New Phytologist

Methods

Efficient gene replacement by CRISPR/Cas-mediated homologous recombination in the model diatom *Thalassiosira pseudonana*

Nigel Belshaw[1], Irina Grouneva[1], Lior Aram[2], Assaf Gal[2], Amanda Hopes[1] and Thomas Mock[1]

[1]School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK; [2]Department of Plant and Environmental Sciences, Faculty of Biochemistry, Weizmann Institute of Science, Rehovot, 7610001, Israel

Authors for correspondence:
Thomas Mock
Email: t.mock@uea.ac.uk

Amanda Hopes
Email: a.hopes@uea.ac.uk

Received: *20 May 2022*
Accepted: *18 October 2022*

New Phytologist (2023) **238:** 438–452
doi: 10.1111/nph.18587

Key words: algae, CRISPR, diatom, genome editing, homology-directed repair, recombination, synthetic biology, *Thalassiosira pseudonana*.

Summary

- CRISPR/Cas enables targeted genome editing in many different plant and algal species including the model diatom *Thalassiosira pseudonana*. However, efficient gene targeting by homologous recombination (HR) to date is only reported for photosynthetic organisms in their haploid life-cycle phase. Here, a CRISPR/Cas construct, assembled using Golden Gate cloning, enabled highly efficient HR in a diploid photosynthetic organism.
- Homologous recombination was induced in *T. pseudonana* using sequence-specific CRISPR/Cas, paired with a dsDNA donor matrix, generating substitution of the silacidin, nitrate reductase and urease genes by a resistance cassette (FCP:*NAT*).
- Up to *c.* 85% of NAT-resistant *T. pseudonana* colonies screened positive for HR by nested PCR. Precise integration of FCP:*NAT* at each locus was confirmed using an inverse PCR approach. The knockout of the nitrate reductase and urease genes impacted growth on nitrate and urea, respectively, while the knockout of the silacidin gene in *T. pseudonana* caused a significant increase in cell size, confirming the role of this gene for cell-size regulation in centric diatoms.
- Highly efficient gene targeting by HR makes *T. pseudonana* as genetically tractable as *Nannochloropsis* and *Physcomitrella*, hence rapidly advancing functional diatom biology, bionanotechnology and biotechnological applications targeted on harnessing the metabolic potential of diatoms.

Introduction

Diatoms represent a highly successful group of unicellular phytoplankton responsible for an estimated 20% of global primary production (Field *et al.*, 1998; Falkowski & Raven, 2013). Due to their silicified cell walls (frustules), diatoms also dominate the ocean's biogenic silicon cycle (Tréguer & De La Rocha, 2013). Many of the traits underpinning their ecological success are the reason why these eukaryotic microbes are desirable for algal biotechnology and fuel production (Bozarth *et al.*, 2009; Levitan *et al.*, 2014; Wang & Seibert, 2017). Their most characteristic traits include fast growth (Ichimi *et al.*, 2012), biomineralisation and complex metabolism derived from (1) successive endosymbiotic events (Bhattacharya *et al.*, 2007; Frommolt *et al.*, 2008; Moustafa *et al.*, 2009; Keeling, 2010), (2) genome hybridisation (Tanaka *et al.*, 2015) and (3) haplotype divergence (Mock *et al.*, 2017; Paajanen *et al.*, 2017). The processes that govern this remarkable ability are largely unknown but are probably shared by many diatoms as they generally outcompete other algal groups when growth conditions become favourable. Their metabolism was shaped by different endosymbiotic events (primary and secondary endosymbiosis) that took place at least 1 billion years ago and contributed various enzymes, isoforms and entire pathways such as the urea pathway (Armbrust *et al.*, 2004; Allen *et al.*, 2011). These 'mix and match' genomes are considered to have contributed to the plasticity of their genomes as well as phenomes and therefore their trait performance landscapes (Messina *et al.*, 2011). More recent evolutionary processes that shaped diatom genomes include genome hybridisation (Tanaka *et al.*, 2015) and haplotype divergence (Mock *et al.*, 2017; Paajanen *et al.*, 2017). The latter is considered to be the consequence of adaptive evolution leading to environment-dependent differential allelic expression (Mock *et al.*, 2017; Paajanen *et al.*, 2017). The correlation between diversifying selection and allelic

differentiation suggests a potential role of the divergent alleles for adaptation to environmental fluctuations in aquatic environments.

To harness the unique biology of diatoms for advancing fundamental research and algal biotechnology, we need to improve our knowledge of molecular factors controlling their evolution, adaptation and metabolism. Genome-enabled studies have revealed genes, regulatory elements and epigenetic factors potentially controlling the expression of diverse phenotypes. It is only recently, however, that genome editing tools such as CRISPR/Cas (Doudna & Charpentier, 2014; Sander & Joung, 2014; Lander, 2016) and TALENs (Gaj *et al.*, 2013) have become available, including in diatoms (Daboussi *et al.*, 2014; Weyman *et al.*, 2015; Hopes *et al.*, 2016; Nymark *et al.*, 2016; Serif *et al.*, 2017), enabling us to improve our understanding of individual genetic factors for the evolution and biology of diatoms. The introduction of gene targeting by high-frequency homologous recombination (HR) allows precise genetic changes via knockin of a particular sequence at a specific locus, rather than just gene knockout. Thus, a proven ability of gene targeting via HR in diatoms will enable improvements in gene and protein engineering, which will be instrumental to advance our understanding of fundamental diatom biology and to facilitate diatom domestication.

Gene targeting via HR has been demonstrated in a number of photosynthetic eukaryotes to date including *Cyanidioschyzon merolae* (Minoda *et al.*, 2004), *Ostreococcus tauri* (Lozano *et al.*, 2014), *Chlamydomonas reinhardtii* (Ferenczi *et al.*, 2017; Greiner *et al.*, 2017; Angstenberger *et al.*, 2020), *Nannochloropsis* sp. (Kilian *et al.*, 2011; Naduthodi *et al.*, 2019), *Phaeodactylum tricornutum* (Daboussi *et al.*, 2014; Weyman *et al.*, 2015; Moosburner *et al.*, 2020), *Physcomitrella patens* (Kamisugi *et al.*, 2005, 2006, 2016), *Arabidopsis* (Schiml *et al.*, 2014; Miki *et al.*, 2018), *Oryza* (Begemann *et al.*, 2017) and *Nicotiana* (Li *et al.*, 2013). Most of these, however, are either haploid and/or have low rates of HR. Highly efficient HR can be seen in two of these species: the moss *P. patens* and the eustigmatophyte *Nannochloropsis* sp. *Nannochloropsis* sp. is haploid, and the dominant phase of the life cycle in *P. patens* is also haploid, which makes backcrossing to establish homozygous transgenic lines obsolete. Targeted and efficient HR in both species can be achieved without introducing nucleases to produce DNA double-strand breaks (DSBs) although high rates of HR are also seen in *Nannochloropsis* sp. with CRISPR–Cas-induced DSBs (Naduthodi *et al.*, 2019). For example, in *Nannochloropsis* sp., Kilian *et al.* (2011) used 1-kb flanking sequences to target genes for HR-induced gene knockout. For all the other photosynthetic eukaryotes, rates of HR are rather low either with or without the help of nucleases. There are two reports of TALEN-mediated gene editing including HR in the marine diatom *P. tricornutum* (Daboussi *et al.*, 2014; Weyman *et al.*, 2015). Both studies tested the occurrence of HR by cotransformation of the TALEN-carrying plasmid and a plasmid with a donor template and reported an efficiency of up to 27% (Daboussi *et al.*, 2014; Weyman *et al.*, 2015). CRISPR-mediated HR has also been reported in *P. tricornutum* using cotransformation with a Cas9-shBle plasmid, sgRNA plasmid and donor plasmid with a biallelic HR efficiency of 17% (Moosburner *et al.*, 2020).

