CellPress
Partner Journal

Molecular Plant
Research Article

The nucleolus as a genomic safe harbor for strong gene expression in *Nannochloropsis oceanica*

Christian Südfeld[1,*], Ana Pozo-Rodríguez[1], Sara A. Manjavacas Díez[1], René H. Wijffels[1,2], Maria J. Barbosa[1] and Sarah D'Adamo[1]

[1]Wageningen University, Bioprocess Engineering, PO Box 16, 6700 AA Wageningen, Netherlands

[2]Faculty of Biosciences and Aquaculture, Nord University, 8049 Bodø, Norway

***Correspondence: Christian Südfeld (christian.sudfeld@wur.nl)**

https://doi.org/10.1016/j.molp.2021.11.003

ABSTRACT

Microalgae are used in food and feed, and they are considered a potential feedstock for sustainably produced chemicals and biofuel. However, production of microalgal-derived chemicals is not yet economically feasible. Genetic engineering could bridge the gap to industrial application and facilitate the production of novel products from microalgae. Here, we report the discovery of a novel gene expression system in the oleaginous microalga *Nannochloropsis* that exploits the highly efficient transcriptional activity of RNA polymerase I and an internal ribosome entry site for translation. We identified the nucleolus as a genomic safe harbor for Pol I transcription and used it to construct transformant strains with consistently strong transgene expression. The new expression system provides an outstanding tool for genetic and metabolic engineering of microalgae and thus will probably make substantial contributions to microalgal research.

Key words: *Nannochloropsis*, gene expression, nucleolus, RNA polymerase I, IRES, aIRES

Südfeld C., Pozo-Rodríguez A., Manjavacas Díez S.A., Wijffels R.H., Barbosa M.J., and D'Adamo S. (2022). The nucleolus as a genomic safe harbor for strong gene expression in *Nannochloropsis oceanica*. Mol. Plant. **15**, 340–353.

INTRODUCTION

Microalgae are at the interface of industrial fermentation and agriculture. Their high protein content, nutritional value, and phototrophic growth rate, together with the prospect of cultivation on non-arable land, make them a potential replacement for traditional agricultural products in food and feed, as well as platforms for the biomanufacturing of a variety of desirable products for the chemical and pharmaceutical sector (Lamminen et al., 2019; Adissin et al., 2020; Sarker et al., 2020). Microalgae naturally produce many compounds of interest, including high-value lipids, vitamins, pigments, and sterols (Levasseur et al., 2020). In particular, oleaginous species, including the emerging model organism *Nannochloropsis*, have attracted interest as potential candidates for the sustainable production of lipids that can be used as biodiesel feedstocks and as a replacement for palm and fish oil (Chisti, 2007; Wijffels and Barbosa, 2010).

However, industrial production of microalgae-derived chemicals is not yet economically feasible due to the high cost of biomass production and insufficient productivities (Ruiz et al., 2016). Productivities could be increased by strain improvement strategies (Ajjawi et al., 2017), but a sophisticated toolbox for genetic engineering is still under development (Sproles et al., 2021). Nevertheless, the toolset currently available has shown to enable the advancement of microalgae with optimized productivities (Ajjawi et al., 2017) and reduced genome size (Wang et al., 2020).

Within the past decade, efficient transformation (Kilian et al., 2011; Vieler et al., 2012), gene silencing (Wei et al., 2017), and gene knockout (Wang et al., 2016; Ajjawi et al., 2017) tools were developed for *Nannochloropsis* spp. Despite this, tools for strong gene expression are still lacking. Only a few transcriptional promoters have been characterized (Poliner et al., 2018a), and gene expression levels are highly variable among different transformant strains (Kang et al., 2015; Grahl et al., 2020). This is, at least in part, due to positional effects that are associated with chromatin structure, if expression constructs (ECs) are integrated into the genome at random (Nguyen and Bosco, 2015). Moreover, chromatin structure is dynamic and subject to regulatory mechanisms that depend on physiological conditions affecting gene expression (Nguyen and Bosco, 2015). Furthermore, transgenes in microalgae are known to be subject to transcriptional silencing (Cerutti et al., 1997; Rasala et al., 2014). As a result, extensive and laborious screening procedures are

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

necessary to select transformant lines that express the transgene to sufficient and detectable levels. Improving the efficiency and reproducibility of existing gene expression systems or the discovery of new ones would greatly contribute to the genetic toolbox available for development of *Nannochloropsis* and other microalgal strains with desirable traits such as improved lipid productivity or enhanced production of novel protein products.

In this study, we report the discovery and development of an unusual chimeric system that confers exceptionally high gene expression by recruiting the cellular rRNA synthesis machinery for transcription.

RESULTS

Insertion of a promoterless cassette in a 25S rRNA gene leads to exceptionally strong expression

To discover novel transcriptional promoters in *N. oceanica*, we generated a mutant library containing random insertions of a promoter-trapping construct (TC) (Figure 1A). The absence of a transcriptional promoter prohibited *GFP* expression unless the cassette was inserted into a gene. A splice acceptor (SA) sequence was included at the 5′-terminus to allow expression in case of insertion into an intron, and an endogenous terminator ($T_{\alpha-tub}$) was added at the 3′-terminus. A control construct (CC) (Figure 1A) carried the promoter of the highest-expressed nuclear gene *VCP1* (Li et al., 2014). Among a total of 48 independent TC transformants that were selected for screening, the majority of strains had increased green fluorescence compared with the wild type (Figure 1B). TC strain #17 clearly stood out, with its GFP fluorescence exceeding that of the CC strain by ~8-fold (Figure 1C and 1D). Remarkably, *GFP* transcript abundance in TC#17 was increased by 135-fold over the control (Figure 1F).

We traced the TC insertion in TC#17 to a 25S rRNA gene using cassette PCR (Figure 1G). In eukaryotes, the 18S, 5.8S, and 25/28S rRNA genes form a single transcriptional unit (cistron) that is transcribed by DNA-directed RNA polymerase I (Pol I). rRNA cistrons are typically arranged as tandem repeats and located in a dedicated part of the genome, i.e. the nucleolus organizer region (NOR). *N. oceanica* contains only four rDNA cistrons in distinct genomic loci according to the current genome assembly (Gong et al., 2020). Genotyping PCR (Supplemental Figure 2) revealed that TC insertion in TC#17 had occurred in the rDNA cistron on chromosome 3 (chr3, position 1,029,753), between 25S rRNA bases 2333 to 2334 (corresponding to bases 2296–2297 of the *Saccharomyces cerevisiae* 25S rRNA), which are located in the loop of helix 71 of the large ribosomal subunit (Figure 1H).

TC insertion within a 25S rRNA gene explained the high *GFP* transcript abundance in TC#17 because rRNA genes are transcribed highly efficiently (Lodish et al., 2000). However, rRNA molecules are not translated into protein and, therefore, lack mRNA features such as a 5′ cap and a poly(A) tail. Consequently, if Pol I was responsible for transcription of the genetic fusion of 25S rDNA and TC in TC#17, the resulting chimeric RNA (chRNA) molecule should not be translated. Puzzled by these findings, we investigated the underlying mechanism of gene expression in this mutant.

Defining the minimal rDNA elements required for GFP expression in TC#17

Ribosome biogenesis in the nucleolus of eukaryotic cells is a complex and rigorously controlled process. The primary Pol I transcript undergoes multiple endo- and exonucleolytic cleavages, among other modifications, to yield mature rRNA molecules (Lodish et al., 2000). During maturation, the intermediates are continuously screened by an RNA surveillance machinery that recognizes aberrant or unstable RNA (Sloan et al., 2014). A hypothetical chRNA in TC#17 would have to escape this surveillance machinery to allow nuclear export and translation into protein. To ascertain which rDNA cistron parts are required for the chRNA to escape the nucleolar RNA surveillance machinery, we designed a number of ECs with a succession of missing elements compared with the full rDNA-TC fusion in TC#17 (EC1, Figure 2A). We obtained colonies with *GFP* expression levels comparable to those of TC#17 for all constructs (Figure 2B), indicating that the length of the EC can be substantially reduced compared with EC1. Strikingly, EC7 transformants could be separated into two distinct clusters, one without GFP fluorescence (GFP^-) and another with TC#17-like GFP fluorescence (GFP^+). Variations in transgene expression levels among different transformant strains are usually related to positional effects, if ECs are inserted into a genome at random, through nonhomologous end joining (NHEJ). However, positional effects do not typically produce the binary phenotypes that we observed for EC7 transformants. The lack of colonies with intermediate GFP fluorescence, therefore, suggested that NHEJ may not be responsible for EC insertion in all EC7 strains.

Strong GFP expression in EC7 transformants depends on cassette integration into the nucleolar DNA

In addition to random insertion by NHEJ, *Nannochloropsis* can integrate a DNA molecule into their genome by homologous recombination (HR) (Kilian et al., 2011). Because EC1-7 shares a substantial amount of sequence with the rDNA cistron on chromosome 3, we hypothesized a double-crossover event at the cassette termini, corresponding to the Pol I promoter and terminator sequences. Therefore, we tested if EC7 had been inserted at this locus for nine GFP^+ and GFP^- transformants through PCR genotyping (Figure 2C). The cassette was present in all transformants, but it had been inserted by HR only in GFP+ transformants (Figure 2D). These findings indicate that Pol I can drive high levels of transgene expression in *N. oceanica*, if a Pol I promoter-containing EC is inserted at the rDNA cistron locus on chromosome 3. Insertion of the same EC in random genomic loci, however, failed to promote strong expression in EC7 GFP^- transformants. To test whether Pol I can drive *GFP* expression from cisgenic promoters within the NOR but not within the rDNA cistron, we inserted EC7 in the vicinity of the chromosome 3 rDNA cistron, 188 nucleotides upstream of the promoter (EC7-vicinal, Figure 2E). Insertion of EC7 adjacent to the rDNA cistron and replacement of the rDNA cistron by HR resulted in similar *GFP* expression levels (Figure 2F). Therefore, Pol I can efficiently initiate transcription from a cisgenic Pol I promoter in *N. oceanica*, if the promoter is integrated within the NOR.

