



# Multiplexed Knockouts in the Model Diatom *Phaeodactylum* by Episomal Delivery of a Selectable Cas9

Mark Andrew Moosburner<sup>1,2\*</sup>, Pardis Gholami<sup>2</sup>, James K. McCarthy<sup>2</sup>, Maxine Tan<sup>1,2</sup>, Vincent A. Bielinski<sup>2</sup> and Andrew E. Allen<sup>1,2\*</sup>

<sup>1</sup> Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, United States, <sup>2</sup> J. Craig Venter Institute, La Jolla, CA, United States

## OPEN ACCESS

### Edited by:

Martin G. Klotz,  
Washington State University,  
United States

### Reviewed by:

Manuel Serif,  
Norwegian University of Science  
and Technology, Norway  
Peter Hegemann Peter,  
Humboldt University of Berlin,  
Germany  
Zahir Ali,  
King Abdullah University of Science  
and Technology, Saudi Arabia

### \*Correspondence:

Mark Andrew Moosburner  
mmoosbur@ucsd.edu  
Andrew E. Allen  
aallen@ucsd.edu

### Specialty section:

This article was submitted to  
Aquatic Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 09 April 2019

**Accepted:** 03 January 2020

**Published:** 28 January 2020

### Citation:

Moosburner MA, Gholami P,  
McCarthy JK, Tan M, Bielinski VA and  
Allen AE (2020) Multiplexed  
Knockouts in the Model Diatom  
*Phaeodactylum* by Episomal Delivery  
of a Selectable Cas9.  
*Front. Microbiol.* 11:5.  
doi: 10.3389/fmicb.2020.00005

Marine diatoms are eukaryotic microalgae that play significant ecological and biogeochemical roles in oceans. They also have significant potential as organismal platforms for exploitation to address biotechnological and industrial goals. In order to address both modes of research, sophisticated molecular and genetic tools are required. We presented here new and improved methodologies for introducing CRISPR-Cas9 to the model diatom *Phaeodactylum tricornutum* cells and a streamlined protocol for genotyping mutant cell lines with previously unknown phenotypes. First, bacterial-conjugation was optimized for the delivery of Cas9 by transcriptionally fusing Cas9 to a selectable marker by the 2A peptide. An episome cloning strategy using both negative and positive selection was developed to streamline CRISPR-episome assembly. Next, cell line picking and genotyping strategies, that utilize manual sequencing curation, TIDE sequencing analysis, and a T7 endonuclease assay, were developed to shorten the time required to generate mutants. Following this new experimental pipeline, both single-gene and two-gene knockout cell lines were generated at mutagenesis efficiencies of 48% and 25%, respectively. Lastly, a protocol for precise gene insertions via CRISPR-Cas9 targeting was developed using particle-bombardment transformation methods. Overall, the novel Cas9 episome design and improved genotyping methods presented here allow for quick and easy genotyping and isolation of *Phaeodactylum* mutant cell lines (less than 3 weeks) without relying on a known phenotype to screen for mutants.

**Keywords:** diatom, gene editing, CRISPR-Cas9, conjugation, multiplex, bioengineering, mutant, *Phaeodactylum*

## INTRODUCTION

Marine diatoms are of biotechnological significance for their capacity to assimilate large amounts of carbon and nitrogen, divide rapidly (Cermeño et al., 2011), thrive at high cell densities, and fix carbon via photosynthesis. Diatoms shunt nutrients toward metabolic processes that produce industrially valuable products such as biodiesel precursors (triglycerides a.k.a. TAGs), polyunsaturated fatty acids (PUFAs), and photosynthetic pigments (Pulz and Gross, 2004; Vinayak et al., 2015). Regardless of their innate ability to produce these high-valued molecules (HVMs),

**Abbreviations:** GA, Gibson assembly; GG, golden gate assembly; *Phaeodactylum*, *Phaeodactylum tricornutum*; RT, room temperature.

increasing production beyond their native capabilities is desirable for reducing production costs, and achieving sustainability goals in industrial agriculture applications (Gao, 2018). Utilization of genetic engineering technologies could shift metabolic outputs of marine diatoms toward these HVMs and improve commercial production where algae have been suggested as biological platforms for industrial feedstock and nutraceutical production (Vinayak et al., 2015). Nevertheless, the development and optimization of such genetic engineering technologies is necessary before they can be safely, easily, and robustly employed.

Genetic engineering technologies, particularly the versatile CRISPR-Cas9 toolbox, have been used for numerous applications in a vast array of organisms (Mali et al., 2013b; Sanders and Joung, 2014; Wang et al., 2016). CRISPR (clustered regularly interspaced short palindromic repeats), a foreign nucleic acid immunity system, is encoded in the genomes of most bacteria and some archaeal species (Deltcheva et al., 2011). The type-II CRISPR system from the bacterium *Streptococcus pyogenes* was first adapted for gene editing using the Cas9 enzyme, which induces blunt-ended double-stranded breaks (DSBs) when guided to a genetic target by a chimeric RNA molecule of the tracrRNA and crRNA called a single-guided RNA (sgRNA) (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013a). Together Cas9 and an sgRNA form a ribo-endonuclease complex *in vivo* in many cell types to edit or mutate genomic targets. In the bacterial immunity system, the sgRNA targeting sequence, a string of ~20 nucleotides (nt), preserves the sequence of a previous immune response, called spacer acquisition (McGinn and Marraffini, 2018); whereas in gene editing applications the nucleotide sequence of 20-nt spacer is designed by the researcher to target a specific gene/target, followed by interrogation and cleavage of a nucleic acid target by the Cas9 endonuclease. The sgRNA guides Cas9 to the target by forming complementary base pairing with the 20 nt at the 5' end of the sgRNA, called the spacer. Once a DSB is induced, the blunt ended breaks are repaired by one of two native DNA repair mechanisms: homology-directed repair (HDR) or non-homologous end joining (NHEJ). HDR relies on a homologous donor to repair the DSB, typically a sister chromosome in eukaryotes, by homologous recombination (LaFave and Sekelsky, 2009). Endogenous HDR can be hijacked in order to introduce, at the DSB, an exogenous donor sequence, containing a user-designed mutation that will result in an edited target locus. For NHEJ, ligation of the blunt ends repairs the DSB, however, this process is error-prone and leads to random insertions or deletions of nucleotides of unpredictable sizes called indels (Davis and Chen, 2013).

CRISPR-Cas9 has been employed in both prokaryotic and eukaryotic organisms but eukaryotes pose particular challenges for CRISPR-Cas9 experiments as most have multiple copies of their chromosomes. Multiple copies of chromosomes provide two or more target loci for Cas9. For instance, in a diploid organism, two individual DNA repair events must occur to produce homozygous or heterozygous mutations at the target loci. For NHEJ-mediated mutagenesis (where no homologous donor is introduced), a heterozygous genotype arises after two separate Cas9 cleavage events and NHEJ-mediated repair that results in two distinct indels at the target loci. A homozygous

genotype can also arise when one indel-containing target locus is used as a homologous donor for the second target loci (located on the sister chromosome of the first target), thereby producing an identical indel mutation. The unpredictability of NHEJ-mediated mutagenesis in diploid organisms therefore requires highly sensitive genotyping methodologies. Many genotyping methods have been used to genotype diploid organisms; among the most common are high-resolution melt curve analysis (HRMC) (Nymark et al., 2016), T7 endonuclease I assay (Guschin et al., 2010; Cho et al., 2013), high-throughput next generation sequencing (Zhou et al., 2014), and manual curation of sequencing reads.

Marine diatom species have been subjected to CRISPR-Cas9 mutagenesis including the model strains: the pennate, *Phaeodactylum tricornutum* (Nymark et al., 2016; Serif et al., 2018; Sharma et al., 2018; Slattery et al., 2018; Stukenberg et al., 2018) and the centric *Thalassiosira pseudonana* (Hopes et al., 2016). To date, NHEJ-mediated mutagenesis has been observed only in *Phaeodactylum* (Nymark et al., 2016; Serif et al., 2018; Slattery et al., 2018) by using one of two genetic transformation methods, micro-particle bombardment or bacterial conjugation, to introduce CRISPR-Cas9. Despite the successes of utilizing CRISPR-Cas9 in generating mutant diatom cell lines, genotyping the generated cell lines has been difficult. *Phaeodactylum*, a diploid organism, requires high-sensitivity screening methods to genotype multiple cell lines in a high-throughput manner. Genotyping methods such as high-throughput melt curve (HRMC) analysis and the T7 endonuclease I assay are commonly used to evaluate mutant genotypes in multi-ploidy organisms. To date, HRMC analysis and CRISPR Activity Factor (CAF) analysis have been used for *Phaeodactylum* genotyping experiments (Nymark et al., 2016; Sharma et al., 2018; Stukenberg et al., 2018). Similarly to CAF, TIDE (Tracking Indels by DEconvolution) sequencing analysis, an open-source software package, deconvolves raw sequencing reads to reveal indels in diploid or multiploidy organisms (Brinkman et al., 2014; Serif et al., 2018). Allele-specific PCR amplification has been used for genotyping *Phaeodactylum* mutants without the need for subcloning (Serif et al., 2018).

Here, TIDE sequencing analysis and manual sequencing curation is used to streamline the genotyping pipeline from colony formation to mutant cell line validation. Both single-gene and two-gene mutant cell lines were produced utilizing conjugation to introduce a selectable-Cas9 and either one or two sgRNAs to *Phaeodactylum*. Also, the T7 endonuclease I assay has also been adapted to quantify the mutagenesis efficiency of individual sgRNAs throughout a population of transformed *Phaeodactylum* prior to genotyping individual colonies.

