Novel endogenous promoters for genetic engineering of the marine microalga *Nannochloropsis gaditana* CCMP526

Margaret Ramarajan\(a\), Michele Fabris\(a, b, *\), Raffaela M. Abbriano\(b\), Mathieu Pernice\(a\), Peter J. Ralph\(a\)

\(a\) University of Technology Sydney, Climate Change Cluster (C3), Faculty of Science, Broadway Campus, Ultimo, NSW, 2007, Australia
\(b\) CSIRO Synthetic Biology Future Science Platform, GPO Box 2583, Brisbane, QLD, 4001, Australia

**ARTICLE INFO**

**Keywords:**
Promoter
Nuclear transformation
Genetic engineering
*Nannochloropsis*
Algae biotechnology

**ABSTRACT**

*Nannochloropsis* is a marine microalga from the *Eustigmatophyceae* stramenopile lineage that has been studied extensively due to a broad range of industrial applications, mostly related to their oil and pigment production. However, tools to genetically engineer members of this group, and therefore further understand and maximise their industrial potential are still limited. In order to expand the potential industrial uses of this organism, several molecular tools, including gene promoters of different strength, are needed. A comprehensive and diverse set of well-characterized promoters is key to a number of genetic engineering and synthetic biology applications, such as the assembly of complex biological functions or entire metabolic pathways.

In this study, we measured the promoter activity of three endogenous constitutive promoters from *N. gaditana* genes EPPSII (*Nga02101*); HSP90 (*Nga00934*); ATPase (*Nga06354.1*) in driving the expression of a Sh ble-mVenus fluorescent reporter fusion protein. Through a combined approach that includes flow cytometry, RT-qPCR and immunoblotting, we profiled the activity of these promoters at both the transcript and protein level. Two promoters HSP90 (*Nga00934*) and EPPSII (*Nga02101*) outperformed the widely used β-tubulin promoter, exhibiting 4.5 and 3.1-fold higher mVenus fluorescence, respectively. A third promoter ATPase (*Nga06354.1*) was also able to drive the expression of transgenes, albeit at lower levels. We show that the new promoters identified in this study are valuable tools, which can be used for genetic engineering and functional genetics studies in *N. gaditana*.

1. Introduction

*Nannochloropsis gaditana* is an *Eustigmatophyceae* (Stramenopiles) marine microalga [1] that has been extensively studied for its potential use in biofuel production [2–4] because it naturally accumulates high levels of triacylglycerol [5]. *N. gaditana* also produces valuable pigments such as violaxanthin, canthaxanthin, astaxanthin and zeaxanthin [6]. Moreover, it is a rich source of polyunsaturated fatty acids such as eicosapentaenoic acid (EPA) [7], hence it is often employed in nutraceutical production [8–10] and as a feedstock in aquaculture [11].

Genetic engineering is a powerful approach for industrial microbial applications, as it enables the editing of existing traits as well as the installation of novel new ones in production organisms. *Nannochloropsis* species are good platform for genetic engineering for several reasons: (1) they have an haploid genome [3,12,13], which simplifies genome editing of non-essential genes as only one chromosome must be modified; (2) their genome is fully sequenced [12–15]; (3) omics datasets are available [16,17]; (4) nuclear transformation is feasible [12,13,18–22]; and (5) homologous recombination (HR) is efficient [21] and can be exploited for targeted gene editing [3]. These advances in phycological research have laid an excellent foundation to engineer this species as an industrially relevant strain to produce high-value compounds.

However, unlike other model algal species that are rapidly entering the field of molecular biotechnology and synthetic biology, such as *Chlamydomonas reinhardtii* and *Phaeodactylum tricornutum* [23,24], there is a less well developed and diversified set of resources to genetically engineer *Nannochloropsis*. A wider array of molecular tools, including gene promoters are needed to enable complex metabolic engineering or synthetic biology strategies in *Nannochloropsis* [25]. A diverse set of promoters of different strength is essential for fine-tuning the expression of transgenes and entire metabolic pathways, and can circumvent issues such as gene silencing due to promoter repetition, as previously reported in higher plants [26–28].

---

**https://doi.org/10.1016/j.algal.2019.101708**

Received 12 June 2019; Received in revised form 24 October 2019; Accepted 24 October 2019

Available online 19 November 2019

2211-9264/ © 2019 Elsevier B.V. All rights reserved.
Several constitutive endogenous promoters have been identified in *Nannochloropsis* species [25] that drive the expression of antibiotic resistance genes (e.g. *Sh ble*) [12] and reporter genes (e.g. firefly luciferase, *Nanolum* from a deep sea shrimp [29], and *Venus* [30]). To date, seven endogenous promoters have been successfully used for transgene expression in *N. gaditana* [12,31]. Of these, three endogenous promoters have been identified in *N. gaditana* strain CCMP526, which regulate the expression of genes putatively annotated as β-tubulin (TUB, *Nga00092*), heat shock protein 70 (HSP, *Nga07210*), and the ubiquitin extension protein (UEP, *Nga02115.1*) [12]. Although these were shown to effectively drive the expression of heterologous genes in *N. gaditana*, they have not been thoroughly profiled and characterized [12,31].