The present study, however, is the first report of highly efficient HR in any diploid photosynthetic organism mediated by CRISPR/Cas. The aim of this work was to show that HR can be induced in a diatom by means of the sequence-specific nuclease Cas9, paired with a donor matrix (exogenous DNA templates with homology to the targeted locus), generating a complete knockout of a target gene in *Thalassiosira pseudonana*. In this study, the silacidin, nitrate reductase (*NitR*) and urease genes were substituted by a resistance cassette (FCP:*NAT*). By extension, this method could also enable targeted gene insertion/substitution in *T. pseudonana* and possibly other diatoms as the versatile Golden Gate cloning system was used (Weber *et al.*, 2011; Belhaj *et al.*, 2013) to make the vector constructs. High-frequency gene targeting by HR will considerably expand the genetic tool set for diatoms and has already made *T. pseudonana* as genetically tractable as *Nannochloropsis* and *Physcomitrella*, hence rapidly advancing diatom functional genomics and biotechnology.

Materials and Methods

Strains and growth conditions

Thalassiosira pseudonana (CCMP 1335) was grown in half salinity (½SOW: 16 g l^{-1}) Aquil seawater medium (Price *et al.*, 1989) supplemented with silica under constant illumination (*c.* 100 µmol photons m^{-2} s^{-1}) at 20°C. Unless stated otherwise, $NaNO_3$ was replaced with NH_4Cl as the nitrogen source in all experiments, for consistency and since replacement of the *NitR* gene prevents growth on nitrate.

Plasmid design and assembly

Two plasmids were used to introduce all components necessary for HR (Fig. 1). (1) The first plasmid expresses Cas9 and either one or two guide RNAs (gRNA) designed to induce one or two DSBs in the target sequence (pAGM4723_Tp:FCP_P:Cas9:T_TpU6P:gRNA). (2) The second plasmid contained a cassette including the nourseothricin resistance gene *NAT*, flanked by two regions 629–868 bp in length, which are homologous to noncoding 5′ and 3′ ends of the silacidin, *NitR* or urease genes to serve as a template (donor matrix) for HR. The second plasmid was assembled from PCR products equipped with *Bpi*I restriction sites and overhangs for directional cloning (pAGM4723_5′-flank_TpFCP:*NAT*_3′flank) to allow cloning by the Golden Gate method. Cloning was carried out according to Belhaj *et al.* (2013) and Weber *et al.* (2011) in a single-step restriction/ligation reaction. Forty femtomoles of each component was combined in a 20-µl reaction with 10 units of *Bsa*I (L1 assembly) or *Bpi*I (L2 assembly) and 10 units of HC T4 DNA ligase in 1× ligation buffer. The reaction was incubated at 37°C for 5 h, 50°C for 5 min and 80°C for 10 min. Five microlitres of the reaction was transformed into 50 µl of NEB 5-alpha chemically competent *Escherichia coli*.

New Phytologist

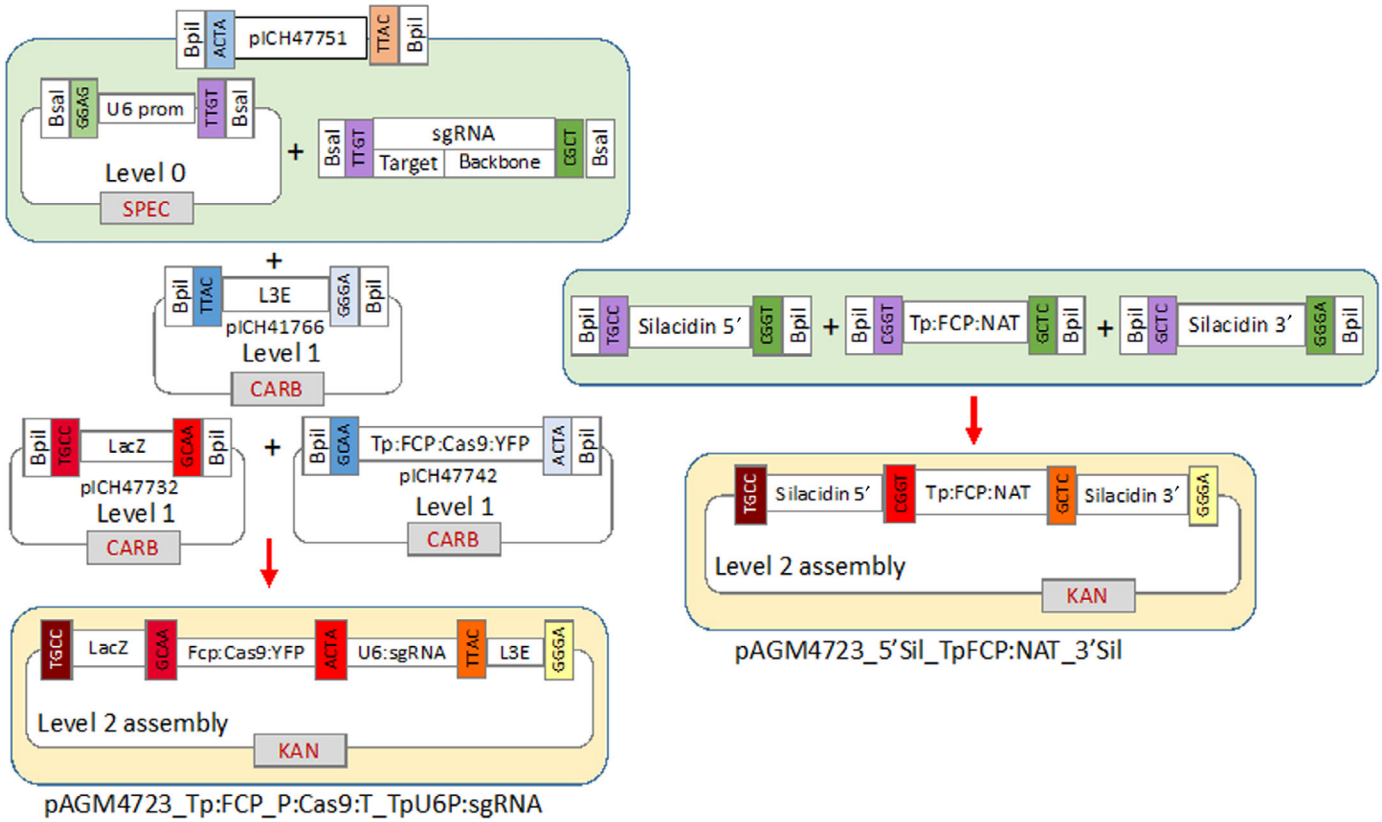

Fig. 1 Overview of modules and overhangs used in Golden Gate assembly of level 2 transformation vectors pAGM4723 Tp:FCP_P:Cas9:T_TpU6P:sgRNA and 5′Sil_TpFCP:*NAT*_3′Sil. Tp:FCP, module flanked by endogenous *Thalassiosira pseudonana* FCP gene promoter and terminator sequences. Plasmids carry antibiotic resistance selection genes against spectinomycin (SPEC), carbenicillin (CARB) and kanamycin (KAN).