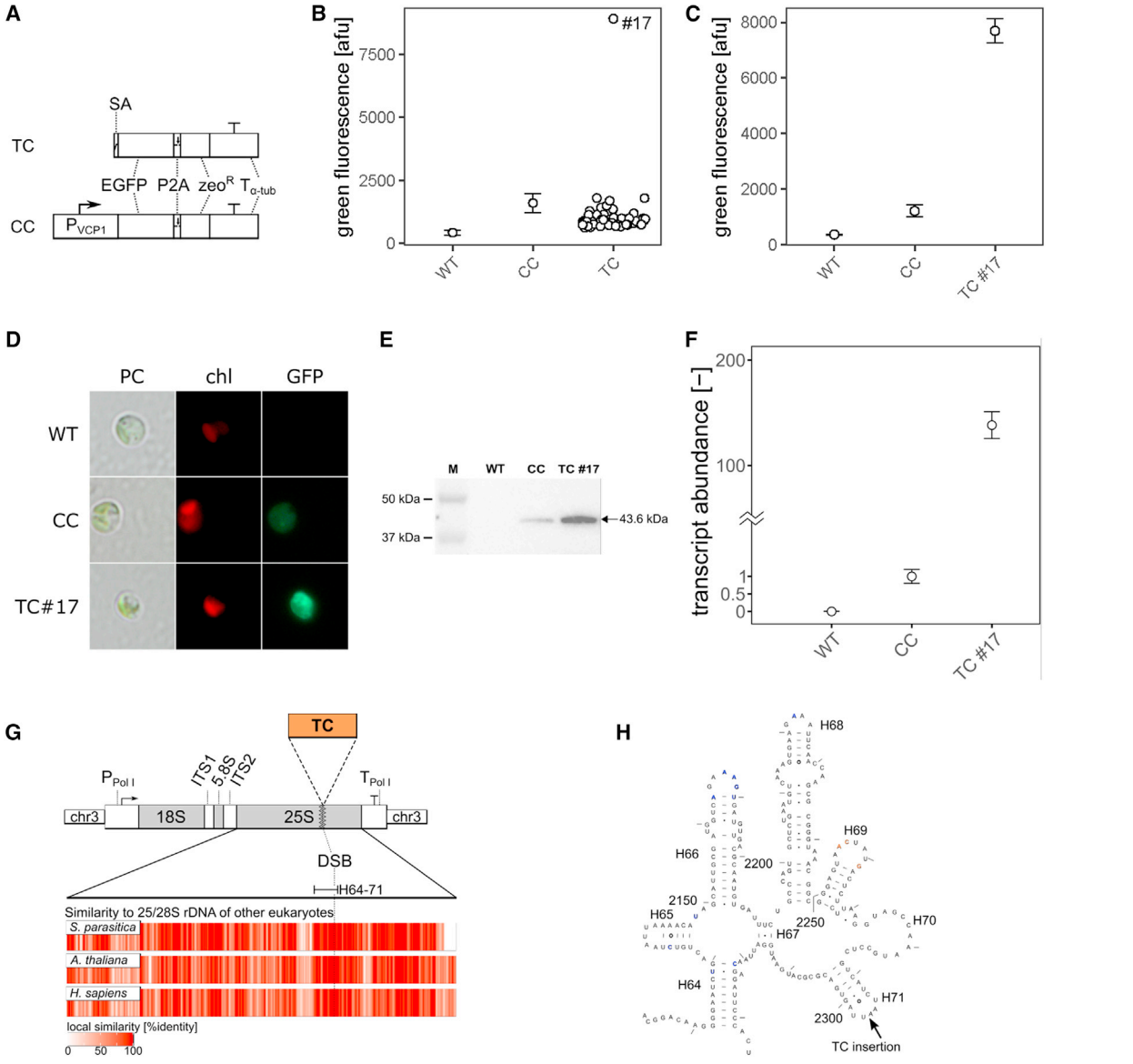

Figure 1. Discovery of an *N. oceanica* transformant with exceptionally strong *GFP* expression.

(A) Schematic of the TC and CC used to transform *N. oceanica* cells. The promoterless TC relies on insertion into an active gene for expression of *GFP* and *zeo*^{R} by "trapping" upstream exons during mRNA splicing, safeguarded by a 5′-terminal SA. Transcription of transgenes from the CC is driven by an endogenous promoter (P_{VCP1}) and terminator ($T_{\alpha-tub}$).

(B) Flow cytometric quantification of GFP fluorescence for 48 TC transformants, compared with a representative CC transformant (Supplemental Figure 1) and WT (n = 3).

(C) TC#17 exceeds GFP fluorescence of CC strains by ∼8-fold. Dot and error bars show the mean ± SD (n = 6) of the median of fluorescence emission distributions, measured by flow cytometry.

(D) Fluorescence microscopy images showing chlorophyll (chl) autofluorescence and cytosolic GFP fluorescence. PC, phase contrast channel.

(E) Western blot using an Anti-GFP antibody on 30 μg of soluble protein separated by SDS-PAGE.

(F) TC#17 showed a ∼135-fold increase in *GFP* transcript abundance relative to a CC strain, measured by RTq-PCR (n = 3).

(G) TC insertion was traced to a 25S rRNA gene of the rDNA cistron on chromosome 3 (chr3) by cassette PCR and subsequent genotyping PCR (Supplemental Figure 2). Insertion occurred at position 2333 of the 25S rDNA. Local similarity between the *N. oceanica* 25S rDNA and 25/28S rDNA of eukaryotes from different kingdoms (*Saprolegnia parasitica*, SAR, *Arabidopsis thaliana*, Plantae, *Homo sapiens*, Metazoa) are displayed as a heat map. The insertion-proximal ∼200 nucleotides, corresponding to 25S rRNA helices 64 to 71, are highly conserved among eukaryotes (∼95% ± 2% local identity, ∼71% ± 6% average identity over full-length 25S rDNA). DSB, double-strand break.

(H) Secondary structure within domain IV of the *S. cerevisiae* 25S rRNA, modified from Xie et al. (2012). Only 10 nucleotides are not conserved between *S. cerevisiae* and *N. oceanica* (highlighted blue). Residues that interact directly with the 18S rRNA in intersubunit bridge B2a are highlighted orange (Liang et al., 2007)

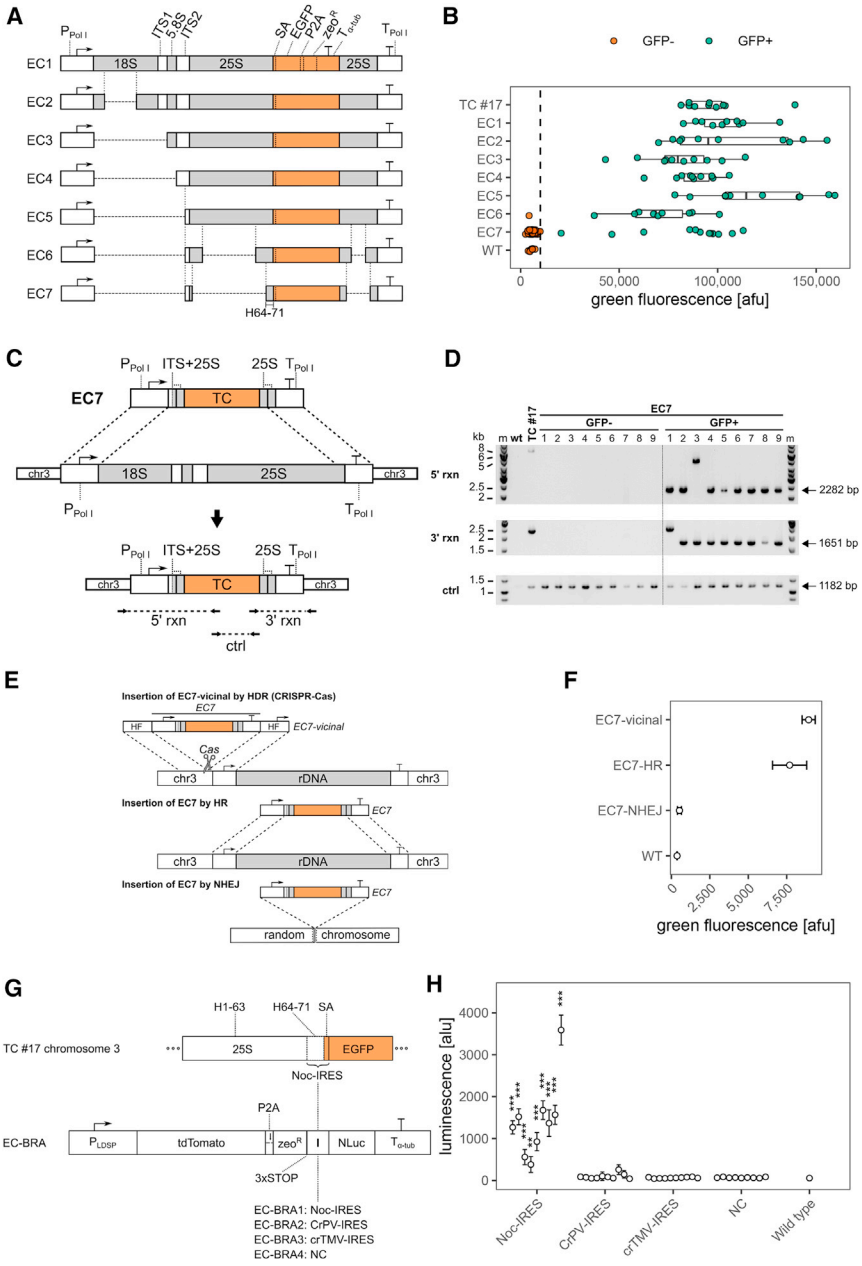

Figure 2. Definition of minimal elements for chimeric gene expression reveals strong positional effects.

(A) Schematic of ECs for deletion mutation. EC1 was amplified as the rDNA cistron from chromosome three of TC#17. Different rDNA parts were removed in EC2-7 to define the elements required for gene expression.

(B) Quantification of transgene expression in EC1-7 strains by flow cytometry. Fluorescence of EC1-6 transformants was comparable to TC#17. For EC7, several transformants showed similarly high fluorescence (GFP$^+$), but most strains did not produce any detectable GFP signal (GFP$^-$).

(C) Possible insertion mechanism of EC7 into chromosome 3 via HR. The linear EC (top) can homologously recombine with the genomic DNA (center) via double crossovers at the Pol I promoter and terminator.

(D) Characterization of EC7 strains by genotyping PCR using primers displayed as solid horizontal arrows in (C). Arrows on the right indicate the expected band size for HR-mediated cassette integration by the mechanism shown in (C). EC7 was inserted by HR only in GFP$^+$ strains.

(E) Schematic of EC7 integration in the vicinity of the rDNA cistron of chromosome 3 (top), compared with cassette insertion by HR (middle) or NHEJ (bottom). EC7-vicinal carried homology flanks (HF) that allowed insertion of 188 nucleotides upstream of the rDNA cistron using CRISPR-Cas technology.

(F) GFP quantification for an EC7-vicinal transformant compared with EC7-HR (GFP$^+$) and EC7-NHEJ (GFP$^-$). Strains carrying EC7 adjacent to or in place of the rDNA cistron show comparable GFP fluorescence. Faithful cassette insertion was verified by PCR.

(G) Schematic of bicistronic reporter construct employed for testing IRES activity of the *GFP*-proximal 5′-UTR sequence. A total of 255 nucleotides upstream of the AUG codon were considered as potential IRES (Noc-IRES). Expression of the NanoLuc luciferase (*NLuc*) gene was possible only by cap-independent translation at the (putative) IRES element.

(H) NLuc signal of EC-BRA transformants. After tdTomato pre-selection, 9 to 10 transformants were screened for luminescence (n = 4 technical replicates). Significant differences to the average of the NC were assessed by Tukey's honestly significant difference test. **p < 0.01; ***p < 0.001.

Translation of the chRNA is facilitated by an internal ribosome entry site

The previous results strongly implicated Pol I in *GFP* expression in highly fluorescent transformants. As Pol I transcripts lack the mRNA features necessary for canonical translation initiation, we hypothesized that chRNAs would undergo translation by non-canonical translation initiation. In eukaryotes, non-canonical translation initiation can occur on RNA molecules harboring an internal ribosome entry site (IRES). IRESs are well-studied *cis*-acting RNA elements that recruit the small ribosomal subunit (SSU) to a transcript via non-canonical RNA-protein interactions (Thompson, 2012). These interactions do not require presence of a 5′ cap because the SSU can bind to the IRES either directly or indirectly, mediated by eukaryotic initiation factors (eIFs) and sometimes additional proteins called IRES *trans*-acting factors (ITAFs).