In this work, we set out to identify and study the activity of novel *N. gaditana* endogenous promoters that could efficiently drive the expression of transgenic constructs. We mined the publicly available transcriptomic data sets of *Nannochloropsis oceanica* CCMP1779 [16] to identify constitutively expressed genes, searched for their orthologs in *N. gaditana*, cloned their upstream putative promoter regions, and profiled their activity in *in vivo* in *N. gaditana* at both the transcript and protein level. Here, we report the isolation and characterization of three novel promoters putatively annotated as heat shock protein 90/ Molecular chaperon HtpG (HSP90, *Nga00934*), extrinsic protein in photosystem II (EPSSII, *Nga02101*), and P-type H+ ATPase (ATPase, *Nga06534.1*). We parameterized the activity of these promoters in vivo with reporter genes, and determined that they drive transgene expression at different levels in *N. gaditana*. We anticipate that these novel promoters represent useful resources to expand the genetic toolkit available for genetic engineering of this microalga, both for functional genetics studies and biotechnological applications.

2. Materials and methods

2.1. Cell culture

*N. gaditana* CCMP526 strain was purchased from the Provasoli-Guillard National Centre for Culture of Marine Phytoplankton, USA. The cells were grown in liquid modified F/2 medium [32] where the seawater component was replaced with Enriched Seawater, Artificial Water (ESAW) [33] and the vitamins concentration was doubled (primary vitamin stock solution: Thiamine · HCl- 400 mg, Biotin- 2 g/L). Axenic cultures were maintained under cool white LED light panel set to a light intensity of 100 μmol photons m⁻² s⁻¹, with a photoperiod of light: dark 12 h: 12 h, at 21 °C in a shaker incubator (Kühner, Switzerland). Batch cultures were maintained in 250 mL Erlenmeyer flasks closed with foam plugs to allow gas exchange. Cultures were checked for contamination every two months by streaking on marine broth agar plates. All transformation and promoter analysis experiments were carried out in the above mentioned conditions, but under continuous light regime.

2.2. Identification of promoters of high expressing genes in *N. gaditana*

To identify genes with constitutive high expression, a published transcriptomic data set of *Nannochloropsis oceanica* CCMP1779 [16] was used, as this closely represents commonly used laboratory growth conditions without nutrient stresses, and such dataset is currently not available for *N. gaditana*. Nine genes with Fragments Per Kilo base of exon per Million fragments mapped (FPKM) values greater than 60 across the different time points were selected (Table 1) and their corresponding protein sequences were extracted from the Michigan State University, *Nannochloropsis oceanica* CCMP1779 resources [34]. The orthologs of these genes were obtained using the web-based N. gaditana BLASTp search tool using default parameters, in the CRIBI Genomics *Nannochloropsi* genome portal [35]. The β-tubulin TUB (*Nga00092*) promoter was selected as a control promoter for this study because of its utility in driving the constitutive expression of transgenes in *N. gaditana* has been demonstrated [12].

2.3. Gene expression analysis by Real-Time quantitative PCR (RT-qPCR)

Cell pellets from 50 mL of early log phase cultures of *N. gaditana* CCMP526 (1 × 10⁷ cells mL⁻¹) were washed with 5 mL 1% PBS, re-pelleted at 3500 g for 10 min., and flash-frozen in liquid nitrogen. Frozen cells were lysed with a tissue lyser (TissueLyser LT, Qiagen, Germany) with 2 RNase-free 1.6 mm stainless steel beads at 50 Hz for 2 min. at 0 °C. Total RNA was extracted using Tri-reagent (Trizol, Thermo Fisher Scientific, USA) and an RNAeasy mini purification kit (Qiagen, Germany) including a gDNA elimination step following manufacturer’s instructions. The RNA quantity and quality was assessed using a spectrophotometer (NanoDrop 2000, Thermo Fisher Scientific, USA) and RNA samples were stored at −80 °C for further RT-qPCR experiments. A total of 280 ng of RNA was used for each sample for cDNA synthesis using iScript™ cDNA clear cDNA Synthesis Kit (Biorad, USA). An Eppendorf epMotion® 5075 (Eppendorf, Germany) robot was programmed to prepare the RT-qPCR reactions in a Hard-Shell® 384-well, thin wall, skirted, white PCR plate (Biorad, USA). The RT-qPCR reaction was carried out in a CFX384 Touch™ Real-Time PCR Detection System (Biorad, USA) with the following conditions: incubation at 95 °C for 10 min, then 44 cycles of 95 °C, 60 °C, 68 °C for 30 s each, followed by a melt curve to verify specific amplification of the target amplicon. RT-qPCR reactions consisted of 10 μL total volume with 4 μL of cDNA and 6 μL of iTaq™ Universal SYBR® Green Supermix (Biorad, USA) together with gene specific primers of final concentration 300 nM (Table S1). RT-qPCR analysis was carried out in technical triplicate using the CFX Manager™ 3.1 software. RT-qPCR efficiency for each primer pair was determined from a calibration curve with a cDNA dilution gradient of 243, 81, 27, 9 and 3 ng and a linear regression model [36]. The corresponding real-time PCR efficiencies were calculated according to the equation described by Radonic et al. [37]. All qRT-PCR displays efficiencies between 92% and 115% (R² of calibration curve > .92), well within a range considered as valid (85%–115%) as previously described [38–41]. RT-qPCR primers used to amplify the selected orthologous genes from *N. gaditana* are given in Table S1 and their primer efficiency, calibration curve’s R², amplicon length and Tm are given in Table S2. A no template control as well as a no reverse transcription control was generated for each primer pair and RT-qPCR run to ensure that the cDNA samples were free of DNA contamination.