Plasmid 1 assembly

The gRNAs were amplified with primers (introducing the specific 20 nt targeting region at the 5′ end) for silacidin (primer 1), *NitR* sgRNA1 (primer 2) and *NitR* sgRNA2 (primer 3) together with primer 5 (Table 1) using plasmid pICH86966_AtU6p_sgRNA_NbPDS (containing the gRNA backbone) as a template (Supporting Information Table S1). PCR was carried out with Phusion DNA Polymerase with 56°C annealing temperature and 1 min elongation step. Each of the purified PCR products for silacidin and *NitR* sgRNA1 were then assembled along with the L0 vector containing the Tp:U6 promoter sequence (Hopes *et al.*, 2016) into L1 vector pICH47751. The purified PCR product for *NitR* sgRNA2 was assembled with the L0 Tp:U6 promoter vector into L1 vector pICH47761. Guide RNA products and subsequent L1 U6:gRNA plasmids for urease (sgRNA 1) were previously produced by Hopes *et al.* (2016; primer 4). L1 modules pICH47732 (lacZ), pICH47742:FCP:Cas9YFP (Hopes *et al.*, 2016), pICH47751:TpU6p:gRNA (silacidin, *NitR sgRNA1* or urease) and linker plasmid pICH41766 (L3E) were then assembled together in the L2 backbone pAGM4723 to generate pAGM4723_Cas9_SIL (silacidin), pAGM4723_Cas9_NitR-1 (*NitR* one sgRNA) and pAGM4723_Cas9_Urease (urease). To create pAGM4723_Cas9_NitR-2 (*NitR* two sgRNAs), L1 modules pICH47732 (lacZ), pICH47742:FCP:Cas9YFP, pICH47751:TpU6p:gRNA (*NitR* sgRNA1), pICH47761: TpU6p:gRNA (*NitR* sgRNA2) and linker plasmid pICH41780 (L4E) were assembled together in the L2 backbone pAGM4723.

For details on sequences, restriction sites and selection markers, see Addgene database, Hopes *et al.* (2017) and Table S1.

Plasmid 2 assembly

5′ and 3′ flanking, noncoding sequences for silacidin (868 and 629 bp), *NitR* (728 and 729 bp) and urease (781 and 731 bp) were amplified from genomic DNA with primer pairs shown in Table 1 introducing overhangs for directional cloning and *Bpi*I restriction sites. The FCP:*NAT* resistance gene cassette with *T. pseudonana*-specific promoter and terminator was amplified with primers 6/7 with vector pICH47732:FCP:*NAT* as template (Hopes *et al.*, 2016). Phusion DNA Polymerase was used and the annealing temperature was 56°C in all cases. After purification, PCR products were assembled into a pAGM4723 backbone using the Golden Gate method as described previously to generate the plasmids pAGM4723_SIL_NAT_donor (silacidin), pAGM4723_NitR_NAT_donor (*NitR*) and pAGM4723_Urease_NAT_donor (urease).

gRNA selection and design

Guide RNA sequences followed the format of NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGCAAGTT

Table 1 Oligonucleotide primers used for PCR amplification for cloning and screening of *Thalassiosira pseudonana* transformants.

Name	Sequence	Primer reference no.
Primers for introducing BsaI/BpiI restriction sites and overhangs for directional cloning		
Silacidin sgRNA_F	AGGTCTCATTGTGTGATGAGTCTGAAGAGTCCGGTTTTAGAGCTAGAAATAGCAAG	1
NitR sgRNA1_F	AGGTCTCATTGTGCCGAAGAGAGTTTGACAGGGTTTTAGAGCTAGAAATAGCAAG	2
NitR sgRNA2_F	AGGTCTCATTGTGTCGAGGTATCTACCGATGGGTTTTAGAGCTAGAAATAGCAAG	3
Urease sgRNA_F	AGGTCTCATTGTGTCGTAATCAAGTATTGCCGGTTTTAGAGCTAGAAATAGCAAG	4
sgRNA_R	TGGTCTCAAGCGTAATGCCAACTTTGTACAAG	5
FCP_P:*NAT*_F	CTCGAAGACGACGGTATCGATAAGCTTGC	6
FCP_P:*NAT*_R	CGGGAAGACTGGAGCAGCTACTACTTTGCT	7
5′ Silacidin_F	AAAGAAGACCATGCCACGCAGTGAGCCACTATC	8
5′ Silacidin_R	ATTGAAGACACACCGTCCAACAAAAAACCACTCT	9
3′ Silacidin_F	TGGGAAGACATGCTCAGTAGAGTTTAAGGCAT	10
3′ Silacidin_R	ACCGAAGACCTTCCCTCTGAAGTGACGGCTGCAG	11
5′ *NitR*_F	CTCGAAGACCTTGCCTACCGAAAAGCTCACGGAGA	12
5′ *NitR*_R	GTTGAAGACATACCGATCCGAGGATAACCATCAAACCAA	13
3′ *NitR*_F	TGCGAAGACATGCTCTTGAGAGAATGTCTCAT	14
3′ *NitR*_R	ACGGAAGACCATCCCCAGAACGGTGGAGCGTA	15
5′ Urease_F	CCGGAAGACGCTGCCCAGGGGTTGTTTTTGCA	16
5′ Urease_R	GGTGAAGACGCACCGTCACCAACAACGCCGA	17
3′ Urease_F	GATGAAGACAAGCTCCAACACCGTTCTTCG	18
3′ Urease_R	CTAGAAGACTGTCCCTTCAAACAAACGAAGAGATA	19
Primers for screening of transformants		
HR5R (NAT 5′)	CGTTGTCGGATACATTCTTGATTGGAG	20
HR3F (NAT 3′)	GGTGTAGGCATGTTGGACACAACGAGGTA	21
Silacidin_locusF	GAAGCTCACATCCACCGTTGTTG	22
Silacidin_locusR	CTTTGAATCTTGCTCAGAGTGCAGAAG	23
Silacidin_HR5F	CGTCAATTGAATTCTTATATCCCATTC	24
Silacidin_HR3R	GCTTCGAGTCTTGAAGACC	25
Silacidin_gene_F	CAACAACCCCACAGCAACTC	26
Silacidin_gene_R	GCCCATGCATCAAACATCATCTAATCTTCAC	27
*NitR*_locusF	GCTACGATCCAACCTCCCACACCTCTA	28
*NitR*_locusR	CGTACAGGTGTCAACGTCTGAGGCT	29
*NitR*_HR5F	CCAACGGCCTCCCACCCTACATCATT	30
*NitR*_HR3R	CCAGGATACACTGATGCAGAAGGTGAC	31
*NitR*_gene_F	CGGATATCAATGTGTTGGAGCACGAGA	32
*NitR*_gene_R	GGACCTTGATACGGAAGACTTGGTTGTTT	33
Urease_locusF	CCATCGATTACCATATCAAACCTCCACTCACA	34
Urease_locusR	GTGTAGCCAAAGAATCTCCAAACCAGGCCAAA	35
Urease_HR5F	GAGGCAGATTTGTCGTCTGGTTCACA	36
Urease_HR3R	GCAATACCCTACTACTCTGAAGAAGAGTCTGAA	37
Urease_gene_F	CAAGGCGTTGATTGCCATGCAAA	38
Urease_gene_R	AACACACTTAAATCAGGAACAGGCAAA	39
Urease iPCR 5′	ACCGATACCAAAATGACCACTCTT	40
Urease iPCR 3′	CCGGAGTACGAGACGACCACGAA	41

AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG CACCGAGTCGGTGC*TTTTTT* where the underlined sequence indicates the gene-specific 20 nt target, the mid-region is the scaffold and the italicised sequence is the 3′ termination signal.

The silacidin gene is composed of extensive repeats. The gRNA target used in the present work, GTGATGAGTCTGAA GAGTCCG, was designed to induce two cuts within the coding region of the gene. It has to be noted, however, that two variants of silacidin (allele variation) are present in the database. Only one of them predicts two identical target sequences matching the seed sequence of the selected gRNA.