A new selectable-Cas9 episome design was used that permits antibiotic selection for the Cas9 gene with the goal of both increasing mutagenesis efficiency and potentially avoiding mixed genotypes forming in generated colonies after conjugation. The P2A self-cleaving peptide was used to transcriptionally fuse Cas9 and the *Phaeodactylum* antibiotic resistant gene, *sh ble* (Kim et al., 2011). By doing so, Cas9 and *sh ble* are co-transcribed under the same promoter followed by cleavage of the translated products, essentially selecting for Cas9 via antibiotic treatment. Lastly, a

robust golden-gate assembly based method was developed to visualize positive and negative *E. Coli* clones that harbor ready-to-conjugate episomes. Overall, the methods presented here allow for easy Cas9-sgRNA episome assembly and streamlined genotyping methodology and isolation of *Phaeodactylum* mutant cell lines in less than 3 weeks without relying on a known phenotype to screen for mutants.

## MATERIALS AND METHODS

### Cell Culturing

A reference strain of *Phaeodactylum* (CCAP-1055/1) was used in all experiments. *Phaeodactylum* was grown at 18°C under white fluorescent lights (50 uE m<sup>-2</sup> s<sup>-1</sup>) and subjected to a diel growth cycle (14 h:10 h; light:dark). Culture medium was artificial sea water (ASW) supplemented with trace metals, essential vitamins, 55 uM of NaPO<sub>4</sub>, and 880 uM of the appropriate nitrogen source (NaNO<sub>3</sub>, NH<sub>4</sub>Cl, or Urea). Cultures, when not used in growth assays, were maintained with chloramphenicol antibiotic (10 mg/L) to keep cultures bacteria-free. Mutant *Phaeodactylum* strains were supplemented with either phleomycin (50 mg/L), zeocin (50 mg/L), or nourseothricin (200 mg/L).

*Escherichia coli* (TOP-10, Life Technologies, Carlsbad, CA) was used for all molecular cloning purposes. The cultures were grown on Luria-Bertani broth or agar and supplemented with the following antibiotics when necessary: ampicillin (100 mg/L), carbicillin (100 mg/L), tetracyclin (10 mg/L), gentamicin (20 mg/L), zeocin (25 mg/L).

### Molecular Biology

Plasmid construction was performed with reference to Sambrook and Russell (2001) unless stated otherwise. Multiple DNA polymerases were used for cloning purposes including Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA, United States), AccuPrime Taq High-Fidelity DNA polymerase (Thermo Fisher Scientific), OneTaq 2X Master Mix with standard buffer (New England Biolabs, Ipswich, MA, United States), and Phire Plant Direct PCR Master Mix (Thermo Fisher Scientific). Enzymatic components to make a 2X GA master mix that included Phusion HF DNA polymerase, T5 exonuclease (Thermo Fisher Scientific), and DNA Taq Ligase (Thermo Fisher Scientific) were purchased separately and mixed in lab. Enzymatic components for Golden-Gate cloning were also purchased separately but only mixed on the day of cloning. They included the type-II restriction enzymes *BsaI*-HF and *BbsI* (New England Biolabs) and a T4 DNA Ligase (Thermo Fisher Scientific).

Primer sequences can be found in the **Supplementary Material**.

### CRISPR/Cas9 Genetic Components

A human-codon optimized Cas9 open reading frame (ORF) containing three nucleus-localization signals (NLS) was produced by the Feng Zhang laboratory (Broad Institute, Cambridge, MA, United States) and purchased from addgene.com [pSpCas9(BB)-2A-Puro (PX459) V2.0, ID:

62988, Ran et al., 2013]. The Cas9 ORF was cloned into a pUC19 backbone using GA cloning (Gibson et al., 2009) using primers Cas9-2A-shble\_1 and P2A\_1 (**Supplementary Material**). For constitutive expression of Cas9, the FcpB/FcpA (Fucoxanthin chlorophyll *a/c* promoter B/Fucoxanthin chlorophyll *a/c* terminator A) (Bailleul et al., 2010) promoter/terminator pair were chosen (FcpB-Cas9).

The HDR-donor plasmid used in Weyman et al. (2014) for TALEN mutagenesis was adapted to a HDR-donor plasmid for targeting *nitrate reductase* (pKO-NR). pKO-NR contained 1 kb homology arms flanking an expression cassette coding for the *sh ble* gene that conferred resistance to the antibiotic phleomycin. The cassette inserted both a *sh ble* cassette and a STOP codon upstream of the *sh ble* promoter (Fucoxanthin chlorophyll *a/c* binding promoter F, FcpF). This was in frame with the targeted coding gene. The primers and plasmid sequence information can be found in Weyman et al. (2014).

The sRNAi promoter (chr2: 28,038-29,083) and terminator (chr2: 29,124-29,424) was amplified from *Phaeodactylum* genomic DNA (promoter primers = sRNAi\_1 and sRNAi\_2; terminator primers = sRNAi\_3 and sRNAi\_4). For Golden Gate cloning, the sRNAi terminator was shortened by PCR to 112 bp to eliminate an internal *BsaI* restriction site. The sRNAi promoter and terminator was cloned from *Phaeodactylum* genomic DNA at the locus of the most highly expressed small, non-coding, intergenic RNA called sRNAi (Rogato et al., 2014; Bielinski et al., 2017). The sequences for the sRNAi promoter and terminator can be found in the **Supplementary Material**.

The sgRNA backbone, which consists of the genetic fusion of the crRNA and tracrRNA (Jinek et al., 2012), was amplified from the same Addgene plasmid that contained Cas9 (primers = sRNAGU1\_F and sRNAGU1\_R). An sgRNA expression cassette for *Phaeodactylum* (sRNAi\_gBbsI) was constructed using a 4-piece GA approach into an *XbaI*-digested pUC19 plasmid containing a bacterial expression cassette for the *bla* gene conferring resistance to ampicillin/carbicillin. The resulting plasmid was then PCR amplified at the 3' end of the sRNAi and the 5' end of the sgRNA backbone (primers = gRNA\_1 and gRNA\_2). This amplicon served as a linearized vector for a 2-piece GA procedure for a spacer insert that contains two flanking *BbsI* restriction sites (**Supplementary Figure S2**). The 100 bp GA insert was constructed by ligating two 60 bp oligos with homology to the sRNAi promoter and sgRNA backbone. The resulting plasmid, sRNAi\_g2XBbsI, was used for *BbsI*-based GG cloning of CRISPR spacers (**Supplementary Figure S2B**).

For all Cas9 target genes, spacer sequences were obtained for *nitrate reductase* (*Phatr3\_J54983*), *glutamine synthetase II* (*GS-2*) (*Gene ID: Phatr3\_J51092*), and *chloroplast glutamate synthase* (*cGOGAT*) (*Gene ID: Phatr3\_J24739*) using the Phyto CRISPR-Ex online package (Rastogi et al., 2016) that passed the built-in criteria of minimizing off-target Cas9 activity.

A GG cloning protocol was modified from the sgRNA expression cassette supplied by Feng Zhang laboratory. In place of the *BbsI*-containing spacer sequence within the sgRNA expression plasmid, a LacZ bacterial expression cassette with flanking *BbsI* sites was cloned (primers = LacZ\_GG\_1 and LacZ\_GG\_2) into the sRNAi expression cassette using GA. The Pt



sgRNA expression plasmid previously produced (sRNAi\_g*BbsI*) was PCR amplified (primers = LacZ\_GG\_3 and LacZ\_GG\_4) and the amplicon was gel extracted and used as a GA vector. The sgRNA cloning vector (sRNAi\_gLacZ) was the assembled plasmid. This plasmid was used in all future sgRNA expression cassette cloning and assembly experiments as a cloning vector (**Supplementary Figure S1B**). An appropriate GG insert for sRNAi\_gLacZ was assembled by ligating two 56-nt oligos together that are complimentary to each other. The oligo primer sequences for all sgRNA target sequences can be found in **Supplementary Material**.

Detailed protocols for spacer cloning into the sRNAi\_gLacZ plasmid (name GG1 cloning) are available at protocols.io (doi: 10.17504/protocols.io.4abgsan).

### Cas9-Episomal Plasmid Cloning

The *Phaeodactylum* episomal plasmid pPtPBR1 was obtained from R. Diner (Diner et al., 2016). First, one *BsaI* restriction site within the *bla* coding region of pPtPBR1 was removed by PCR amplification (primers = pBR322-Amp\_1 and pBR322-Amp\_2) and reassembly to produce plasmid pPtPBR1 (*BsaI*-). pPtPBR1 was then used as a template for GA cloning of a Cas9 ORF and a GG cloning *BsaI* acceptor site for sgRNA expression cassette cloning (**Supplementary Figure S2**). Cas9 was PCR amplified twice using one forward primer (Cas9-2A-sh ble\_1) and two reverse primers (P2A\_1 and P2A\_2) so to add GA homology to the FcpF promoter (5' of Cas9) and the sh ble ORF (3' of Cas9). The double PCR reaction added a GSG-2A peptide DNA coding sequence to transcriptionally fuse hCas9 and sh ble (**Supplementary Figure S2**, GSG-P2A nucleotide sequence is in **Supplementary Material**). Next, a red-fluorescent protein bacterial expression cassette (mRFP) was obtained and PCR amplified (primers = Cas9-2A\_sh ble-RFP\_1 and 2) to produce GG acceptor sites. The remaining two GA pieces were PCR amplified from pPtPBR1 (*BsaI*-) to insert two GG cloning *BsaI* sites between the *bla* and *tet* expression cassettes and flanking mRFP, one using primers Cas9-2A-sh ble\_2 and 3 (5' - sh ble - Pt\_Centromere - pMB1 - bla - 3') and two using primer Cas9-2A-sh ble\_6 and 7 (5' - tet - OriT - FcpF - 3'). After cloning and *E. coli* transformation, the correctly assembled episome, pBR\_Cas9-2A-sh ble, 2XBsaI-mRFP(AT) (**Supplementary Figure S3A**), was sequence verified.

