, W., Mittag, M., Kroth, P. G., & Biichel, C. (2010). Characterization of a trimeric light-harvesting complex in the diatom *Phaeodactylum tricornutum* built of FcpA and FcpE proteins. *Journal of Experimental Botany*, *61*, 3079–3087.

Kadono, T., Miyagawa-Yamaguchi, A., Kira, N., Tomaru, Y., Okami, T., Yoshimatsu, T., Hou, L., Ohama, T., Fukunaga, K., Okauchi, M., Yamaguchi, H., Ohnishi, K., Falciatore, A., & Adachi, M. (2015). Characterization of marine diatom-infecting virus promoters in the model diatom *Phaeodactylum tricornutum*. *Scientific Reports*, *5*, 18708.

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P. R., Noddings, C. M., Brunson, J. K., Valas, R. E., Deerinck, T. J., Jablanovic, J., Gillard, J. T. F., Beeri, K., Ellisman, M. H., Glass, J. I., Hutchison III, C. A., Smith, H. O., Venter, J. C., Allen, A. E., Dupont, C. L., & Weyman, P. D. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Comm*, *6*, 6925.

Kersey, P. J., Allen, J. E., Christensen, M., Davis, P., Falin, L. J., Grabmueller, C., Hughes, D. S. T., Humphrey, J., Kerhornou, A., Khobova, J., Langridge, N., McDowall, M. D., Maheswari, U., Maslen, G., Nuhn, M., Ong, C. K., Paulini, M., Pedro, H., Toneva, I., … Staines, M. (2014). Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Research*, *42*, D546–D552.

Kinsella, R. J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., & Flicek, P. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database*, *2011*, bar030.

Kwon, A. T., Arenillas, D. J., Worsley Hunt, R., & Wasserman, W. (2012). oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets. *G3*, *2*, 987–1002.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P., & Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements, and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, *30*(1), 325–327.

Maheswari, U., Jabbari, K., Petit, J.–L., Porcel, B. M., Allen, A. E., Cadoret, J.–P., De Martino, A., Heijde, M., Kaas, R., La Roche, J., Lopez, P. J., Martin-Jézéquel, V., Meichenin, A., Mock, T., Parker, M. S., Vardi, A., Armbrust, E. V., Weissenbach, J., Katinka, M., & Bowler, C. (2010). Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biology*, *11*, R85.

Maheswari, U., Montsant, A., Goll, J., Krishnasamy, S., Rajyashri, K. R., Patell, V. M., & Bowler, C. (2005). The diatom EST database. *Nucleic Acids Research*, *33*(suppl_1), D344–D347.

Mathieu-Rivet, E., Kiefer-Meyer, M. C., Vanier, G., Ovide, C., Burel, C., Lerouge, P., & Bardor, M. (2014). Protein N-glycosylation in eukaryotic microalgae, and its impact on the production of nuclear expressed biopharmaceuticals. *Frontiers Plant Science*, *5*, 359.

Niu, Y.-F., Yang, Z.-K., Zhang, M.-H., Zhu, C.-C., Yang, W.-D., Liu, J.-S., & Li, H.-Y. (2012). Transformation of diatom *Phaeodactylum tricornutum* by electroporation and establishment of inducible selection marker. *BioTechniques*, *52*.

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., & Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Scientific Reports*, *6*, 24951.

Ohno, N., Inoue, T., Yamashiki, R., Nakajima, K., Kitahara, Y., Ishibashi, M., & Matsuda, Y. (2012). CO_2-cAMP-responsive cis-elements targeted by a transcription factor with CREB/ATF-like basic zipper domain in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiology*, *158*, 499–513.

Okumura, T., Makiguchi, H., Makita, Y., Yamashita, R., & Nakai, K. (2007). Melina II: a web tool for comparisons among several predictive algorithms to find potential motifs from promoter regions. *Nucleic Acids Research*, *35*, W227–W231.