The gRNA target for the urease gene, GTCGTAATCAAG TATTGCCG, targets a single location within the gene. Both of the NitR gRNAs targeted a single location within the gene, with the use of one sgRNA target, GCCGAAGAGAGTTTGA CAGG, expected to induce a single DSB and use of two sgRNAs targets, GCCGAAGAGAGTTTGACAGG and GTCGAGG TATCTACCGATGG expected to induce a deletion of 183 nt. Only guide RNAs with no predicted off-target locations within the *T. pseudonana* genome were chosen; off-target predictions were carried out using RGEN Cas-OFFinder (Bae *et al.*, 2014). Guide RNAs were designed according to Hopes *et al.* (2016, 2017).

Transformation of *T. pseudonana*

Biolistic particle delivery transformation was carried out according to Poulsen *et al.* (2006) and Hopes *et al.* (2017).

For each shot, 5×10^7 cells from the exponential growth phase were collected onto a filter (47 mm diameter 1.2 μm Whatman

New Phytologist

Isopore; Merck Life Science UK Ltd, Gillingham, UK) and placed on the top of a 1.5% agar, ½SOW Aquil medium Petri dish. One microgram of each plasmid (Tp:Cas9:sgRNA and 5′flank:FCP:*NAT*:3′flank) was used for coating 3 mg of M10 (0.7 μm diameter) tungsten particles (Bio-Rad) per shot. A 7 cm flight distance and 1350 psi rupture discs were used. After transformation, each filter was immediately placed into 25 ml of fresh ½SOW Aquil medium and cells were left to recover under standard growth conditions for 24 h. The following day, 5×10^6 cells were spread onto selective 0.8% (w/v) ½SOW Aquil agar plates with 100 μg ml^{-1} nourseothricin. An aliquot of 10^6 cells ml^{-1} was supplemented with nourseothricin and grown as a liquid culture. Colonies appeared on selective plates after 8–12 d.

Screening of colonies

Colonies were restreaked onto fresh selective plates and grown in liquid medium for DNA isolation. DNA was then used as a template for PCR analysis using MyTaq DNA polymerase (Bioline; Meridian Bioscience, London, UK). A nested PCR approach was adopted using primers shown in Figs 2 and 3(a) and detailed in Table 1. Initially, primer pairs targeting the genomic loci outside the 5′ and 3′ regions used for inducing HR were used to amplify the silacidin, *NitR* and urease loci irrespective of whether HR had occurred. The products of these PCRs were used as templates in secondary PCRs (two for each gene) with primers nested to those used in the primary PCR but still 5′ or 3′ to the flanking regions used in the 5′flank:FCP:*NAT*:3′flank plasmid, together with primers specific for the FCP:*NAT* gene cassette (Fig. 2). Specific amplicons from these PCRs signified HR-mediated replacement of silacidin, *NitR* or urease. In addition, a fragment of each gene including putative Cas9 cut site(s) was amplified in further nested PCRs with primers listed in Table 1. The different potential outcomes from the PCR screening of the transformants are shown in Fig. 3(b) and include no gene replacement (lane 1), monoallelic gene replacement with wild-type (WT) retained copy (lane 2), monoallelic gene replacement with edited retained copy (lane 3) and biallelic gene replacement (lane 4).

Confirmation of HR by inverse PCR

The site of integration of the FCP:*NAT* cassette into the genome was investigated using inverse PCR (iPCR). Homologous recombination transformant and WT genomic DNA (1 μg) was digested with the restriction enzymes MfeI (silacidin), BstBI (*NitR*) and KpnI (urease) selected to cut upstream and downstream of the 5′ and 3′ regions used to induce HR. KpnI also cut the FCP:*NAT* cassette; two iPCRs were therefore performed to amplify the regions upstream and downstream of the site of integration for the urease transformants. The restricted DNA was circularised with T4 DNA ligase and used as template in PCRs with primers 20/21 for silacidin and *NitR* and 20/40 and 21/41 for urease upstream and downstream regions, respectively (Table 1). The resulting amplicons were separated by gel electrophoresis, excised and cloned into pCR4-TOPO-TA (Invitrogen) for sequencing.

Phenotyping silacidin transformants based on cell-size estimation

In accordance with expectations from earlier experiments on silacidin knock-down mutants that showed larger cell size/valve diameter in transformant lines (Kirkham *et al.*, 2017), we used three different approaches to compare cell sizes of selected mutant silacidin lines to WT. All measurements were done on cells during mid-exponential growth phase. (1) Culture aliquots were measured with a Z2 Coulter counter (Beckman Coulter, High Wycombe, UK). The cell size was determined as the maximum value of the Gaussian distribution of cell sizes in each measurement. (2) Forward scattering of light by single cells was measured with an Eclipse iCyt flow 531 cytometer (Sony Biotechnology Inc., Champaign, IL, USA), equipped with 405 and 488 nm solid state air-cooled lasers, both with 25 mW on the flow cell and with standard 533 filter set-up. The forward light scattering is proportional to the size of the cell. (3) Direct measurements of the diameter of the valves were done on scanning electron microscope (SEM) images. Sample aliquots were

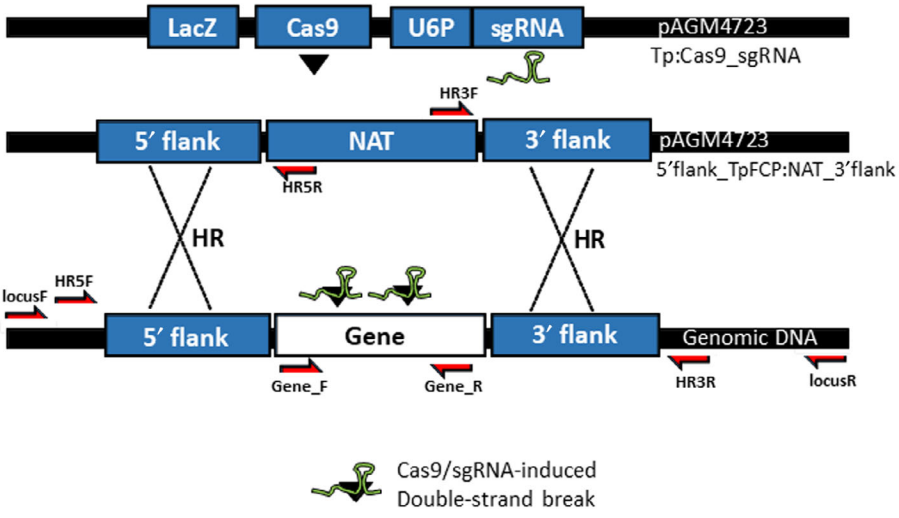

Fig. 2 Overview of homologous recombination mechanism between donor matrix and target genomic locus in *Thalassiosira pseudonana*. Vector Tp:Cas9_sgRNA expressed Cas9 nuclease and sgRNA to target gene. Vector 5′flank_TpFCP:*NAT*:3′flank carried the exogenous homology matrix containing FCP:*NAT* resistance cassette and noncoding 5′/3′ flanking regions. Red arrows depict primers used in the screening strategy by PCR, labels correspond to Table 1.