GFP protein in TC#17 had the same size as conventionally expressed GFP (Figure 1E), suggesting that translation on the chRNA initiated on or in the vicinity of the *GFP* AUG codon. Most IRES types facilitate translation from an initiation codon that is located close to or at their 3′-boundary (Thompson, 2012). An IRES in the chRNA would therefore likely be positioned immediately upstream of *GFP*. By analyzing the shortest functional construct, EC7, we considered the *GFP* proximal 255

nucleotides, corresponding to the 25S rDNA bases 2130 to 2333 and the splice acceptor (SA) element of the TC, as putative IRES (Noc-IRES, Figure 2G). We then verified IRES activity of Noc-IRES *in vivo*, using a bicistronic reporter assay. Following pre-selection for similar expression of the 5′-cistron, 9/9 transformant strains containing the Noc-IRES (EC-BRA1) showed expression of the 3′-cistron (Figure 2H). This is evidence for an IRES element that is functional in a microalga. Furthermore, we found that the well-studied cricket paralysis virus IGR IRES (Hodgman and Jewett, 2014), and crucifer-infecting tobamovirus IRES (Dorokhov et al., 2002) were inactive in *N. oceanica*.

The proximal 3′-UTR of the chRNA enhances gene expression by an unknown mechanism

Based on the previous results, we concluded that the rDNA cistron on chromosome 3 of *N. oceanica* is a genomic safe harbor for transgene expression by Pol I and that uncapped chRNA is efficiently translated from the Noc-IRES. Next, we studied the impact of removing different EC elements located on the 3′-side of *GFP* (Figure 3A). Deletion constructs were inserted into the genome of *N. oceanica* at the safe harbor site, and EC insertion was confirmed by PCR genotyping for all transformants discussed hereafter. ECT1 resembled the previous shortest construct EC7, but it was missing the 25S rDNA sequences downstream of *GFP*. *GFP* expression from ECT1 was similar to EC7 (Figure 3B). Likewise, deletion of the Pol I terminator (ECT2) did not affect *GFP* expression compared with EC7, indicating that the α-tubulin (Pol II) terminator can substitute for the Pol I terminator.

Surprisingly, deletion of the α-tubulin terminator ($T_{\alpha-tub}$), which originally was part of the TC, caused an 86% ($p < 0.001$) decrease in fluorescence of ECT3 strains compared with EC7 strains (Figure 3A and 3B). In eukaryotes, Pol II terminators are required for transcriptional arrest and polyadenylation of nascent mRNAs. Poly(A) tails are essential for mRNA export, stability, and translation, including cap-independent translation initiation at several IRESs (Bradrick et al., 2007). However, Pol II terminators should not be able to cause polyadenylation of Pol I transcripts because polyadenylation requires the C-terminal domain of the Pol II enzyme complex (Hirose and Manley, 1998). Nevertheless, we investigated a possible link between polyadenylation of chRNA and *GFP* expression, by substituting $T_{\alpha-tub}$ with a different endogenous Pol II terminator (T_{LDSP}) and a poly(A)-encoding tract in ECT4 and ECT5, respectively. The poly(A) tract in ECT5 was flanked on its 3′-boundary by a self-cleaving HDV ribozyme sequence, to facilitate formation of a free poly(A) tail on the chRNA by endoribonucleolytic cleavage (Poliner et al., 2018b). Transformants carrying ECT4 at the safe harbor site had a decreased *GFP* expression, whereas GFP fluorescence of ECT5 transformants was higher than for ECT3, but 79% ($p < 0.001$) decreased compared with EC7. Therefore, $T_{\alpha-tub}$ appears to improve gene expression by a mechanism that is unrelated to polyadenylation.

The 3′-UTRs of mRNAs can contain *cis*-acting elements that are involved in regulation of post-transcriptional processing, transcript degradation, nucleocytoplasmic transport, and translation initiation through interaction with specific RNA binding proteins (Matoulkova et al., 2012). As the α-tubulin terminator increased gene expression by a mechanism unrelated to polyadenylation, it is likely that it contains one or more *cis*-acting elements. *GFP* transcript abundance in ECT3 transformants was decreased by 53% ($p < 0.001$) compared with ECT1 and by 33% ($p = .039$) compared with ECT2 (Figure 3C), suggesting that the $T_{\alpha-tub}$ element increases chRNA stability. However, *GFP* fluorescence for ECT3 strains was decreased by 84% ($p < 0.001$) compared with ECT2 strains, which exceeds the 33% difference in transcript abundance. Therefore, the most significant effect of $T_{\alpha-tub}$ on gene expression may be related to chRNA translation or nucleocytoplasmic transport.

The hybrid IRES requires the 25S rRNA part and splice acceptor element for activity

We designed additional deletion constructs to narrow down functionally relevant elements of the Noc-IRES. Deletion of the 25S rRNA part of the IRES (ECT2Δ25S, Figure 3D) decreased *GFP* expression of transformants by 97% ($p < 0.001$) compared with full-length Noc-IRES-containing cassettes (Figure 3E). Consequently, the 25S rRNA part is essential for high activity of this IRES. The residual GFP levels in ECT2Δ25S indicate that the splice acceptor (SA) element may have an intrinsic ability to drive cap-independent translation, albeit at low efficiency.

Transformants carrying ECT2 with a deletion of either the SA element (ECT2ΔSA) or of the entire Noc-IRES (ECT2ΔIRES) had no detectable GFP fluorescence (Figure 3E). The role of the SA element for cap-independent translation from the Noc-IRES may be linked to the polypyrimidine tract that it contains because polypyrimidine tracts are essential elements in several natural and artificial IRESs (Mitchell, 2005). The complete loss of GFP fluorescence in ECT2ΔIRES transformants further shows that the α-tubulin terminator is not able to drive effective cap-independent translation of the chRNA independently of an IRES. Therefore, the mechanism by which $T_{\alpha-tub}$ enhances gene expression is likely distinct from the mechanisms linked to 3′-cap-independent translation enhancer elements of plant viruses (Truniger et al., 2017). Instead, $T_{\alpha-tub}$ may function as a translation enhancer element only in conjunction with an IRES. A similar mechanism is known for foot and mouth disease virus and Hepatitis C virus, where 3′-UTR elements stimulate cap-independent translation from IRESs located in the 5′-UTR (Serrano et al., 2006; Song et al., 2006; García-Nuñez et al., 2014).

The Pol I promoter is indispensable for strong expression at the safe harbor

Substitution of the Pol I promoter with an endogenous Pol II promoter (Figure 3F) caused a 95% decrease ($p < 0.001$) in GFP fluorescence of ECT2-P_{LDSP} transformants (Figure 3G). GFP fluorescence of ECT2-P_{LDSP} with faithful cassette insertion at the safe harbor site was comparable to values obtained for random insertion of this cassette in the genome (data not shown), indicating that the rDNA cistron on chromosome 3 may also be used for gene expression by Pol II, albeit not necessarily with an advantage over insertion at nucleoplasmic loci.

Establishing a screening pipeline for transformants with strong gene expression

Strikingly, gene expression with this novel system was strong enough to easily distinguish *N. oceanica* transformant colonies producing a suitable fluorescent reporter protein on agar plates,

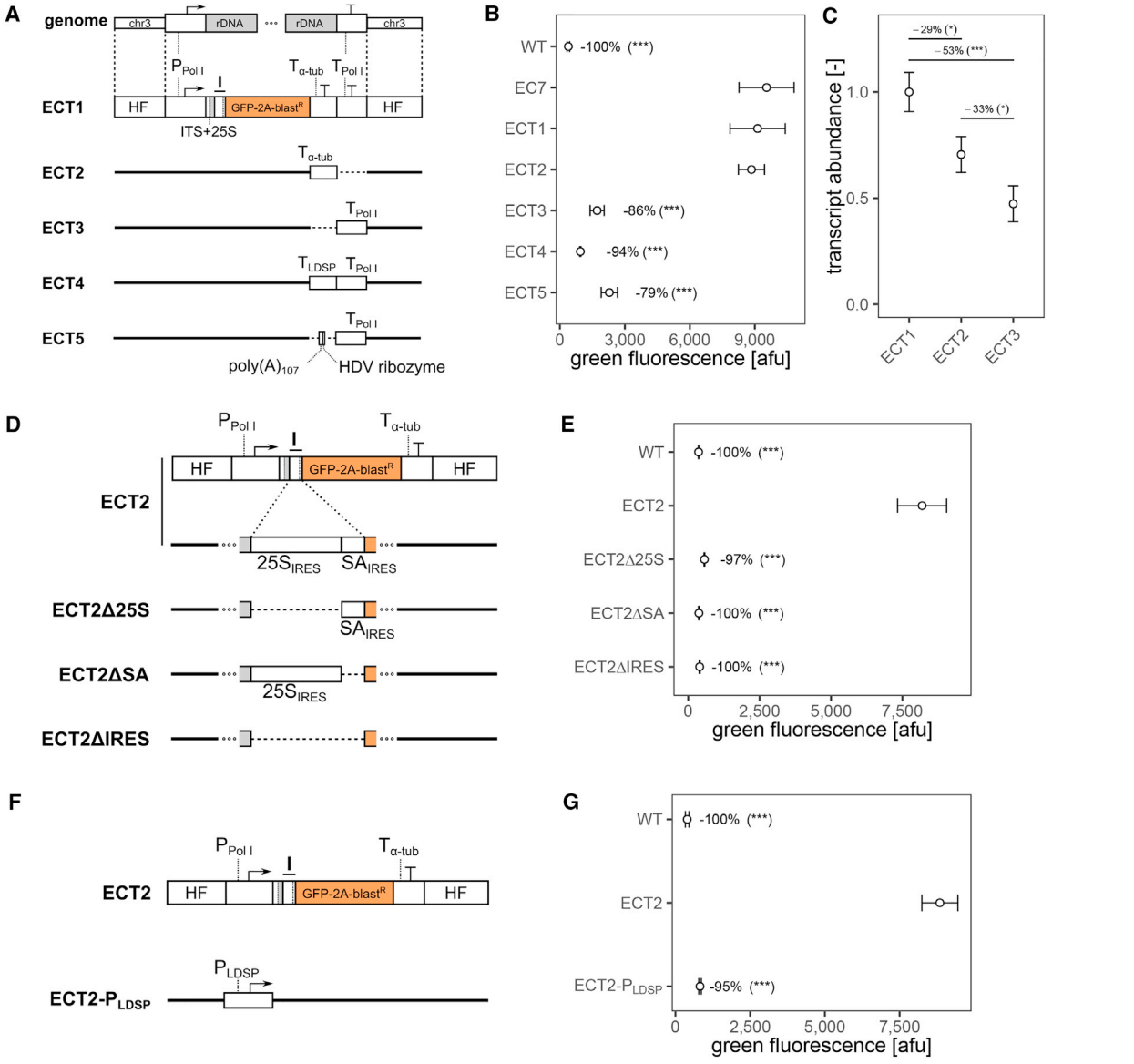

Figure 3. Defining regulatory elements required for strong chimeric gene expression at the safe harbor site.

