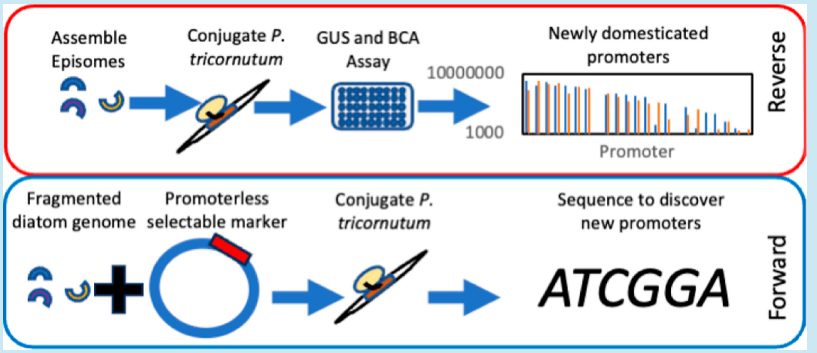
ACS Synthetic Biology

pubs.acs.org/synthbio
Research Article

Validating a Promoter Library for Application in Plasmid-Based Diatom Genetic Engineering

Erin A. Garza, Vincent A. Bielinski, Josh L. Espinoza, Kona Orlandi, Josefa Rivera Alfaro, Tayah M. Bolt, Karen Beeri, Philip D. Weyman, and Christopher L. Dupont*

Cite This: https://doi.org/10.1021/acssynbio.3c00163
Read Online



ABSTRACT: While diatoms are promising synthetic biology platforms, there currently exists a limited number of validated genetic regulatory parts available for genetic engineering. The standard method for diatom transformation, nonspecific introduction of DNA into chromosomes via biolistic particle bombardment, is low throughput and suffers from clonal variability and epigenetic effects. Recent developments in diatom engineering have demonstrated that autonomously replicating episomal plasmids serve as stable expression platforms for diverse gene expression technologies. These plasmids are delivered via bacterial conjugation and, when combined with modular DNA assembly technologies, provide a flexibility and speed not possible with biolistic-mediated strain generation. In order to expand the current toolbox for plasmid-based engineering in the diatom *Phaeodactylum tricornutum*, a conjugation-based forward genetics screen for promoter discovery was developed, and application to a diatom genomic DNA library defined 252 *P. tricornutum* promoter elements. From this library, 40 promoter/terminator pairs were delivered via conjugation on episomal plasmids, characterized in vivo, and ranked across 4 orders of magnitude difference in reporter gene expression levels.

KEYWORDS: *diatom, genetic engineering, promoter characterization, parts registry, episomal gene expression, forward genetics*

INTRODUCTION

At least half of Earth's primary productivity occurs in the ocean through microalgae that is too small to see with the naked eye. Diatoms (*Bacillariophyceae*) account for 40% of marine productivity through the combined effect of hundreds of thousands of different species[1] and are also a focus in the fast-growing field of algal biotechnology.[2−4] However, a majority of the previous studies on genetic engineering of diatoms and algae were limited to biolistic-mediated random chromosomal integration of expression cassettes to generate transgenic lines.[5−7] While effective, strains generated by this method can demonstrate off-target effects and varying transcriptional activity, increasing the complexity of screening and analysis.[8,9]

Recent progress in phytoplankton synthetic biology has led to the development of artificial chromosomes (hereafter referred to as "episomes") that replicate autonomously in both the pennate diatom *Phaeodactylum tricornutum* and the centric diatom *Thalassiosira pseudonana*.[10] Episomes are delivered via bacterial conjugation and can be stably maintained in vivo as plasmids up to 100 kb in size due to the presence of a centromere like sequence on the vector.[10,11] Recent studies have demonstrated that these plasmids are amenable for regulation of transgene expression via both constitutive and inducible promoter systems and are also excellent expression platforms to deliver technologies like CRISPR and synthetic gene expression pathways for compounds such as vanillin and terpenoids[12−14] by eliminating or reducing many previous issues observed with biolistics-based strain development approaches. Additionally, an episomal delivery system allows for regulatory elements to be

Received: March 21, 2023

designed as individual parts for compatibility with newer high-throughput DNA assembly methods and generation of diatom-specific gene expression part libraries.[15,16]

Assembling and validating a gene expression library requires the collection and curation of a well-understood, annotated, and calibrated set of gene expression parts (toolkits). Cloning methods in *Saccharomyces cerevisiae* and *Escherichia coli* have flourished due in part to large repositories of validated genetic parts used for cloning like Biobricks.[17] The list of functional promoter elements described for *P. tricornutum* has been expanding over the years since the publication of the genome sequence.[18−21] Siaut and co-workers[22] combined the legacy P_{h4} and P_{fcpB} promoters into the Gateway vectors, providing low- and high-expression promoters for deployment in diatoms. The nitrate reductase (Phatr3_J54983) 5′-UTR element (P_{nr}), originally characterized in *Cylindrotheca fusiformis*, has been the workhorse for inducible gene expression in *P. tricornutum*, and other nitrogen-responsive promoters have since been described.[23−26] However, a recent study described a detailed analysis of a small set of new diatom promoters delivered via episomes and characterized activity under different stages of the growth cycle using fluorescent reporter genes.[21] Phosphate limitation has also been utilized to drive gene expression from a *phoA*-type alkaline phosphatase 5′-UTR.[27] Unfortunately, promoter induction via starvation by essential nutrients like nitrate and phosphate may have detrimental effects on cell health and result in global transcriptional reprogramming in diatoms.[28]

Gene expression toolkits have been reported using diatom viral promoters[29] and studies on the effects of iron deprivation, light cycle, and CO_2 availability on gene expression from *P. tricornutum* promoters have identified DNA binding motifs for a small subset of bZIP transcription factors.[30−32] Small-molecule-induced gene expression tools for diatoms are rare but are expanding with the recent report of reversible gene expression via promoters induced by beta-estradiol and digoxin.[14] Overall, the results of these studies have been the creation of a small, but useful, set of transcriptional elements with limited control over gene expression.

A recent study demonstrated that diatom episomal plasmids, at least in this first *CEN-ARS-HIS* centromeric iteration, show long-term stability and functionality issues when expressing toxic gene products.[33] These results suggest that chromosomal integration[34,35] or the creation of a new and more stable version of the diatom episomal plasmid is required when engineering potentially toxic genes or large synthetic pathways that can stress the cells. However, in the context of discovery and validation of genetic regulatory parts for use in cloning toolkits, the application of these first-generation plasmids for the production of transgenic diatom strains analyzed for expression of nontoxic reporter enzymes or subjected to DNA sequencing may remain valid.[36]

Although *P. tricornutum* has a high-quality genome assembly and multiple RNaseq data sets are available,[21,28,37−41] in vivo characterization of transcriptional activity for a large set of promoter elements is lacking. In this study, a workflow was developed for in vivo screening of a *P. tricornutum* gDNA fragment library to probe for transcriptional activity by applying the episome to identify new promoter elements. We also analyzed and ranked, based on transcriptional output of a marker enzyme, a set of 40 promoter/terminator pairs cloned directly from the *P. tricornutum* genome in order to expand the current diatom genetic engineering toolkit. These pairs were cloned flanking the β-glucuronidase (*GUS*) reporter gene, which provides a clean signal-to-noise readout in *P. tricornutum*.[42,43] Presumably, the control of *GUS* gene transcription was due solely to the presence of these promoter/terminator pairs. Plasmids were constructed in which each promoter was paired with either native or non-native terminators, with "native" describing the DNA sequence found immediately downstream of the stop codon of a predicted protein in the diatom genome. These pairings were carried out in order to determine if the presence of different terminator sequences affect *GUS* expression levels when combined with specific promoters on episomal plasmids. The characterized promoters and terminators, which modulate gene expression over 4 orders of magnitude, were made compatible with the uLoop library, a DNA assembly system method that utilizes an open-access library of validated parts for efficient, reproducible DNA assembly.[15,16]

RESULTS

GUS Activity Analysis from Biolistics- and Conjugation-Derived Diatom Clones. Previous studies reporting clone-to-clone variability of transgene expression cassettes upon integration into the *P. tricornutum* genome when introduced via biolistics are available.[6,9,44] We chose to test (1) if episome-driven expression of the marker gene *GUS* and in vivo biochemical activity were similar to gene expression from a genomically integrated strain, (2) if the levels of GUS activity between biological replicates from the same transformation or conjugation differed when assayed, and (3) the possibility of setting upper and lower limits on *GUS* expression from episomal expression cassettes.

Figure 1 shows the results of blind screening transformant *P. tricornutum* cell lines generated via biolistics. In this experiment, a diatom expression vector (PB-fcpB) containing a GUS expression cassette (P_{fcpB}-*GUS*-T_{fcpA}), based on previous studies,[42] was introduced into *P. tricornutum* cells using standard methodologies.[22] For this experiment, clones were randomly isolated from selection plates, cultured with antibiotics on 1/2-L1 liquid medium, and clarified cell lysates probed for levels of *GUS* activity (DOI: 10.17504/protocols.io.bbexijfn). Consistent with our past experience and previous reports,[8,45] 5 of the 15 colonies recovered from particle bombardment experiments completely lacked *GUS* activity, while the remaining 10 isolates exhibited varying levels of GUS activity between them, highlighting the inefficiency of the particle bombardment process due to random and possibly incomplete chromosomal integration of the expression cassette.

In order to assess promoter activity via reporter enzyme assays, receiver plasmids were designed and built using the Gateway cloning technology[46] with the intention of analyzing the expression level of previously reported diatom promoters in our system. Exconjugant strains of *P. tricornutum* were also generated via bacterial conjugation using pDEST receiver plasmids and one of four different promoters driving *GUS* expression (P_{fcpB}, P_{nr}, P_{p49}, and P_{h4}). These plasmids were labeled pFcpB-DEST, pH4-DEST, p49202-DEST, and pNR-DEST, Supporting Information Table S1. Briefly, these "destination" vectors contained any of the four legacy diatom promoters and the fucoxanthin-chlorophyll a−c binding protein A (Phatr3_J18049) terminator (T_{fcpA}), flanking att recombination sites, and the *ccdB* selection marker. An entry vector containing the *GUS* coding sequence was generated to

https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX

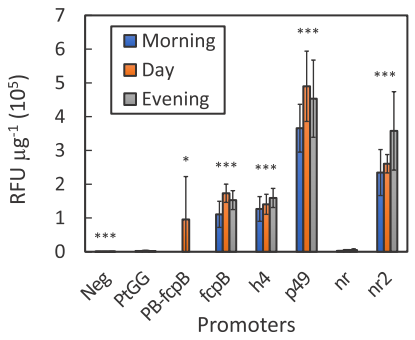


Figure 1. Validation of GUS marker gene expression as a reporter for promoter activity when driven from a diatom episome. Neg is the negative control strain containing a genomically integrated selection marker gene but lacking the GUS gene. PtGG represents the hairpin forming DNA sequence-GUS-T_{fcpA} cassette delivered via pDEST and not expected to drive protein expression, whereas PB-fcpB represents a strain containing a P_{fcpB}-*GUS*-T_{fcpA} cassette integrated into the genome via particle bombardment with the error bars representing 10 independently picked transformants. The other lines were generated by conjugation and refer to the promoter upstream of the GUS reporter as follows: p_{49} = hypothetical protein (Phatr3_J49202), nr = nitrate reductase (Phatr3_J54983), h4 = histone H4 (Phatr3_J34971), and fcpB = fucoxanthin-chlorophyll a−c binding protein B (Phatr3_J25172). P_{nr} represents pairing of the *nr* 5′UTR with the *fcpA* terminator, while P_{nr2} utilized the native 3′UTR. Error bars for the conjugative vectors represent the standard deviation from assaying three (3) exconjugants in parallel, while the error bar for PB-fcpB is the result of assaying 10 biological replicates. Asterisks indicate a *t*-test *P*-value ≤0.05 (*) or ≤0.001 (***) when comparing the expression of each cell line to PtGG. For the conjugation strains, the difference in the expression of P_{fcpB} vs P_{h4}, P_{p49}, and P_{nr2} was analyzed, and in agreement with transcriptomic data, the relative strength of this promoter set was ranked as $P_{p49} > P_{nr2} > P_{fcpB} = P_{h4}$. For the conjugation lines, samples of liquid cultures were taken at three different time points during the day cycle; within a promoter, the expression levels did not fluctuate in a statistically significant manner over the course of the day for this set of plasmids (*t*-test, $p > 0.05$). Cells were cultured in the presence of 8.8 mM NO_3. Raw RFUs were normalized to total protein content in each assayed lysate.