Next, primers were designed to PCR amplify up to six sgRNA expression cassettes (cloning and construction described above) and to add *BsaI*-restriction and cleavage sites to the 5' and 3' ends of the sRNAi promoter/terminator. GG cloning is an optimal assembly platform when cloning more than one sgRNA expression cassette. For inserting one sRNAi\_sgRNA cassette, primers GG-gRNA1-F and sgRNA(GG)-2 were used. For two sgRNAs, primers GG-gRNA1-F and R were used for sgRNA #1 and primers GG-gRNA3-F and sgRNA(GG)-2 were used for sgRNA #2. Between the two sgRNAs was cloned a LacZ expression cassette, which, when induced with IPTG activates beta-galactidose that digests Xgal to produce blue *E. coli* colonies (**Supplementary Figure S3C**). The sgRNA expression cassette(s) and LacZ amplicons was mixed at double the molar

concentration of the *BsaI*-digested pBR\_Cas9-2A-sh ble, 2XBsaI-mRFP(AT) vector and carried through the GG cloning reaction with *BsaI*-HF and a T4-ligase (**Supplementary Figure S1C**). The final sgRNA-LacZ-sgRNA episomal plasmid map can be seen in **Supplementary Figure S3B**.

Detailed protocols for sgRNA cloning into the Cas9-2A-sh ble episome (named GG2 cloning) is available at protocols.io (doi: 10.17504/protocols.io.4acgsaw).

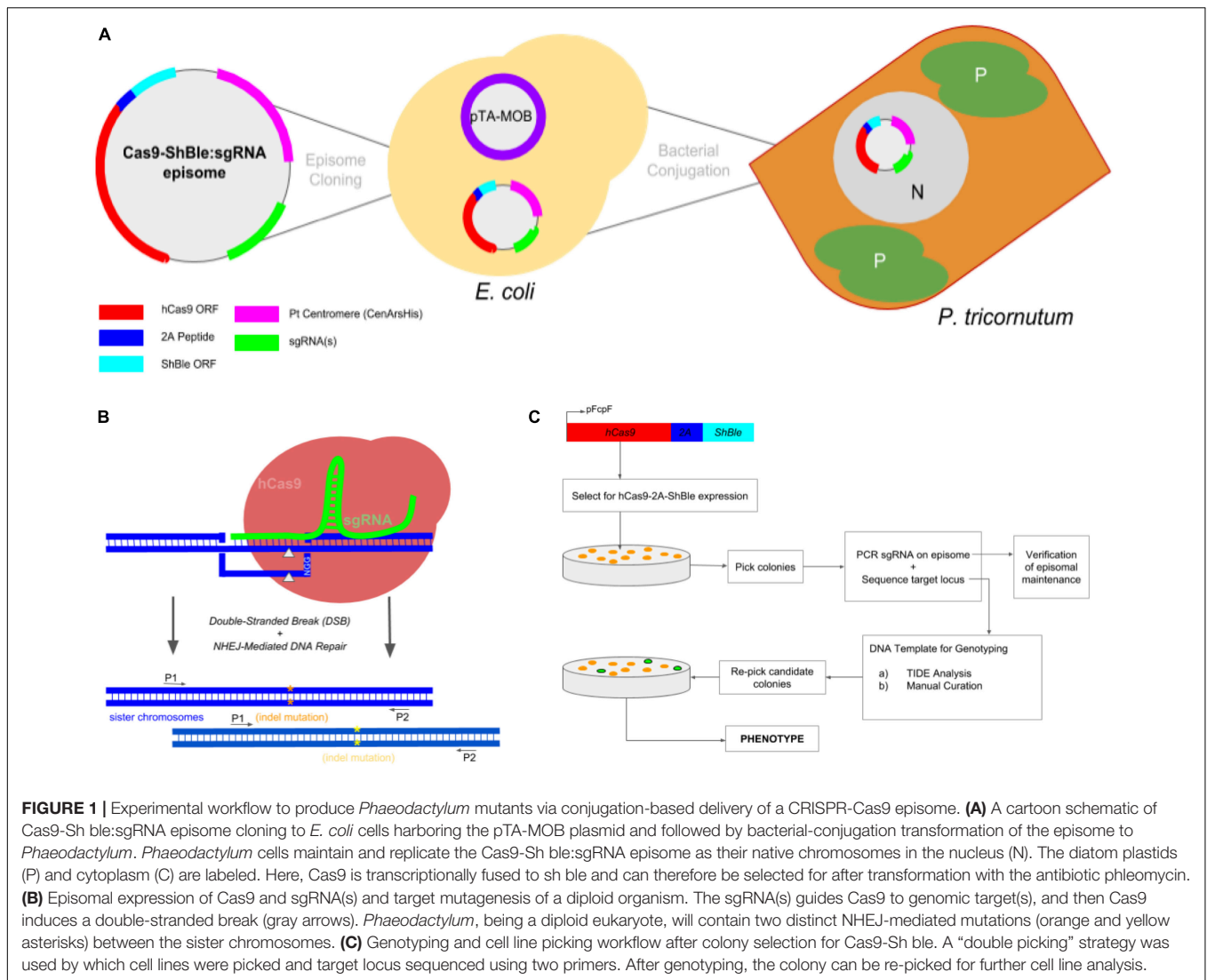
## *Phaeodactylum tricornutum* Genetic Transformations

### Bacterial-Mediated Conjugation

Bacterial-mediated conjugation was used to introduce CRISPR components to *Phaeodactylum*. Diner et al., 2016 provided detailed methodology for introducing episomal plasmids to *Phaeodactylum* in a high-throughput manner. For all transformations, 100  $\mu$ L of dense *Phaeodactylum* (1e8 cells/mL) was plated on conjugation-based solid agar medium (NH<sub>4</sub>/NO<sub>3</sub>-ASW, 1% agar, 5% LB in NH<sub>4</sub>-supplemented ASW) in 6-well cell culture plates and incubated for 4 days under light (18:6) and at 18°C. Prior to transformation, pBR\_Cas9-2A-sh ble, sRNAi-sgRNA episomes were transformed into recipient *E. coli* harboring the conjugation plasmid pTA-MOB (Karas et al., 2015; Diner et al., 2016) and selected on agar plates for both the cargo and conjugation plasmids (**Figure 1**). For a transformation control (positive for colonies not Cas9 cleavage), the episome pBR\_Cas9-2A-sh ble, 2XBsaI\_gRFP1 was used where Cas9 expression would still occur without an accompanying sgRNA. For a negative control, a similar episomal plasmid was built harboring the NAT gene that confers resistance to nourseothricin and not phleomycin. Under phleomycin selection, the transformed cells would die. Transformation efficiency for the delivery of Cas9-2A-sh ble:sgRNA episomes was 500–750 transformants per  $1 \times 10^8$  phleomycin-resistant cells.

The day before transformation, 4 random colonies from the selection plates were used to inoculate 3 mL of LB containing selection (Amp, Tet, Gent). After growing overnight, 200  $\mu$ L of culture was used to inoculate 12.5 mL of LB. The culture was grown for 3–4 h, until an OD-600 nm of 0.8–1 was reached, pelleted, and suspended in 100 $\mu$ L of SOC media. Then, the *E. coli* was pipetted into a corresponding well containing *Phaeodactylum* and mixed either by a spreading loop or by rotating the 6-well plate such that the *E. coli* covered the entire lawn of *Phaeodactylum*. After *E. coli* was added to all wells and allowed to dry under a vacuum PCR hood, the 6-well plate was placed in the dark at 30°C for 90 min. The plate was then placed back at 18°C and exposed to light for 48 h.

Finally, each well was scraped using a sterile cell spreader, collected in a 2 mL Epi tube then re-plated on 100 mm selection plates containing the appropriate selective media. The selection plates were then allowed to grow for 10–14 days or until colonies appeared. For the T7 Endonuclease I Assay reaction, half of the scraped *Phaeodactylum* cell lines (~400  $\mu$ L) was resuspended in 200 mL NH<sub>4</sub>-supplemented ASW media supplemented with phleomycin and chloramphenicol. The cells were then grown under light and at 18°C for 4 days, passed into 200 mL fresh



**FIGURE 1 |** Experimental workflow to produce *Phaeodactylum* mutants via conjugation-based delivery of a CRISPR-Cas9 episome. **(A)** A cartoon schematic of Cas9-Sh ble:sgRNA episome cloning to *E. coli* cells harboring the pTA-MOB plasmid and followed by bacterial-conjugation transformation of the episome to *Phaeodactylum*. *Phaeodactylum* cells maintain and replicate the Cas9-Sh ble:sgRNA episome as their native chromosomes in the nucleus (N). The diatom plastids (P) and cytoplasm (C) are labeled. Here, Cas9 is transcriptionally fused to sh ble and can therefore be selected for after transformation with the antibiotic phleomycin. **(B)** Episomal expression of Cas9 and sgRNA(s) and target mutagenesis of a diploid organism. The sgRNA(s) guides Cas9 to genomic target(s), and then Cas9 induces a double-stranded break (gray arrows). *Phaeodactylum*, being a diploid eukaryote, will contain two distinct NHEJ-mediated mutations (orange and yellow asterisks) between the sister chromosomes. **(C)** Genotyping and cell line picking workflow after colony selection for Cas9-Sh ble. A “double picking” strategy was used by which cell lines were picked and target locus sequenced using two primers. After genotyping, the colony can be re-picked for further cell line analysis.

media, grown for an additional 3 days, then finally pelleted and flash frozen.

Detailed protocols for bacterial-conjugation in *Phaeodactylum* are publicly available at protocols.io (doi: 10.17504/protocols.io.5pvg5n6).