Pinto, F. L., & Lindblad, P. (2010). A guide for in-house design of template-switch-based 5′ rapid amplification of cDNA ends systems. *Analytical Biochemistry*, *397*(2), 227–232.

Poulsen, N., & Kroger, N. (2005). A new molecular tool for transgenic diatoms: control of mRNA and protein biosynthesis by an inducible promoter-terminator cassette. *FEBS Journal*, *272*(13), 3413–3423.

Rastogi, A., Maheswari, U., Dorrell, R. G., Vieira, F. R. J., Maumus, F., Kustka, A., McCarthy, J., Allen, A. E., Kersey, P., Bowler, C., & Tirichine, L. (2018). Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Scientific Reports*, *8*, 4834.

Rayko, E., Maumus, F., Maheswari, U., Jabbari, K., & Bowler, C. (2010). Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytologist*, *188*, 52–66.

Saade, A., & Bowler, C. (2009). Molecular tools for discovering the secrets of diatoms. *BioScience*, *59*, 757–765.

Sanders, P. R., Winter, J. A., Barnason, A. R., Rogers, S. G., & Fraley, R. T. (1987). Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. *Nucleic Acids Research*, *15*, 1543–1558.

Shemesh, Z., Leu, S., Khozin-Goldberg, I., Didi-Cohen, S., Zarka, A., & Boussiba, S. (2016). Inducible expression of *Haematococcus* oil globule protein in the diatom *Phaeodactylum tricornutum*: association with lipid droplets and enhancement of TAG accumulation under nitrogen starvation. *Algal Research*, *18*, 321–331.

Siaut, M., Heijde, M., Mangogna, M., Montsant, A., Coesel, S., Allen, A., Manfredonia, A., Falciatore, A., & Bowler, C. (2007). Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene*, *406*, 23–35.

Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., Arnaiz, O., Awedh, M. H., Baldock, R., Barbiera, G., Bardou, P., Beck, T., Blake, A., Bonierbale, M., Brookes, A. J., Bucci, G., Buetti, I., Burge, S., Cabau, C., … Kasprzyk, A. (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Research*, *43*, W589–W598.

Tanaka, Y., Nakatsuma, D., Harada, H., Ishida, M., & Matsuda, Y. (2005). Localization of soluble beta-carbonic anhydrase in the marine diatom *Phaeodactylum tricornutum*. Sorting to the chloroplast and cluster formation on the girdle lamellae. *Plant Physiology*, *138*, 207–217.

Tomaru, Y., Shirai, Y., Suzuki, H., Nagasaki, T., & Nagumo, T. (2008). Isolation and characterization of a new single-stranded DNA virus infecting the cosmopolitan marine diatom *Chaetoceros debilis*. *Aquatic Microbial Ecology*, *50*, 103–112.

Tomaru, Y., Takao, Y., Suzuki, H., Nagumo, T., Koike, K., & Nagasaki, K. (2011). Isolation and characterization of a single-stranded DNA virus infecting *Chaetoceros lorenzianus* Grunow. *Applied and Environmental Microbiology*, *77*, 5285–5293.

Tomaru, Y., Toyoda, K., Kimura, K., Hata, N., Yoshida, M., & Nagasaki, K. (2012). First evidence for the existence of pennate diatom viruses. *ISME Journal*, *6*, 1445–1448.

Zaslavskaia, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R., & Apt, K. E. (2000). Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. *Journal of Phycology*, *36*, 379–386.

How to cite this article: Erdene-Ochir, E., Shin, B.-K., Huda, M. N., Lee, E. H., Song, D.-G., Jung, C., & Pan, C.-H. (2021). Characterization of endogenous promoters of *GapC1* and *GS* for recombinant protein expression in *Phaeodactylum tricornutum*. *MicrobiologyOpen*, 10, e1239. https://doi.org/10.1002/mbo3.1239