For quantitative real-time PCR analysis of *N. gaditana* transformants, a total of 205 ng of RNA was used for each sample for cDNA synthesis using iScript™ cDNA clear cDNA Synthesis Kit (Biorad, USA). Data analysis was done in single threshold mode. RT-qPCR primers used to amplify *mVenus* and *β actin* genes are given in Table S1 and their primer efficiency was 103.8% with a R2 of 0.994. The starting quantity (SQ) of *mVenus* was normalized using *β actin* SQ (which has been used as a housekeeping gene in previous studies [42]) to give the relative abundance of *Sh ble-mVenus* mRNA in all the samples analyzed.

2.4. Construction of plasmids pNaga4.mVenus and pNaga4.empty

Two transformation vectors, pNaga4.empty and pNaga4.mVenus (Figs. S1 and S2, respectively), were designed in silico using the NEBuilder Assembly Tool v1.12 18 software. Phusion High Fidelity DNA polymerase (New England Biolabs, USA) was used to PCR amplify DNA fragments and the pUC19 backbone as per manufacturer’s protocol using primers listed in Table S1. A PureLink Quick Gel Extraction kit (Invitrogen, Life Technologies, USA) was used to purify the vector and PCR insert fragments from 1% agarose gels. The following fragments were assembled into the pUC19 backbone using the Gibson Assembly™ Master Mix (New England Biolabs, USA): TUB (*Nga00092*) promoter region, *Sh ble* gene, His-tag, multi-cloning site, Foot and Mouth Disease Virus (FMDV) self-cleaving 2A fragment, *mVenus* gene, and TUB (*Nga00092*) terminator region.
DH5-alpha Escherichia coli electrocompetent cells (New England Biolabs, USA) were used for DNA cloning, and were grown in Luria broth (Millers LB base) medium (Invitrogen, Life Technologies, USA) with 100 μg mL⁻¹ ampicillin. Plasmid DNA was isolated using the Zippy Plasmid Miniprep kit (Zymo Research, USA). The pNaga4.mVenus vector sequence was confirmed using restriction digestion using EcoRI, BamHI, and HindIII followed by DNA sequencing (Macrogen, South Korea).

2.4.1. Cloning of the putative promoter regions into pNaga4.mVenus

Five genes (Nga01608, Nga21005, Nga02101, Nga00934, Nga06354.1) with the highest relative mRNA levels measured using RT-qPCR were selected. The upstream putative promoter regions (600−900 bp) of these genes were analyzed using PlantCare [43] and Softberry TSSP [44] software and were amplified from N. gaditana genome using specifically designed primers with PvuII genome digestion using

<table>
<thead>
<tr>
<th>N. gaditana</th>
<th>N. oceanica</th>
<th>Abbreviation</th>
<th>Functional annotation</th>
<th>Percent identity</th>
<th>Protein size (aa)</th>
<th>E-value</th>
<th>Bit score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nga02101</td>
<td>NamCCOMP1779</td>
<td>EPPSII</td>
<td>extrinsic protein in photosystem ii</td>
<td>90.62%</td>
<td>64</td>
<td>4.00E-28</td>
<td>119</td>
</tr>
<tr>
<td>Nga06354.1</td>
<td>NamCCOMP1779</td>
<td>ATPase</td>
<td>H+-transporting ATPase</td>
<td>85.00%</td>
<td>260</td>
<td>3.00E-133</td>
<td>471</td>
</tr>
<tr>
<td>Nga01608</td>
<td>NamCCOMP1779</td>
<td>HIP</td>
<td>hypothetical protein</td>
<td>85.53%</td>
<td>159</td>
<td>3.00E-81</td>
<td>296</td>
</tr>
<tr>
<td>Nga21005</td>
<td>NamCCOMP1779</td>
<td>G3P</td>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
<td>93.63%</td>
<td>377</td>
<td>0</td>
<td>716</td>
</tr>
<tr>
<td>Nga00164</td>
<td>NamCCOMP1779</td>
<td>HS20</td>
<td>heat shock protein hsp20</td>
<td>76.32%</td>
<td>152</td>
<td>5.00E-66</td>
<td>246</td>
</tr>
<tr>
<td>Nga21085</td>
<td>NamCCOMP1779</td>
<td>FATPase</td>
<td>F-type H+-transporting ATPase subunit beta</td>
<td>93.09%</td>
<td>666</td>
<td>0</td>
<td>1224</td>
</tr>
<tr>
<td>Nga06853</td>
<td>NamCCOMP1779</td>
<td>FCCP</td>
<td>light harvesting protein</td>
<td>89.13%</td>
<td>184</td>
<td>2.00E-88</td>
<td>321</td>
</tr>
<tr>
<td>Nga20972</td>
<td>NamCCOMP1779</td>
<td>AMT</td>
<td>ammonium transporter</td>
<td>87.12%</td>
<td>489</td>
<td>0</td>
<td>760</td>
</tr>
</tbody>
</table>