produce episomes with differing legacy promoters driving *GUS* expression. Three of the promoter elements (P_{fcpB}, P_{h4}, P_{nr}, or Phatr3_J25172, Phatr3_J54983, and Phatr3_J34971) were previously utilized for in vivo expression studies[22,23,42,47] and have been the promoters of choice for gene expression in *P. tricornutum*. P_{p49} is the 500 bp region upstream of Phatr3_49202, a gene consistently observed to be one of the most highly expressed genes in transcriptomic data sets from *P. tricornutum*. A DNA sequence, named "PtGG", was employed as a negative control for transcriptional read-through with no expected promoter activity in vivo. This spacer consists of inverted DNA repeats flanked by multiple cloning sites and should not drive gene expression in *P. tricornutum* (Supporting Information Figure S1 and Table S2). All plasmids utilized the *fcpA* terminator[22] at the 3′ end of the reporter gene; therefore, the only difference between all of the lines tested was the composition of the 5′-UTR sequences upstream of *GUS* and the method of plasmid delivery (particle bombardment vs conjugation). The *fcpA* terminator was selected since it is considered a "legacy" terminator as it has been the main terminator used in expression cassettes by numerous laboratories for decades.[6,9,10,19,42] It should be noted that the reported sequence of T_{fcpA}, while efficient at transcriptional termination, also contains a portion of the 3′ end of Phatr3_J18049 CDS. The nucleotide sequences for these described DNA elements can be found in Supporting Information Table S3.

Unlike with particle bombardment, we consistently observed *GUS* activity for each picked exconjugant colony (Figure 1). The promoters in the pDEST vectors exhibited a ranked order of *GUS* activity (i.e., $P_{fcpB} = P_{h4} < P_{nr2} < P_{p49}$, *t*-test, see legend), while we did not observe *GUS* activity within cell lines generated using the PtGG-1 "null" promoter (Figure 1). P_{h4} has a similar expression level as P_{fcpB} (Supporting Information Table S4), while P_{p49} drives nearly 5× higher *GUS* activity. This data demonstrates that the diatom episome is capable of delivering reproducible gene expression results across different clonal lines during different periods of the light cycle. We measured little to no *GUS* activity in the P_{nr} exconjugants that were transformed with an episome containing a P_{nr}-*GUS*-T_{fcpA} cassette (average of 200 RFU, close to negative control values) in the presence of nitrate (Figure 1). However, when we swapped out T_{fcpA} with T_{nr} (the native 3′-UTR of the *P. tricornutum* nitrate reductase gene), exconjugants transformed with the P_{nr}-*GUS*-T_{nr} cassette yielded the second highest RFU count in our assay (Figure 1). This data suggests that in the pDEST-derived backbones, the native T_{nr} was required for proper transcription/translation of *GUS* when under control of P_{nr}.

It is important to note that the generation of the pDEST-based expression vectors utilized the Gateway technology,[46] which relies on recombination at attachment (att) sites flanking the *GUS* coding sequence. These att sites, while useful, are short stretches of DNA sequences that lie in between the CDS and regulatory elements in the final plasmid constructs and are not present in the PB-fcpB plasmid. It is possible that the presence of these att sites influences gene regulation and may contribute to some of the differences in *GUS* activity observed between the strains generated with different transformation methods. To further test whether expression is dependent on specific terminator sequences or if the lack of expression seen in the P_{nr}-*GUS*-T_{fcpA} construct was due to the way the cassette was built, a new set of plasmids were constructed (pEG vectors) using Gibson assembly[48] (Supporting Information Figure S2). This allowed for the elimination of excess att recombination sequences and the addition of a consensus Kozak sequence (5′-GGGGCCACC-3′) immediately before the 5′ end of the *GUS* gene in order to aid in translation initiation. The result was that the pEG vector containing the P_{nr}-*GUS*-T_{fcpA} construct showed expression levels similar to the native pEG T_{nr} construct. This suggests that the native terminator is not required for the expression of the nitrate reductase gene when the space between the gene and terminator is minimized. In this case, removal of the att sites likely altered the spacing between the promoter and start of the gene and the end of the gene and terminator, allowing for higher gene expression. This also shows that while the pDEST-derived episomes are functional within diatoms, they do not explore the potential role of 3′-UTRs in regulating transgenic diatom gene expression, as they all use the common transcriptional terminator part T_{fcpA}.

Episome-Based Forward Genetics Screen for Promoter Identification. A diatom episomal plasmid was designed and utilized as part of a forward genetics-based promoter discovery screen by ligating sheared *P. tricornutum* genomic DNA upstream of the ShBle coding region that

https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX

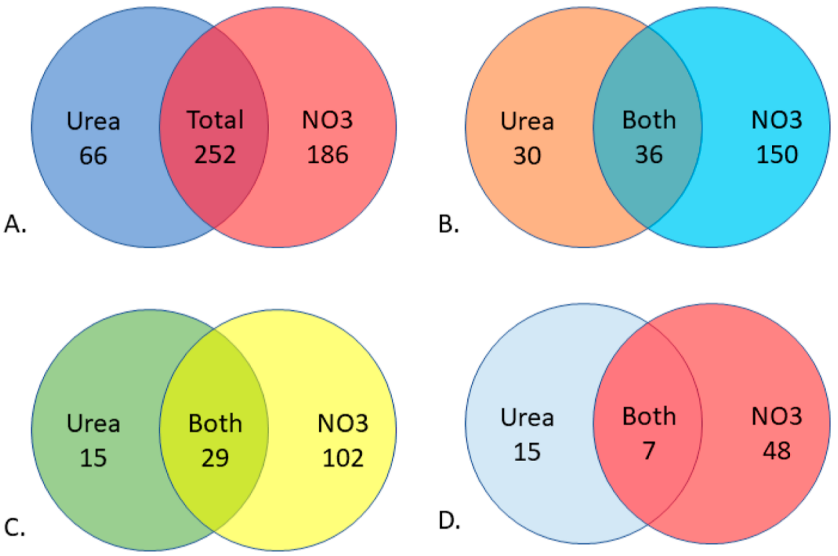


Figure 2. Venn diagrams representing the results of the forward genomic library promoter screen sequencing. (A) 252 sequencing hits were obtained from the forward genetic screening using the diatom conjugative plasmid driving the phleomycin resistance gene when exconjugants were isolated using the medium containing different nitrogen sources (nitrate vs urea). Of these 252 hits, 186 were recovered from NO_3-grown exconjugants, while 66 were recovered from exconjugants selected using urea as the nitrogen source. (B) When mapped to the Phatr3 genome assembly, it was shown that 150 unique hits were recovered from NO_3-grown colonies and 30 unique hits from colonies selected on urea. 36 duplicate hits were recovered from both media selections (total of 72). (C) Based on the criteria outlined in the text, 175 of the 252 sequencing hits were classified as "promoters" with 102 hits from NO_3 and 15 from urea-selected colonies. Twenty-nine promoters were identified as duplicates (found in both media conditions) for a total of 146 unique promoters identified in this screen. (D) 77 sequencing hits were labeled as "possible", defined as DNA fragments recovered from colony selection not fitting the criteria for a canonical promoter but still able to drive expression of ShBle in vivo. NO_3-selected colonies produced 48 hits, while urea-selected colonies produced 15 hits unique to that selection condition. Seven duplicate (14 in all) unclear hits were found to overlap, for a total of 70 hits.

confers resistance to phleomycin (pDEST-ShBle, Supporting Information Figure S3) to generate a plasmid library. We hypothesized that library fragments that contained a promoter in the proper spacing and orientation would drive expression of the ShBle gene and provide phleomycin resistance during selection of exconjugants. The gDNA library consisted of DNA fragments 2−5 kb in size with a diversity that covered approximately 0.9× of the *P. tricornutum* genome (see Materials and Methods for details). We transferred the library and control plasmids into *P. tricornutum* via pTA-Mob-mediated conjugation[49] and selected for exconjugants on 1/2 L1 medium containing 20 μg mL^{-1} phleomycin and either 8.8 mM nitrate or 4.4 mM urea as the nitrogen source (i.e., 8.8 mM total nitrogen). Both treatments yielded hundreds of colonies (see Materials and Methods), and episomes were extracted en masse from the resulting *P. tricornutum* exconjugants. Isolated episome pools were passaged through *E. coli* to eliminate *P. tricornutum* DNA not associated with the promoters, and the plasmids were re-extracted from bacteria en masse. The purified episomes were then prepared for Illumina sequencing using predefined primer binding sites that amplify the region upstream of the ShBle cassette while adding sequencing adapters. For both nitrate and urea-selected data sets, no chloramphenicol-resistant *E. coli* colonies were obtained, indicating that spiked-in control and nonfunctional library episomes were completely removed during selection. A diagram detailing the workflow for this screen can be found in Supporting Information Figure S4.

One important quality control checkpoint was to ensure that the sequencing reads recovered from the exconjugant diatoms matched the regions of the published *Phaeodactylum* genome sequence. This would support the conclusion that in vivo expression of the ShBle antibiotic resistance marker was due to the presence of a transcriptionally active region isolated from the diatom gDNA library. Approximately 1 million reads per library were obtained by Illumina NextSeq. A detailed explanation of sequence read mapping and pairing can be found in the Materials and Methods section (below). Briefly, reads were mapped to a combined sequence set comprising the *P. tricornutum* chromosome-scale scaffolds, the unscaffolded assemblies, mitochondrial and chloroplast genomes, and the negative control template (pPtProEXP-23). Only uniquely matching reads were mapped, and mapped reads where a top strand and a bottom strand read were separated by 2−5 kb were considered paired.