APPENDIX 1

TABLE A1 Primers used in this study

Names	Sequence (5'–3')	RE	Description	References
GapC1-500_F	CATATGGGAATTGAAGCAATCCATTTTGG	NdeI	Genomic DNA PCR	
GapC1-1086_F	CATATGTTTACTGTGTAAGTATGGGGAC	NdeI	Genomic DNA PCR	
GapC1_R	GAATTCGATGGAGTCAAAAAAGAAAGTAG	EcoRI	Genomic DNA PCR	
GS–501_F	CATATGATCACAGAAGCGGCAAAGTTCC	NdeI	Genomic DNA PCR	Erdene-Ochir et al. (2016)
GS–996_F	CATATGTGGTGCCGTTGATGCCGTGG	NdeI	Genomic DNA PCR	Erdene-Ochir et al. (2016)
GS_R	GAATTCGCTTGGAAGTTTGGGATGTGG	EcoRI	Genomic DNA PCR	Erdene-Ochir et al. (2016)
GFP_F	GAATTCATGGTGAGCAAGGGCGAGGAG	EcoRI	pPha-T1-*gfp* PCR	Erdene-Ochir et al. (2016)
GFP_R	GGATCCTTACTTGTACAGCTCGTCCATGC	BamHI	pPha-T1-*gfp* PCR	Erdene-Ochir et al. (2016)
dT-Long-P_R	GGCCACGCGTACTAGTGAATTCT$_{17}$		Adapter for 3' RACE	
short-P_R	GGCCACGCGTACTAGTGAATTC		3'-RACE and cloning	
TSO_F	GTCGCACGGTCCATCGCAGCAGTCACAG$_5$		Template-switch oligonucleotide	Pinto & Lindblad (2010)
GSP-GS_R	GATGGCCCAATCAAAGACAGCC		5'-UTR of GS	
U-SENSE_F	GTCGCACGGTCCATCGCAGC		5'-UTR of GS	Pinto & Lindblad (2010)
nGSP-GS_R	AGGTATTGGTCGGCAATCTTTCC		5'-UTR of GS	
CIP1_F	CATATGTACGTAGAATCCTACG	NdeI	Genomic DNA PCR	Kwon et al. (2012)
fcpA_F	CATATGGGGCTGCAGGACGCAATGG	NdeI	Genomic DNA PCR	
pPha-T1-Multi-B_R	ACTCCCAACTGTTCGTGCACCATG		Genomic DNA PCR	

Abbreviation: RE, restriction enzyme.