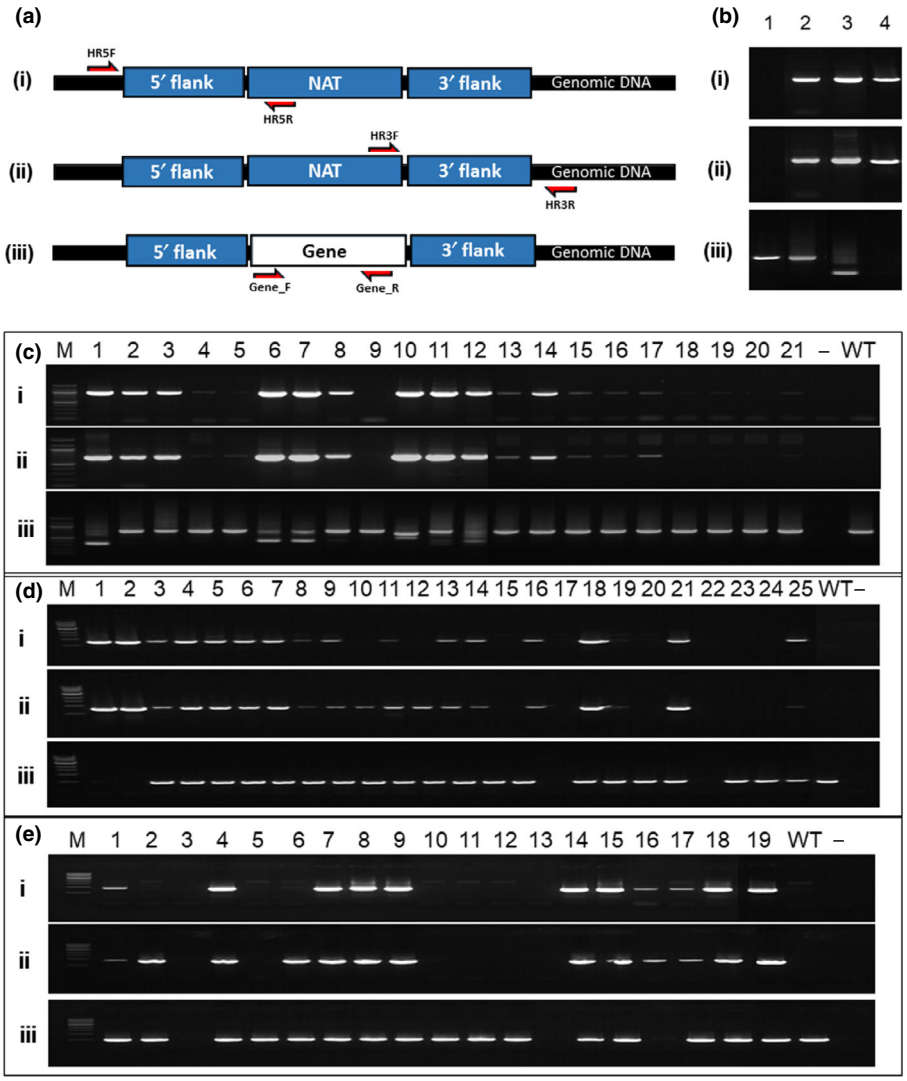

Fig. 3 Screening *Thalassiosira pseudonana* transformants for homologous recombination (HR) and gene deletion by PCR. Primers target genomic region up/downstream of 5′/3′ flanking regions used for the HR construct in combination with FCP:*NAT* resistance gene cassette ((a) (i) HR5F/HR5R and (ii) HR3F/HR3R, respectively) and a region of each gene's coding sequences ((a) (iii) Gene_F/Gene_R). (b) Potential screening outcomes, no gene replacement (lane 1), monoallelic gene replacement with wild-type retained copy (lane 2), monoallelic gene replacement with edited retained copy (lane 3) and biallelic gene replacement (lane 4). (c) (i) Homologous recombination at silacidin 5′, primer pair 24/20, (ii) HR at silacidin 3′, primer pair 21/25 and (iii) silacidin gene region, primer pair 26/27. (d) (i) Homologous recombination at *NitR* 5′, primer pair 30/20, (ii) HR at *NitR* 3′, primer pair 21/31 and (iii) *NitR* gene region, primer pair 32/33. (e) (i) Homologous recombination at urease 5′, primer pair 36/20, (ii) HR at urease 3′, primer pair 21/37 and (iii) urease gene region, primer pair 38/39. M, 100-bp or 1-kb size ladder; –, no template control; WT, wild-type.

dried on Isopore™ membrane filters (Merck Life Science UK Ltd), coated with gold and imaged with a SUPRA55 SEM (Zeiss).

Phenotyping nitrate reductase and urease transformants based on the utilisation of nitrate and urea

Selected *NitR* and urease mutant lines and WT cells were grown in ½SOW Aquil medium containing NH_4Cl as the nitrogen source until the cells reached stationary phase. The cells were transferred to nitrogen-free ½SOW Aquil medium and incubated for 48 h. 5×10^5 cells were used to inoculate ½SOW Aquil medium containing either $NaNO_3$ (1 mmol l^{-1}), $NaNO_2$ (1 mmol l^{-1}) or urea (0.5 mmol l^{-1}). Cell numbers were measured using a Beckmann Multisizer (Meritics Ltd, Leighton Buzzard, UK) 4E with a 100 µm aperture tube. Briefly, 1 ml of culture was diluted with 19 ml of salt water (33 PSU) in a coulter vial and mixed by inversion. The coulter counter was set to take three measurements from the vial, and an average cell density was calculated.

Results

Eighteen out of 21 *T. pseudonana* transformant colonies from the cotransformation with pAGM4723_Cas9_SIL and pAGM4723_SIL_NAT_donor targeting the silacidin locus screened positive for HR (Fig. 3c(i,ii)). The entire silacidin locus was amplified with primer pair 22/23 (Table 1) in 21 transformants capable of growth on selective medium. These amplicons were used as templates for two further nested PCRs with primer pairs 24/20 and 21/25 (Table 1). The primers were designed to cover the transition from the 3′/5′ end of the silacidin noncoding sequence to the resistance cassette contained within the exogenous HR donor matrix (Fig. 2); therefore, the production of specific amplicons indicates that the NAT resistance cassette has integrated precisely at the silacidin locus by HR within the flanking sequences. The results showed that correct, targeted HR had taken place at both 3′ and 5′ ends between the WT locus and the donor sequence in 85% of cases (Fig. 3c(i,ii)). The WT was used as negative control because the amplified combination of up/downstream noncoding silacidin sequences

interrupted by a selective marker does not occur naturally (Fig. 2).

Screening for the presence of silacidin coding sequences (PCR with primer pair 26/27) showed that the WT-sized sequence was present in 18 transformants and absent in three (Fig. 3c(iii) lanes 1, 10 and 12). The PCR with transformants 1 and 10 (S1 and S10) together with S6 and S7 produced shorter bands (Fig. 3c (iii) lanes 1, 6, 7, 10). S12 produced a smear suggesting nonspecific amplification (Fig. 3c(iii) lane 12). Attempts to clone and sequence a specific silacidin fragment failed, which together with the positive amplification of NAT at the silacidin locus (Fig. 3c(i, ii) lane 12) suggests that the silacidin gene is absent in this transformant and has been replaced by NAT. Sequencing data for the shorter bands from transformants S1, S6, S7 and S10 together with the WT-sized band in S2 confirmed that the smaller products were all truncated versions of silacidin and not nonspecific products, while the S2 amplicon was identical to the WT sequence (Fig. S1). This was expected because previous studies in diatoms showed either small deletions after targeting two sites (Hopes *et al.*, 2016) or larger deletions (Nymark *et al.*, 2016) targeting one site, following Cas9 nuclease activity. Furthermore, in plants, other repair outcomes were also shown to be possible (Schiml *et al.*, 2014).

Taken together, the above results revealed that in some cases (e.g. transformants S6 and S7), three possible outcomes of targeted gene editing were detected per clone (HR-mediated replacement of silacidin by NAT as revealed by PCR amplicons in Fig. 3c(i,ii), WT silacidin and truncated silacidin as shown by two bands in Fig. 3c(iii)). Because *T. pseudonana* is a diploid, only two outcomes are expected to occur per cell, affecting either allele. This suggests that some transformant colonies were mosaic, that is contained a mixed population of cells with differing genotypes.