(A) Design of ECs to study the requirement for the transcriptional terminators. ECT1 resembled the minimal functional construct EC7, but carried no 25S rDNA sequences in the 3′-UTR. The Pol I and Pol II terminator were deleted in ECT2 and 3, respectively. In ECT4 and 5, the Pol II terminator was replaced with a different Pol II terminator and a poly(A)-encoding tract flanked by the HDV ribozyme, respectively. When available, at least three transformants were analyzed (Supplemental Figure 3A) before choosing a representative strain.

(B) Flow cytometric GFP quantification for representative transformants carrying constructs shown in (A). The mean ± SD (n = 6) are shown together with changes relative to an EC7 transformant. EC7, ECT1 and ECT2 promote comparable levels of *GFP* expression, whereas deletion of the Pol II terminator drastically decreased expression, which was not restored by using a different Pol II terminator or poly(A) tract.

(C) *GFP* transcript abundance in transformants, measured by RTq-PCR. The mean ± SD (n = 3) normalized to the abundance in the ECT1 transformant are shown together with relative changes between groups and significance levels.

(D) Schematic of ECs with partial or full Noc-IRES deletion. Six transformants were analyzed by flow cytometry (Supplemental Figure 3B).

(E) Flow cytometric GFP quantification for representative strains carrying constructs shown in (D).

(F) Schematic of EC carrying substitution of the Pol I promoter with a Pol II promoter.

(G) GFP expression levels of an ECT2-P_{LDSP} transformant, quantified by flow cytometry. (E and G) The mean ± SD (n = 6) are shown together with relative changes compared with ECT2. (B, C, E, and G) Faithful EC insertion at the safe harbor was verified by PCR for all presented strains. Significant differences were assessed by Tukey's honestly significant difference test. *p < 0.05, **p < 0.01, ***p < 0.001.

regardless of pigment autofluorescence. We used this to construct a "landing pad" (LP) strain (LP-tdTomato) that expresses the fluorophore *tdTomato* from a modified version of ECT2, inserted at the safe harbor (Figure 4A). When LP-tdTomato is transformed with an LP-targeted EC containing a gene of interest, faithful EC insertion at the LP will result in replacement of the *tdTomato* cassette, rendering loss of tdTomato fluorescence a marker for correct insertion. A resistance

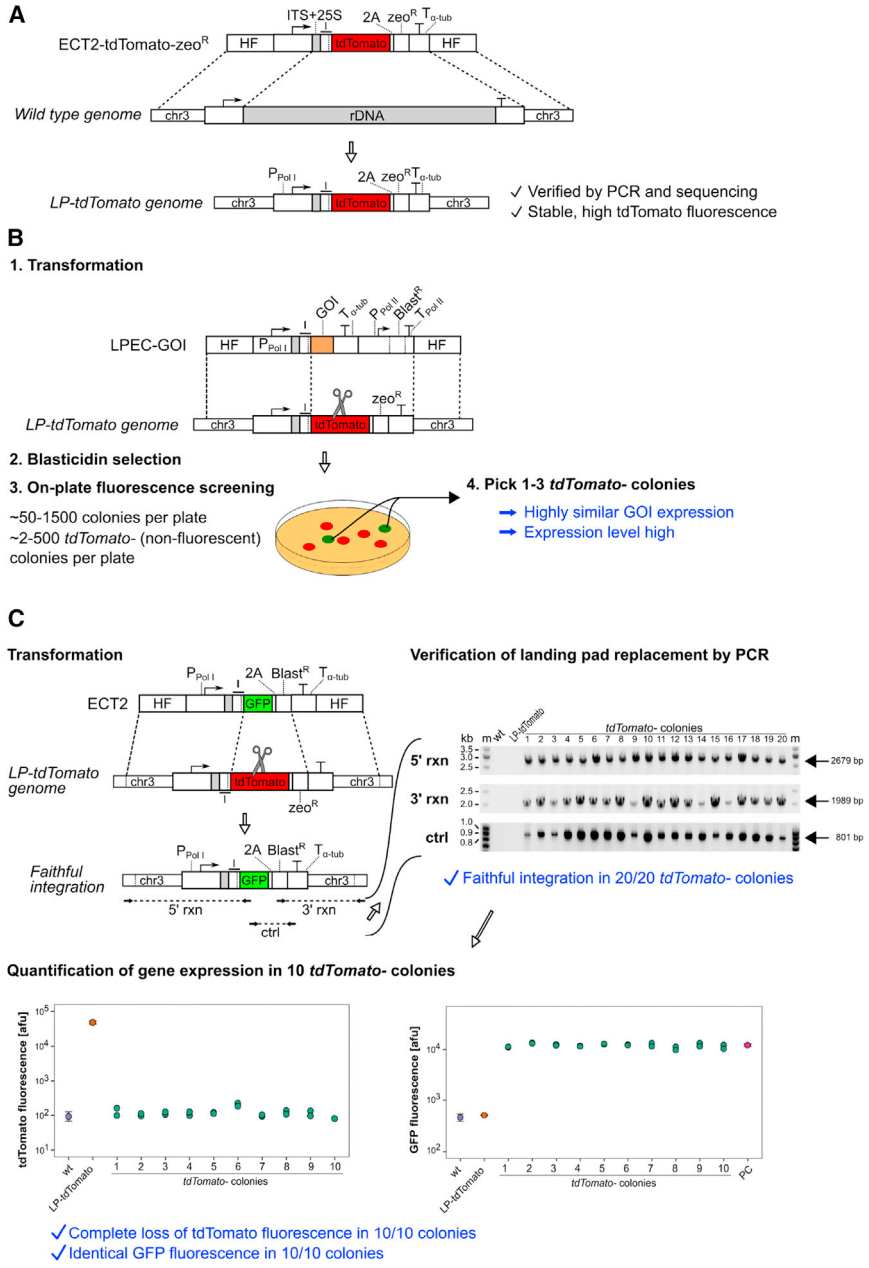

Figure 4. Development of a screening pipeline for transformants with faithful cassette integration.

(A) Construction of a landing pad strain, expressing *tdTomato* and *zeo*^{R} at the safe harbor site. Faithful integration of a modified version of ECT2 was verified by PCR. The tdTomato fluorescence of this strain is detectable on agar plates even for small colonies.

(B) Example for the construction of transformants expressing a gene of interest at the safe harbor site. The landing pad strain is transformed with a cassette targeting the landing pad (LPEC-GOI) containing the gene of interest (GOI) under control of the Pol I promoter, the Noc-IRES (I) and $T_{\alpha-tub}$, flanked on their 3′-side by a Pol II-driven *Blast*^{R} cassette. Subsequently, transformants carrying LPEC-GOI are selected for antibiotic resistance. Faithful EC integration at the landing pad results in replacement of the *tdTomato* gene. Therefore, screening for tdTomato$^-$ colonies (i.e., colonies with no detectable tdTomato fluorescence) by on-plate fluorescence imaging allows selection of transformants with faithful EC insertion.

(C) Validation of the screening pipeline using ECT2. After transformation of the landing pad strain with ECT2, transformants were selected for Blasticidin resistance and screened for loss of tdTomato fluorescence. Faithful ECT2 integration was found in 20/20 tdTomato$^-$ colonies, analyzed by PCR. PCR primers are displayed as solid straight arrows in the schematic. Flow cytometry analysis confirmed loss of tdTomato fluorescence and revealed identical GFP expression levels in 10/10 tdTomato$^-$ colonies. Fluorescences of biological duplicates are shown separately for tdTomato$^-$ colonies. Values of controls are shown as mean ± SD (n = 3).

gene (other than *zeo*^{R}) can either be added to the gene of interest using a 2A-encoding sequence or as a second cistron, under control of a Pol II promoter (Figure 4B).

To validate this novel screening pipeline, we transformed the LP strain with the optimized, minimum functional construct ECT2 (Figure 4C) and screened transformants for loss of on-plate tdTomato fluorescence; 20/20 tdTomato$^-$ colonies showed faithful EC insertion and had lost resistance to Zeocin (Supplemental Figure 4). Flow cytometry analysis confirmed complete loss of tdTomato fluorescence in 10/10 colonies and revealed highly similar GFP fluorescence for all strains. Growth analysis revealed an only minor impact on growth, with the maximum specific growth rate of the LP strain being 15% reduced compared with the wild type (Supplemental Figure 6). It remains to be tested whether this reduction is a consequence of recombinant protein synthesis or of functional replacement of an rDNA cistron.

Antibody production in *N. oceanica* using the new gene expression system

To put the new gene expression system to the test, we engineered strains to produce a camelid type V_HH antibody. We transformed the landing pad strain LP-tdTomato with an EC (Figure 5A) encoding a his-tagged anti-GFP V_HH. After antibiotic selection, we selected three colonies without detectable on-plate tdTomato fluorescence and found faithful EC integration at the landing pad in all strains. We purified V_HH protein from one of the strains by affinity chromatography (Figure 5B) and tested its ability to efficiently bind GFP by indirect ELISA (Figure 5C). The relationship between ELISA signal and V_HH concentration was reliably described by a Michaelis-Menten model (R^2 = 0.995, Figure 5C). Simultaneously, we tested antibody titer of microalgal raw extracts for all three strains (Figure 5D). Using the purified V_HH as a calibration standard (Figure 5C), we calculated the V_HH concentration in the raw extracts (Figure 5E). V_HH accounted for 7.98% to 8.55% of total extracted protein, showcasing the potential of the new