2.5. N. gaditana transformation

As the transformation of N. gaditana CCMP526 strain was not successful using previously published electroporation protocols [12,21,45], multiple standardization experiments (Table S3) were carried out, and some modifications to the culture conditions and pellet preparation were made as described below. N. gaditana CCMP526 strain cultures were grown to early log phase (approx. 2 × 10⁶ cells mL⁻¹) at 21 °C in continuous light (100 μmol photons m⁻² s⁻¹) using 50% initial culture inoculum in modified F/2 medium, containing half the concentration of NaCl, and sub-cultured three times after 48 h of growth. 300 mL of culture was taken and centrifuged at 4000g for 15 min. at 4 °C. The cell pellet was washed three times in ice-cold 374 mM sorbitol and re-suspended to a final concentration of 2.5 × 10⁹ cells mL⁻¹ in 375 mM sorbitol. Concentrated culture (200 μL) was transferred to a 0.2 mm electroporation cuvette and 2 μg of linearized (using PvuII or Ndel) transformation vector was added. The cuvette was incubated on ice for 10 min. before electroporation. A Gene Pulser Xcell™ Electroporation System (Biorad, USA) was used to electroporate the samples at 2.2 KV, 600 Ω, and 50 μF. The pulse duration observed was 18–25 ms. The cells were transferred to 5 mL modified F/2-Si medium containing half the concentration of NaCl in a 50 mL tube and incubated for 48 h in a shaker incubator at 21 °C under continuous low light (70 μmol photons m⁻² s⁻¹) for recovery. The algal cell cultures were centrifuged at 3500 g for 10 min. at 4 °C, plated onto half-salt modified F/2-Si plates with zeocin (5 μg mL⁻¹) and incubated for 3 weeks until appearance of colonies.

To evaluate the transformation efficiency among constructs, the growth stage selected for cell harvest (log phase, OD750 0.3, approx. 1 × 10⁷ cells mL⁻¹), the final cell concentration (2−2.5 × 10⁷ cells mL⁻¹), the sorbitol wash process, and electroporation conditions were kept constant. The variables being the linearized plasmid amount (2−4.5 μg/electroporation) and the promoter constructs used. The transformation efficiencies were determined using the formula:

\[
\text{Transformation efficiency} = \frac{\text{(Colonies)/(Cells plated out)}}{\text{(Amount of plasmid µg)}}
\]

Individual transformant colonies were placed into a Falcon® 96 Well Clear Assay plate, round bottom (Corning, USA) containing half-salt modified F/2-Si containing zeocin (5 μg mL⁻¹) and incubated at 21 °C under 100 μmol photons m⁻² s⁻¹ light intensity for 4–5 days and subcultured thrice. Analysis of cells was performed by using a flow cytometer (CytoFLEX S, Beckman Coulter Life sciences, USA) to check for mVenus fluorescence using a 488 nM laser and 525/40 optical filter. PCR was performed with mVenus specific primers (Table S1) on three transgenic cell lines per promoter construct that possessed the strongest mVenus signal, to show the presence of the reporter gene (mVenus) in the genome of transformed N. gaditana using the PCR conditions as per manufacturer’s protocol. β-actin primers (Table S1) were used as an internal positive control (Fig. S3).
2.6. Flow cytometry analysis of \textit{N. gaditana} transformants

The selected \textit{N. gaditana} clones expressing \textit{Sh ble-mVenus} under the control of EPPSII (Nga02101); HSP90 (Nga00934) and ATPase (Nga06354.1) promoters were placed into a Falcon® 96 Well Clear Assay plate (200 μL), round bottom (Corning, USA) in triplicate. Flow cytometric analysis of cells was performed once every day to measure mVenus fluorescence using a 488 nM laser and 525/40 optical filter and chlorophyll measurement using the 690/50 band pass filter. An empty vector control was used to compare between the chlorophyll and mVenus fluorescence. The gating used to select the mVenus positive cells is shown in Fig. S4. The mean mVenus fluorescence intensity and the percentage of cells expressing mVenus fusion protein was calculated using \textit{N. gaditana} wild-type strain and \textit{N. gaditana} transformant without mVenus protein as negative controls. The net mVenus fluorescence was calculated as a product of mVenus intensity/cell and number of cells expressing mVenus fusion protein.

2.7. Immunoblotting

\textit{Nannochloropsis} cell cultures (30 mL) were harvested by centrifugation at 4000 g for 15 min., supernatant was removed, and pellet washed in 1 mL phosphate buffer saline (PBS). The cells were centrifuged again at 4000 g for 10 min. and the supernatant discarded. The pellet was re-suspended in lysis buffer (50 mM Tris HCl, 400 mM NaCl, 0.5% tween20 at pH 8) with 1X protease inhibitor cocktail (Sigma-Aldrich, USA) and frozen in liquid nitrogen for one hour. The samples were thawed and 200 μg of 2 mm glass beads were added. A TissueLysyer LT (Qiagen, Germany) was used at 50 Hz for 5 min. to lyse the cells. The samples were centrifuged at 10,000 g for 10 min. at 4 °C, and the clear supernatant was collected. The concentration of soluble crude protein was estimated using BCA reagent (Thermo Fisher Scientific, USA) following the manufacturer’s guidelines. As the samples were dilute, the crude protein extracts were concentrated by lyophilization and re-suspension in 25 μL of PBS.