### Micro-Particle Bombardment Genetic Transformation

Micro-particle bombardment genetic transformation (Falciatore et al., 1999) was used to introduce CRISPR components to *Phaeodactylum*. First, *Phaeodactylum* was pelleted during exponential growth in liquid culture (NO<sub>3</sub>-supplemented ASW), suspended, and plated onto agar growth plates at a cell concentration of 3e8 cells/mL (400 uL). Next, 24 ug of DNA (8 ug each plasmid; FcpB-Cas9, sRNAi-gNR-B, pKO-NR) were hybridized to tungsten beads and introduced to the plated *Phaeodactylum* at high velocity using the PDS-1000/He Biolistic® Particle Delivery System (Bio-Rad, Hercules, CA). The transformed *Phaeodactylum* plates were then allowed to recover for 48 h at 18°C and in constant darkness. After

recovery, *Phaeodactylum* was re-plated on agar plates containing NH<sub>4</sub>-supplemented ASW with *phleomycin* to select for the presence of the pKO-NR plasmid. The selection plates were then grown under normal conditions for 21–28 day, or until *Phaeodactylum* colonies formed and were visible. Experimental controls were also performed where either FcpB-Cas9 and pKO-NR were co-transformed or pKO-NR was transformed alone.

### T7 Endonuclease I Assay for Determining Cas9 Mutagenesis Efficiency

The T7 endonuclease I Assay (New England Biolabs) was used to quantify the *in vivo* cleavage efficiency for multiple clonal *Phaeodactylum* populations for each sgRNA target. Frozen cell pellets obtained following liquid selection (as described above) were slowly thawed on ice. Genomic DNA was then extracted from the pelleted cells using the Plant DNAzol Reagent (Thermo Fisher Scientific). The product-supplemented protocol was used; however, the pulverization step, the first step, was skipped because diatom pellets can be easily resuspended. Genomic DNA

was then extracted from the cell pellet. 200  $\mu$ L of genomic DNA at concentration ranging from 200 to 500 ng/ $\mu$ L was recovered. A total of 200 ng of genomic DNA was used as a PCR template to start the T7 endonuclease I assay. Six separate amplifications were performed (only 25X amplification cycles) and pooled prior to the T7 assay. For the *NR* locus, primers NR-gene-1 and NR-KO-2\* were used. This amplification product was then re-amplified with nested primers NR-HRM-A and NR-HRM-B. The nested-amplicons were then PCR cleaned and concentrated and used as a DNA input for the heteroduplex hybridization. This input is referred to as “Amplicon” in **Figure 2** and the hybridized heteroduplex is referred to as “Heteroduplex.” T7 endonuclease I was then added to the hybridized product to cleave all heteroduplex DNA. All reaction conditions and efficiency calculations (as shown in **Figure 2**) were supplied by New England Biolabs (protocol website<sup>1</sup>). Genome editing efficiency was determined using the Agilent 4200 TapeStation bioanalyzer and Agilent High-Sensitivity D1000 screen chip and assay. The fraction cleaved value was determined by examining the peak intensities of the high-sensitivity gel of the T7 product compared to the uncut Heteroduplex product. Fraction cleaved was then an input for the following equation to calculate the percentage of gene modification:

$$\% \text{ gene modification} = 100 * (1 - (1 - \text{fraction cleaved})^{1/2}).$$

## Genotyping *Phaeodactylum* Exconjugants

*Phaeodactylum* cells selected on solid media were grown for 10–14 day or until visible colonies formed. Colonies were picked with pipette tips directly into 5 $\mu$ L dilution buffer supplied with the Direct Plant PCR Kit (Thermo Fisher Scientific) and incubated at RT for 15 min. The buffered colony was then pipette mixed and 2 $\mu$ L were used as a DNA template for two PCR reactions. The first, common for all sgRNAs, used primers Ars\_Seq and V2\_insert\_2 (**Supplementary Figure S4A**, black arrows) to amplify from the Pt\_Centromere to and through the sgRNA expression cassette to check for correct episomal maintenance and sgRNA presence (**Supplementary Figure S4B**). For *nitrate reductase*, primers NR-gene-1 and NR-KO-2\* were used to amplify a 641 bp region around the target loci for gNR-A and gNR-B. The Direct Plant PCR Kit was used for both PCR reactions for 40 cycles. The resulting PCR amplicons were used for genotyping.

## TIDE Sequencing Analysis

TIDE (Tracking of Indels by DEconvolution, Brinkman et al., 2014) sequence analysis is an online-based software<sup>2</sup> that analyzes raw sequencing data to find one or multiple indels mutations at a Cas9 target site. For the purposes of this study, TIDE was used to analyze indels resulting from Cas9 targeting of *nitrate reductase* in *Phaeodactylum* by sgRNAs gNR-A and gNR-B (**Figure 3**). An example of the TIDE analysis output for a

wild-type sequence is shown in **Figure 3A**. 80 colonies for gNR-A and 64 colonies for gNR-B were picked directly into PCR mix (protocol described above). 200 ng of the PCR amplicons were then directly sequenced using two primers, NR-HRM-A and NR-gene-2, for paired-end reads (**Figure 2A**). *Phaeodactylum* transformed with the Cas9-2A-sh ble vector without an sgRNA was used as a control. TIDE analysis requires the input of (1) the nuclease species used (SpCas9), (2) the 20 nt spacer target, (3) an ABI (.abi) file for a reference sequence (control Cas9-2A-sh ble), and (4) an ABI (.abi) file for a sample sequence. The sample sequence is compared to the control sequence for significant differences between the sequence electropherograms. Certain criteria were identified and selected for when analyzing the TIDE output. First, a total efficiency value (in percentage) of 50% was chosen because it signified that the target locus was mutated at a high rate. Second, the *p*-value of each output peak must be <0.001. Third, only colonies with distinct genotypes, based on the criteria of only 1 or 2 predominant indels each with greater than 10% of the total sequencing reads, were selected as candidate mutant lines. Lastly, for the cell lines that passed each of these criteria, results were verified by conducting TIDE analysis of the paired-end read of the same sample. If both paired-end TIDE analyses pass the criteria and identified identical indels, the sample was chosen as a mutant cell line candidate. Each sequencing read was manually analyzed, individually, using the open-sourced TIDE web-tool.

## Nitrate Phenotype Assay for *Phaeodactylum* Mutant Cell Line Candidates

NR-KO *Phaeodactylum* cell lines were identified by TIDE and subjected to the NR phenotype assay where *Phaeodactylum* mutants of *nitrate reductase* are grown on media supplemented with nitrate and ammonium or urea, separately. The same NR assay was used for mutant cell lines produced via micro-particle bombardment. Mutants with a dysfunctional NR gene are unable to grow on nitrate media while proliferating on ammonium. First, all 40 colonies (for each sgRNA) were re-picked into liquid media supplemented with ammonium. The cultures (including four control cultures) were grown under normal conditions until mid-log phase and then passed into nitrate and ammonium media at a concentration of 1e4 cells/mL. Cell density was measured daily (2:00pm PST) for 4 days and growth between the two condition was compared. The cell lines that could grow on ammonium-supplemented ASW and could not grow on nitrate-supplemented ASW were considered NR mutant cell lines.