APPENDIX 2

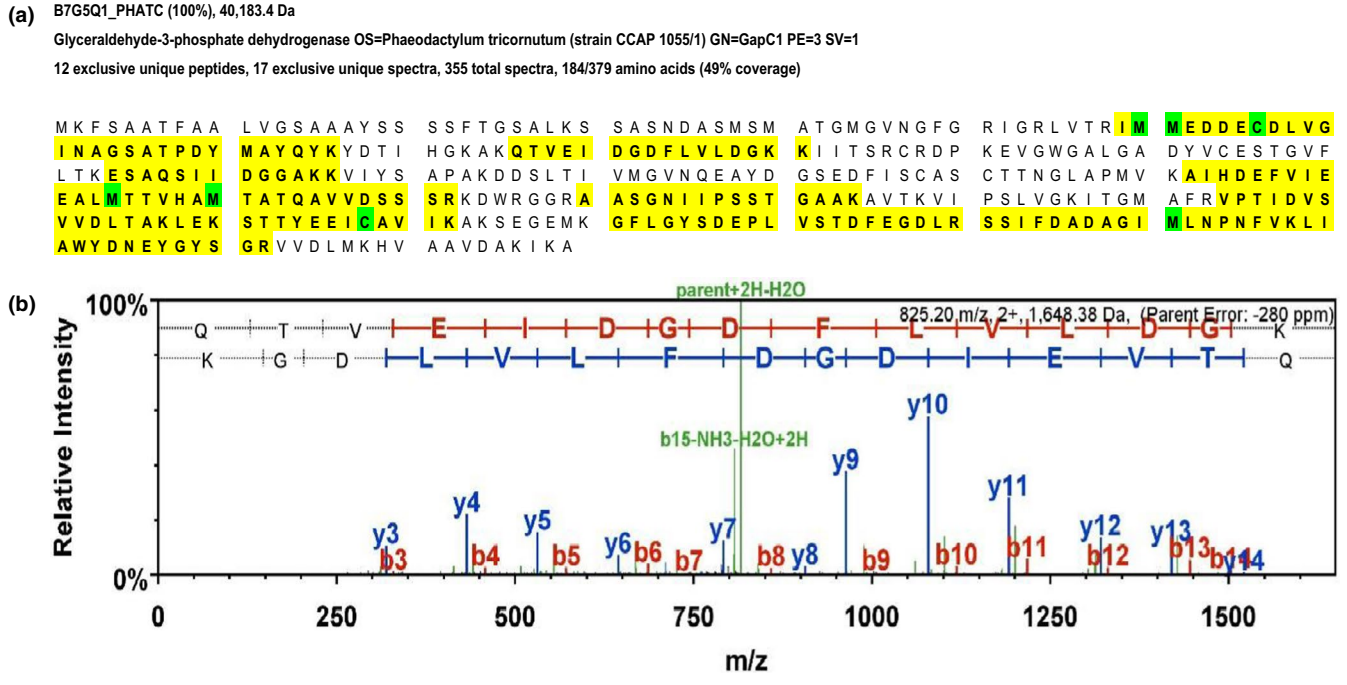
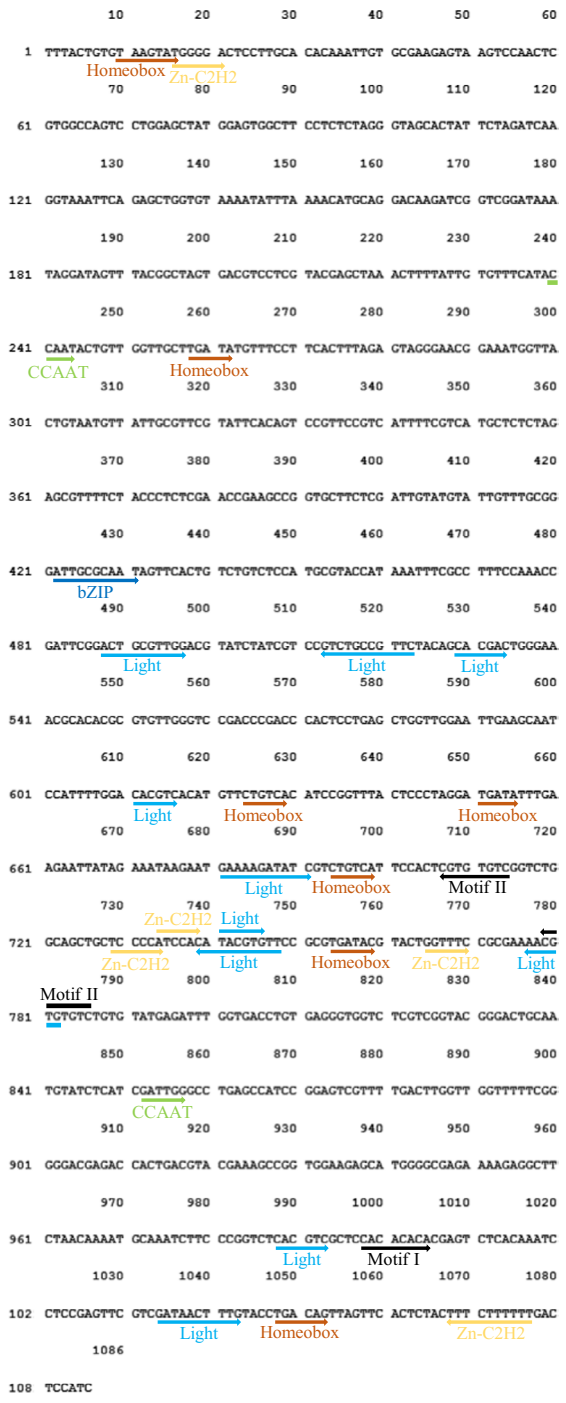