Although all paired sequencing reads (hereafter referred to as a "paired read") yielded colonies in the functional conjugation/selection assay, we set criteria for labeling a paired read as a probable promoter: (1) the 3′ end of the paired read fell within a 700-bp range of the start codon of a downstream CDS predicted by the Phatr3 assembly including hypothetical proteins, (2) the associated CDS is downstream of the paired read on the same DNA strand, and (3) allowed for an extension of a sequencing read 3′ end beyond the predicted start codon (ATG) of a CDS (200 bp). In total, we recovered 252 paired reads with 186 recovered from exconjugant colonies selected on NO_3 and 66 paired reads from colonies selected on urea (Figure 2 and Supporting Information Tables S5 and S6). Analysis of this set showed 150 unique paired reads recovered from NO_3-grown colonies and 30 unique paired reads from colonies selected on urea, while 36 duplicates were recovered from both media selections (total of 72). When we applied the criteria outlined above, 175 of the 252 paired reads were classified as "promoters" with 102 called

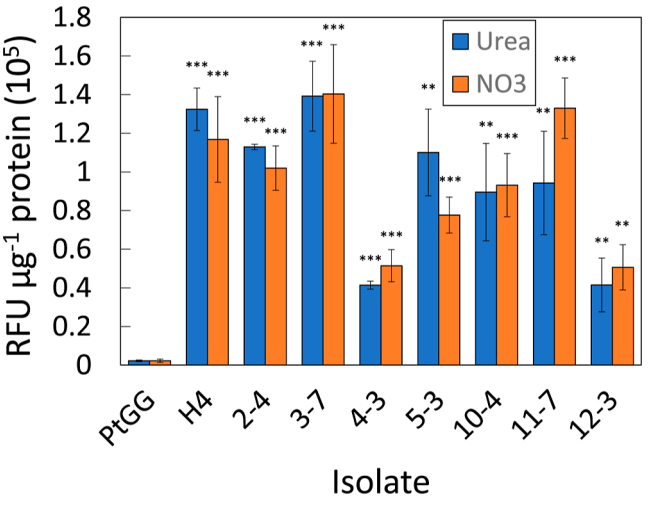
https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX

from NO_3 and 15 from urea-selected colonies. Twenty-nine promoters were identified as duplicates (found in both media conditions) for a total of 146 unique promoters identified in this screen (Supporting Information Table S5). In addition, 77 paired reads, called "unclear", are defined as DNA fragments recovered from colony selection not fitting the criteria for a canonical promoter but still able to drive expression of ShBle in vivo (Supporting Information Table S6). NO_3-selected colonies produced 48 paired reads, while urea-selected colonies produced 15 paired reads unique to that selection condition. Seven duplicate (14 in all) possible paired reads were found to overlap, for a total of 70 paired reads. It is important to note that no reads mapping to the *GUS* sequence in pPtProEXP-23 were obtained in the final sequencing of the plasmids.

Functional analysis and binning of the ORFs downstream of positive paired reads showed many pertained to basic cellular functions (as defined by KOG classifications) such as energy production and metabolism (6); transcription and translation (15); protein trafficking and proteasome functions (15); signal transduction (9); transport of ions, amino acids, and other small molecules (13); and DNA replication and cell cycle control (5). An additional 8 paired reads were classified as "general function only" due to unclear annotation of the associated CDS.

A large number (52) of paired reads mapped near coding regions labeled hypothetical proteins that are part of the "diatom-only" set of genes yet to be assigned functions[50] and were therefore not assigned KOG classifications. Importantly, none of the recovered paired reads mapped to the chloroplast or mitochondrial genomes, which are genetically prokaryotic, bearing 70S ribosomes and promoters regulated in a bacterial manner.[50,51] We believe this speaks to the selectivity of our library screening vector for eukaryotic, polII-like promoters in the nucleus, as previous studies have demonstrated that the diatom episomes, similar to those used in these experiments, interact with histones.[52] In addition, this genomic screen identifies regions of the genome not previously appreciated for their transcriptional activity and highlights the ability of diatoms to utilize nonpromoter regions of the genome for driving transcription under selective pressure (i.e., phleomycin resistance).

Validation of Sequencing Hit Fragments Identified Using the Genomic Library Approach. Because the library screening vector required transcription of the ShBle gene for exconjugant survival, we chose to address the theory that some paired reads, especially those not mapping near annotated genes or in low gene density regions of the genome, were recovered due to illicit or random transcription during conjugation due to the selective pressure of antibiotic resistance. A set of these putative types of pseudopromoter elements were validated by fusing them to *GUS* rather than antibiotic resistance expression. We chose seven *E. coli* colonies obtained after transforming the extracted episome library from *P. tricornutum*. The promoter fragment upstream of the ShBle gene was amplified from the episome in each *E. coli* colony and cloned in front of the *GUS* gene using the Gateway system into a vector utilizing the same *fcpA* 3′-UTR fragment as pShBle-DEST. These episomes were conjugated into *P. tricornutum*, exconjugants isolated, and levels of GUS enzyme activity were determined when cells were grown in the presence of either 880 μM nitrate or 440 μM urea (Figure 3). The clones screened demonstrated the ability to drive *GUS* expression when evaluated by enzyme activity (*t*-test vs PtGG, $p < 0.05$, see the legend) (Supporting Information Table S7), many close to the same levels as detected in cells expressing P_{h4}-driven *GUS* (Figure 3). The expression of the *GUS* marker demonstrates that the expression of phleomycin resistance in the large-scale screen was not due solely to the need for survival under antibiotic selection. The ability for library-generated paired reads to drive the expression of a second transgene (*GUS*) suggests that they are bona fide promoter elements, and all seven mapped well within the genome according to our criteria. Analysis of the presumed in vivo regulated ORFs showed that at least two clones were a part of the "diatom-only" class of genes,[50] and HMMer analysis[53] identified a predicted N-terminal signal sequence (clone C2) or a transmembrane domain (clone C4) but no activity-associated domains. Four other ORFs were associated with the remaining clones. Two of the colonies possessed the same 5′-UTR fragment when analyzed by Sanger sequencing (clones C5 and C10) and are upstream of Phatr3_J17683, annotated as *sybA*, a putative synaptobrevin part of the protein trafficking network. The remaining clones, C3, C11, and C12, were associated with a putative RNA-splicing factor, casein kinase II



Colony	Recovered Library Hit Element
C2-4	NO3_035: Phatr3_EG01598, hyptothetical protein
C3-7	NO3_023: Phatr3_J10122, RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)
C4-3	NO3_127: Phatr3_EG00559, hypothetical protein
C5-3	NO3_012: Phatr3_J17683, Synaptobrevin||Regulated-SNARE-like domain
C10-4	NO3_012: Phatr3_J17683, Synaptobrevin||Regulated-SNARE-like domain
C11-7	NO3_080: Phatr3_J12614, Casein kinase II regulatory subunit
C12-3	NO3_172: Phatr3_EG01908, E1-E2 ATPase, haloacid dehalogenase-like hydrolase, cation transporting ATPase, C-terminus, 7 TM domains

Figure 3. Validation of the forward genomics promoter screen workflow by cloning and analysis of GUS expression relative to the negative PtGG construct. In the top panel, the activities of seven putative *P. tricornutum* promoters were quantified based on *GUS* expression when cells were grown in L1 medium with either nitrate or urea as the sole nitrogen source. PtGG represents a negative control strain containing a plasmid harboring the PtGG DNA fragment cloned in front of *GUS* with no expected promoter activity. The h4 promoter is used as a positive control with known promoter activity. The error bars represent the standard deviation of three biological replicates for each cell line. Asterisks represent a *t*-test *P*-value ≤0.01 (**) or ≤0.001 (***) when comparing the expression to PtGG. Bottom panel, table of hit IDs for colonies tested with the GUS assay.

https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX

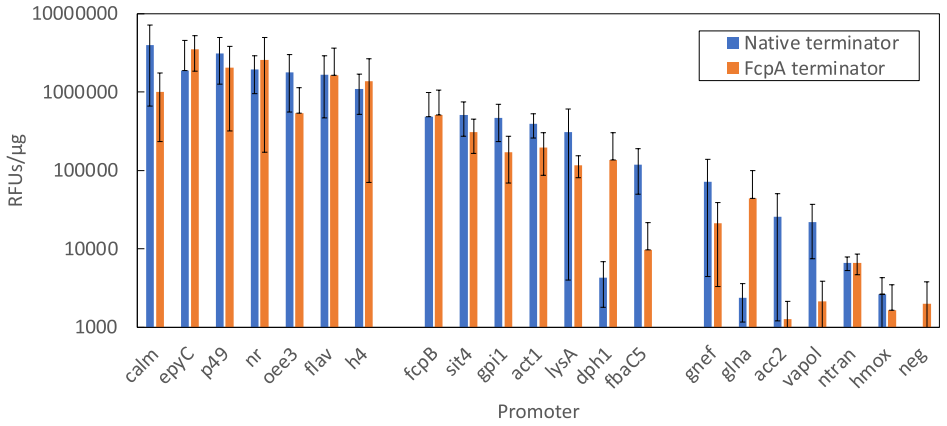


Figure 4. Expression profiles of pEG constructs active across a range of 4 orders of magnitude comparing the presence of native versus non-native terminators with putative promoters. Normalized beta-glucuronidase activity for diatom exconjugants expressing synthetic promoter−reporter−terminator constructs on episomes. The GUS activity is determined using a fluorescence assay on protein extracts and normalized to protein concentrations. Error bars are for 3−6 biological replicates with 1−2 technical replicates. Neg represents a negative control strain containing a plasmid harboring the PtGG DNA fragment cloned in front of GUS with no expected promoter activity. Mapping of promoter names to full protein IDs is provided in Supporting Information Table S6. As presenting 361 individual *t* tests would make this figure unreadable, we have included this information as a separate heatmap in Figure 5.

β subunit, and an E1-E2 ATPase with seven transmembrane domains, respectively. The results of this screen demonstrate the usefulness of the episome as a backbone for discovery and validation of new regulatory elements for use in diatom genetic engineering.

Domesticating and Characterizing New Diatom Promoter/Terminator Pairs. Based on the result of the library promoter screen and previous transcriptome analyses, 16 predicted promoters were selected for validation and characterization using the previously described pEG-GUS reporter constructs. When designing primers, we chose an agnostic approach to promoter cloning, dictated by the genomic structure of diatom protein coding regions. The criteria for DNA cloning was a maximum of 1500 bp upstream of the gene initiation codon or 5′-UTR (if present) for the promoter fragment and 200 bp downstream of the stop codon for the terminator. When another gene was present within 1500 bp upstream of the gene of interest, the DNA fragment clone consisted of the intergenic region, regardless of the size. We chose to exclude the CDS or predicted 5′-UTR of other genes on the same strand, if possible, in order to isolate the activity of the promoter of interest. Therefore, some putative promoters and terminators are significantly shorter than 1500 and 500 bp, respectively.

The rationale behind the diatom regulatory parts chosen for testing (Supporting Information Table S8) via the pEG system included previously reported promoter/terminators to act as benchmarks (P_{nr}, P_{fcpB}, P_{h4}, P_{fbac5}, and P_{glna}) and housekeeping, metabolic, structural, and presumed constitutively active diatom genes identified via transcriptomic analysis.[28,54,55] Although 16 additional promoters were selected based on transcriptomic data that suggested potential promoter activation in response to changes in CO_2, iron availability, nitrogen source, growth stage, or cell cycle, testing promoter activity in these conditions is beyond the scope of this paper as the purpose of this study is promoter identification. Additional testing will still need to be performed to determine promoter response to various growth conditions.