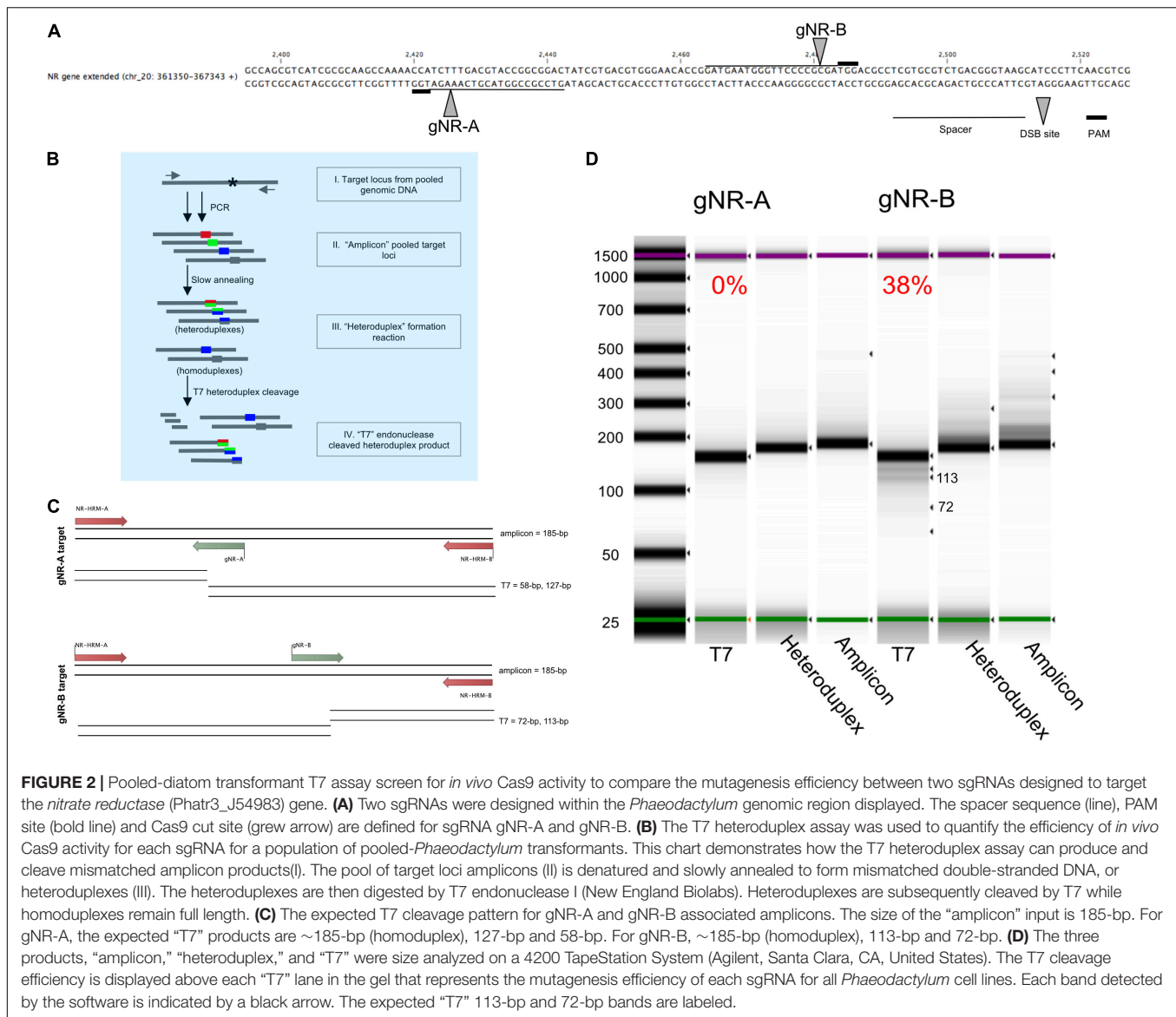
## RESULTS

### CRISPR-Cas9 Mutagenesis Workflow in *Phaeodactylum* Using Selectable-Cas9 Episome

The general workflow from episomal exchange to *Phaeodactylum* colony genotyping is shown in **Figure 1**. The *Phaeodactylum* episome was built with a transcriptionally fused Cas9 and sh

<sup>1</sup><https://www.neb.com/protocols/2014/08/11/determining-genome-targeting-efficiency-using-t7-endonuclease-i>

<sup>2</sup><https://tide.deskgen.com/>

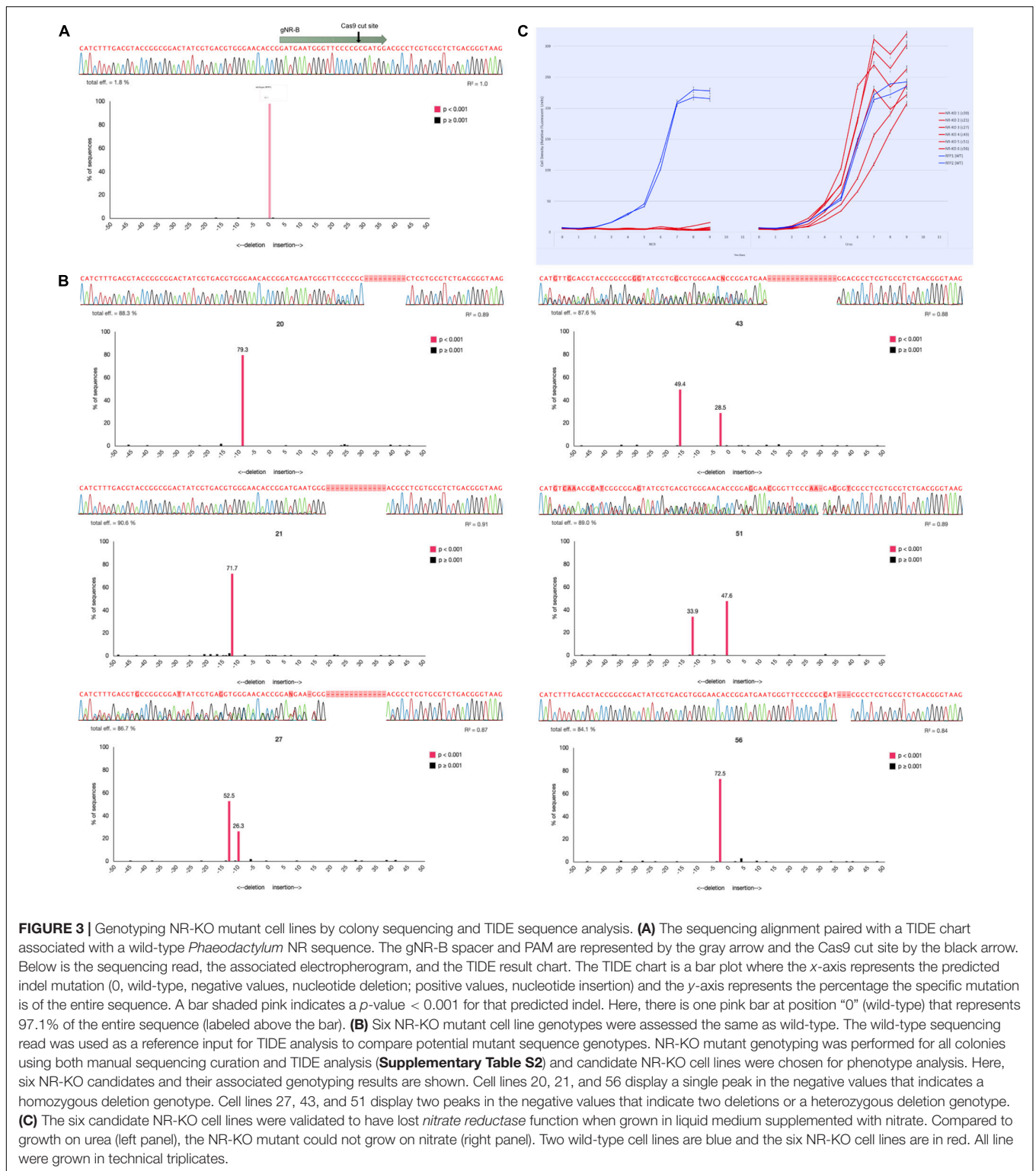


**FIGURE 2 |** Pooled-diatom transformant T7 assay screen for *in vivo* Cas9 activity to compare the mutagenesis efficiency between two sgRNAs designed to target the *nitrate reductase* (Phatr3\_J54983) gene. **(A)** Two sgRNAs were designed within the *Phaeodactylum* genomic region displayed. The spacer sequence (line), PAM site (bold line) and Cas9 cut site (grew arrow) are defined for sgRNA gNR-A and gNR-B. **(B)** The T7 heteroduplex assay was used to quantify the efficiency of *in vivo* Cas9 activity for each sgRNA for a population of pooled-*Phaeodactylum* transformants. This chart demonstrates how the T7 heteroduplex assay can produce and cleave mismatched amplicon products (I). The pool of target loci amplicons (II) is denatured and slowly annealed to form mismatched double-stranded DNA, or heteroduplexes (III). The heteroduplexes are then digested by T7 endonuclease I (New England Biolabs). Heteroduplexes are subsequently cleaved by T7 while homoduplexes remain full length. **(C)** The expected T7 cleavage pattern for gNR-A and gNR-B associated amplicons. The size of the "amplicon" input is 185-bp. For gNR-A, the expected "T7" products are ~185-bp (homoduplex), 127-bp and 58-bp. For gNR-B, ~185-bp (homoduplex), 113-bp and 72-bp. **(D)** The three products, "amplicon," "heteroduplex," and "T7" were size analyzed on a 4200 TapeStation System (Agilent, Santa Clara, CA, United States). The T7 cleavage efficiency is displayed above each "T7" lane in the gel that represents the mutagenesis efficiency of each sgRNA for all *Phaeodactylum* cell lines. Each band detected by the software is indicated by a black arrow. The expected "T7" 113-bp and 72-bp bands are labeled.

ble, an antibiotic selectable marker gene, expression cassette that allows Cas9 to be selected for after transformation with the eukaryote antibiotic phleomycin. Additionally, a red-fluorescent protein bacterial expression cassette (mRFP) was cloned into the episome with flanking *BsaI* restriction enzyme sites that, when digested, leave unique 4-bp overhangs for golden-gate cloning of sgRNA constructs (**Supplementary Figure S1**). A hierarchical cloning strategy was used to build and clone one *Phaeodactylum* sgRNA expression cassette in two golden-gate assembly steps (**Supplementary Figure S2**). Negative selection against RFP was used for cloning an individual sgRNA expression cassette where negative *E. coli* colonies were visibly red. For cloning more than one sgRNA, a LacZ bacterial expression cassette was cloned with and between the sgRNAs that turns *E. coli* colonies with positively assembled episomes blue in the presence of IPTG and Xgal. The red-white-blue screening method helped expedite episome cloning in general (**Supplementary Figure S3**).

After construction and sequence verification, a Cas9-sh ble:sgRNA episome was transformed into electro-competent *E. coli* cells harboring a pta-MOB plasmid (Karas et al., 2015; Diner et al., 2016) that has been shown to be required for efficient conjugation exchange of the episome to *Phaeodactylum* (**Figure 1A**). The episome then resides in the *Phaeodactylum* nucleus, separate from the *Phaeodactylum* chromosomes as an artificial chromosome, where expression of Cas9 and sgRNA(s) occurs. Like many CRISPR/Cas9 gene targeting experiments in organisms with diploid genomes, the Cas9-sgRNA ribo-endonuclease complex is guided to both gene alleles where distinct NHEJ-mediated mutation may occur (**Figure 1B**). While a heterozygous genotype (two NHEJ-mediated mutations) was expected, a homozygous genotype was also expected where one mutated allele can serve as an HDR donor template to repair the second allele (HDR) or where micro-homology loci upstream and downstream of the cut





site anneal following strand re-sectioning, referred to as MMEJ (Microhomology-Mediated End Joining; Wang and Xu, 2017).

Since both hetero- and homozygous mutations were expected, the genotyping workflow was conceived such that resulting colonies could be screened following selection for Cas9-2A-sh

ble expression (**Figure 1C**). Colonies were first picked and added directly to PCR mix to amplify the sgRNA expression cassette on the episome harbored within *Phaeodactylum* (**Supplementary Figure S4**). Only colonies that contained a correctly sized amplicon were chosen for genotyping. While



Cas9 expression was not verified for each individual colony, a Cas9-Venus (E-YFP) fusion was visualized by confocal microscopy to ensure localization in the *Phaeodactylum* nucleus (**Supplementary Figure S5**).