The fusion protein Sh ble-mVenus was detected by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting. The protein concentration in each sample was normalized to 20 μg prior to loading. The proteins were separated by 4–15% SDS-PAGE using a Tris/glycine/SDS buffer (Biorad, USA) and blotted onto a PVDF membrane (Biorad, USA) using the Trans-Blot® Turbo™ Transfer System (Bio-rad, USA). The protein blot was incubated with mouse anti-His monoclonal antibody (Bio-rad, USA) diluted 1:1,000 in 5% (w/v) skim milk followed by rabbit anti-mouse IgG conjugated to horseradish peroxidase (Bio-rad, USA) diluted 1:3,000 in 5% (w/v) skim milk. The signals were detected with the chemiluminescence detection system, (Clarity Western ECL, Biorad, USA) according to the manufacturer’s instructions and documented using the ChemiDoc™ XRS + System (Biorad, USA).

2.8. Statistical analysis

The fluorescence intensity of the fusion protein Sh ble-mVenus, which expression was driven by the experimental promoter constructs EPPSII (Nga02101); HSP90 (Nga00934) and ATPase (Nga06354.1) was compared to the fluorescence intensity of Sh ble-mVenus driven by the TUB (Nga00992) promoter construct. Paired t-tests were performed on flow cytometric data and quantitative RT-PCR data to estimate the significant difference in relative promoter activity between the (TUB) positive control and each of the three experimental constructs.

3. Results and discussion

3.1. Identification of constitutive promoter candidates in \textit{N. gaditana}

Endogenous promoters for constitutive and efficient expression of transgenes are generally obtained from non-coding regions upstream of a transcription initiation site of genes [46], that are associated with high and steady expression throughout most experimental conditions. Thus, genome-wide transcriptomic analysis represent a particularly useful resource to identify such candidate genes based on their expression patterns. Since transcriptomic data for \textit{N. gaditana} are limited, we mined a publicly available \textit{N. oceanica} transcriptomic dataset [16] to identify a set of putative constitutive promoters. This dataset was selected as it closely represents a batch cultivation growth condition free of other environmental stresses, such as nutrient depletion. We identified nine genes (Table 1) with Fragments per Kilo base of exon per Million fragments mapped (FPKM) values greater than 60 across all time points (Supp. data file 1). The putative annotation of most of these genes suggested that they could be involved in essential cellular functions such as photosynthesis, protein maintenance, ion transport and nutrient acquisition [Table 1, 45–64]. Promoters of homolog genes encoding heat-shock proteins (HSPs) and adenosinetrinophosphatases (ATPases) have previously been used in mammalian cell lines [65] and in higher plants [66] for constitutive or cell-specific gene expression [67]. Similarly, glyceraldehyde-3-phosphate dehydrogenase (G3P) promoters have been used for protein expression in yeast and plants [68–70]. As these functions are generally highly conserved among different organisms, we performed a database search on sequence similarity to identify potential orthologs in the \textit{N. gaditana} genome (Table 1).

To ensure that the selected genes were actively expressed in \textit{N. gaditana} cultures, we performed real time quantitative PCR (RT-qPCR) on mid-log phase samples as a representative condition commonly used for transgene expression [29,71]. All tested genes were actively transcribed in the cell, and five genes with the lowest Cq value, which is inversely proportional to the transcript copy number in the sample, were selected for further analysis (Fig. S5).

The upstream intergenic regions of EPPSII (Nga02101); HSP90, (Nga00934); ATPase (Nga06354.1); HYP (Nga01608); and G3P (Nga21005) were extracted using the \textit{N. gaditana} CCMP526 genome browser [14]. Since computational tools to analyse the structure of algal promoter are not available, we used plant-specific tools PlantCARE [43] and Softberry TSSP [44,72] to predict transcription factor binding sites and cis-regulatory elements such as CAAT and TATA boxes (Supp. data file 1), and their position with respect to the transcription start site (TSS). All promoter sequences that were analyzed had at least one predicted promoter region (Supp. data file 1). The computational results showed the presence of CAAT-boxes, common cis-acting elements in promoter and enhancer regions; TATA-boxes, which are core promoter elements generally 30 bp upstream of the start codon; and G-boxes, cis-acting regulatory elements involved in light responsiveness in plants [73–76]. Other putative transcription factor (TF) binding site like MYC (TF of the basic helix-loop-helix-leucine zipper family) [77], MYB (TF family involved in biotic and abiotic stresses, development, differentiation, metabolism, defense in plants) [78,79] and the stress response element (STRE) [80] were also found in all the tested promoters, whereas the promoter-specific Spl TF recognition site (Spl is a protein that binds to GC box promoter elements to selectively activate mRNA synthesis in mammalian cells) [81,82] was only found in EPPSII (Nga02101) and HSP90 (Nga00934) gene promoters. Based on these promoter motif predictions, a 600 bp–900 bp of intergenic sequence was selected as a putative promoter for each gene (Table S4).