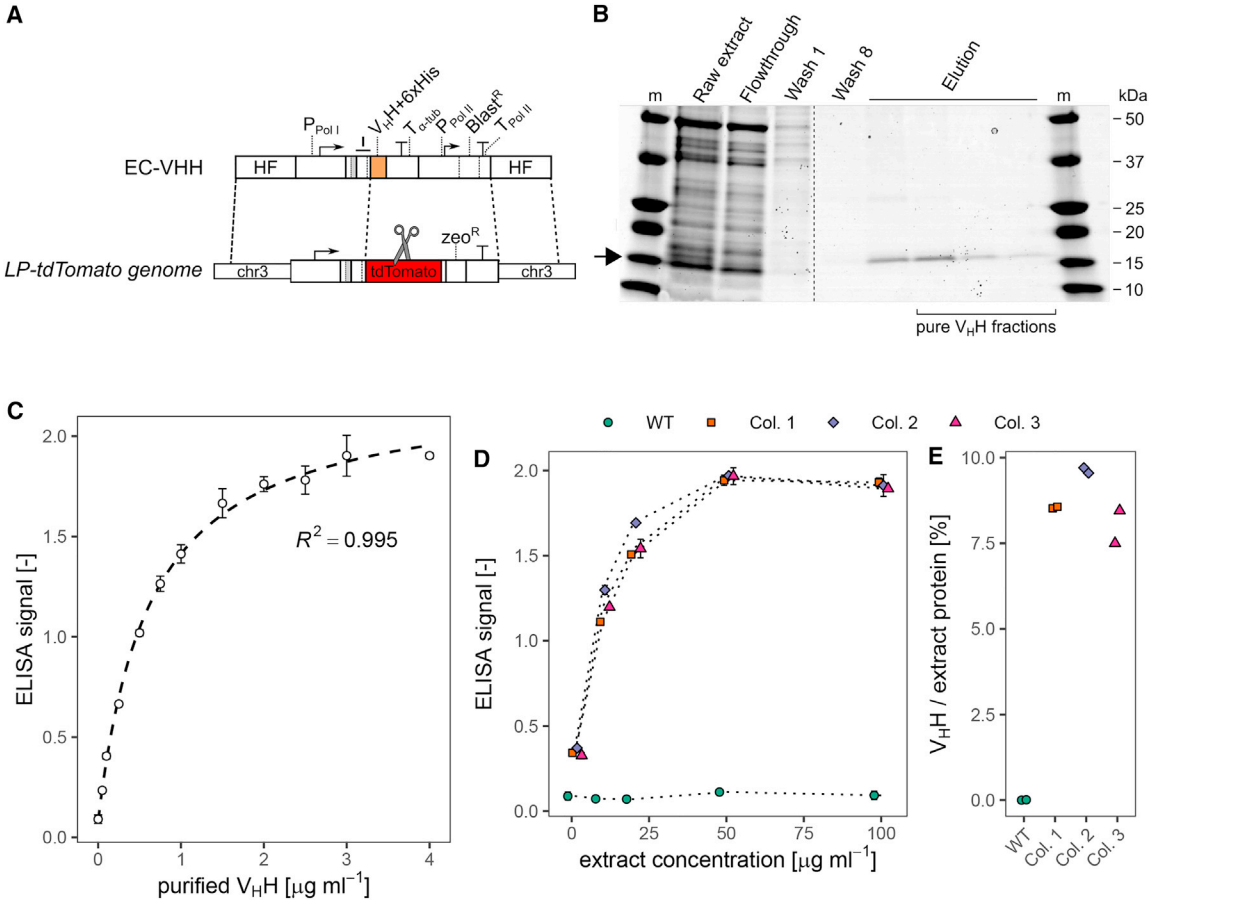

Figure 5. Expression of a V_HH antibody using the novel expression system.

(A) Design of an EC encoding a V_HH with a C-terminal his_6-tag. The cassette was inserted into the safe harbor site of the landing pad strain LP-tdTomato. Faithful EC integration was verified for three *tdTomato*-strains by PCR.

(B) Affinity-chromatography purification of V_HH from protein extracts of an EC-VHH transformant, analyzed by SDS-PAGE.

(C–E) Pure V_HH fractions were pooled and used as standard for quantification of antibody activity in microalgal protein extracts. (C) Calibration curve of indirect ELISA using purified V_HH. The relationship between OD_{450} values and V_HH concentration was excellently explained by a Michaelis-Menten fit. (D) ELISA signal of microalgal protein extract from three independent EC-VHH tdTomato$^-$ transformant colonies (Col. 1–3). Protein extracts showed the same Michaelis-Menten-like response as purified V_HH. (E) Fraction of V_HH per protein in microalgal extracts. Fractions were calculated using the nonlinear regression-derived Michaelis-Menten fit of purified V_HH (C) and ELISA signal of microalgal protein extract at 1 µg mL^{-1} (D).

gene expression system to produce high levels of functional protein in *N. oceanica*.

DISCUSSION

In this study, we developed a novel system for gene expression in *N. oceanica* that relies on highly efficient Pol I for transcription and on an IRES element for translation. Prior studies have shown that IRESs can be used to overcome translational block of Pol I-transcribed molecules in mammalian cells, yeast, and plants (Palmer et al., 1993; Ghoshal et al., 2004; Oem et al., 2007; Wen et al., 2008; Komarova et al., 2012; Sims et al., 2019), but no previous reports exist for microalgae. Importantly, we found that high efficiency of the Pol I + IRES expression system depended on vector insertion into the NOR of *N. oceanica*.

The discovery of a chimeric IRES with activity in a microalga was rather serendipitous. Several elements of this IRES may be involved in ribosomal recruitment in *N. oceanica*. Polypyrimidine tract-binding protein (PTB) is the most ubiquitous ITAF and it is required for activity of several viral, cellular, and artificial IRESs (Mitchell, 2005; Martinez-Salas et al., 2018). Previous studies have shown that artificial IRESs harboring only a polypyrimidine tract with the PTB recognition sequence can recruit the ribosome in animal cells (Mitchell, 2005; Spriggs et al., 2005). In this context, removal of the polypyrimidine tract-containing splice acceptor part of the Noc-IRES abolished gene expression (Figure 3D and 3E), suggesting that this element and possibly interaction with PTB are essential for Noc-IRES activity. The 25S rRNA part of the Noc-IRES may also play a role in ribosome recruitment, as it is a highly conserved region of the ribosome, i.e., helices 64 to 71 of the large subunit (LSU, Figure 1G and 1H). Helix 69 (H69) is part of the intersubunit bridge B2a that connects the LSU to the decoding center of the small subunit (SSU) at the heart of the ribosome during translation (Yusupov et al., 2001). It has been shown that deletion of H69 entirely prohibits subunit joining in the absence of tRNA, indicating that B2a is the most essential connection between the ribosomal subunits (Ali et al., 2006).

Given the strong interaction of H69 with helix 44 of the 18S rRNA in B2a, the H69 hairpin structure of the Noc-IRES may directly recruit the SSU to the transcript via base-pairing with complementary sequences of the 18S rRNA, analogous to the interaction during subunit joining. Similarly, several viral and cellular IRESs recruit the SSU through complementarity to the 18S rRNA (Chappell et al., 2000; Malygin et al., 2013; Angulo et al., 2016).

Both parts of the chimeric IRES are essential for strong activity, but further studies are necessary to elucidate the mechanism of ribosomal recruitment. Identification of IRES-interacting proteins by RNA-centric methods would help to check for interaction of the Noc-IRES with ITAF proteins such as PTB. At the same time, further deletion studies may help to define the minimal functional IRES, whereas constructs carrying targeted nucleotide mutations would be required to validate an interaction of the 25S part of the Noc-IRES with SSU. Thorough understanding of the *modus operandi* of this IRES might make it possible to use its structural blueprint for designing artificial IRESs for other organisms.

Previous studies had identified several transcriptional promoters that enable gene expression in *N. oceanica*; however, positional effects and modulation of gene expression due to physiological stimuli limit the application of commonly used promoters, and necessitate screening large numbers of transformants (Supplemental Figure 1). The identification of the nucleolus as a genomic safe harbor in *N. oceanica*, coupled with a new chimeric gene expression system, enables simple manufacturing of microalgal transformants with highly consistent transgene expression. This novel tool represents an outstanding asset for genetic and metabolic engineering of an industrially attractive microalga (Li et al., 2019). Moreover, it can be used to transform *Nannochloropsis* into a viable source of functional (aquaculture) feed (Sarker et al., 2018, 2020; Adissin et al., 2020). A list of potential strategies includes tailoring the algal proteome to dietary requirements (Herman and Schmidt, 2016), enrichment for bioactive proteins, including antimicrobial peptides and immunostimulants (Falco et al., 2012), or design of a feed-based vaccine (Clarke et al., 2013; Herman and Schmidt, 2016). At the same time, it will be interesting to transfer this type of gene expression system to other microalgae and eukaryotes in general.

METHODS

Media and strains

N. oceanica IMET1 was a kind gift from Prof. Jian Xu (Qingdao Institute for Bioenergy and Bioprocess Technology, Chinese Academy of Sciences). The organism was cultivated using artificial sea water (ASW) containing 419.23 mM NaCl, 22.53 mM Na_2SO_4, 5.42 mM $CaCl_2$, 4.88 mM K_2SO_4, 48.21 mM $MgCl_2$, and 20 mM HEPES at pH 8, supplemented with 2 mL L^{-1} of commercial nutribloom plus (Necton, Portugal) growth media (ASW-NB) in an HT Multitron Pro (Infors Benelux, Netherlands) orbital shaker unit operated at 25°C, 90 rpm, 0.2% CO_2 enriched air and illuminated with warm-white fluorescent light bulbs with an intensity of 150 μmol m^{-2} s^{-1} with a 16:8-h diurnal cycle. Growth comparison of the landing pad strain and wild type was carried out at 225 μmol m^{-2} s^{-1} using an Algem HT24 photobioreactor (Algenuity, UK) that was placed inside an HT Multitron Pro operated at conditions as described above.

The landing pad strain *N. oceanica* IMET1-LP-tdTomato can be requested from Wageningen University, by contacting the corresponding author. The strain was not tested for mycoplasma contamination.

Plasmid construction

Plasmids were constructed using either restriction cloning or Gibson assembly technique. For restriction cloning, different DNA elements were designed with terminal recognition sites for type IIS restriction enzyme Eco31I. Fragments were amplified via PCR (Q5 polymerase, NEB #M0492) according to manufacturer instructions, column or gel-purified (Thermo Fisher Scientific #K0831), and 65 ng of the backbone was mixed with inserts, typically in molar ratios of 1:2, supplemented with T4 DNA ligase (Thermo Fisher Scientific #EL0011) and the corresponding buffer as well as with Eco31I (Thermo Fisher Scientific #FD0293). The mixture was incubated for six cycles of 5 min, 37°C and 5 min, 16°C and finally for an additional 10 min, 37°C and 5 min, 65°C and transformed into chemically competent *Escherichia coli* TOP10. Competent cells were created using *Mix & Go E. coli* transformation kit (Zymo Research #T3001). For Gibson assemblies, we used NEBuilder HiFi DNA Assembly mix (New England Biolabs) according to manufacturer instructions, typically with 25 nucleotide overlap. The cloning vector employed for all experiments was pUC19 (GenBank accession number M77789.2). *N. oceanica* genomic sequences including promoters, terminators, splice acceptor, rDNA sequences, and homology flanks were amplified from genomic DNA of *N. oceanica* IMET1 via PCR. The bleomycin resistance gene of *Streptoalloteichus hindustanus* (*zeo*^{R}, GenBank accession number A31898.1) was amplified from pPtPuc3 (addgene #62863) which was a kind gift from Hamilton Smith. The *GFP* sequence was codon harmonized and synthesized by Integrated DNA Technologies, Inc. (Coralville, USA). The viral P2A linker sequence was codon optimized and synthesized together with *GFP*. *tdTomato* was amplified from pCSCMV:tdTomato (addgene #30530). *Blast*^{R} and the NanoLuc luciferase gene were amplified from pNOC-ARS-CRISPR-BlastR, which was a kind gift from Mihris Naduthodi.

Genomic DNA was extracted from exponentially growing cultures using Phire Plant Direct PCR (Thermo Scientific #F160) dilution buffer. Approximately 1×10^7 cells were pelleted (15 min) and resuspended in 100 μL of dilution buffer, frozen at −20°C for 20 min, boiled (90°C for 10 min and 95°C for 5 min) and pelleted (10 min) again. The supernatant was used as template for PCR. For genotyping PCR, Phire polymerase (Thermo Scientific #F160) was used according to manufacturer instructions.