NAT-resistant colonies from the cotransformation with pAGM4723_NitR_NAT_donor and either pAGM4723_Cas9_NitR-1 or pAGM4723_Cas9_NitR-2 expressing one or two gRNAs targeting *NitR* were screened for HR using nested PCR as before for silacidin but with primers 28–33 replacing primers 22–27, respectively (primers shown in Table 1). Fig. 3(d) shows representative gels from the screening of some transformants; bands in Fig. 3(d(i,ii)) indicate the presence of NAT in the *NitR* locus suggesting the replacement of at least one allele of *NitR* by NAT, whereas the presence or absence of a band in Fig. 3(d(iii)) indicates the retention or loss of the *NitR* gene, respectively. Therefore, monoallelic HR-mediated gene replacement is indicated by bands produced by all three PCRs (e.g. Fig. 3d lane 3), whereas biallelic gene replacement would lead to bands in Fig. 3 (d(i,ii)) only with no amplicon in Fig. 3(d(iii)) (e.g. Fig. 3b lane 4). The absence of a product from any of the three PCRs (e.g. Fig. 3d(i–iii) lane 22) indicates a lack of amplifiable DNA from the transformant probably resulting from a failure to grow. Twenty out of 54 colonies from the transformation with a single gRNA and 23 out of 46 transformants obtained with two gRNAs were positive for HR, respectively (data not shown). Further PCRs for a fragment of the *NitR* gene (primers 32 and 33, Table 1) showed that the *NitR* gene was absent in seven and two transformants with either one or two gRNAs, respectively (data not shown). Taken together, this indicates that 13 of the 54 (24%) transformants had one allele of *NitR* replaced and seven of the 54 (13%) had both alleles replaced when one gRNA was expressed. Transformation with the plasmid expressing two gRNAs led to 21 (46%) and 2 (4%) of the 46 transformants containing monoallelic or biallelic replacement of *NitR*, respectively. This slight increase in the overall efficiency of HR from 37% to 50% with two gRNAs may simply result from the increased probability of a Cas9-induced DSB with two gRNAs compared with only one.

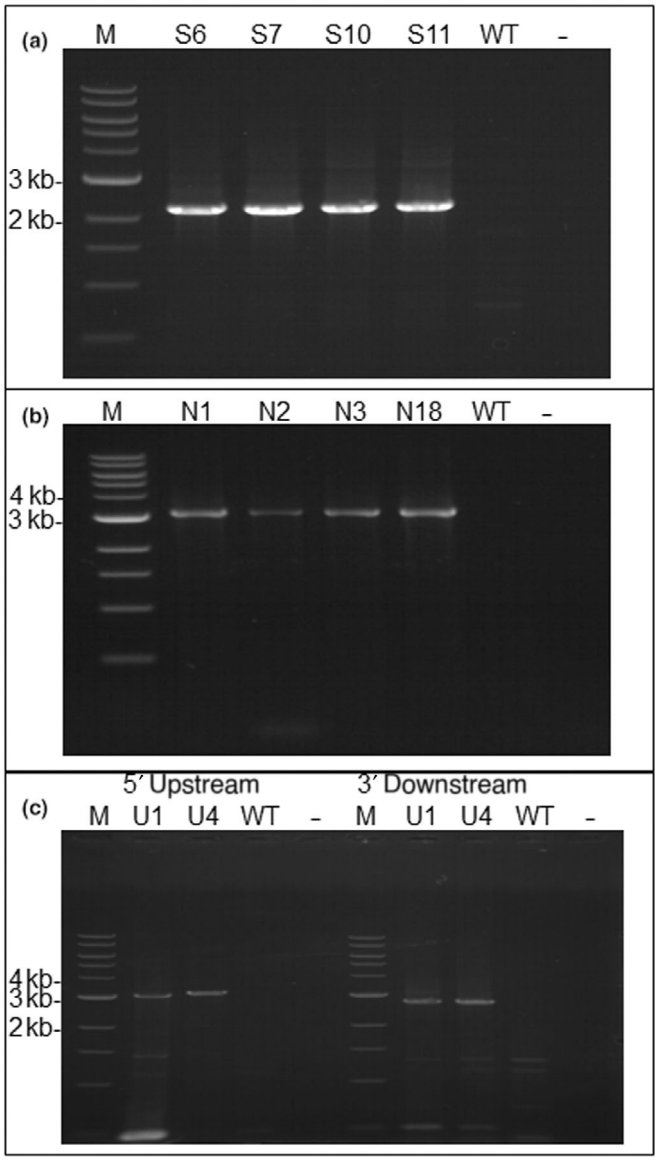

Fig. 4 Confirmation of homologous recombination-mediated integration of FCP:*NAT* resistance gene cassette at target loci in *Thalassiosira pseudonana* by inverse PCR. PCR was performed with divergent primers to the FCP:*NAT* resistance gene cassette and restriction enzyme digested and ligated (circularised) genomic DNA from selected silacidin (a), *NitR* (b) and urease (c) transformants. M, 1-kb size ladder; –, no template control; WT, wild-type.

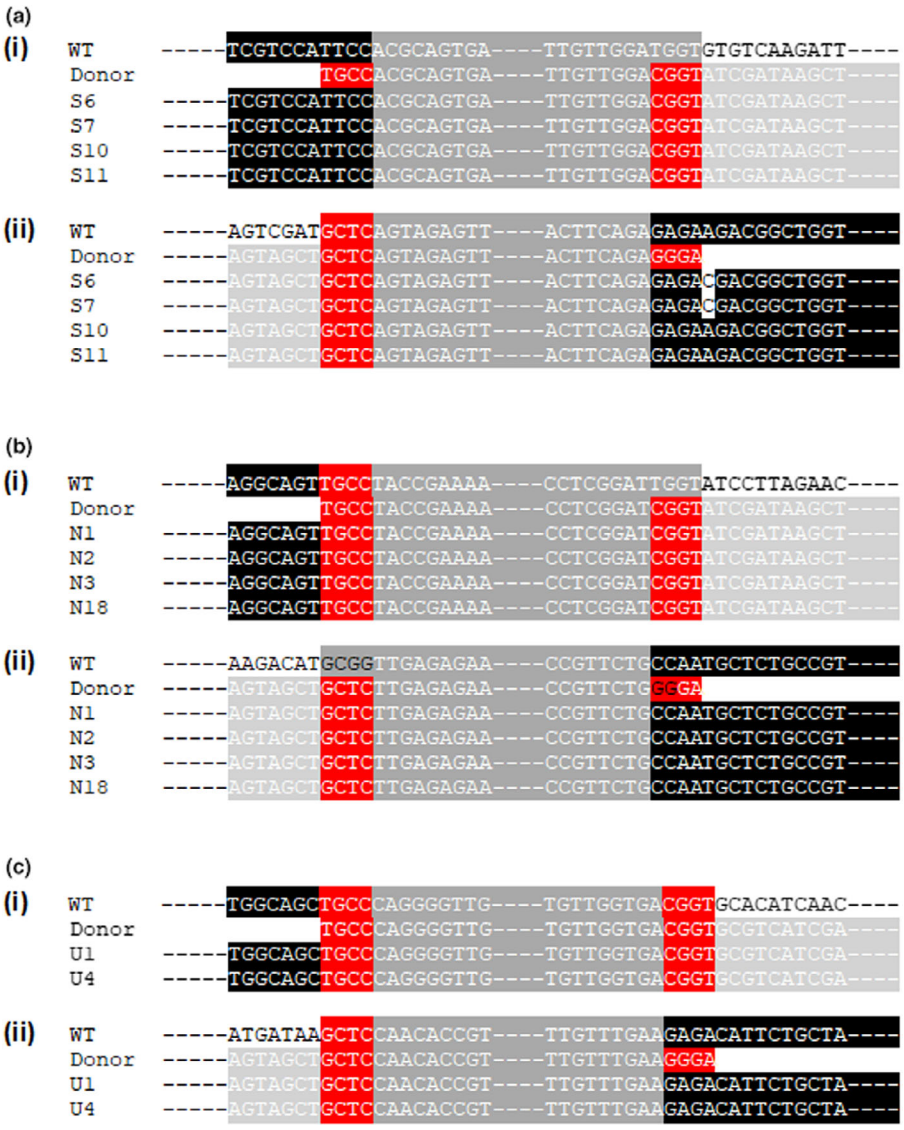

Fig. 5 DNA sequences of the 5′ and 3′ ends of the flanking regions for the inverse PCR amplicons from each of the *Thalassiosira pseudonana* transformants aligned with DNA sequences of the wild-type (WT) and targeting plasmids. (a) Silacidin, (b) *NitR* and (c) urease. (i) 5′ flanking regions and (ii) 3′ flanking regions. Black shading, sequences upstream/downstream of the 5′/3′ flanking regions; light grey shading, sequences of the 5′/3′ flanking regions; dark grey shading, sequences of the FCP:*NAT* resistance cassette; red shading, overhang sequences for directional cloning.