3.2. Transformation efficiencies vary with selected promoter region

To evaluate whether the selected putative promoter regions in \textit{N. gaditana} are able to drive the expression of transgenes, EPPSII (Nga02101), HSP90 (Nga00934), ATPase (Nga06354.1), HYP (Nga01608) and G3P (Nga21005) putative promoter regions were cloned upstream of a \textit{Sh ble-mVenus} gene fusion (Fig. S2) in pNaga4.mVenus expression vector and introduced in \textit{N. gaditana} by
electroporation. Three out of five vectors containing EPPSII, (Nga02101), HSP90 (Nga00934) and ATPase (Nga06354.1) yielded transformant colonies, whereas three separate transformation attempts were unsuccessful for the remaining constructs (HYP, Nga01608 and G3P, Nga21005). Failure to generate antibiotic resistant colonies may be due to low or no expression of the fusion protein Sh ble – mVenus, possibly as a result of the misidentification of the putative promoter region or the absence of essential regulatory elements in the cloned sequence [83].

A previous study in N. gaditana showed that endogenous promoters improved the transformation efficiency of this species [12]. In this study, the six promoter constructs used yielded different transformation efficiencies (Table S5). The highest transformation efficiency was achieved using the EPPSII (Nga02101) promoter, which resulted in an efficiency of $3 \times 10^{-8}$ colonies cells$^{-1}$ μg DNA$^{-1}$. The lowest was observed using the HSP90 (Nga00934) promoter, which resulted in an efficiency of $3 \times 10^{-9}$ colonies cells$^{-1}$ μg DNA$^{-1}$. The control TUB (Nga00092) promoter achieved a transformation efficiency of $7 \times 10^{-9}$ colonies cells$^{-1}$ μg DNA$^{-1}$. Radakovits et al. observed a transformation efficiency of $12.5 \times 10^{-6}$ colonies cells$^{-1}$ μg DNA$^{-1}$ for TUB (Nga00092) [12]. Since each promoter directly drives the expression of the selectable marker Sh ble, it is plausible that the wide range of transformation efficiencies observed is due to intrinsic features of the cloned promoter regions that affect strength and expression profiles.

3.3. mVenus expression varies among and within transformant lines

Clones that maintained antibiotic resistance after two months of subculturing were considered stable and were screened by flow cytometry for mVenus fluorescence. Forty N. gaditana clones carrying EPPSII (Nga02101) and ATPase (Nga06354.1) promoters and ten clones transformed using HSP90 (Nga00934) promoters, due to low transformation efficiency, were evaluated (Fig. S6). As expected in case of random chromosomal integration of transgenes, the level of mVenus fluorescence varied largely among transformants, presumably due to multiple factors, including: i) position effects due to random chromosomal transgene integration and influence of neighboring regulatory elements on transgene expression [84–86]; ii) number of intact
transgene integrations [73]; iii) integrative fragmentation or DNA re-arrangement, including deletion or translocation, which can also influence the growth and fitness of the host [87–89]; and iv) various forms of epigenetic and homology-dependent gene silencing [90,91].

All the transformant colonies screened were positive for mVenus fluorescence, confirming that the selected upstream regions for the EPPSII (Nga02101); ATPase (Nga06354.1); and HSP90 (Nga00934) genes are able to serve as functional promoter sequences for transgene expression in N. gaditana. However, the clonal population did not show uniform mVenus expression (Fig. S 4). Heterogeneous transgene expression has been observed in different microalgae species and seems to be a common consequence of random chromosomal integration of the expression cassette [92,93]. Similar phenomena occur in mammalian model systems, where the transgene expression was found to be heterogeneous due to silencing position effects [84]. These can result in multiple cell populations with various levels of expression within the same clonal line [94]. For example, in mice the proportion of cells expressing a transgene can vary widely between lines of mice carrying the same transgene [95,96]. Variations within clonal lines can also be non-genetic in nature and dependent on other factors, such as unequal cell division, cell age and cell cycle stage in non-synchronized cultures [97,98].

Monitoring transgene expression in individual cells within a clonal population gives an accurate account of cellular response, rather than a bulk measurement [99]. To address this, the percentage of cells in a population expressing mVenus was used to calculate a net fluorescence intensity. Three independent transgenic cell lines per construct with the highest net mVenus fluorescence were selected for further experiments.

3.4. Multi-level analysis of promoter activity

To profile the activity of the three promoters throughout the growth cycle and to select time points for further analysis, we performed a small-scale growth experiment in multiwell plates to measure mVenus fluorescence over a period of 10 days. The net fluorescent intensity, as a proxy for fusion protein expression, was calculated in triplicate for each independent transgenic line, using the product of mVenus intensity/cell and the number of cells expressing mVenus (Fig. 1).

For all 10 days tested, net mVenus fluorescence intensities for EPPSII (Nga02101) and HSP90 (Nga00934) transformants were significantly higher than those measured for TUB (Nga00092) transformants. Averaged EPPSII (Nga02101) putative promoter activity was 4.8 fold higher, and the HSP90 (Nga00934) putative promoter activity was 2.7 fold higher than the TUB (Nga00092) control promoter. As maximum expression of mVenus fusion protein was observed between Day 4 and 6 (early-mid log phase) of culture (Fig. 1), the time point (Day 5) was used for further evaluation of reporter expression based on mRNA and protein analysis.