Transformation of *N. oceanica*

Transformation of *N. oceanica* was carried out as described previously (Südfeld et al., 2021), based on the electroporation protocol from Vieler et al. (2012). Cells were plated on ASW-NB agar (1%) plates, supplemented with 5 μg zeocin mL^{-1} for selection of zeocin-resistant cells or 100 μg Blasticidin S mL^{-1} for selection of Blasticidin S-resistant cells. Plates were incubated at 25°C and 80 μmol m^{-2} s^{-1} for 3 to 4 weeks before transformant colonies were transferred to liquid media containing antibiotic. Depending on the desired analytical method, transformants were either cultivated in microplates (48 wells or 96 wells) or shake flasks for several days. Constructs TC, CC, and EC1-7 were used to transform wild-type *N. oceanica* culture, whereas other constructs were inserted into the genome of a parental *tdTomato*-expressing strain, assisted by CRISPR-Cas technology.

Transformation of *N. oceanica* using Cas12a ribonucleoprotein

Targeted gene insertion in the NOR of chromosome 3 was achieved using CRISPR-Cas technique with a ribonucleoprotein (RNP)-based approach, as described previously (Naduthodi et al., 2019). Purified FnCas12a and CRISPR RNAs (crRNA) were assembled *in vitro* and co-transformed with editing template to facilitate homology directed repair-based insertion of the ECs. crRNAs were designed using CHOPCHOPv2 (Labun et al., 2016). This method was employed for construction of transformants carrying EC7-vicinal, ECT1-5, and ECT2 derivatives, and EC-VHH. To facilitate straightforward screening and selection of transformant colonies with faithful EC insertion, the parental strain for transformation with ECT1-5 and ECT2 derivatives carried a modified version of

Channel	Fluorophore	Bandwidth [nm]	Detector gain
Forward scatter	-	488	2
Side scatter	-	488	22%
FL1	GFP	510 ± 10	45%
FL2	tdTomato	585 ± 15	45%
FL5	chlorophyll a	720 ± 30	40%

Table 1. Detector wavelengths and target fluorophore for flow cytometry analysis on an SH800S instrument

EC7 (EC7-tdTomato) and the parental strain for transformation with EC-VHH or ECT2 (during validation of screening pipeline) carried a modified version of ECT2 (ECT2-tdTomato-zeoR) in place of the rDNA cistron 3.1. EC7-tdTomato and ECT2-tdTomato-zeoR were constructed by transformation of wild-type *N. oceanica* as described above, and colonies were screened for faithful EC insertion by fluorescence quantification and PCR. crRNA for transformation with EC7-vicinal was oCS249 and crRNA for transformation with all other constructs was a 1:1 mixture of oCS250 and oCS251, both targeting different positions of the *tdTomato* gene. HR efficiencies that were obtained for CRISPR-Cas-assisted transformations (i.e., the fraction of faithful cassette insertions over total number of screened colonies) were 0.03 ± 0.01 for oCS249, and 0.28 ± 0.17 for oCS250 + oCS251.

Analysis of gene expression

Flow cytometry analysis

Expression of GFP was quantified by measuring single-cell fluorescence through flow cytometry analysis using different devices. Exponentially growing *N. oceanica* cultures were diluted to ~4×10^6 cells mL^{-1} with ASW prior to analysis. Analyses for EC1-7 transformants were carried out using an Attune NxT flow cytometer (Invitrogen, USA) equipped with lasers of 405 nm and 488 nm wavelengths. Singlet cells were selected by appropriate gating in the 488 nm forward and side scatter channels and only cells with a minimum chlorophyll a autofluorescence of 10,000 arbitrary fluorescence units (afu) in the red channel (detection at 710 ± 25 nm, excitation at 405 nm) were considered for statistical analysis. GFP signal was measured at 530 ± 15 nm with blue excitation. Detector gains were set to 350 mV (forward and side scatter), 400 mV (710 ± 25 nm) and 500 mV (530 ± 15 nm). All other flow cytometry analyses were carried out using an SH800S (Sony Biotechnology, USA) instrument equipped with a 70 μm nozzle microfluidic chip and lasers of 488 nm and 561 nm wavelengths. Detector wavelengths for different channels are shown in Table 1. A minimum of 50,000 events were screened per sample and only singlet events with a minimum chlorophyll a fluorescence of ~7,500 afu were considered for statistical analysis. Gating for singlet events was done as previously described (Südfeld et al., 2020).

Fluorescence microscopy

Fluorescence microscopy was carried out on fresh samples of exponentially growing cultures, using an Olympus IX71 inverted system microscope with a ×100 magnification objective, a U-MWIBA2 filter cube (excitation 460–490 nm, emission 510–550 nm) for detection of GFP fluorescence, and a U-MWIG3 filter cube (excitation 530–550 nm, emission 515+ nm) for detection of chlorophyll fluorescence. Exposure settings were kept identical for different samples and chosen to accommodate the strong fluorescence of TC #17 without over-exposure.

RTq-PCR analysis

Transcript abundance was quantified using RTq-PCR. RNA was extracted from exponentially growing cultures using the E.Z.N.A. plant RNA kit (Omega Bio-tek #R6827) according to manufacturer instructions with the following modifications: ~3×10^8 cells were harvested by centrifugation (4,000 × *g* for 5 min at 4°C), the pellet was snap-frozen in liquid N_2 and resuspended in 500 μL of RB buffer prepared with fresh 2-mercaptoethanol. The suspension was transferred to a bead beater tube (MP Biomedicals #116914500) and bead beat for 2 × 60 s at 4,000 rpm with a 120 s pause between cycles with a Precellys 24 homogenizer (Bertin Technologies). After centrifugation (2,500 × *g*, 1 min) the supernatant was transferred to a Homogenizer Mini Column and processed according to the "Standard Protocol." An on-column DNase I digestion step was conducted using RNase-free DNase Set I (Omega Bio-tek #E1091) for 25 min. After elution in nuclease-free (NF) H_2O, RNA concentration was measured with a NanoDrop device and integrity of RNA was monitored by separating 500 ng of RNA with agarose gel electrophoresis using a 1.25% agarose gel and an RNA denaturation step (10 min at 70°C in a 66% [v/v] formamide solution) prior to gel loading. RNA was subjected to a second DNase I digestion using TURBO DNA-free Kit (Thermo Fisher Scientific #AM1907) with a 30 min digestion. Absence of DNA was verified by using 50 ng of RNA as PCR template with Taq polymerase (Thermo Fisher Scientific #K1081).

cDNA was synthesized by subjecting typically 700 ng of total RNA to reverse transcription (New England Biolabs #M0253S) with random hexamer primers according to manufacturer instructions. cDNA libraries were diluted 1:48 with NF H_2O and subjected to qPCR on a CFX96 Real-Time PCR detection system (Bio-Rad laboratories) using SYBR Select Master Mix (Thermo Fisher Scientific #4472903) according to manufacturer instructions with 200 nM primer concentrations in technical triplicate. Primer efficiencies were calculated from a standard curve with purified PCR product at concentrations between 1×10^3 and 1×10^6 template copies μL^{-1}. *GFP* transcript abundance was quantified relative to *Actin* as reference gene.

Western blot

Protein presence and size was verified using western blot technique. Soluble protein was extracted from exponentially growing *N. oceanica* cultures; ~1.5×10^9 cells were pelleted (2,500 × *g*, 5 min) and resuspended in 500 μL 0.075 mM Tris buffer (pH of 8). The suspension was bead beat 3 × at 2,500 rpm for 20 s with a 120 s pause between cycles, using Lysing Matrix E (#116914500, MP Biomedicals) with a Precellys 24 homogenizer (Bertin Technologies). Subsequently, the tubes were frozen at −20°C for 90 min, thawed at 20°C, and pelleted (15,000 × *g*, 5 min). The protein-containing supernatant was transferred to fresh tubes and protein contents were quantified using a modified Lowry assay (DC Protein Assay, Bio-Rad #5000116) with a BSA calibration standard. An amount of 45 μg of soluble protein was mixed with 5 × Laemmli reagent, heated to 85°C for 3 min, and separated by SDS-PAGE on 4% to 15% TGX protein gels (Bio-Rad #5678084) with TGS running buffer for 40 min at 200 V. Immediately after, proteins were transferred to PVDF membranes (Bio-Rad #162-0177) using a Criterion Blotter (Bio-Rad) at 50 V for 60 min with pre-cooled Towbin buffer. Membranes were blocked with TBS-T+1% skim milk powder (Bio-Rad #170-6404), incubated with a GFP antibody (500 × diluted, Thermo Fisher Scientific #14-6674-82) on a rocking shaker for 1.5 h at RT and then overnight at 4°C, washed thrice with TBS-T, incubated for 2 h with a horseradish peroxidase-conjugated secondary antibody at RT (2,000 × diluted, Thermo Fisher Scientific #A10551) and washed thrice again. Chemiluminescence was detected using a Chemi-Doc XRS+ system and enhanced chemiluminescence substrate (Thermo Fisher Scientific #34096). After detection, gels and membranes were

Coomassie-stained to confirm appropriate protein separation and equal blotting efficiencies across samples.

Luminescence assays

Nanoluc activity was determined using a modified version of the protocol reported by Poliner and colleagues (2018a). Briefly, NanoLuciferase substrate (Promega) was diluted 10,000 × in ASW. Microalgal cultures were diluted in ASW to a concentration of 2×10^7 cells/mL and 100 μL of cell suspension was transferred to a 96-well microtiter plate. For bicistronic assays, tdTomato fluorescence was measured using a CLARIOstar Plus plate reader (BMG LABTECH GmbH), to verify that transgene transcription was similar in all transformant strains before addition of chemiluminescence substrate. Subsequently, 100 μL of substrate-containing ASW was added to each well, the plate was agitated for 20 s and luminescence was measured at 470 ± 40 nm.

On-plate fluorescence screening

On-plate tdTomato fluorescence of transformant colonies was visualized using PathoScreen (PhenoVation Life Sciences, The Netherlands) fluorescence imager. The device captured RFP and chlorophyll fluorescence and color images with a 6-megapixel camera. The RFP channel and the dedicated Data Analysis (Version 5.4.7 beta-64b) software (PhenoVation Life Sciences, The Netherlands) was used to select colonies with or without tdTomato fluorescence.

Purification of V_HH antibody

V_HH antibody was purified from algal extracts as follows: 6×10^9 exponentially growing cells of an EC-VHH transformant strain were harvested by centrifugation (3,000 × *g*, 10 min, 4°C), resuspended in 500 μL of ice-cold TBS-T, supplemented with 2 μL mL^{-1} of protease inhibitor (Merck #P9599, TBS-T-PI) and bead beat 3 × at 2,500 rpm for 20 s with a 120 s pause between cycles, using Lysing Matrix D (#116913500, MP Biomedicals). Debris was pelleted by centrifugation (15,000 × *g*, 10 min, 4°C) and the supernatant was transferred to fresh tubes and kept on ice. Protein concentration was quantified using modified Lowry procedure as described above. His-tagged V_HH was purified from the extract by Ni-NTA spin column purification (Thermo Fisher Scientific #88224) according to manufacturer instructions with the following modifications. Five wash steps were carried out using 700 μL of 25 mM imidazole in PBS and each time the flowthrough was collected in separate tubes. Then, three additional wash steps were conducted with 700 μL of 40, 50, and 70 mM imidazole in PBS and also collected separately. Last, the elution was done in four separate steps with 200, 200, 500, and 200 μL of 250 mM imidazole in PBS used in steps 1, 2, 3, and 4, respectively. The raw extract, and all flowthrough, wash and elution fractions were then analyzed by SDS-PAGE. Briefly, samples were mixed with 5 × Laemmli reagent, heated to 95°C for 5 min and 30 μL were separated on a 4% to 15% TGX protein gel with TGS running buffer for 35 min at 200 V. The gel was stained with Coomassie Brilliant Blue G-250 (Bio-Rad #1610787) overnight and destained in dH_2O with multiple changes over 1 d. Purification and SDS-PAGE analysis were done for three biological replicates. Elution fractions that showed no visible bands other than the V_HH band after destaining were pooled, and the elution buffer was replaced with TBS-T-PI using Amicon Ultra Centrifugal Filter Units (Merck #UFC500308) with three washing steps of 15-min centrifugation at 14,000 × *g*. All centrifugations were carried out at 4°C. Then, V_HH solutions of triplicates were pooled, adjusted to 300 μL total volume and protein concentration was measured by Lowry method.