The methods described above were used for the following two applications: single-gene mutagenesis and two-gene mutagenesis.

## Single Gene Mutagenesis

Two sgRNAs were individually cloned into the Cas9-sh ble episome by golden-gate assembly (cloned episome product can be visualized in **Supplementary Figure S2A**). The sgRNAs were designed to mutate the *nitrate reductase* (NR) gene in *Phaeodactylum*. NR was picked as a genomic target because the function of NR has been well studied in *Phaeodactylum* and NR knockout cell mutants exhibit an easily screenable growth phenotype of cell death when supplemented with nitrate as a sole nitrogen source (McCarthy et al., 2017). The two sgRNAs were designed to target regions 55-bp apart so that the same amplification and sequencing primers could be used for genotyping efforts.

## Comparative sgRNA-Effectiveness Assessment of Cas9-Episome Transformed *Phaeodactylum* Population

To quantify the mutagenesis efficiency for both sgRNAs population-wide, the T7 endonuclease assay was used (**Figure 2**). While the T7 assay has been typically used to genotype a clonal cell line, this method was useful to quantify the mutagenesis efficiency of all cells after transformation and phleomycin-selection of Cas9-2A-sh ble. This was done with the intention to identify sgRNAs that may produce low rates of mutagenesis and subsequently discard them. The sgRNAs were designed to target the NR locus 55-bp apart (**Figure 2A**). This allowed both targeted populations (one for gNR-A and one for gNR-B) to be subjected to the same T7 assay parameters (**Figure 2B**). The T7 enzyme cleaves mismatched nucleotides and should cut at Cas9 target loci with NHEJ-mediated mutations. The expected band sizes after T7 digestion was 127-bp and 58-bp for gNR-A and 113-bp and 72-bp for gNR-B (**Figure 2C**). Homoduplexes, annealed bands that perfectly match during “heteroduplex” formation (**Figure 2B**), was expected to be observed in the T7 migration gel because biallelic knockouts that arise from MMEJ-mediated repair may, in theory, anneal to each other and avoid T7 cleavage. It was observed that gNR-A insufficiently targeted the NR locus (0% mutagenesis efficiency) while gNR-B mutated the NR locus at 38% (**Figure 2D**).

By employing this method prior to colony genotyping, sgRNAs with low rates of mutagenesis throughout the population can be ignored which saves on reagent costs and labor. Henceforth, gNR-A was discarded from future targeting experiments due to its low mutagenesis efficiency compared to gNR-B.

## Picking and Genotyping *Phaeodactylum*\_NR-KO Mutants

TIDE (Tracking Indels by DEconvolution) sequence analysis was used to genotype *Phaeodactylum* cell lines (**Figure 3**). To do so,

each sequence of the gNR-B target locus was first curated and analyzed manually. Next, one of two paired-end sequencing files were entered into the TIDE software and compared to a wild-type sequence for each colony. TIDE outputs a bar plot with predicted indel (insertion or deletion) mutations based on the inputted electropherogram compared to the wild-type. In **Figure 3A**, a wild-type NR sequence read is shown above the TIDE result plot that compares wild-type to wild-type. The TIDE plot displays one bar at *x*-axis position “zero” and with a *y*-axis value of 97.1% (displayed above bar). A pink colored bar also indicates the predicted indel was statistically significant ( $p < 0.001$ ). This result indicates that there is a zero-nucleotide indel mutation at the predicted Cas9 cut site that comprises 97.1% of the total sequencing reads. Clearly, the TIDE plot validates the sequencing read as a wild-type genotype.

**Figure 3B** shows six NR knockout mutant genotypes for cell lines 20, 21, 27, 43, 52, and 56. Cell line 20, for example, appears to have a bi-allelic mutation that is homozygous because the TIDE chart shows one pink peak that corresponds to a deletion mutation of 9-bp. Cell line 43, for example, appears to have a bi-allelic mutation that is heterozygous because the TIDE charts shows two pink peaks corresponding to a deletion genotype of 3-bp (28.5%) and 15-bp (49.4%). All six NR-KO genotypes display either homozygous or heterozygous genotypes by their TIDE charts. They also lack any peak at the “zero” position that would indicate that there was a wild-type contamination. Furthermore, cell lines with TIDE charts with only 1 or 2 indel peaks are shown here; additional peaks would indicate that the cell lines were not clonal and need to be re-streaked to find a clonal mutant.

Manual sequencing curation and TIDE analysis was performed for 80 gNR-A colonies and 64 gNR-B colonies. 2/80 (2.5%) gNR-A colonies and 26/64 (40%) gNR-B colonies contained a Cas9-mediated mutation. Here, the Cas9-sgRNA mutagenesis efficiencies of 0% and 28% calculated using T7 (**Figure 2**) reflects the efficiencies observed using TIDE.

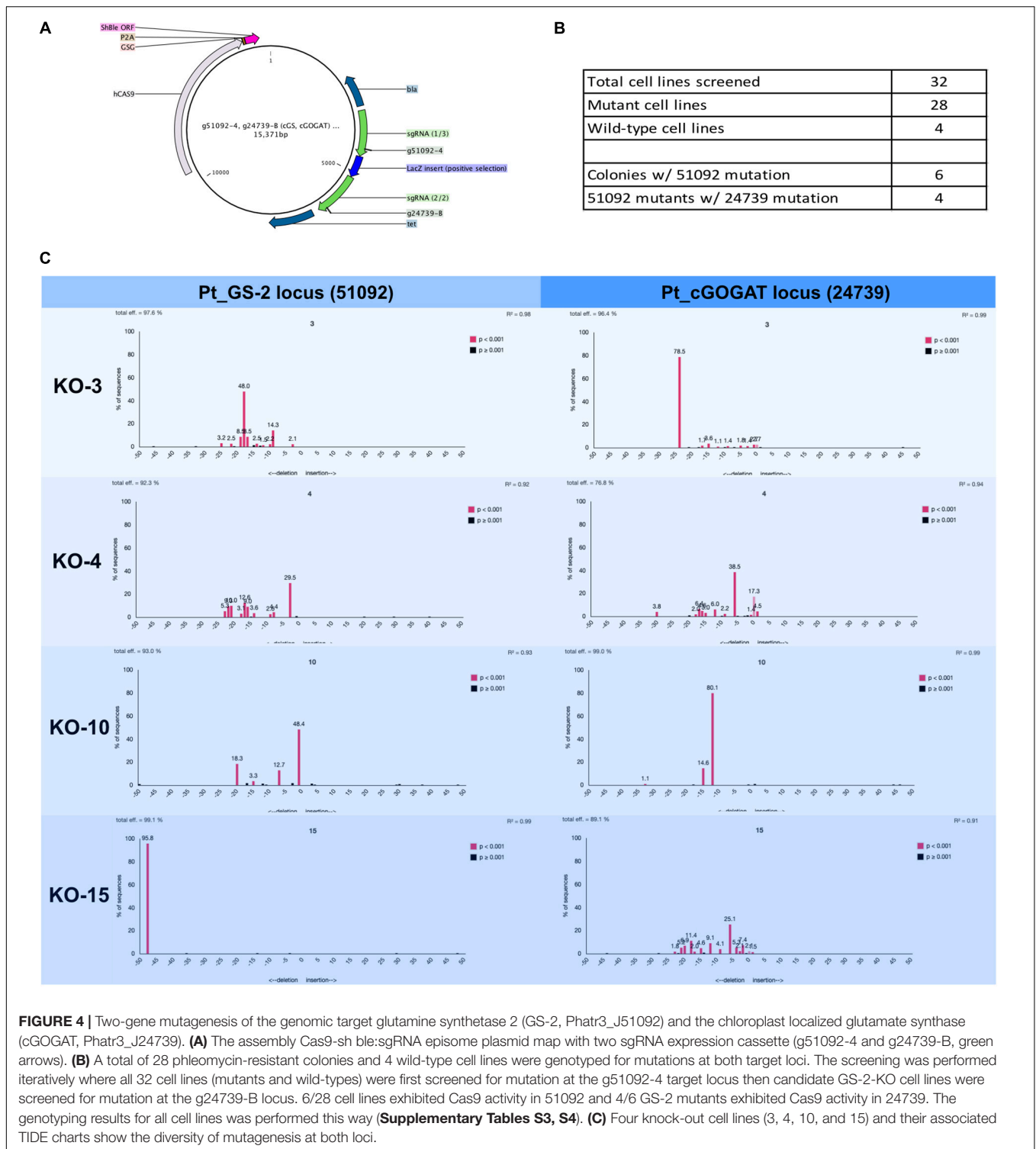
Lastly, the nitrate growth assay was used to correlate the TIDE genotyping efforts to the expected NR mutant phenotype. The six mutant cell lines identified using TIDE did not grow on nitrate media compared to growth on ammonium media (**Figure 3C**).

## Two-Gene Mutagenesis

The Cas9-2A-sh ble episome was assembled to harbor two sgRNA, one targeting glutamine synthetase 2 (GS-2, Gene ID: Phatr3\_J51092) and one targeting a chloroplast-localized glutamate synthase (cGOGAT, Phatr3\_J24739). The episome was assembled by golden-gate cloning and built to harbor both sgRNA expression cassette flanking a LacZ bacterial expression cassette (**Figure 4A** and **Supplementary Figure S3**).