To assess transcript and transgenic protein abundance, three independent transgenic cell lines expressing the constructs described above were scaled-up to include RNA and protein extraction. Clonal lines were cultivated in batch mode over a period of 5 days while keeping the culture parameters consistent with the previous pilot experiment. The growth curves (Fig. 2b) of all the transgenic and wild-type N. gaditana lines are shown in Fig. 2b. Growth rate in the exponential phase was calculated for both transgenic and wild-type cell lines, and no significant difference was observed among all transgenic lines except EPPSII, which was found to have a significantly higher exponential growth rate compared to wild type N. gaditana (Fig. S7, Table S6). However, wild type cells had a higher initial growth rate (from day 0–1) and maintained a higher maximum cell density than all transgenic lines throughout the experiment (Fig. 2b). Differences in growth between wild-type and transgenic lines suggests that random genomic integration may have had detrimental effects to algal fitness due to random insertional mutagenesis, such as genetic rearrangements and disruption of relevant genetic loci [87–89,100].
mVenus fluorescence was tracked daily to follow the expression level of the fusion protein in vivo, until Day 4 (Figs. 2a, S8). A temporal expression pattern similar to the initial pilot experiment was observed, with increasing net mVenus fluorescence as the cultures approached exponential phase (Figs. 1c and 2a). The exponential phase promoter activity of EPPSII (Nga02101) again surpassed the TUB (Nga00092) control promoter, (3.1 fold higher), although it was higher (4.8 fold) in the pilot experiment. The activity of HSP90 (Nga00934) putative promoter was 4.5 fold higher than the TUB (Nga00092) control promoter, while previously we observed only a 2.7 fold efficiency in multiwell plates. The discrepancies observed in the two experiments may be due to the marked environmental differences associated with cultivation in shake flasks and microowell plates, even though the culture parameters were kept consistent.

Day 5 was chosen as suitable mid-exponential phase time-point to evaluate the activity of the tested promoters, both at the transcript and protein level. The relative transcriptional promoter activity for each promoter construct was tested by quantification of transgene transcripts on Day 5 using RT-qPCR. At the transcriptional level, the average relative promoter activities measured were 1.7 and 1.8-fold higher than that of the control promoter TUB (Nga00092) for ATPase (Nga06354.1) and HSP90 (Nga00934), respectively (Fig. 2c). Conversely, no significant difference between the average relative promoter activity of TUB (Nga00092) and the EPPSII (Nga02101) promoter was observed (Fig. 2c).

To evaluate the activity of the promoters at the protein level, we evaluated the abundance of the recombinant Sh ble-mVenus fusion protein in the same samples. The immunoblot analysis on protein samples harvested on Day 5 revealed a 47 kDa band corresponding to the Sh ble-mVenus fusion protein, in transformants carrying the HSP90 (Nga00934) promoter and EPPSII (Nga02101) promoter (Fig. 2d), the two promoters associated with highest net mVenus fluorescence (Fig. 2a) based on flow cytometry. No 47 kDa band was observed in transgenic cell lines containing the ATPase (Nga06354.1) promoter and the control TUB, (Nga00092) promoter. This may be due to low protein production, which may have sufficient protein to convey antibiotic resistance and for detection by flow cytometry, but not enough protein to be detected in an immunoblot. The flow cytometer used in this study has a sensitivity of <30 molecules of equivalent soluble fluorochrome in the FITC channel (CytoFLEX S, Beckman Coulter Life sciences,USA) whereas the Clarity Western ECL substrate used for chemiluminescence based immunoblot detection can detect as low as 0.6 ng protein in a sample (Biorad, USA). As expected, no bands were detected in lanes corresponding to the empty vector control and wild-type samples.

Transcript levels (Fig. 2c), mVenus expression, and recombinant fusion protein detected by immunoblot (Fig. 2d) were all detected in EPPSII (Nga02101) and HSP90 (Nga00934) clonal lines. The same cannot be said for ATPase (Nga06354.1) and TUB (Nga00092) clonal lines as no 47 kDa protein band was observed to correlate with mRNA abundance observed. As stated earlier, reasons for this discrepancy can include low protein abundance or protein degradation. As described in various studies, mRNA abundance does not always correlate with protein level [101–103]. A number of factors, can interfere with transgene expression, such as position effects or post-transcriptional/post-translational silencing, both of which can be a consequence of random transgene integration [85,86,88,104]. These largely uncharacterised mechanisms could have plausibly affected Sh ble-mVenus protein concentrations in transgenic cell lines. Hence, high mRNA levels observed can be sometimes misleading, which makes it key to study promoter activity with an integrated approach including both the transcriptional and translational level.

Finally, no apparent relationship was observed between transformation efficiency and full-scale transcript and protein abundance for each promoter. The EPPSII (Nga02101) promoter yielded the highest transformation efficiency and had the second highest expression of the fusion reporter protein by mVenus fluorescence. In contrast, the HSP90
(Nga00934) promoter was associated with the strongest expression levels both at the transcript and protein level, despite being associated with poor transformation efficiency. The ATPase (Nga06354.1) promoter was the least efficient of the three at driving transgene expression, despite of having high transformation efficiency.