FIGURE A1 LC-MS/MS analysis of the GapC1 protein (a) Sequence coverage of the GapC1 protein according to LC-MS/MS analysis. The yellow highlighted sequences represent peptide sequences found in the LC-MS/MS analysis (49% coverage). Green highlighted sequences represent potential oxidation sites. (b) Mass spectra of the GapC1 protein



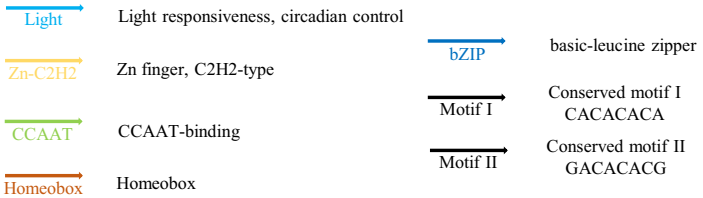
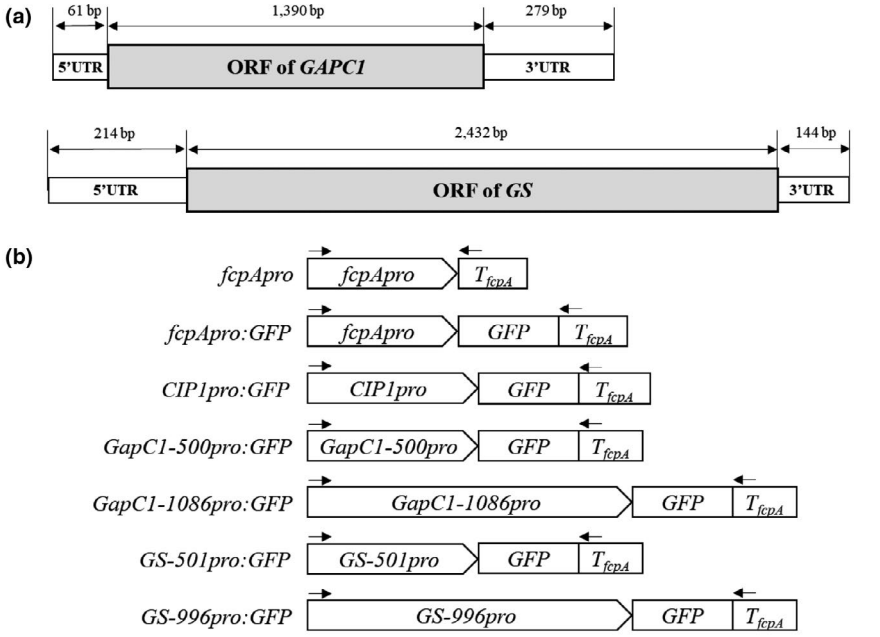
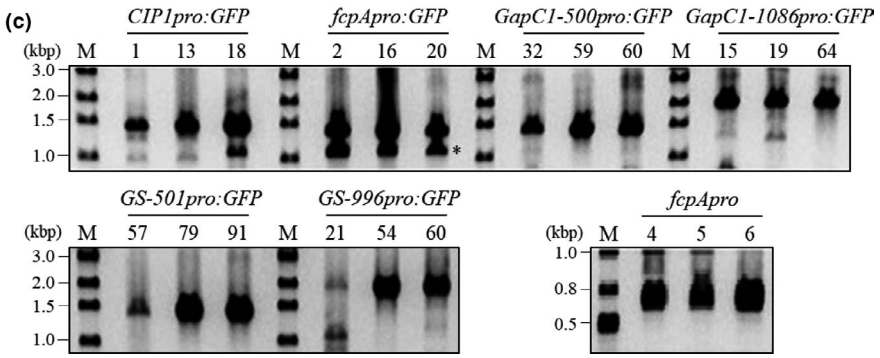