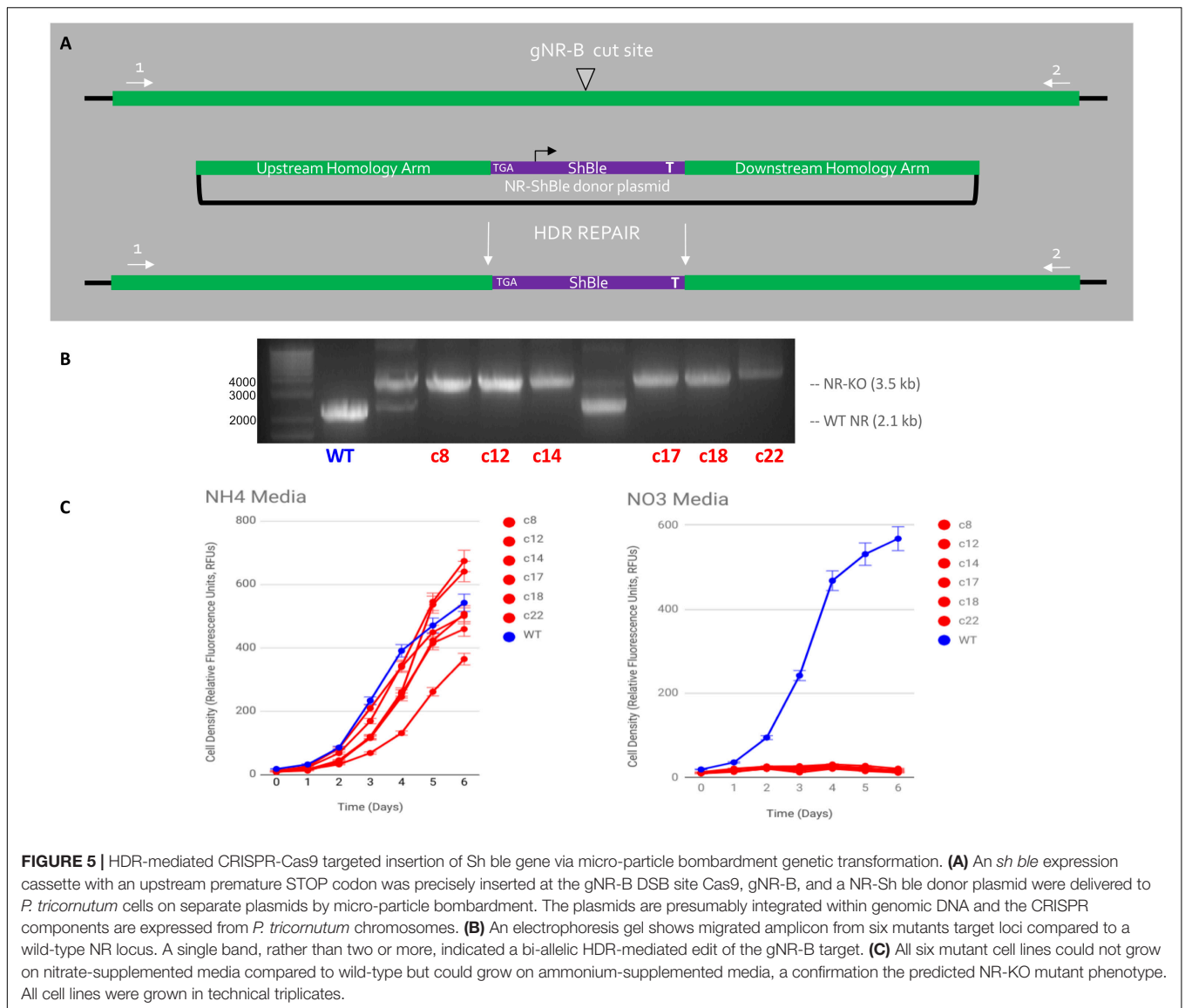
## TIDE Genotyping of *Phaeodactylum* Two-Gene Mutant Cell Lines

A total of 28 transformed *Phaeodactylum* colonies and 4 wild-type colonies were picked and genotyped using manual sequencing curation and TIDE sequencing analysis (**Figure 4**). The genotyping was performed successively. First, all 28 colonies were screened for mutation at the GS-2 target locus. Second,



GS-2 KO candidate colonies were screened for mutation at the cGOGAT target locus. Therefore, genotyping information is not provided for all 28 colonies for the cGOGAT target locus. The genotyping results using manual curation and TIDE analysis for each colony is provided in **Supplementary Table S3** (GS-2) and **Supplementary Table S4** (cGOGAT).

A total of 6/28 (21%) colonies contained a Cas9-mediated mutation(s) at the GS-2 target locus and 4/6 GS-2 mutant candidates also contained Cas9-mediated mutations at the cGOGAT target locus (**Figure 4B**). **Figure 4C** shows the TIDE plots for each double-knock-out cell line (3, 4, 10, and 15) for both target loci (GS-2 on the left, cGOGAT on the right).



Some mutations appear bi-allelic and homozygous, such as cell line 15 at the GS-2 locus (48-bp deletion) and cell line 3 at the cGOGAT locus (24-bp deletion). Only cell line 10 at locus cGOGAT displayed a bi-allelic, heterozygous mutation of a 15-bp deletion and 12-bp deletion. The remaining genotypes were not as clearly defined and have multiple (more than two) peaks in their TIDE chart and/or they contain a trace of a wild-type sequence (both examples are shown in cell line 4 at the cGOGAT locus).

### HDR-Mediated Gene Editing of NR in *Phaeodactylum* by Micro-Particle Bombardment

Six cell lines were produced via HDR-mediated gene editing of the NR gene in *Phaeodactylum* (Figure 5). It should be noted that, although attempted, HDR-mediated mutagenesis was not observed when the HDR-donor was encoded on

the Cas9-sh ble episome (results not shown). All six cell lines exhibited the genotype (Figure 5B) and phenotype (Figure 5C) of having an HDR-mediated insertion of the sh ble expression cassette at the gNR-B target locus. An efficiency of 17% for producing a bi-allelic, HDR-mediated Pt\_NR mutant was achieved.

### DISCUSSION

CRISPR-Cas9 mutagenesis was performed in the model diatom *P. tricornutum* specifically to streamline the use of bacterial-conjugation transformation. First, the episomal delivery system for Cas9 was re-engineered so Cas9 could be selected for by phleomycin-antibiotic pressure. The 2A peptide transcriptionally fused Cas9 and the selective gene, sh ble, to ensure Cas9 could be selected for after conjugation. Previous attempts to deliver Cas9 in an expression cassette separate from the sh ble cassette resulted

in low mutagenesis efficiency when colonies were not initially screened for a predicted phenotype, such as NR phenotype. Although not reported here, previously published experiments that delivered a non-selectable Cas9 via bacterial conjugation in *Phaeodactylum* resulted in mutagenesis efficiencies ranging from 25 to 33% using one sgRNA target (Sharma et al., 2018). Here, mutagenesis efficiencies ranged from 2 to 40% when targeting the nitrate reductase gene. Despite reporting a higher mutagenesis efficiency when using gNR-B, mutagenesis efficiencies cannot be directly compared when targeting two different genomic loci.

It has been observed that only 10% of transformed colonies contained a full Cas9 expression cassette, drastically lowering the mutagenesis efficiency (results not shown). It has been previously reported that the *Phaeodactylum* episome, naked, only retains its full plasmid sequence after transfer from *E. coli* to *Phaeodactylum* at 30% (Karas et al., 2015; Diner et al., 2016). Such a Cas9 episome design would not be ideal for targets without a previously known phenotype. The 2A peptide proved valuable in increasing the efficiency of Cas9 delivery to *Phaeodactylum* and in identifying mutant cell lines prior to checking for a phenotype response. Due to a transcriptional fusion to *sh ble*, it was presumed that the full Cas9 coding region was retained in all transformed *Phaeodactylum* transformants, though PCR amplification of the Cas9 ORF was not performed post-conjugation.

Although not thoroughly quantified, selection for Cas9 seems to alleviate previously reported problems of producing non-clonal *Phaeodactylum* that led to inevitable re-streaking, waiting for clonal cell lines to appear, and re-picking (Weyman et al., 2014). TIDE sequence analysis provided a predictive genotype for each cell line that can be interpreted as homozygous, heterozygous, or mixed genotype. For gNR-B, 26/64 cell lines exhibited Cas9 activity while 6 exhibited either a homo- or heterozygous genotype by TIDE analysis (Figure 3C), which can be interpreted to mean that those cell lines were clonal. However, 20/64 lines that passed the same efficiency criteria (in methods) contained more than 2 peaks in the TIDE chart or contained a trace wild-type sequence and therefore they were not considered clonal. Clonality was inherently tested for, also, when the NR-KO cell lines were subjected to a nitrate growth assay. If a wild-type cell contaminated the culture there would be growth on nitrate. Nevertheless, none of the six NR-KO cell lines grew on nitrate.

While TIDE sequence analysis is a useful tool to quickly screen raw sequencing reads for Cas9-induced indels. Here, TIDE was useful in identifying cell lines that contained targeted Cas9 activity and in inferring genotype, however, it is not recommended to use TIDE to confirm a cell lines genotype. Rather, the gold standard of sub-cloning the target locus into a TOPO vector followed by sequencing is recommended after TIDE analysis. It should be noted that TIDE analysis is complementary to CRISPR/Cas9 activity factor (CAF) analysis that was also designed to deconvolve diploid genotypes after CRISPR/Cas9 targeting (Stukenberg et al., 2018).

Re-streaking *Phaeodactylum* mutant is also recommended after TIDE analysis. The TIDE chart generated for gNR-B cell lines (Figure 3) suggest that the genotypes are either homozygous or heterozygous and that the cell lines are clonal.

Nevertheless, TIDE is not the gold-standard for genotyping and assessment of clonality. As shown in Figure 4, not all TIDE plots indicate clonality and therefore re-streaking would be necessary to isolate double-knock-out mutant cell lines. All supplemental tables outline the genotyping efforts performed in this study. The last column for each table indicates the predicted genotype of the colony and suggests whether or not streaking the culture is recommended.