Figure 4 shows the results of the screening of conjugation-generated *P. tricornutum* assay lines using the pEG plasmids, with control lines included. We also compared the expression of each promoter terminator pair against all other native or non-native pairs; as 361 individual *t* tests would make Figure 4 unreadable, we have included these tests as heatmaps in Figure 5. Exconjugant strains of *P. tricornutum* with episomes containing promoter−*GUS*−terminator constructs were grown on nitrate as the sole carbon source with a 14:10 diel cycle in the presence of 20 μg mL^{-1} phleomycin. In total, 20 different promoters were tested with a total of 40 promoter−*GUS*−terminator combinations. The resulting expression profiles show that this new tool set greatly increases the diversity of expression profiles relative to the "legacy" tool set; the new one spans four rather than 2 orders of magnitude in expression level. Overall, the entirety of the cell lines and all tested promoter/terminator pairings generated with pEG plasmids expressed the *GUS* reporter gene and did not display general growth defects. Three clusters of promoter activity emerged from the plotted data which were classified as the "low", "medium", and "high" sets. This ranking was based on the observed in vivo GUS activity measured from the clarified lysates of transgenic lines.

The "high" expression set was defined as promoters displaying an average RFU μg^{-1} value above 1,000,000 and includes seven of the promoters tested (P_{calm}, P_{epyC}, P_{p49}, P_{nr}, P_{oee3}, P_{flav}, and P_{h4}), with at least one promoter/terminator pair displaying an average RFU μg^{-1} value of over 1,000,000. Of this set, P_{oee3} (oxygen-evolving enhancer protein 3; Phatr3_J54499) and P_{calm} (calmodulin-dependent protein kinase II; Phatr3_J39236) showed the greatest levels of expression when paired with their native terminators, as there was about a 3.3−4-fold decrease in expression when T_{fcpA} was used. Additionally, P_{epyC} showed extremely high variability when it was paired with its native terminator, indicating that if this promoter was selected for expression that it should be paired with a non-native terminator to achieve more stable expression. The opposite trend is seen when the non-native *fcpA* terminator is paired with P_{nr}, P_{oee3}, P_{flav} (flavodoxin; Phatr3_J23658), and P_{h4}, as the variability increases with the loss of the corresponding native terminator. This suggests that terminators can affect expression levels in *P. tricornutum* for

https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX

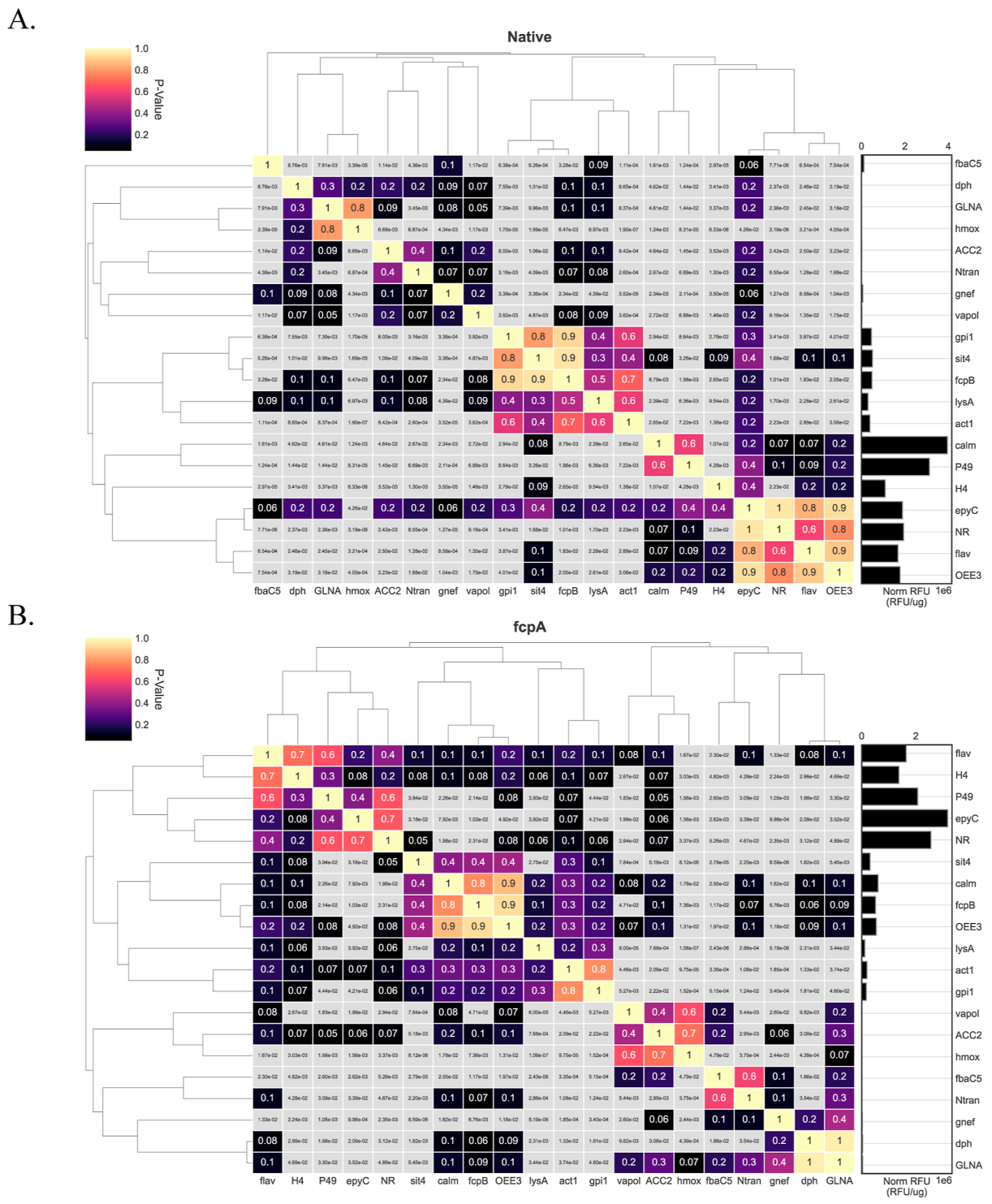


Figure 5. *P*-values from *t* tests for native and *fcpA* terminator pairs. (A) All vs all heatmap of the *p*-values for the promoter-native terminator pairs. (B) All vs all heatmap of the *p*-values for the promoter–*fcpA* terminator pairs. Comparisons that yielded a statistically significant *p*-value are highlighted in gray. The column on the right-hand side represents the mean expression value.

some promoters, but not all, possibly through regulatory elements or a gene loop that regulates transcription.[56]

The "medium" expressing promoter set includes seven of the tested promoters (P_{fcpB}, P_{sit4}, P_{gpi1}, P_{act1}, P_{lysA}, P_{dph1}, and P_{fbaC5}). These were found to have mid-range levels of expression and at least one promoter/terminator pair had an average RFU μg^{-1} value between 100,000 and 1,000,000 (Figure 4). P_{gpi1} (glucose-6-phosphate isomerase; Phatr3_J23924) and P_{fbaC5} (fructose bisphosphate aldolase; Phatr3_J41423) displayed higher levels of expression when paired with their native terminators. A decrease in expression when pairing these promoters with T_{fcpA} implies that the native terminators were in some way able to increase the overall expression of *GUS* and resulted in a change in epigenetic regulation when the native terminator was exchanged. When paired with T_{fcpA}, P_{dph1} (sensory transduction histidine kinase; Phatr3_J54330) out-

https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX

performed its native terminator with a 32-fold increase in expression. The increased expression from the P_{dph1}-*GUS*-T_{fcpA} cassette could be due to the loss of a regulatory element that was present in the native terminator that allowed for a promoter/terminator interaction that would modulate expression. Based on the standard deviation (Supporting Information Table S9), the P_{lysA}-*GUS*-T_{lysA} construct showed more variability in expression than the other mid-range promoters. The *lysA* gene is a diaminopimelate decarboxylase (Phatr3_J21592) gene involved in lysine biosynthesis, so the high variation in expression can be attributed to dynamic transcriptional regulation of this pathway.[28,57] Determining the dynamic range of expression for each promoter/terminator pair is important so that appropriate pairings can be used in engineered constructs. For example, certain transcriptional units may be required to express at a consistently high rate. In this case, the P_{fcpB}-T_{fcpB} terminator pair would be a more ideal candidate due to its stability, over pairings like P_{lysA}-T_{lysA}, which have a noisier expression profile. It should also be noted that the P_{act1}-*GUS*-T_{act1} construct was found to have a large deletion in the terminator region resulting in a 460 bp truncation. Further testing is required to determine the effects of the full native terminator sequence.

In regard to the "low" expressing promoter set, these appear to be the most affected by differing terminator elements when the *GUS* gene is expressed. Only "Ntran" (Phatr3_EG02608), which encodes a putative nitrite transporter, does not show dramatic changes in promoter activity when the terminator is changed. This group displayed an average GUS activity RFU μg^{-1} value under 100,000 and consisted of six promoters: P_{gnef}, P_{glna}, P_{acc2}, P_{vapol}, P_{ntran}, and P_{hmox}. The native and T_{fcpA} terminator constructs for P_{hmox} (heme oxygenase; Phatr3_J5902), the native terminator construct for P_{glna} (glutamine synthetase; Phatr3_J22357), and the T_{fcpA} constructs for P_{acc2} (acetyl-CoA carboxylase; Phatr3_ J55209) and P_{vapol} (vacuolar polyphosphate accumulation protein; Phatr3_J47434) had the lowest overall expression rates out of all the promoter/terminator pairs that were tested, with average RFU μg^{-1} values under 5000. The expression levels for P_{hmox} were consistently so low that it had a similar expression pattern to the negative control (vector contains the same PtGG promoter sequence). Both of the P_{ntran} terminator constructs are also considered low-expressing with average GUS activity values of only 6600 RFU μg^{-1}. There was also a large increase in GUS activity observed when P_{glna} was paired with T_{fcpA} (18.6-fold increase over the native terminator), while an increase in GUS activity was observed in P_{acc2} and P_{vapol} constructs paired with their native terminators (20.3- and 10.4-fold increase, respectively). However, the expression of P_{gnef} (guanine nucleotide exchange factor; Phatr3_J41365), P_{glna}, and P_{acc2} native terminator constructs is highly variable, which may be due to the dynamic transcriptional variability during the day/night cycle.[28]

To better illustrate the differences in promoter expression between promoters and promoter/terminator pairs, *P*-values based on *t* tests were calculated comparing all of the promoter/native terminator pairs (Figure 5A) and all of the promoter/*fcpA* terminator pairs (Figure 5B) to each other. These heatmaps help highlight which promoters show statistical significance from one another and can aid in choosing a promoter for an engineered expression cassette. As can be seen in comparing Figure 5A,B, promoters like P_{fbaCS}, P_{vapol}, and P_{p49} are more statistically different from the other promoters whether they are paired with their native terminator or the *fcpA* terminator, while promoters P_{h4}, P_{nr}, P_{oee3}, P_{flav}, P_{calm}, and P_{acc2} should be paired with their native terminators for more reliable expression. Promoters P_{epyC} and P_{lysA} show a lot more variation in their native configurations and should be paired with T_{fcpA} if they are going to be used. For more information on replicates and calculated values like mean, standard error, degrees of freedom, and *t*-statistics, see Supporting Information Table S10.