This study identified three novel endogenous promoters that efficiently and constitutively drive the expression of reporter transgenes. Due to random chromosomal integration, the observed expression levels were variable among independent cell lines carrying the same transgenic construct. As mentioned, copy number and integration loci are key factors influencing transgene expression [85,86,104,105]. These factors cannot be controlled using the current transgenic method but other strategies, such as targeted knock-in using homologous recombination, are key factors influencing transgene expression [85,86,104,105].

In this work, promoters were profiled in the experiments carried out in batch mode under continuous light and nutrient replete conditions and not under variable culture conditions such as heat and light stress. In other organisms, the responsiveness of putatively similar promoters varies in stress conditions. For example, the heat shock proteins (HSPs) are molecular chaperones functionally involved in fundamental cellular processes and are induced by heat shock and other stimulants such as heavy metals, nitric oxide, hormones, microbial infections etc [108].

The orthologue of HSP90 (Nga00934) in E. coli (hsp8) is regulated by temperature and an increase in gene expression was observed in response to temperature under different growth conditions [48]. In humans, there are two major cytoplasmic isoforms of Hsp90: Hsp90α and Hsp90β which are generally inducible and constitutive, respectively [57]. In higher plants, the expression of some genes in the HSP90 gene family are affected by temperature and heavy metal stress, while others are constitutive [109,110].

Similarly, expression of EPPSII (Nga02101) may also vary under specific conditions. In higher plants, EPPSII (PsbQ) is a part of the oxygen-evolving complex of photosystem II (PSII) and is required for the PSII reaction center accumulation under low light [60,61]. PsbQ is vital for photoautotrophic growth under low light conditions and may be upregulated during this state. In Arabidopsis, the expression of this gene was found to be light-inducible and organ specific [111]. The behaviour of these promoters under light, temperature, and heavy metal stress in N. gaditana may improve transgene expression further but are beyond the scope of this paper. These are avenues that could be explored for future research.

4. Conclusions

Complex genetic engineering and synthetic biology resources are rapidly advancing and being applied to microalgae, widening their industrial potential. However, limitations include 1) a limited number of promoters experimentally shown to drive transgene expression; 2) a
narrow repertoire of functional promoters with limited diversity in terms of expression range and inducibility; and 3) a lack of knowledge of specific regulatory elements that control transcription in *Nannochloropsis* that would allow for predictable expression and rational design of promoter regions. To date, several studies involving gene overexpression have been conducted in *Nannochloropsis* using endogenous promoters [25] but in-depth promoter analysis done in other model microalgae like *Phaeodactylum* [112–116] and *Chlamydomonas* [117] is still missing with *Nannochloropsis* sp.

In this work we identified and profiled 3 endogenous promoters HSP90 (Nga00934); EPPSI (Nga02101) and ATPase (Nga06354.1) and we demonstrated that they can efficiently drive the expression of transgenes at different levels in *N. gaditana*, expanding the current genetic toolbox for this industrially relevant algal species. The availability of multiple promoters with similar strength, such as HSP90 (Nga00934) and EPPSI (Nga02101), is a key requisite in synthetic biology. For example, when designing and assembling multi-genome constructs, the repeated use of the same promoter might incur in genetic re-arrangements during the assembly phase in *E. coli* and/or *S. cerevisiae*, as well after the transformation in *Nannochloropsis* sp, given that all these organisms have highly efficient homologous recombination machineries. In other scenarios, such as the use of gene-stacking and combinatorial assembly of metabolic pathways, it is often required to express transgenes at different levels in order to reach the optimal balance necessary for maximizing the product yield. In these cases, the availability of weak, constitutive promoters such as ATPase (Nga06354.1) is relevant. In conclusion, the new promoters characterized here broaden the repertoire of molecular tools that can be used for genetic engineering of this species, both for biotechnological applications such as metabolic engineering or recombinant protein production, and to better understand the biology of *Nannochloropsis*.

**Author contributions**

M.R. performed the experiments; M.F. and P.J.R. conceived the research project; M.F. and M.P. contributed in data analysis; M.R. wrote the manuscript; all authors edited and reviewed the manuscript.

**Statement of informed consent, human/animal rights**

No conflicts, informed consent, human or animal rights applicable.

**Declaration of Competing Interest**

The authors declare no conflict of interest.

**Acknowledgements**

This research was supported by UTS: C3 funding. MF is supported by a CSIRO Synthetic Biology Future Science Fellowship, co-funded by CSIRO and the University of Technology Sydney. The authors would like to thank Kun Xiao for technical assistance with flow cytometry, M.Ramarajan, et al. [1] D.J. Hibberd, Notes on the taxonomy and nomenclature of the algal classes Eustigmatophyceae and Tribophyceae (synonym Xanthophyceae), Bot. J. Linn. Soc. 82 (1981) 93–119, https://doi.org/10.1111/j.1095-8339.1981.tb00954.x.


[37] V.V. Solovyev, I.A. Shahmuradov, A.A. Salamov, Identification of promoter re-


[40] T. Bricker, M. Bourry, The plant plasma membrane proton pump ATPase: a highly

[41] V.V. Solovyev, I.A. Shahmuradov, A.A. Salamov, Identification of promoter re-

[42] L. Mcalisters, M.J. Holland, Differential expression of the three yeast