Lastly, the new Cas9 episome design decreased the time to produce and validate mutant cell lines. The time to produce visible colonies was 10–14 days, genotyping was 3 days, and phenotype screening was 10 days after re-picking. In total, the production of 6 *Phaeodactylum* mutants for the NR gene took 3–4 weeks compared to 4–8 weeks when using particle bombardment (Sharma et al., 2018). In a rapidly developing research landscape, optimization of time to produce mutant cell lines should not be a time-limiting step when there are questions regarding biological processes, ecological relevance, or biotechnology utility is an important goal.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the MM Frontiers 2-27-29 Data Availability Document [https://docs.google.com/spreadsheets/d/1snw\\_Jao9sdCDy2Y7cC3tP33cNzZ0xTxOQTdHUEaARME/edit#gid=2069237253](https://docs.google.com/spreadsheets/d/1snw_Jao9sdCDy2Y7cC3tP33cNzZ0xTxOQTdHUEaARME/edit#gid=2069237253). Plasmid designed and mentioned in this manuscript will be made publicly available at the plasmid repository addgene.com.

## AUTHOR CONTRIBUTIONS

MM was responsible for all of the molecular cloning, genetic transformations, genotyping assays, and writing. VB helped in plasmid and episome design, specifically for the conjugation protocols and implementation of the 2A peptide, and other intellectual contributions. PG performed the phenotype assays. MT designed golden gate primers for the episome. JM helped in daily tasks and manuscript writing. AA contributed to intellectual design and scope of the study. All authors contributed to the editing and feedback for this manuscript.

## FUNDING

This study was supported by the Gordon and Betty Moore Foundation Grants GBMF3828 and GBMF5006 (AA), National Science Foundation Grants NSF-OCE-1756884 and NSF-MCB-1818390 (AA), United States Department of Energy Genomic Science Program Grant DE-SC0018344 (AA), and National Aeronautics and Space Administration Grant 15-EXO15\_2-0158.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.00005/full#supplementary-material>



## REFERENCES

- Bailleul, B., Rogato, A., de Martino, A., Coesel, S., Cardol, P., Bowler, C., et al. (2010). An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. *Proc. Nat. Acad. Sci.* 107, 18214–18219. doi: 10.1073/pnas.1007703107
- Bielinski, V. A., Bolt, T. M., Dupont, C. L., and Weyman, P. D. (2017). Episomal tools for RNAi in the diatom *Phaeodactylum tricornutum*. *PeerJ Preprints* 5:e2907v1. doi: 10.7287/peerj/preprints.2907v1
- Brinkman, E. K., Chen, T., Amendola, M., and Steensel, B. V. (2014). Easy quantification assessment of genome editing by sequence trace decomposition. *Nucleic Acid Res.* 42:e168. doi: 10.1093/nar/gku936
- Cermeño, P., Lee, J.-B., Wyman, K., Schofield, O., Falkowski, P. G. (2011). Competitive dynamics in two species of marine phytoplankton under non-equilibrium conditions. *Mar. Ecol. Prog. Ser.* 429, 19–28. doi: 10.3354/meps09088
- Cho, S. W., Kim, S., Kim, J. M., and Kim, J. S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 230–232. doi: 10.1038/nbt.2507
- Cong, L., Ran, F., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143
- Davis, A., and Chen, D. (2013). DNA double strand break repair via non-homologous end-joining. *Transl. Cancer Res.* 2, 130–143.
- Deltcheva, E., Chylinski, K., Sharma, C., Gonzales, K., Chao, Y., Pirzada, Z., et al. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602–607. doi: 10.1038/nature09886
- Diner, R. E., Bielinski, V., DuPont, C., Allen, A. E., and Weyman, P. (2016). Refinement of the diatom episome maintenance sequence and improvement of conjugation-based DNA delivery methods. *Front. Bioeng. Biotechnol.* 4:65. doi: 10.3389/fbioe.2016.00065
- Falciatore, A., Casotti, R., LeBlanc, C., Abrescia, C., and Bowler, C. (1999). Transformation of nonselectable reporter genes in marine diatoms. *Mar. Biotechnol.* 1, 239–251. doi: 10.1007/pl00011773
- Gao, C. (2018). The future of CRISPR technologies in agriculture. *Nat. Rev. Mol. Cell Bio.* 19, 275–276. doi: 10.1038/nrm.2018.2
- Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. I. I., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. doi: 10.1038/nmeth.1318
- Guschin, D. Y., Waite, A. J., Katibah, G. E., Miller, J. C., Holmes, M. C., Rebar, E. J., et al. (2010). A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* 649, 247–256. doi: 10.1007/978-1-60761-753-2\_15
- Hopes, A., Nekrasov, V., Kamoun, S., and Mock, T. (2016). Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant Methods* 12:49.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-Guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829
- Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife* 2:e00471. doi: 10.7554/eLife.00471
- Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P., Noddings, C., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 21:6925. doi: 10.1038/ncomms7925
- Kim, J. H., Lee, S. R., Li, L. H., Park, H. J., Park, J. H., Lee, K. Y., et al. (2011). High cleavage efficiency of a 2A peptide derived from *Porcine Teschovirus-1* in human cell lines, Zebrafish and Mice. *PLoS One* 6:e18556. doi: 10.1371/journal.pone.0018556
- LaFave, M., and Sekelsky, J. (2009). Mitotic recombination: why? when? how? where?. *PLoS Genet.* 5:e1000411. doi: 10.1371/journal.pgen.1000411
- Mali, P., Esvelt, K., and Church, G. (2013a). Cas9 as a versatile tool for engineering biology. *Nat. Methods* 10, 957–963. doi: 10.1038/nmeth.2649
- Mali, P., Yang, L., Esvelt, K., Aach, J., Guell, M., DiCarlo, J., et al. (2013b). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. doi: 10.1126/science.1232033
- McCarthy, J. K., Smith, S. R., McCrow, J. P., Tan, M., Zheng, H., Beeri, K., et al. (2017). Nitrate reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in model pennate diatom. *Plant Cell* 29, 2047–2070. doi: 10.1105/tpc.16.00910
- McGinn, J., and Marraffini, L. A. (2018). Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat. Rev. Microbiol.* 17, 7–12. doi: 10.1038/s41579-018-0071-7
- Nymark, M., Sharma, A. K., Sparstad, T., Bones, A., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951. doi: 10.1038/srep24951
- Pulz, O., and Gross, W. (2004). Valuable products from biotechnology of microalgae. *Appl. Microbiol. Biotechnol.* 65, 635–648. doi: 10.1007/s00253-004-1647-x
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome Engineering using the CRISPR-Cas 9 system. *Nat. Protoc.* 8, 2281–2308. doi: 10.1038/nprot.2013.143
- Rastogi, A., Murik, O., Bowler, C., and Tirichine, L. (2016). PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing. *BMC Bioinform.* 17:261. doi: 10.1186/s12859-016-1143-1
- Rogato, A., Richard, H., Sarazin, A., Voss, B., Cheminant Navarro, S., Champeimont, R., et al. (2014). The diversity of small non-coding RNAs in the diatom *Phaeodactylum tricornutum*. *BMC Genom.* 15:698. doi: 10.1186/1471-2164-15-698
- Sambrook, J. F., and Russell, D. W. (eds) (2001). *Molecular Cloning: A Laboratory Edition*. New York, NY, NY: Cold Spring Harbor Laboratory Press.
- Sanders, J., and Joung, J. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355. doi: 10.1038/nbt.2842
- Serif, M., Dubois, G., Finoux, A. L., Teste, M. A., Jallet, D., and Daboussi, F. (2018). One-step generation of multiple gene knock-outs in the diatom *Phaeodactylum tricornutum* by DNA-free genome editing. *Nat. Comm.* 9:3924. doi: 10.1038/s41467-018-06378-9
- Sharma, A. K., Nymark, M., Sparstad, T., Bones, A. M., and Winge, P. (2018). Transgene-free genome editing in marine algae by bacterial conjugation - comparison with biolistics CRISPR/Cas9 transformation. *Sci. Rep.* 8:14401. doi: 10.1038/s41598-018-32342-0
- Slattery, S., Diamond, A., Wang, H., Therrian, J. A., Lant, J. T., Jazey, T., et al. (2018). An expanded plasmid-based genetic toolbox enables Cas9 genome editing and stable maintenance of synthetic pathways in *Phaeodactylum tricornutum*. *ACS Syn. Biol.* 7, 328–338. doi: 10.1021/acssynbio.7b00191
- Stukenberg, D., Zauner, S., Dell'Aquila, G., and Maier, U. G. (2018). Optimizing CRISPR/Cas9 for the diatom *Phaeodactylum tricornutum*. *Front. Plant Sci.* 9:740. doi: 10.3389/fpls.2018.00740
- Vinayak, V., Manoylov, K., Gateau, H., Blanckhaert, V., Herauld, J., Pencreac'h, G., et al. (2015). Diatom milking: a review and new approaches. *Mar. Drugs* 13, 2629–2665. doi: 10.3390/md13052629
- Wang, H., La Russa, M., and Qi, L. S. (2016). CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* 85, 227–264. doi: 10.1146/annurev-biochem-06015-014607
- Wang, H., and Xu, X. (2017). Microhomology-mediated end joining: new players join the team. *Cell Biosci.* 7:6. doi: 10.1186/s13578-017-0136-8
- Weyman, P., Beeri, K., Lefebvre, S., Rivera, J., McCarthy, J., Heuberger, A., et al. (2014). Inactivation of *Phaeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol.* 13, 460–470. doi: 10.1111/pbi.12254
- Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509, 487–491. doi: 10.1038/nature13166

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Moosburner, Gholami, McCarthy, Tan, Bielinski and Allen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.