In order to test the long-term stability of the conjugated episomes, a subset of pEG-engineered cell lines was grown via serial transfer over a 4 month period and rescreened to determine expression stability. As shown in Figure 6, P_{h4}-*GUS*-T_{h4}, P_{nr}-*GUS*-T_{nr}, P_{dph1}-*GUS*-T_{fcpA}, and P_{gnef}-*GUS*-T_{fcpA} all had statistically indistinguishable expression patterns (*t*-test, *P*-value >0.05, within promoter $t = 0$ vs $t = 4$) (Supporting Information Table S11) after being tested 4 months later. The lack of gene expression changes suggests that the episomes are resistant to epigenetic modifications, while the ability to recover the episomes using a plasmid prep shows that they are unlikely to undergo chromosomal integration at a high frequency. This suggests that the episomes are stably maintained over time without being lost, mutated, or chromosomally integrated, which further supports the use of conjugatable episomes for exogenous gene expression.

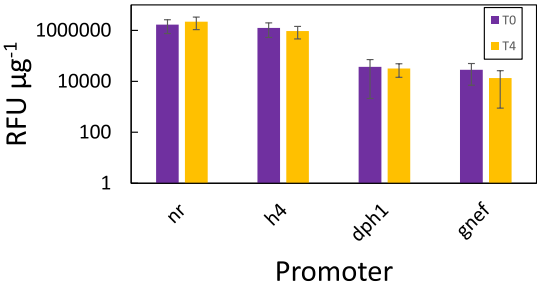


Figure 6. Expression of GUS from conjugated diatom episomes is stable over an extended period of culturing. *P. tricornutum* strains harboring plasmids with two high (P_{nr} and P_{h4}) (promoter/native terminator), one medium (P_{dph1}), and one low (P_{gnef}) (promoter/*fcpA* terminator) expressing promoter/terminator pairs driving the GUS marker gene were analyzed initially (*T0*) and after a period of 4 months (*T4*). No difference in expression strength (*t*-test, $P > 0.05$, within promoter $t = 0$ vs $t = 4$ tested) or loss of plasmid occurred when the cells were cultured appropriately and in the presence of antibiotic selection (phleomycin, 20 μg/mL). The error bars represent the standard deviation of expression of 3 biological replicates for each cell line.

Creating a Diatom Toolkit for the uLoop Library. The promoters and terminators characterized in this study have been incorporated into the uLoop system library for open-access distribution.[15] Genetic parts in the uLoop repository have been sequence-validated and characterized so that others may freely order parts for their own experiments. This will not only increase the genetic repertoire of *P. tricornutum* parts but will allow researchers to easily use these exact parts/sequences in their own studies resulting in more consistent, reproducible results among individuals and across research groups. Although the parts that have been added to the system have currently only been tested in a limited set of conditions, they are capable of controlling gene expression over multiple orders of

https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX

magnitude allowing for more fine-tuned control and expanding the current genetic toolkit for *P. tricornutum*.

CONCLUSIONS

The data herein reinforce recent studies reporting the high efficacy and reproducibility of enzyme expression in episomal plasmid-engineered *P. tricornutum* strains. Twenty diatom promoters, many new, were characterized via expression of the GUS reporter enzyme and rank-ordered for use in designing genetic engineering strategies. Data was also presented, suggesting that the activity of some promoters in *P. tricornutum* may be influenced by the combined terminator sequence, suggesting that caution should be applied when designing expression modules. In addition, we demonstrate that an episomal plasmid can function as a platform for diatom gene regulatory element discovery via in vivo screening of a genomic DNA library when combined with next-generation sequencing technology.

METHODS

Strains and Growth Conditions. *P. tricornutum* CCAP1055/1 was cultured in L1 medium[10] supplemented with 8.8 mM nitrate or 4.4 mM urea at 50 μmol m^{-2} s^{-1} light, atmospheric levels of CO_2, 18 °C, and a 14:10 day/night cycle. Diatom agar plates consisted of one-part L1 medium (sterilized) containing 8.8 mM nitrate combined with one-part 2% autoclaved agar. When necessary, media was supplemented with phleomycin (Invivogen) at a concentration of 20 μg mL^{-1}. L1 agar used for conjugations was made by making a mixture consisting of 2% agar (50%), L1 media (45%), and LB broth (5%). Conjugation agar plates were supplemented with ampicillin (50 μg mL^{-1}) and poured into 12-well plates. Phleomycin was used at a concentration of 20 μg mL^{-1} for selection of diatom exconjugants and maintenance of transgenic lines in liquid medium. EPI300 *E. coli* (Lucigen) were utilized for DEST vector and library assemblies. EPI300 strains harboring the pTA-Mob vector were used for conjugation of diatom cells. NEB5α (New England Biolabs) cells were used for construction of the pBR322-based plasmids. *E. coli* strains were grown on Luria−Bertani broth or agar supplemented with the following antibiotics as necessary: ampicillin (50 μg mL^{-1}), tetracycline (10 μg mL^{-1}), or gentamycin (20 μg mL^{-1}).

Molecular Biology Methods and Colony PCR Screening. Gibson assemblies were carried out as previously described.[48] *E. coli* clones and *P. tricornutum* exconjugants were subjected to colony PCR screening using OneTaq polymerase (NEB) or SapphireAmp Fast polymerase (Takara). Bacterial clones were amplified for 25 cycles, whereas for screening of diatom exconjugants, the total number of cycles was increased to 30. The full protocol can be found at dx.doi.org/10.17504/protocols.io.hnzb5f6.

Conjugation Protocols. Bacterial conjugation of diatom library screening cultures was carried out as previously reported,[10] and conjugations for generation of promoter testing and validation lines were carried out in a multiwell format.[11] Three to six colony PCR-positive isolates from each conjugation were saved and tested for each construct.

Plasmid Construction. The *GUS* coding region and the *fcpA* terminator fragment were amplified from vectors PB-fcpB (previously PtRNAi-2c) and PtRNAi-3.[58] Two other fragments were also amplified from the vector PtRNAi-3. One containing the *ampR* gene through the *oriT* region, whereas the other contained the *tetR* gene, including the *CEN6-ARSH4-HIS3* region, to the beginning of the ShBle cassette (driven by *P. tricornutum* P_{fcpF}). Other promoter and terminator fragments were amplified from the *P. tricornutum* genome. All fragments were generated using PrimeStar polymerase (Takara) and purified using the QiAQuick cleanup kit (Qiagen). The fragments were then assembled via the Gibson method and validated using Sanger sequencing. Completed vectors were transformed into *E. coli* containing the mobilization helper plasmid, pTA-Mob.[49] These strains were then used to conjugate *P. tricornutum*. Primers and plasmids used in this study are listed in Supporting Information Table S12.

GUS Activity Assay. There were two methods that were followed for completing the GUS assays. The pDEST-based vectors were grown in 5 mL 1/2-L1 liquid medium to an approximate density of 3−4 × 10^6 cells mL^{-1}. Cells were then harvested and subjected to three freeze/thaw cycles. The clarified lysates were used in the GUS assay, and fluorescence was measured using a FlexStation3 microplate reader (Molecular Devices) with excitation/emission settings of 360 nm excitation/440 nm emission (cutoff at 435 nm). The complete pDEST GUS assay protocol can be found at protocols.io (dx.doi.org/10.17504/protocols.io.hefb3bn).

Leftover lysate volumes (10 μL) from the GUS assay were used in a Pierce BCA protein assay (Thermo Fisher) following the manufacturer's protocol. The absorbance (562 nm) was read on a FlexStation 3 microplate reader. The amount of total protein added to each assay well was calculated and used for normalization, and final concentrations were determined as RFU μg^{-1} total protein. Lysate-free wells using only GUS extraction buffer were added as blanks for each plate reader assay.

For the pEG vectors, a high-throughput GUS assay was developed. Cultures were grown in 5 mL L1 to a density of at least 1 × 10^6 cells mL^{-1}. For each culture, 250 μL was transferred to a 96-well plate. Cells were harvested and lysed using bacterial protein extraction reagent (B-PER, Thermo Fisher). Clarified lysates were subjected to the GUS assay, and fluorescence was determined using a microplate reader as previously described. Samples were diluted with additional buffer as needed and were normalized by total protein as determined by the BCA assay. The full high-throughput version of the protocol can also be found on protocols.io (dx.doi.org/10.17504/protocols.io.bbexijfn).

Testing of the Legacy Promoter Set. The regions encompassing the *fcpB*, *h4*, *p49*, and *nr* promoters were amplified from templates PtRNAi-3, -8, -9, and -11, respectively, using primers designed with overhangs for Gibson assembly into the entry vector PtGG-1. The entry vectors were assembled using the Gibson method, validated by Sanger sequencing and assigned the names PtPro-8, -11, -12, and -13 (P_{fcpB}, P_{h4}, P_{p49}, and P_{nr}, respectively). These entry vectors were used in the LR Clonase reaction (LifeTech) to recombine the fragments in front of the *GUS* coding region in pDEST-GUS by following the manufacturer's protocol. These clones were validated using restriction digest analysis to determine proper recombination and assigned the names PtProEXP27, -29, -30, and -31 (P_{fcpB}, P_{h4}, P_{p49}, and P_{nr}, respectively). In order to test P_{nr} with its native terminator, the *GUS* CDS was amplified from PB-fcpB using primers (*GUS*-PtGG-1 + *GUS*-PtGG-2) and assembled into PtGG-1 to make PtGG-GUS. This entry vector was then applied in the LR reaction to PtRNAi-11 to

https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX

generate PtProEXP-40. A negative control vector consisting of the PtGG-1 spacer region was also recombined into pGUS-DEST to serve as a random DNA sequence-negative control. Expression plasmids were transformed into EPI300-pTA-Mob and conjugated into *P. tricornutum* using the multiwell protocol.[11] After 10 days of selection, colonies were patched onto fresh selective agar plates and verified via colony PCR screening. Three PCR-positive colonies were selected and moved forward for screening of GUS enzyme activity.

Construction of DNA Parts for the uLoop Library. All of the promoters and terminators that were characterized in this study were domesticated as level 0 parts for the uLoop library as previously described[15] (Supporting Information Table S5). Each part was sequence-verified via Sanger sequencing before being deposited. The full protocol can be found at dx.doi.org/10.17504/protocols.io.yxnfxme.

Construction of pShBle-DEST-lib. Plasmid pShBle-DEST was designed with Gateway destination recombination sites upstream of a promoterless bleomycin/phleomycin resistance gene (ShBle) from *Streptoalloteichus hindustanus*. To construct this plasmid, pPtPBR1[11] was amplified such that the promoter driving the ShBle marker was removed. This backbone contains the *S. cerevisiae CEN6-ARSH4-HIS3* fragment to promote episomal maintenance in *P. tricornutum* and an *oriT* from plasmid RP4/RK2 for bacterial conjugation via the pTA-Mob helper plasmid. This sequence was amplified as two fragments using primers ptPRO-004 and ptPRO-005 for the first fragment and primers ptPRO-005 and ptPRO-006 for the second fragment. The Gateway destination cassette region (attR1-CmR-CcdB-attR2) was amplified by PCR from pfcpB-DEST using primers ptPRO-003 and ptPRO-004, and this product was assembled with the two fragments from pPtPBR1 such that the Gateway destination cassette was located upstream from the ShBle gene where the promoter would have been in pPtPBR1. A Kozak initiation signal (5′-GGGGCCACC-3′) was installed upstream of the ShBle start codon to facilitate protein translation. As a negative control for library selection, a 1 kb region of the *GUS* coding sequence (nucleotides 400−1400) was cloned using primers ptPRO-108 and ptPRO-109 from pDEST-GUS and inserted in a reverse (3′−5′) orientation into the entry vector to generate PtGG-bwGUS. This vector was utilized in the LR reaction with pShBle-DEST-lib to create negative control plasmid PtProEXP-23.