FIGURE A2 *In silico* analysis of *cis*-acting elements in *GapC1* promoter. The upstream sequence of the *GapC1* was analyzed by PlantCARE

FIGURE A3 *In silico* analysis of *cis*-acting elements in the *GS* promoter. The upstream sequence of the *GS* gene was analyzed by PlantCARE





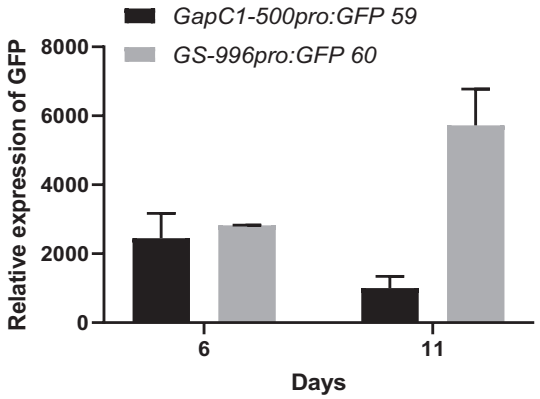


FIGURE A4 Identification of the 5' UTRs of *GS* and *GapC1* genes and selection of transformants. (a) Schematic representation of the *GS* and *GapC1* mRNA structures. (b) Vector constructs used for the transformation of *P. tricornutum*. Arrows indicate the primers used for PCR analysis. (c) Transgenes are amplified by PCR from transformant genomic DNA. The numbers show three independent transgenic lines generated by each construct. Asterisks show nonspecific PCR products. M, molecular size marker

FIGURE A5 Relative levels of GFP transcript in Condition 1. The levels of GFP mRNA at the mid-log (day 6) and stationary (day 11) phases of cultivation in *GapC1pro:GFP* and *GSpro:GFP* transgenic lines. GFP expression levels were normalized to *TBP* expression. Data are expressed as the mean ± SD of three replicates

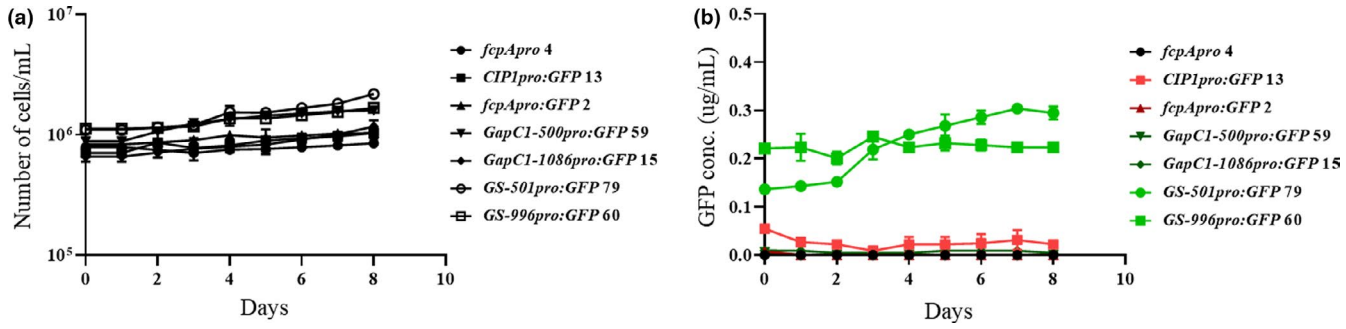


FIGURE A6 Growth curves of all transgenic lines and levels of GFP expression in Condition 3. (a) Growth curves of *P. tricornutum* cultures. All transgenic lines were cultivated for 8 days. (b) Levels of GFP fluorescence in cell lysates of *GapC1pro:GFP* and *GSpro:GFP* transgenic lines were measured by a fluorometer