ELISA

Soluble extracts were prepared for three EC-VHH transformant strains as described above, using $\sim6\times10^8$ cells per sample. After protein quantification, extracts were diluted to 200, 100, 50, 20, 10, and 1 μg protein mL^{-1} in TBS-T-PI. Purified V_HH was diluted to 4.00, 3.00, 2.50, 2.00, 1.50, 1.00, 0.75, 0.50, 0.25, 0.10, and 0.05 μg mL^{-1} in TBS-T-PI and used as a calibration standard in quantitative ELISA. For this, microtiter plates were coated with GFP and blocked with skim milk as follows. Recombinant purified GFP (Thermo Fisher Scientific #A42613) was resuspended in TBS (pH 8.0) at 2.5 μg mL^{-1}, and 50 μL was transferred to each well of a MediSorp 96-well microtiter plate (Thermo Fisher Scientific #467320). Control wells were treated with TBS only. The plate was sealed and incubated at RT in the dark for 2 h. The solution was removed and the plate was blocked by addition of 170 μL of 1% skim milk powder (Bio-Rad #170-6404) in TBS per well. After 1 h incubation at RT in the dark, blocking solution was removed and wells were washed several times with TBS-T. Then, 50 μL of all V_HH calibration standard dilutions and of all microalgal extract dilutions were transferred to separate wells in duplicate and incubated at RT for 1 h in the dark. Unbound molecules were washed off and secondary anti-camelid V_HH antibody conjugated to HRP was added (Abcam #ab112786, 10,000 × diluted in 0.5% skim milk powder in TBS-T). The plate was sealed and incubated at RT in the dark for 1 h. Solutions were removed and wells were washed several times with TBS-T. Then, 90 μL of colorigenic TMB substrate (Thermo Fisher Scientific #10301494) was added to each well and incubated at RT in the dark for exactly 10 min. The reaction was stopped by addition of 90 μL of 0.16 M H_2SO_4 and optical density was measured at 450 nm. A dose-response curve was created for the calibration samples and fitted with a 3-parameter Michaelis-Menten regression model using the drc package (Ritz et al., 2015) of R statistical computing software (R Team, 2018). This model was used to calculate the V_HH concentration in microalgal extracts at 1 μg mL^{-1}.

Cassette PCR and genotyping procedure for TC #17

Tracing of the insertion site in TC transformant strain #17 was carried out using a MmeI-based version of cassette PCR as reported before (Südfeld et al., 2021), with the exceptions that the double-stranded DNA adapters were assembled from oligos oCS089 and oCS090, and that different PCR primers were used for nested PCRs (listed in Supplemental Table 1). PCR products were purified and sequenced in order to reveal the nucleotide sequence surrounding the TC by BLAST analysis against the *N. oceanica* genome. This analysis implicated four possible genomic loci, corresponding to 25S rRNA genes. All four potential loci were tested for presence of the cassette by genotyping PCR with primer pairs that were unique for each locus (Supplemental Figure 2).

Analysis of conserved regions of *N. oceanica* 25S rRNA

The 25/28S rRNA sequences of several eukaryotes (*Gallus gallus*, accession number KT445934; *Arabidopsis thaliana*, JSAD01000007; *Brassica napus*, CCCW010027254; *Saprolegnia parasitica*, ADCG02001922; *Nannochloropsis gaditana*, AFGN01000272; *S. cerevisiae*, LIUU01000096; *Galdieria sulphuraria*, ADNM02000184; *Schizosaccharomyces pombe*, AY048171; *Phytophthora nicotianae*, LNFO01001540; *Drosophila persimilis*, AAIZ01013212) from different clades were retrieved from the SILVA database (Quast et al., 2013) or the NCBI (*Homo sapiens*, NT_167214.1) and aligned together with the 25S rRNA sequence from *N. oceanica* using ClustalΩ with standard parameters. Sequence regions from organisms other than *N. oceanica* without a homologous region in the *N. oceanica* 25S rRNA were removed from the multiple sequence alignment for visualization purposes. Local similarities of a subset of sequences compared with the *N. oceanica* 25S rRNA were then calculated as percent of identical nucleotides in a sliding 15-nucleotide window. The similarities were converted to a color-gradient and plotted as a heatmap using R statistical software. The local and global similarities mentioned in the caption of Figure 1 were calculated as mean ± SD of identities of all abovementioned organisms compared with *N. oceanica*.

Statistical data treatment

All data processing and statistical analysis was done using R statistical software (R Team, 2018). No statistical predetermination of sample size was applied. Sample sizes were chosen based on experience. Significant main effects were evaluated by two-way ANOVA. Significant ANOVAs were followed by Tukey's honestly significant difference (HSD)

test to compare means of multiple groups. Differences were considered significant in case of $p < 0.05$. For flow cytometry analysis, the median of the fluorescence distribution in a fluorescence channel of interest was considered as average cellular fluorescence for a respective sample. In figures representing flow cytometry analyses, individual points correspond to the median fluorescence value of at least 50,000 cells. For statistical analysis, the mean of the median wild-type fluorescence was substracted from each median. Then, the means of wild-type-corrected median single cell fluorescences were used to compare gene expression between different groups.

Oligonucleotides and gene fragments

All oligonucleotides and gene fragments used in this study are listed in Supplemental Tables 1 and 2, respectively.

Data availability

All data relevant for interpretation of this study are presented in the article and supplementary material. Any further information is available from the corresponding author upon reasonable request.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at *Molecular Plant Online*.

FUNDING

This work was part of the Dutch Research Council (NWO) Building Blocks of Life programme (grant number 737.016.007).

AUTHOR CONTRIBUTIONS

C.S. conceptualized the study, developed the methodology, performed investigation and formal analysis, wrote the original draft, and reviewed and edited the manuscript. A.P.R. conceptualized the study, developed the methodology, and performed investigation. S.A.M.D. performed study investigation and developed the methodology. R.H.W. performed project administration and supervision, reviewed and edited the manuscript, and acquired funding. M.J.B. performed project administration and supervision, and reviewed and edited the manuscript. S.D. conceptualized the study, performed project administration and supervision, and reviewed and edited the manuscript.

ACKNOWLEDGMENTS

We are grateful to J. van der Oost for insightful and helpful scientific discussions. We thank M. Naduthodi for providing purified Cas proteins, and T. van der Lee for expert assistance with on-plate fluorescence screenings. We further thank W. Wu and R. Joosten for excellent technical assistance with flow cytometry, C. Llavata Peris for expert assistance with RTq-PCR, and L. Maas for expert assistance with fluorescence microscopy. C.S., R.H.W., M.J.B., and S.D. are employees of Wageningen University. Wageningen University has filed a patent related to the contents of this study, with C.S., R.H.W., M.J.B., and S.D. listed as inventors. The remaining authors declare no conflict of interest.

Received: June 30, 2021
Revised: September 28, 2021
Accepted: November 8, 2021
Published: November 11, 2021

REFERENCES

Adissin, T.O., Manabu, I., Shunsuke, K., Saichiro, Y., Moss, A.S., and Dossou, S. (2020). Effects of dietary Nannochloropsis sp. powder and lipids on the growth performance and fatty acid composition of larval and postlarval kuruma shrimp, *Marsupenaeus japonicus*. Aquac. Nutr. **26**:186–200.

Ajjawi, I., Verruto, J., Aqui, M., Soriaga, L.B., Coppersmith, J., Kwok, K., Peach, L., Orchard, E., Kalb, R., Xu, W., et al. (2017). Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a single transcriptional regulator. Nat. Biotechnol. **35**:647–652.

Ali, I.K., Lancaster, L., Feinberg, J., Joseph, S., and Noller, H.F. (2006). Deletion of a conserved, central ribosomal intersubunit RNA bridge. Mol. Cell **23**:865–874.

Angulo, J., Ulryck, N., Deforges, J., Chamond, N., Lopez-Lastra, M., Masquida, B., and Sargueil, B. (2016). LOOP IIId of the HCV IRES is essential for the structural rearrangement of the 40S-HCV IRES complex. Nucleic Acids Res. **44**:1309–1325.

Bradrick, S.S., Dobrikova, E.Y., Kaiser, C., Shveygert, M., and Gromeier, M. (2007). Poly(A)-binding protein is differentially required for translation mediated by viral internal ribosome entry sites. RNA **13**:1582–1593.

Cerutti, H., Johnson, A.M., Gillham, N.W., and Boynton, J.E. (1997). Epigenetic silencing of a foreign gene in nuclear transformants of Chlamydomonas. Plant Cell **9**:925–945.

Chappell, S.A., Edelman, G.M., and Mauro, V.P. (2000). A 9-nt segment of a cellular mRNA can function as an internal ribosome entry site (IRES) and when present in linked multiple copies greatly enhances IRES activity. Proc. Natl. Acad. Sci. U S A **97**:1536–1541.

Chisti, Y. (2007). Biodiesel from microalgae. Biotechnol. Adv. **25**:294–306.

Clarke, J.L., Waheed, M.T., Lössl, A.G., Martinussen, I., and Daniell, H. (2013). How can plant genetic engineering contribute to cost-effective fish vaccine development for promoting sustainable aquaculture? Plant Mol. Biol. **83**:33–40.

Dorokhov, Y.L., Skulachev, M.V., Ivanov, P.A., Zvereva, S.D., Tjulkina, L.G., Merits, A., Gleba, Y.Y., Hohn, T., and Atabekov, J.G. (2002). Polypurine (A)-rich sequences promote cross-kingdom conservation of internal ribosome entry. Proc. Natl. Acad. Sci. U S A **99**:5301–5306.

Falco, A., Martinez-Lopez, A., Coll, J.P., and Estepa, A. (2012). The potential for antimicrobial peptides to improve fish health in aquaculture. In Infectious Disease in Aquaculture (Elsevier), pp. 457–479.

García-Nuñez, S., Gismondi, M.I., König, G., Berinstein, A., Taboga, O., Rieder, E., Martínez-Salas, E., and Carrillo, E. (2014). Enhanced IRES activity by the 3′UTR element determines the virulence of FMDV isolates. Virology **448**:303–313.

Ghoshal, K., Majumder, S., Datta, J., Motiwala, T., Bai, S., Sharma, S.M., Frankel, W., and Jacob, S.T. (2004). Role of human ribosomal RNA (rRNA) promoter methylation and of methyl-CpG-binding protein MBD2 in the suppression of rRNA gene expression. J. Biol. Chem. **279**:6783–6793.