Promoter Screen Library Generation. We developed a general genomic library construction method in which the *P. tricornutum* genome was fragmented by sonication using a Covaris S2, ligated with Illumina adapters, and assembled into a vector (p*GUS*-lib) prepared with homology sequences to the Illumina sequences (dx.doi.org/10.17504/protocols.io.hfb-b3in). For library assembly, plasmid pShBle-DEST was amplified using primers ptPRO-001 and ptPRO-002 to give a 7 kb product that omitted the Gateway destination cassette and added homology sequences to the Illumina adapters. The Illumina adapter-ligated samples were assembled into the prepared vector using Gibson assembly and transformed directly into EPI300 cells containing the pTA-Mob conjugative plasmid. Total library diversity was approximately 11,000 unique *E. coli* colonies. Test screening 20 *E. coli* colonies from the library by colony PCR indicated unique insert sizes at the expected size (2−5 kb) and no empty vector. Plates containing the colonies were flooded with LB medium, scraped, and cells stored in 15% glycerol at −80 °C. With an average estimated insert size of 3.5 kb, total library size was estimated to be approximately 38.5 Mb. This is approximately 0.9× coverage of the *P. tricornutum* genome, assuming a genome size per cell of 43 Mb, including 27.4 Mb for the nuclear genome, ∼100× of the 0.118 Mb chloroplast genome, and ∼50× of the 0.077 Mb mitochondrial genome.

Library Selection in *P. tricornutum*. The library was revived from frozen stocks by pooling and growing overnight in LB broth containing Amp/Tet/Gent (100/10/20 μg mL^{-1}) at 37 °C. The next day, this culture was diluted 1:50 into 50 mL of LB-Amp/Tet/Gent and grown for several hours until the OD_{600} was roughly 0.6. For library selection conjugation experiments, two internal controls were included. First, an EPI300 strain lacking pTA-Mob but carrying a plasmid bearing a chloramphenicol resistance marker (pCMR)[59] cultured in parallel with the library was added as 5% of the *E. coli* cell volume during conjugation. This control was designed to assess the persistence of nonconjugated plasmid on agar plates after selection and efficacy of *DpnI* treatment on samples. Second, a strain bearing pTA-Mob and a plasmid with a nonfunctional promoter sequence in front of ShBle (PtProEXP-23) was added at 5% of the *E. coli* cell volume during conjugation (final ratio 90:5:5 of the library and controls). This PtProEXP-23 control was designed to assess any background appearance of nonfunctional conjugated plasmid in the final sequencing data. For both nitrate and urea selected data sets, no chloramphenicol-resistant *E. coli* colonies were obtained, indicating that nonfunctional plasmids were completely removed during selection and *DpnI* purification. Similarly, no reads mapping to the *GUS* sequence in pPtProEXP-23 were obtained in the final sequencing of the plasmids.

Standard conjugation conditions were performed, and a total of eight conjugations per library experiment were carried out. After a recovery period of 48 h, the plates were flooded with L1 medium, and the cells scraped into a total volume of 600 μL. 200 μL of resuspended cells was then replated onto a single 1/2 L1-P20 plate containing 8.8 mM nitrate as the nitrogen source (total of 24 selection plates) and incubated under standard conditions. From these plates, approximately 1000 colonies were obtained. Under these conditions, approximately 20,000−30,000 colonies would have been expected from a control pPtPBR1 conjugation[11] resulting in an efficiency of finding a functional promoter of 0.03−0.05. These colonies were picked with an inoculating loop into 10 mL L1 medium with 8.8 mM nitrate as the nitrogen source and propagated for at least 1 week. The resulting cells were harvested via centrifugation and frozen at −80 °C until episome extraction.

Episome extraction from *P. tricornutum* was performed as previously described.[10] After extraction, episomes derived from *P. tricornutum* were further purified from trace DNA carried over from the conjugation by treatment with *DpnI* restriction endonuclease. Because DNA originating from dead *E. coli* contained the bacterial adenosine methylation mark on the *DpnI* recognition sequence (GATC) and is not expected to be found in DNA from *P. tricornutum*, all bacterial plasmid DNA would be digested, resulting in highly enriched episomal DNA from *P. tricornutum*. After purification of digestion reactions, the plasmids were electroporated into *E. coli* EPI300 and plated on LB-Amp/Tet (50/10 μg mL^{-1}). This final retransformation served to purify away any trace chromosomal DNA from *P. tricornutum* so that the only diatom sequences in the resulting plasmids would be from putative promoter

https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX

regions driving ShBle on the episome. Plasmids were purified from ∼15,000−20,000 *E. coli* colonies resulting in 10−20× coverage of the original ∼1000 *P. tricornutum* colonies obtained from selection on nitrate. An identical procedure was followed to select for promoters that functioned when selection for the episomes in *P. tricornutum* was performed with urea as the main nitrogen source. Approximately 300 *P. tricornutum* colonies were obtained from selection of the library on urea and 20,000−30,000 *E. coli* colonies, almost 100× coverage of the original 300 *P. tricornutum* colonies.

Library Sequencing and Analysis. Plasmids containing episomes with functional *P. tricornutum* promoters that were extracted from exconjugants and passaged through *E. coli* were sequenced using a directed tagmentation approach adapted from IS-seq.[60] A tagmentation library was prepared using the purified plasmid using the Nextera XT library preparation kit to insert randomly throughout the plasmid. Amplification of the tagmented library using the standard barcoded Illumina P7 adapter primer and a custom primer (ptPRO-009 or ptPRO-010) derived from either the upstream or downstream (closest to the ShBle start site) Illumina sites. Thus, each of the two pools of plasmids (nitrate or urea) were each amplified with either the P5dwnstrmP or P5upstrmP primer plus the standard P7 primer resulting in four libraries (an upstream location and downstream cloning junction library for each of the nitrate and urea libraries). Approximately 1 million reads per library were obtained by Illumina NextSeq Reads that were mapped to a combined sequence set comprising the *P. tricornutum* chromosome-scale scaffolds, the unscaffolded assemblies (i.e., "bottom drawer" or BD), the mitochondrial and chloroplast genomes, and the negative control template (pPtProEXP-23). Only uniquely matching reads were mapped.

Direction of the inset in a genomic context (plus- or minus-strand) was determined by the following method. We observed that in a given library (i.e., from a single primer site, P5dwnstrmP or P5upstrmP), reads from both sides of a given library member were obtained. This was likely due to the presence of a short sequence (5′gctcttccgatc) that was common to both primers. However, reads mapping to one side of the promoter were always in vast overabundance over the other. We reasoned that the PCR amplification was improved with the full-length primer but occurred at a low level when binding to the partial small sequence. This difference in mapped read abundance on either side of the library member correlated with the direction. Thus, with the sequencing library prepared from the P5dwnstrmP primer (closest to the ShBle "downstream" position), the side of the mapped library element with greater read abundance was assigned as the downstream side (i.e., the promoter went in the direction toward the larger number of reads). This strategy was empirically verified using the seven promoter elements that were sequenced with Sanger DNA sequencing.

Mapped reads where a top strand and a bottom strand read were separated by 2−5 kb were considered paired. The coordinates of the reads corresponding to the boundaries of the cloned promoter library element were extracted along with the scaffold and combined with the read abundance data for each read location to determine the orientation of the cloned promoter relative to the genome. Rare situations were handled as follows: first, from time to time, mapped read peaks were observed without a pair. This was most likely related to problems with read mapping, and these reads were discarded. Second, occasionally, a peak of very low read count (20−80) was observed between two peaks that were spaced appropriately and each contained 10-fold more reads than the small intervening peak. In this case, the low count peak was discarded. A third unusual situation was when two sets of peaks overlapped. In this case, the first encountered forward read was paired with the first encountered reverse read. In order to evaluate the genomic positions of the recovered library elements, the paired reads were mapped onto the *P. tricornutum* genome (including "bottom drawer", chloroplast, and mitochondrial sequences) and paired reads containing a 3′ end within 500 bp of an initiator methionine of a predicted ORF were considered positive.

Promoter Library Validation. Seven *E. coli* colonies containing recovered episomes were selected for the validation of the promoter fragment transcriptional ability. These library fragments were cloned in front of the *GUS* open reading frame and then mobilized into *P. tricornutum* to measure GUS activity as a proxy for promoter strength under a nonsurvival condition. The fragments were amplified by colony PCR from the library vector using the primers ptPRO-010 and ptPRO-011 which bind to the Illumina library spacer sequences flanking the promoter fragment. The amplified fragments were cleaned up with SPRI-beads and assembled into PtGG-1 to build promoter Gateway entry vectors. After transformation of the entry vectors into *E. coli*, clones were screened by colony PCR and the positive plasmids were purified and sequenced. The purified entry vectors containing the promoter fragments were then transferred to plasmid p*GUS*-DEST using the Gateway LR reaction (Life Technologies) and subsequently mobilized into *P. tricornutum*. The resulting *P. tricornutum* colonies were screened by colony PCR to verify that the promoter−*GUS* junction was present, and cells were grown in L1 medium containing either urea (440 μM) or nitrate (880 μM) as the sole nitrogen source. The GUS assay was performed to measure the promoter activity in cells grown in each nitrogen condition.

Statistical Analysis. Both two-sided and alternative hypothesis *t* tests were calculated using SciPy v1.8.0.[61] Clustermaps were plotted using Seaborn v0.11.2.[62] Pairwise *t*-test *p*-values were transformed to dissimilarity measures using 1 − *P*-value and transformed into linkage matrices for hierarchical clustering using SciPy and Soothsayer.[63]

Public Deposition of DNA and Plasmid Sequences. All library sequencing reads were deposited on NCBI as a BioProject (PRJNA942336) with the following accession numbers SRR23750734, SRR23750735, SRR23750736, and SRR23750737 for down- or upregulation on growth with urea or nitrate. Gateway cloning plasmid backbones were deposited to the GenBank database under the following accession numbers: pGUS-DEST (OR228473) and pShBle-DEST (OR228472). The promoter and terminator sequences used to make the pEG vectors were deposited to the Addgene database, and sequence IDs and additional information can be found in Supporting Information Table S8.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.3c00163.