Table 2 Cell size and valve diameter of *Thalassiosira pseudonana* wild-type (WT) and silacidin transformant lines (S1, S6, S7, S10, S11 and S12).

Colony no.	Coulter counter Cell size (μm) (average)	FACS Forward scatter (au) (average)	SEM Valve diameter (μm) (average)
WT	4.70 ± 0.01	386 ± 3	3.3 ± 0.7
S1	5.44 ± 0.006***	424 ± 2*	na
S6	5.40 ± 0.006***	478 ± 14***	na
S7	5.08 ± 0.017***	397 ± 13	na
S10	5.40 ± 0.04***	437 ± 12**	na
S11	5.49 ± 0.05***	455 ± 14***	na
S12	6.10 ± 0.02***	625 ± 23***	4.0 ± 0.5

FACS, fluorescence-activated cell sorting; SEM, scanning electron microscopy. Average values of three biological and three technical replicates. SEM: average of 150 cells. na, not applicable. The statistical test used was ANOVA followed by Dunnett's *post-hoc* test. Significance: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Values ± SD.

Cloning and sequencing of *NitR* amplicons from two transformants showed N3, generated using only one gRNA, contained a 32-bp deletion producing a frameshift mutation that leads to amino acid changes from position 342 to 349 followed by a premature stop codon generating a truncated form of nitrate reductase (Fig. S2). N18, produced with two gRNAs, contained a 183-bp deletion leading to the in-frame deletion of amino acids 350–410 (Fig. S2). The deletion resulted from the precise Cas9-induced excision as the target sites for the two gRNAs are separated by 183 bp.

Eighty-eight NAT-resistant colonies from the cotransformation with pAGM4723_Cas9_Urease and pAGM4723_Urease_NAT_donor targeting the urease gene were screened by PCR as before but with primers 34–39 replacing primers 22–27, respectively (primers shown in Table 1). Fig. 3(e) shows representative gels from the screening of some transformants; bands in Fig. 3(e(i,ii)) indicate the presence of NAT in the urease locus suggesting the replacement of at least one allele of urease by NAT, whereas the presence or absence of a band in Fig. 3(e(iii)) indicates the retention or loss of the urease gene, respectively. Therefore, monoallelic HR-mediated gene replacement is indicated by bands produced by all three PCRs (e.g. Fig. 3e lane 4), whereas biallelic gene replacement would lead to bands in Fig. 3(e(i,ii)) only with no amplicon in Fig. 3(e(iii)) (e.g. lane 16). Forty-four

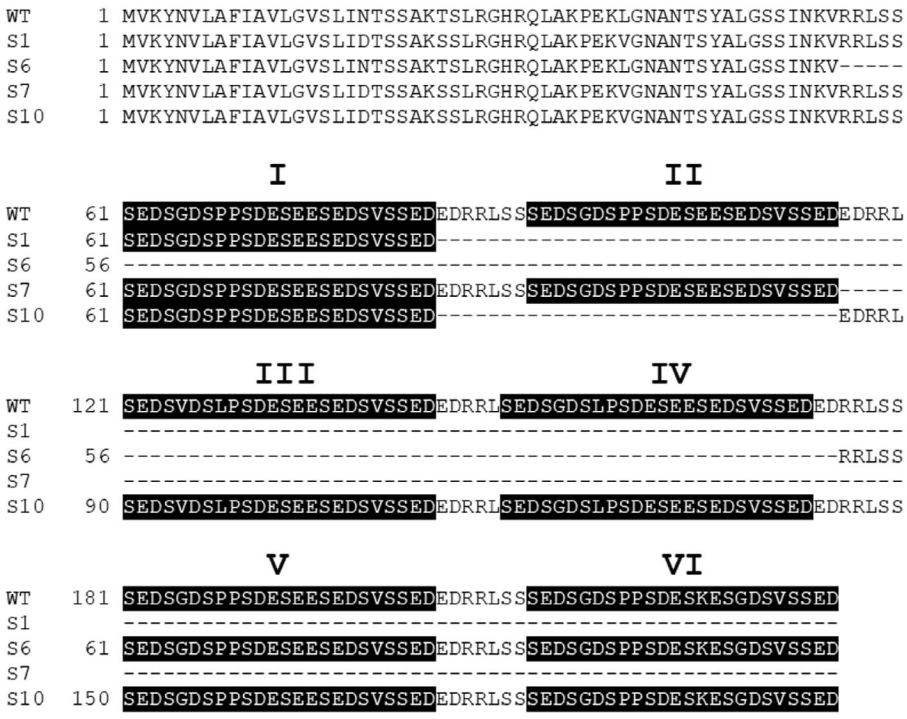

Fig. 6 Predicted protein sequences of *Thalassiosira pseudonana* silacidin for the truncated gene fragments from the silacidin transformants aligned with the wild-type (WT) sequence. Black shading, protein sequence of the repetitive motif. I–VI indicates the number of repetitive motifs.

transformants were positive for HR indicated by the PCR-detected presence of NAT at the urease locus. A region of the urease gene was amplified by PCR in all but one of the HR-positive transformants (Fig. 3e(iii) lane 16) indicating that monoallelic replacement of urease occurred in 49% of transformants but only one transformant contained a biallelic knockout of urease; however, this transformant did not grow when passaged either onto agar plate or in liquid medium suggesting the knockout of the urease gene significantly impacted cell viability.

Urease amplicons from two transformants were cloned and sequenced. The amplicon from transformant U4 matched the WT sequence but that from transformant U1 contained a 26-bp deletion in exon 2, which upon translation would lead to a frameshift mutation after 13 amino acids and a premature stop after 48 amino acids. However, a putative alternative start codon is present in exon 3, which is retained in this transformant; translation from this would lead to an N-terminal truncation of the first 87 amino acids of the urease protein (Fig. S3).