Gong, Y., Kang, N.K., Kim, Y.U., Wang, Z., Wei, L., Xin, Y., Shen, C., Wang, Q., You, W., Lim, J., et al. (2020). The NanDeSyn database for Nannochloropsis systems and synthetic biology. Plant J. **104**:1736–1745. https://doi.org/10.1111/tpj.15025.

Grahl, I. de, Rout, S.S., Maple-Grødem, J., and Reumann, S. (2020). Development of a constitutive and an auto-inducible high-yield expression system for recombinant protein production in the microalga *Nannochloropsis oceanica*. Appl. Microbiol. Biotechnol. **104**:8747–8760.

Herman, E.M., and Schmidt, M.A. (2016). The potential for engineering enhanced functional-feed soybeans for sustainable aquaculture feed. Front. Plant Sci. **7**:440.

Hirose, Y., and Manley, J.L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature **395**:93–96.

Hodgman, C.E., and Jewett, M.C. (2014). Characterizing IGR IRES-mediated translation initiation for use in yeast cell-free protein synthesis. New Biotechnol. **31**:499–505.

Kang, N.K., Lee, B., Shin, S.E., Jeon, S., Park, M.S., and Yang, J.W. (2015). Use of conditioned medium for efficient transformation and cost-effective cultivation of *Nannochloropsis salina*. Bioresour. Technol. **181**:231–237.

Kilian, O., Benemann, C.S.E., Niyogi, K.K., and Vick, B. (2011). High-efficiency homologous recombination in the oil-producing alga Nannochloropsis sp. Proc. Natl. Acad. Sci. U S A **108**:21265–21269.

Komarova, T.V., Schwartz, A.M., Makarov, A.A., and Dorokhov, Y.L. (2012). A new viral vector exploiting RNA polymerase I-mediated transcription. Biochemistry **77**:532–538.

Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B., and Valen, E. (2016). CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Res. **44**:W272–W276.

Lamminen, M., Halmemies-Beauchet-Filleau, A., Kokkonen, T., Jaakkola, S., and Vanhatalo, A. (2019). Different microalgae species as a substitutive protein feed for soya bean meal in grass silage based dairy cow diets. Anim. Feed Sci. Technol. **247**:112–126.

Levasseur, W., Perré, P., and Pozzobon, V. (2020). A review of high value-added molecules production by microalgae in light of the classification. Biotechnol. Adv. **41**:107545.

Li, J., Han, D., Wang, D., Ning, K., Jia, J., Wei, L., Jing, X., Huang, S., Chen, J., Li, Y., et al. (2014). Choreography of transcriptomes and lipidomes of Nannochloropsis reveals the mechanisms of oil synthesis in microalgae. Plant Cell **26**:1645–1665.

Li, D.W., Balamurugan, S., Yang, Y.F., Zheng, J.W., Huang, D., Zou, L.G., Yang, W.D., Liu, J.S., Guan, Y., and Li, H.Y. (2019). Transcriptional regulation of microalgae for concurrent lipid overproduction and secretion. Sci. Adv. **5**:eaau3795.

Liang, X.H., Liu, Q., and Fournier, M.J. (2007). rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. Mol. Cell **28**:965–977.

Lodish, H., Berk, A., and Zipursky, S. (2000). Processing of rRNA and tRNA. In Molecular Cell Biology (W. H. Freeman), p. 1184.

Malygin, A.A., Kossinova, O.A., Shatsky, I.N., and Karpova, G.G. (2013). HCV IRES interacts with the 18S rRNA to activate the 40S ribosome for subsequent steps of translation initiation. Nucleic Acids Res. **41**:8706–8714.

Martinez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J., and Embarek, A.M. (2018). Insights into structural and mechanistic features of viral IRES elements. Front. Microbiol. **8**:2629.

Matoulkova, E., Michalova, E., Vojtesek, B., and Hrstka, R. (2012). The role of the 3′ untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biol. **9**:563–576.

Mitchell, S.A. (2005). Identification of a motif that mediates polypyrimidine tract-binding protein-dependent internal ribosome entry. Genes Dev. **19**:1556–1571.

Naduthodi, M.I.S., Mohanraju, P., Südfeld, C., D'Adamo, S., Barbosa, M.J., and Van Der Oost, J. (2019). CRISPR-Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae *Nannochloropsis oceanica* IMET1. Biotechnol. Biofuels **12**:66.

Nguyen, H.Q., and Bosco, G. (2015). Gene positioning effects on expression in eukaryotes. Annu. Rev. Genet. **49**:627–646.

Oem, J.K., Xiang, Z., Zhou, Y., Babiuk, L.A., and Liu, Q. (2007). Utilization of RNA polymerase I promoter and terminator sequences to develop a DNA transfection system for the study of hepatitis C virus internal ribosomal entry site-dependent translation. J. Clin. Virol. **40**:55–59.

Palmer, T.D., Miller, A.D., Reeder, R.H., and Mcstay, B. (1993). Efficient expression of a protein coding gene under the control of an RNA polymerase I promoter. Nucleic Acids Res. **21**:3451–3457.

Poliner, E., Farré, E.M., and Benning, C. (2018a). Advanced genetic tools enable synthetic biology in the oleaginous microalgae Nannochloropsis sp. Plant Cell Rep. **37**:1383–1399. https://doi.org/10.1007/s00299-018-2270-0.

Poliner, E., Takeuchi, T., Du, Z.Y., Benning, C., and Farré, E.M. (2018b). Nontransgenic marker-free gene disruption by an episomal CRISPR system in the oleaginous microalga, *Nannochloropsis oceanica* CCMP1779. ACS Synth. Biol. **7**:962–968.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. **41**:D590–D596.

Rasala, B.A., Chao, S.S., Pier, M., Barrera, D.J., and Mayfield, S.P. (2014). Enhanced genetic tools for engineering multigene traits into green algae. PLoS One **9**:e94028.

Ritz, C., Baty, F., Streibig, J.C., and Gerhard, D. (2015). Dose-response analysis using R. PLoS One **10**:e0146021. https://doi.org/10.1371/journal.pone.0146021.

RC Team. (2018). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing), p. 63.

Ruiz, J., Olivieri, G., De Vree, J., et al. (2016). Towards industrial products from microalgae. Energy Environ. Sci. **9**:3036–3043.

Sarker, P.K., Kapuscinski, A.R., Bae, A.Y., Donaldson, E., Sitek, A.J., Fitzgerald, D.S., and Edelson, O.F. (2018). Towards sustainable aquafeeds: evaluating substitution of fishmeal with lipid-extracted microalgal co-product (*Nannochloropsis oculata*) in diets of juvenile Nile tilapia (*Oreochromis niloticus*). PLoS One **13**:e0201315.

Sarker, P.K., Kapuscinski, A.R., McKuin, B., Fitzgerald, D.S., Nash, H.M., and Greenwood, C. (2020). Microalgae-blend tilapia feed eliminates fishmeal and fish oil, improves growth, and is cost viable. Sci. Rep. **10**:1–14.

Serrano, P., Pulido Rodriguez, M., Sáiz, M., and Martínez-Salas, E. (2006). The 3′ end of the foot-and-mouth disease virus genome establishes two distinct long-range RNA–RNA interactions with the 5′ and region. J. Gen. Virol. **87**:3013–3022.

Sims, S., Michaelsen, K., Burkhard, S., and Fraefel, C. (2021). In vitro comparison of the internal ribosomal entry site activity from rodent hepacivirus and pegivirus and construction of pseudoparticles. Adv. Virol. https://doi.org/10.1155/2021/5569844.

Sloan, K.E., Bohnsack, M.T., Schneider, C., and Watkins, N.J. (2014). The roles of SSU processome components and surveillance factors in the initial processing of human ribosomal RNA. RNA **20**:540–550.

Song, Y., Friebe, P., Tzima, E., Junemann, C., Bartenschlager, R., and Niepmann, M. (2006). The hepatitis C virus RNA 3'-untranslated region strongly enhances translation directed by the internal ribosome entry site. J. Virol. **80**:11579–11588.

Spriggs, K.A., Mitchell, S.A., and Willis, A.E. (2005). Investigation of interactions of polypyrimidine tract-binding protein with artificial internal ribosome entry segments. Biochem. Soc. Trans. **33**:1483.

Sproles, A.E., Fields, F.J., Smalley, T.N., Le, C.H., Badary, A., and Mayfield, S.P. (2021). Recent advancements in the genetic engineering of microalgae. Algal Res. **53**:102158.

Südfeld, C., Hubáček, M., D'Adamo, S., Wijffels, R.H., and Barbosa, M.J. (2020). Optimization of high-throughput lipid screening of the microalga *Nannochloropsis oceanica* using BODIPY 505/515. Algal Res. **53**:102138.

Südfeld, C., Hubáček, M., Figueiredo, D., Naduthodi, M.I.S., Oost, J. van der, Wijffels, R.H., Barbosa, M.J., and D'Adamo, S. (2021). High-throughput insertional mutagenesis reveals novel targets for enhancing lipid accumulation in *Nannochloropsis oceanica*. Metab. Eng. **66**:239–258.

Thompson, S.R. (2012). Tricks an IRES uses to enslave ribosomes **20**:558–566.

Truniger, V., Miras, M., and Aranda, M.A. (2017). Structural and functional diversity of plant virus 3′-cap-independent translation enhancers (3′-CITEs). Front. Plant Sci. **8**:2047.

Vieler, A., Wu, G., Tsai, C.H., Bullard, B., Cornish, A.J., Harvey, C., Reca, I.B., Thornburg, C., Achawanantakun, R., Buehl, C.J., et al. (2012). Genome, functional gene annotation, and nuclear transformation of the *Heterokont oleaginous* alga *Nannochloropsis oceanica* CCMP1779. PLoS Genet. **8**:e1003064.

Wang, Q., Lu, Y., Xin, Y., Wei, L., Huang, S., and Xu, J. (2016). Genome editing of model oleaginous microalgae Nannochloropsis spp. by CRISPR/Cas9. Plant J. **88**:1071–1081.

Wang, Q., Gong, Y., He, Y., Xin, Y., Lv, N., Du, X., Li, Y., Jeong, B.R., and Xu, J. (2020). Genome engineering of Nannochloropsis with large deletions for constructing microalgal minigenomes. bioRxiv https://doi.org/10.1101/2020.10.08.332478.

Wei, L., Xin, Y., Wang, Q., Yang, J., Hu, H., and Xu, J. (2017). RNAi-based targeted gene knockdown in the model oleaginous microalgae *Nannochloropsis oceanica*. Plant J. **89**:1236–1250.

Wen, L., Liang, D.S., Wu, L.Q., Cai, F., Pan, Q., Long, Z.G., Dai, H.P., Xia, K., and Xia, J.H. (2008). Efficient protein expression from the endogenous RNA polymerase I promoter using a human ribosomal DNA targeting vector. Biochem. Biophysical Res. Commun. **367**:846–851.

Wijffels, R.H., and Barbosa, M.J. (2010). An outlook on microalgal biofuels. Science **329**:796–799.

Xie, Q., Wang, Y., Lin, J., Qin, Y., Wang, Y., and Bu, W. (2012). Potential key bases of ribosomal RNA to Kingdom-specific spectra of antibiotic susceptibility and the possible archaeal origin of eukaryotes. PLoS One **7**:e29468.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. Science **292**:883–896.