SBOL schematic of plasmid PtGG-1; SBOL schematics of the pEG plasmid; SBOL schematic of plasmid pShBle-DEST; and workflow for construction, forward

https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX

screening, and sequencing of *Phaeodactylum* genomic library for new promoter discovery (PDF)

Addgene deposition information for PtGG and pDEST vectors, sequences of non-native or synthesized DNA elements used in this study, sequences of legacy diatom DNA elements used in this study, statistical analysis of p-DEST promoter expression vs PtGG, positive MiSeq hits for large-scale promoter screen, possible MiSeq hits for large-scale promoter screen, statistical analysis of putative promoter expression vs PtGG, promoters and terminators used in this study, standard deviations and variances for pEG vectors, *t*-test and *p*-value statistics for comparing promoter expression of pEG vectors, statistical analysis of promoter expression over time, and primers and plasmids used in this study (XLSX)

AUTHOR INFORMATION

Corresponding Author

Christopher L. Dupont − *J. Craig Venter Institute, La Jolla, California 92037, United States;* orcid.org/0000-0002-0896-6542; Email: cdupont@jcvi.org

Authors

Erin A. Garza − *J. Craig Venter Institute, La Jolla, California 92037, United States;* orcid.org/0000-0001-6988-4885

Vincent A. Bielinski − *J. Craig Venter Institute, La Jolla, California 92037, United States;* orcid.org/0000-0002-8107-8752

Josh L. Espinoza − *J. Craig Venter Institute, La Jolla, California 92037, United States*

Kona Orlandi − *J. Craig Venter Institute, La Jolla, California 92037, United States;* Present Address: Department of Molecular Biology, University of Oregon, Eugene, OR 97403

Josefa Rivera Alfaro − *J. Craig Venter Institute, La Jolla, California 92037, United States;* Present Address: Department of Biology, San Diego State University, San Diego, CA 92182.; orcid.org/0009-0002-4924-6581

Tayah M. Bolt − *J. Craig Venter Institute, La Jolla, California 92037, United States;* Present Address: Plant Biology Graduate Group, University of California, Davis, CA 95616.

Karen Beeri − *J. Craig Venter Institute, La Jolla, California 92037, United States;* Present Address: Vantage, Vanderbilt University, Nashville, TN 37232.

Philip D. Weyman − *J. Craig Venter Institute, La Jolla, California 92037, United States;* Present Address: Andes Ag, Inc., Emeryville, CA 94608.; orcid.org/0000-0001-5787-0291

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.3c00163

Author Contributions

E.A.G. and V.A.B. contributed equally to this work. V.A.B., K.B., C.L.D., and P.D.W. conceived of experiments and approaches. E.A.G., K.O., V.A.B., J.R.A., T.M.B., K.B., and P.D.W. performed the experiments. E.A.G., V.A.B., J.L.E., J.R.A., T.M.B., K.B., C.L.D., and P.D.W. analyzed the data. V.A.B., E.A.G., C.L.D., and P.D.W. wrote the manuscript.

Funding

Funding for this work was provided by the Gordon and Betty Moore Foundation (GBMF5007.1 to P.D.W. and C.L.D., GBMF5007.2 to C.L.D.).

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Behrenfeld, M. J.; Halsey, K. H.; Boss, E.; Karp-Boss, L.; Milligan, A. J.; Peers, G. Thoughts on the evolution and ecological niche of diatoms. *Ecol. Monogr.* **2021**, *91*, No. e01457.

(2) Naghshbandi, M. P.; Tabatabaei, M.; Aghbashlo, M.; Aftab, M. N.; Iqbal, I. Metabolic Engineering of Microalgae for Biofuel Production. *Methods Mol. Biol.* **2019**, *1980*, 153−172.

(3) Fabris, M.; George, J.; Kuzhiumparambil, U.; Lawson, C. A.; Jaramillo-Madrid, A. C.; Abbriano, R. M.; Vickers, C. E.; Ralph, P. Extrachromosomal Genetic Engineering of the Marine Diatom *Phaeodactylum tricornutum* Enables the Heterologous Production of Monoterpenoids. *ACS Synth. Biol.* **2020**, *9*, 598−612.

(4) Hamilton, M. L.; Haslam, R. P.; Napier, J. A.; Sayanova, O. Metabolic engineering of *Phaeodactylum tricornutum* for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids. *Metab. Eng.* **2014**, *22*, 3−9.

(5) Apt, K. E.; Kroth-Pancic, P. G.; Grossman, A. R. Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol. Gen. Genet.* **1996**, *252*, 572−579.

(6) Zaslavskaia, L. A.; Lippmeier, J. C.; Kroth, P. G.; Grossman, A. R.; Apt, K. E. Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. *J. Phycol.* **2001**, *36*, 379−386.

(7) Materna, A. C.; Sturm, S.; Kroth, P. G.; Lavaud, J. First Induced Plastid Genome Mutations In An Alga With Secondary Plastids: Psba Mutations In The Diatom *Phaeodactylum Tricornutum* (Bacillariophyceae) Reveal Consequences On The Regulation Of Photosynthesis. *J. Phycol.* **2009**, *45*, 838−846.

(8) George, J.; Kahlke, T.; Abbriano, R. M.; Kuzhiumparambil, U.; Ralph, P. J.; Fabris, M. Metabolic Engineering Strategies in Diatoms Reveal Unique Phenotypes and Genetic Configurations With Implications for Algal Genetics and Synthetic Biology. *Front. Bioeng. Biotechnol.* **2020**, *8*, 513.

(9) Angstenberger, M.; Krischer, J.; Aktaş, O.; Büchel, C. Knock-Down of a ligIV Homologue Enables DNA Integration via Homologous Recombination in the Marine Diatom *Phaeodactylum tricornutum*. *ACS Synth. Biol.* **2019**, *8*, 57−69.

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

(11) Diner, R. E.; Bielinski, V. A.; Dupont, C. L.; Allen, A. E.; Weyman, P. D. Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods. *Front. Bioeng. Biotechnol.* **2016**, *4*, 65.

(12) Slattery, S. S.; Diamond, A.; Wang, H.; Therrien, J. A.; Lant, J. T.; Jazey, T.; Lee, K.; Klassen, Z.; Desgagné-Penix, I.; Karas, B. J.; Edgell, D. R. An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in *Phaeodactylum tricornutum*. *ACS Synth. Biol.* **2018**, *7*, 328−338.

(13) Jaramillo-Madrid, A. C.; Abbriano, R.; Ashworth, J.; Fabris, M.; Pernice, M.; Ralph, P. J. Overexpression of Key Sterol Pathway Enzymes in Two Model Marine Diatoms Alters Sterol Profiles in *Phaeodactylum tricornutum*. *Pharmaceuticals* **2020**, *13*, 481.

(14) Kassaw, T. K.; Paton, A. J.; Peers, G. Episome-Based Gene Expression Modulation Platform in the Model Diatom *Phaeodactylum tricornutum*. *ACS Synth. Biol.* **2022**, *11*, 191−204.

(15) Pollak, B.; Matute, T.; Nuñez, I.; Cerda, A.; Lopez, C.; Vargas, V.; Kan, A.; Bielinski, V.; von Dassow, P.; Dupont, C. L.; Federici, F.