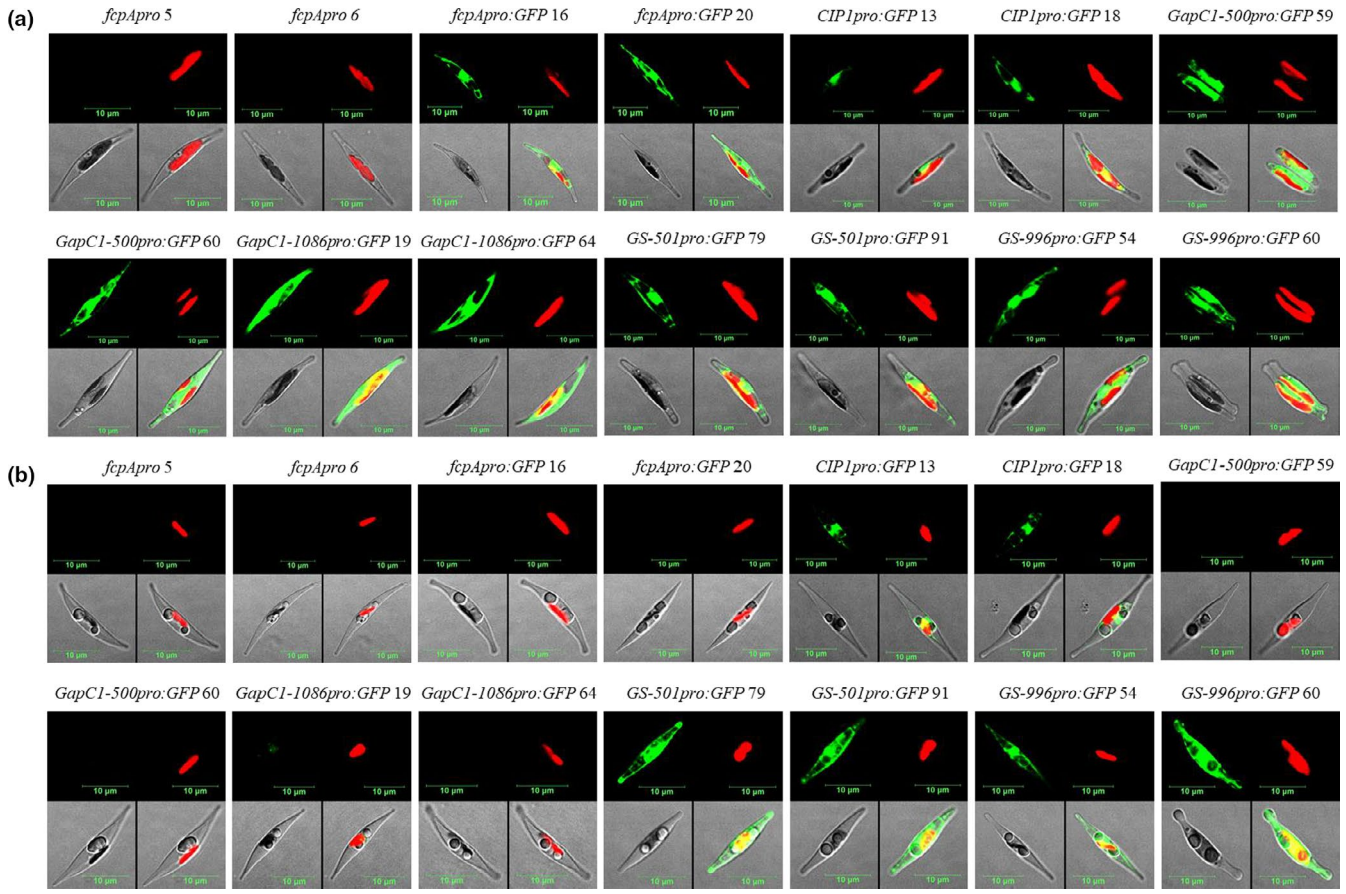


FIGURE A7 Subcellular localization of GFP in transgenic lines. GFP fluorescence and chlorophyll fluorescence in transgenic lines at the (a) mid-log and (b) stationary phases and visualized by confocal microscopy. The numbers on the images show two independent transgenic lines generated by each construct. Scale bars = 10 μm