Universal loop assembly: open, efficient and cross-kingdom DNA fabrication. *Synth. Biol.* **2020**, *5*, ysaa001.
(16) Pollak, B.; Cerda, A.; Delmans, M.; Álamos, S.; Moyano, T.; West, A.; Gutiérrez, R. A.; Patron, N. J.; Federici, F.; Haseloff, J. Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytol.* **2019**, *222*, 628−640.
(17) Shetty, R. P.; Endy, D.; Knight, T. F., Jr. Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2008**, *2*, 5.
(18) Yoshinaga, R.; Niwa-Kubota, M.; Matsui, H.; Matsuda, Y. Characterization of iron-responsive promoters in the marine diatom *Phaeodactylum tricornutum*. *Mar. Genomics* **2014**, *16*, 55−62.
(19) Seo, S.; Jeon, H.; Hwang, S.; Jin, E.; Chang, K. S. Development of a new constitutive expression system for the transformation of the diatom *Phaeodactylum tricornutum*. *Algal Res.* **2015**, *11*, 50−54.
(20) Erdene-Ochir, E.; Shin, B.-K.; Kwon, B.; Jung, C.; Pan, C.-H. Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from *Phaeodactylum tricornutum*. *Sci. Rep.* **2019**, *9*, 9941.
(21) Windhagauer, M.; Abbriano, R. M.; Ashworth, J.; Barolo, L.; Jaramillo-Madrid, A. C.; Pernice, M.; Doblin, M. A. Characterisation of novel regulatory sequences compatible with modular assembly in the Diatom *Phaeodactylum Tricornutum*. *Algal Res.* **2021**, *53*, 102159.
(22) Siaut, M.; Heijde, M.; Mangogna, M.; Montsant, A.; Coesel, S.; Allen, A.; Manfredonia, A.; Falciatore, A.; Bowler, C. Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* **2007**, *406*, 23−35.
(23) Hempel, F.; Bullmann, L.; Lau, J.; Zauner, S.; Maier, U. G. ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Mol. Biol. Evol.* **2009**, *26*, 1781−1790.
(24) Miyagawa, A.; Okami, T.; Kira, N.; Yamaguchi, H.; Ohnishi, K.; Adachi, M. Research note: High efficiency transformation of the diatom *Phaeodactylum tricornutum* with a promoter from the diatomCylindrotheca fusiformis. *Phycol. Res.* **2009**, *57*, 142−146.
(25) Shemesh, Z.; Leu, S.; Khozin-Goldberg, I.; Didi-Cohen, S.; Zarka, A.; Boussiba, S. Inducible expression of Haematococcus oil globule protein in the diatom *Phaeodactylum tricornutum*: Association with lipid droplets and enhancement of TAG accumulation under nitrogen starvation. *Algal Res.* **2016**, *18*, 321−331.
(26) Poulsen, N.; Kröger, N. A new molecular tool for transgenic diatoms. *FEBS J.* **2005**, *272*, 3413−3423.
(27) Lin, H.-Y.; Shih, C.-Y.; Liu, H.-C.; Chang, J.; Chen, Y.-L.; Chen, Y.-R.; Lin, H.-T.; Chang, Y.-Y.; Hsu, C.-H.; Lin, H.-J. Identification and characterization of an extracellular alkaline phosphatase in the marine diatom *Phaeodactylum tricornutum*. *Mar. Biotechnol.* **2013**, *15*, 425−436.
(28) Smith, S. R.; Gillard, J. T. F.; Kustka, A. B.; McCrow, J. P.; Badger, J. H.; Zheng, H.; New, A. M.; Dupont, C. L.; Obata, T.; Fernie, A. R.; Allen, A. E. Transcriptional Orchestration of the Global Cellular Response of a Model Pennate Diatom to Diel Light Cycling under Iron Limitation. *PLoS Genet.* **2016**, *12*, No. e1006490.
(29) Kadono, T.; Miyagawa-Yamaguchi, A.; Kira, N.; Tomaru, Y.; Okami, T.; Yoshimatsu, T.; Hou, L.; Ohama, T.; Fukunaga, K.; Okauchi, M.; Yamaguchi, H.; Ohnishi, K.; Falciatore, A.; Adachi, M. Characterization of marine diatom-infecting virus promoters in the model diatom *Phaeodactylum tricornutum*. *Sci. Rep.* **2015**, *5*, 18708.
(30) Ohno, N.; Inoue, T.; Yamashiki, R.; Nakajima, K.; Kitahara, Y.; Ishibashi, M.; Matsuda, Y. CO(2)-cAMP-responsive cis-elements targeted by a transcription factor with CREB/ATF-like basic zipper domain in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiol.* **2012**, *158*, 499−513.
(31) Schellenberger Costa, B.; Sachse, M.; Jungandreas, A.; Bartulos, C. R.; Gruber, A.; Jakob, T.; Kroth, P. G.; Wilhelm, C. Aureochrome 1a is involved in the photoacclimation of the diatom *Phaeodactylum tricornutum*. *PLoS One* **2013**, *8*, No. e74451.
(32) Rayko, E.; Maumus, F.; Maheswari, U.; Jabbari, K.; Bowler, C. Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol.* **2010**, *188*, 52−66.
(33) Diamond, A.; Diaz-Garza, A. M.; Li, J.; Slattery, S. S.; Merindol, N.; Fantino, E.; Meddeb-Mouelhi, F.; Karas, B. J.; Barnabé, S.; Desgagné-Penix, I. Instability of extrachromosomal DNA transformed into the diatom *Phaeodactylum tricornutum*. *Algal Res.* **2023**, *70*, 102998.
(34) Slattery, S. S.; Wang, H.; Giguere, D. J.; Kocsis, C.; Urquhart, B. L.; Karas, B. J.; Edgell, D. R. Plasmid-based complementation of large deletions in Phaeodactylum tricornutum biosynthetic genes generated by Cas9 editing. *Sci. Rep.* **2020**, *10* (1), 13879.
(35) Pampuch, M.; Walker, E. J. L.; Karas, B. J. Towards synthetic diatoms: The *Phaeodactylum tricornutum* PT-syn 1.0 project. *Curr. Opin. Green Sustainable Chem.* **2022**, *35*, 100611.
(36) George, J.; Kahlke, T.; Abbriano, R. M.; Kuzhiumparambil, U.; Ralph, P. J.; Fabris, M. Metabolic engineering strategies in diatoms reveal unique phenotypes and genetic configurations with implications for algal genetics and Synthetic Biology. *Front. Bioeng. Biotechnol.* **2020**, *8*, 8.
(37) Valenzuela, J.; Mazurie, A.; Carlson, R. P.; Gerlach, R.; Cooksey, K. E.; Peyton, B. M.; Fields, M. W. Potential role of multiple carbon fixation pathways during lipid accumulation in *Phaeodactylum tricornutum*. *Biotechnol. Biofuels* **2012**, *5*, 40.
(38) McCarthy, J. K.; Smith, S. R.; McCrow, J. P.; Tan, M.; Zheng, H.; Beeri, K.; Roth, R.; Lichtle, C.; Goodenough, U.; Bowler, C. P.; Dupont, C. L.; Allen, A. E. Nitrate Reductase Knockout Uncouples Nitrate Transport from Nitrate Assimilation and Drives Repartitioning of Carbon Flux in a Model Pennate Diatom. *Plant Cell* **2017**, *29*, 2047−2070.
(39) Smith, S. R.; Dupont, C. L.; McCarthy, J. K.; Broddrick, J. T.; Oborník, M.; Horák, A.; Füssy, Z.; Cihlář, J.; Kleessen, S.; Zheng, H.; McCrow, J. P.; Hixson, K. K.; Araújo, W. L.; Nunes-Nesi, A.; Fernie, A.; Nikoloski, Z.; Palsson, B. O.; Allen, A. E. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* **2019**, *10*, 4552.
(40) Rastogi, A.; Maheswari, U.; Dorrell, R. G.; Vieira, F. R. J.; Maumus, F.; Kustka, A.; McCarthy, J.; Allen, A. E.; Kersey, P.; Bowler, C.; Tirichine, L. Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci. Rep.* **2018**, *8*, 4834.
(41) Giguere, D. J.; Bahcheli, A. T.; Slattery, S. S.; Patel, R. R.; Browne, T. S.; Flatley, M.; Karas, B. J.; Edgell, D. R.; Gloor, G. B. Telomere-to-telomere genome assembly of *Phaeodactylum tricornutum*. *PeerJ* **2022**, *10*, No. e13607.
(42) De Riso, V.; Raniello, R.; Maumus, F.; Rogato, A.; Bowler, C.; Falciatore, A. Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* **2009**, *37*, No. e96.
(43) Harada, H.; Nakatsuma, D.; Ishida, M.; Matsuda, Y. Regulation of the expression of intracellular beta-carbonic anhydrase in response to CO2 and light in the marine diatom Phaeodactylum tricornutum. *Plant Physiol.* **2005**, *139*, 1041−1050.
(44) Chu, L.; Ewe, D.; Río Bártulos, C.; Kroth, P. G.; Gruber, A. Rapid induction of GFP expression by the nitrate reductase promoter in the diatom *Phaeodactylum tricornutum*. *PeerJ* **2016**, *4*, No. e2344.
(45) Sharma, A. K.; Nymark, M.; Sparstad, T.; Bones, A. M.; Winge, P. Transgene-free genome editing in marine algae by bacterial conjugation - comparison with biolistic CRISPR/Cas9 transformation. *Sci. Rep.* **2018**, *8*, 14401.
(46) Walhout, A. J. M.; Temple, G. F.; Brasch, M. A.; Hartley, J. L.; Lorson, M. A., van den Heuvel, S.; Vidal, M. [34] GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. *Methods in Enzymology*; Thorner, J., Emr, S. D., Abelson, J. N., Eds.; Academic Press, 2000; p 575.
(47) Devaki, B.; Grossman, A. R. Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* **1993**, *21*, 4458−4466.
(48) Gibson, D. G.; Benders, G. A.; Axelrod, K. C.; Zaveri, J.; Algire, M. A.; Moodie, M.; Montague, M. G.; Venter, J. C.; Smith, H. O.; Hutchison, C. A., 3rd One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20404−20409.

https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX

(49) Strand, T. A.; Lale, R.; Degnes, K. F.; Lando, M.; Valla, S. A New and Improved Host-Independent Plasmid System for RK2-Based Conjugal Transfer. *PLoS One* **2014**, *9*, No. e90372.
(50) Bowler, C.; Allen, A. E.; Badger, J. H.; Grimwood, J.; Jabbari, K.; Kuo, A.; Maheswari, U.; Martens, C.; Maumus, F.; Otillar, R. P.; Rayko, E.; Salamov, A.; Vandepoele, K.; Beszteri, B.; Gruber, A.; Heijde, M.; Katinka, M.; Mock, T.; Valentin, K.; Verret, F.; Berges, J. A.; Brownlee, C.; Cadoret, J.-P.; Chiovitti, A.; Choi, C. J.; Coesel, S.; De Martino, A.; Detter, J. C.; Durkin, C.; Falciatore, A.; Fournet, J.; Haruta, M.; Huysman, M. J. J.; Jenkins, B. D.; Jiroutova, K.; Jorgensen, R. E.; Joubert, Y.; Kaplan, A.; Kröger, N.; Kroth, P. G.; La Roche, J.; Lindquist, E.; Lommer, M.; Martin−Jézéquel, V.; Lopez, P. J.; Lucas, S.; Mangogna, M.; McGinnis, K.; Medlin, L. K.; Montsant, A.; Secq, M.-P. O.; Napoli, C.; Obornik, M.; Parker, M. S.; Petit, J.-L.; Porcel, B. M.; Poulsen, N.; Robison, M.; Rychlewski, L.; Rynearson, T. A.; Schmutz, J.; Shapiro, H.; Siaut, M.; Stanley, M.; Sussman, M. R.; Taylor, A. R.; Vardi, A.; von Dassow, P.; Vyverman, W.; Willis, A.; Wyrwicz, L. S.; Rokhsar, D. S.; Weissenbach, J.; Armbrust, E. V.; Green, B. R.; Van de Peer, Y.; Grigoriev, I. V. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **2008**, *456*, 239−244.
(51) McFadden, G. I. Primary and secondary endosymbiosis and the origin of plastids. *J. Phycol.* **2001**, *37*, 951−959.
(52) Diner, R. E.; Noddings, C. M.; Lian, N. C.; Kang, A. K.; McQuaid, J. B.; Jablanovic, J.; Espinoza, J. L.; Nguyen, N. A.; Anzelmatti, M. A.; Jansson, J.; Bielinski, V. A.; Karas, B. J.; Dupont, C. L.; Allen, A. E.; Weyman, P. D. Diatom centromeres suggest a mechanism for nuclear DNA acquisition. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E6015−E6024.
(53) Potter, S. C.; Luciani, A.; Eddy, S. R.; Park, Y.; Lopez, R.; Finn, R. D. HMMER web server: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W200−W204.
(54) Allen, A. E.; Dupont, C. L.; Oborník, M.; Horák, A.; Nunes-Nesi, A.; McCrow, J. P.; Zheng, H.; Johnson, D. A.; Hu, H.; Fernie, A. R.; Bowler, C. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **2011**, *473*, 203−207.
(55) Erdene-Ochir, E.; Shin, B.-K.; Huda, M. N.; Kim, D. H.; Lee, E. H.; Song, D.-G.; Kim, Y.-M.; Kim, S. M.; Pan, C.-H. Cloning of a novel endogenous promoter for foreign gene expression in *Phaeodactylum tricornutum*. *Appl. Biol. Chem.* **2016**, *59*, 861−867.
(56) Barrett, L. W.; Fletcher, S.; Wilton, S. D. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.* **2012**, *69*, 3613−3634.
(57) Bielinski, V. A.; Brunson, J. K.; Ghosh, A.; Moosburner, M. A.; Garza, E. A.; Fussy, Z.; Bai, J.; McKinnie, S. M. K.; Moore, B. S.; Allen, A. E.; Almo, S. C.; Dupont, C. L. The *Phaeodactylum tricornutum* Diaminopimelate Decarboxylase Was Acquired via Horizontal Gene Transfer from Bacteria and Displays Substrate Promiscuity. *bioRxiv* **2020**.
(58) Bielinski, V. A.; Bolt, T. M.; Dupont, C. L.; Weyman, P. D. Episomal tools for RNAi in the diatom *Phaeodactylum tricornutum*. *PeerJ* **2017**, *5*, No. e2907v1.
(59) Shi, Y.-L.; Weiland, M.; Li, J.; Hamzavi, I.; Henderson, M.; Huggins, R. H.; Mahmoud, B. H.; Agbai, O.; Mi, X.; Dong, Z.; Lim, H. W.; Mi, Q.-S.; Zhou, L. MicroRNA expression profiling identifies potential serum biomarkers for non-segmental vitiligo. *Pigm. Cell Melanoma Res.* **2013**, *26*, 418−421.
(60) Wright, M. S.; Mountain, S.; Beeri, K.; Adams, M. D. Assessment of Insertion Sequence Mobilization as an Adaptive Response to Oxidative Stress in *Acinetobacter baumannii* Using IS-seq. *J. Bacteriol.* **2017**, *199*, 008333-16.
(61) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nat. Methods* **2020**, *17* (3), 261−272.
(62) Waskom, M. Seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6* (60), 3021.
(63) Espinoza, J. L.; Dupont, C. L.; O'Rourke, A.; Beyhan, S.; Morales, P.; Spoering, A.; Meyer, K. J.; Chan, A. P.; Choi, Y.; Nierman, W. C.; Lewis, K.; et al. Predicting antimicrobial mechanism-of-action from transcriptomes: A generalizable explainable artificial intelligence approach. *PLoS Comput. Biol.* **2021**, *17* (3), No. e1008857.

https://doi.org/10.1021/acssynbio.3c00163
ACS Synth. Biol. XXXX, XXX, XXX−XXX