Inverse PCR is a technique used to amplify unknown DNA sequences adjacent to a region where the DNA sequence is known. It involves restriction enzyme digestion of genomic DNA with an enzyme that cuts outside the known sequences, followed by circularisation of the digested DNA fragments with DNA ligase. PCR with divergent primers to the known sequences then amplifies DNA sequences immediately upstream and downstream of the known DNA sequences. Here, we used iPCR to identify sites of integration of the FCP:*NAT* resistance cassette into the genome of the *T. pseudonana* transformants. Genomic DNA from silacidin transformants S6, S7, S10 and S11, *NitR* transformants N1, N2, N3 and N18 and urease transformants U1 and U4 was digested with MfeI, BstBI and KpnI restriction enzymes, respectively. There are no sites for MfeI and BstBI digestion within the FCP:*NAT* cassette, but there is a single site for KpnI. Based on the *T. pseudonana* genome sequence, MfeI is predicted to cut 76-bp upstream and 482-bp downstream of the silacidin 5′ and 3′ flanking regions, respectively. BstBI is predicted to cut 129-bp upstream and 1460-bp downstream of the *NitR* 5′ and 3′ flanking regions, respectively. KpnI is predicted to cut 2267-bp upstream and 1715-bp downstream of the urease 5′ and 3′ flanking regions, respectively. Following circularisation, PCRs with primers 20/21 for silacidin and *NitR*, 20/40 for the urease upstream region and 21/41 for the urease downstream region (Table 1) generated the amplicons shown in Fig. 4. The predicted sizes of the amplicons following the integration of the FCP:*NAT* resistance cassette at each target are 2199 bp for silacidin, 3195 bp for *NitR*, 3181 bp for the urease upstream region and 2795 bp for the urease downstream region. Fig. 4 shows that the amplicons were only obtained with transformant DNA, with no amplification with digested and circularised WT DNA, and corresponded to the predicted sizes suggesting the FCP:*NAT* resistance cassette had integrated into each target locus. The presence of unique bands for each PCR suggests a single site of integration of the FCP:*NAT* resistance cassette in each of the transformants. Cloning and sequencing of the amplicons confirmed the accurate integration of the FCP:*NAT* resistance cassette into each target locus leading to the replacement of each target gene (Fig. S4). Several of the primers, listed in Table 1, used to amplify the flanking regions for silacidin, *NitR* and urease modified the 5′ and 3′ ends to introduce compatible overhangs for directional cloning into the vector containing the FCP:*NAT* resistance cassette. The sequences of the iPCR products showed that, for each of the genes, the modified sequences in the flanking regions adjacent to the FCP:*NAT* cassette were

retained, whereas those at the 5′ end of the 5′ flanks and at the 3′ end of the 3′ flanks have been lost and reverted to WT sequences (Fig. 5) providing confirmation that the FCP:*NAT* resistance cassette has integrated at each locus via HR within the 5′ and 3′ flanking regions.

Phenotyping of selected transformants with HR at the silacidin gene showed that transformant cells were significantly larger than WT (Table 2). This was most pronounced in transformant S12 where the biallelic replacement of the silacidin gene was identified (Fig. 3c(i)). Quantification of all silacidin PCR products, including truncated versions, from WT and HR transformants S1, S6, S7, S10 and S11 gave a significant negative correlation (Fig. 7a, $R^2 = 0.8$, $P = 0.007$) between the silacidin amplicon size, given in base pairs, and cell size measured by fluorescence-activated cell sorting scatter. For a more detailed analysis of functionality, truncated versions of the silacidin gene were translated into the corresponding amino acid sequence. The resulting polypeptides were aligned with the original WT protein sequence (Fig. 6). This showed that, if expressed and translated, the truncated silacidins retained a different number of repetitive peptide motifs following the overall amino acid sequence: (SS) SEDSXDSXPSDESEESEDSVSSED(ED). Of the six copies of this motif present in the WT protein, only a single copy was retained in transformant S1, two copies in S6 and S7 and five remaining copies in S10. The difference in cell size between WT and HR transformants was caused by differences in the diameter of the valves, which are overlapping sections of the cell walls also known as thecae (Fig. 7).

Nitrate reductase catalyses the reduction in nitrate to nitrite and is, therefore, a critical step in the nitrogen assimilation pathway that is essential for growth on nitrate as the sole nitrogen source. The effects of the biallelic replacement of *NitR* by HR and also the monoallelic replacement, together with Cas9-induced mutations of the remaining copy on the growth of *T. pseudonana* on nitrate- or nitrite-containing medium, were compared. Fig. 8(a) shows that, as expected, transformant cell growth on nitrite was not significantly affected. However, none of the transformants exhibited cell growth in nitrate-containing medium indicating that nitrate reductase had been inactivated in each of the transformants including N3 and N18 containing Cas9-induced mutations in the remaining copy of *NitR*.

Urease catalyses the hydrolysis of urea into carbon dioxide and ammonia, thereby allowing the utilisation of urea as the sole nitrogen source. The HR-mediated replacement of one copy of the urease gene (transformant U4) led to a significant decrease in the rate of cell growth on both urea and nitrate, which was further significantly diminished in transformant U1, which contained a Cas9-induced deletion in the remaining copy of the urease gene leading to the predicted expression of an N-terminal-truncated version of the urease enzyme (Fig. 8b).

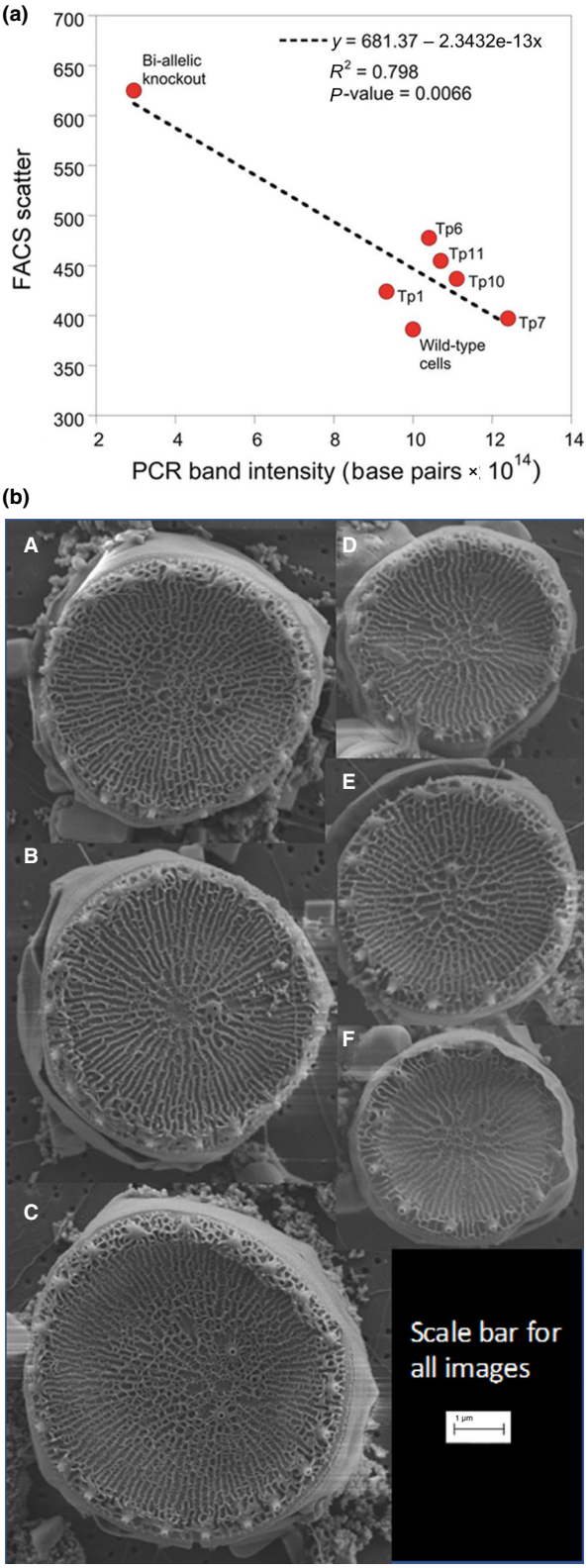

Fig. 7 Linking genotype to phenotype in *Thalassiosira pseudonana* silacidin transformant clones. (a) Correlation between cell size (measured by fluorescence-activated cell sorting (FACS)) and base pair per silacidin amplicon from PCR with primer pair 26/27. (b) Representative scanning electron microscope images of *T. pseudonana* valves. Individual cells from homologous recombination (HR) clone S12 (A–C) and wild-type cells (D–F). Bar, 1 µm.

Discussion

Although the applicability of CRISPR/Cas9 for genome editing in *T. pseudonana* has recently been demonstrated (Hopes

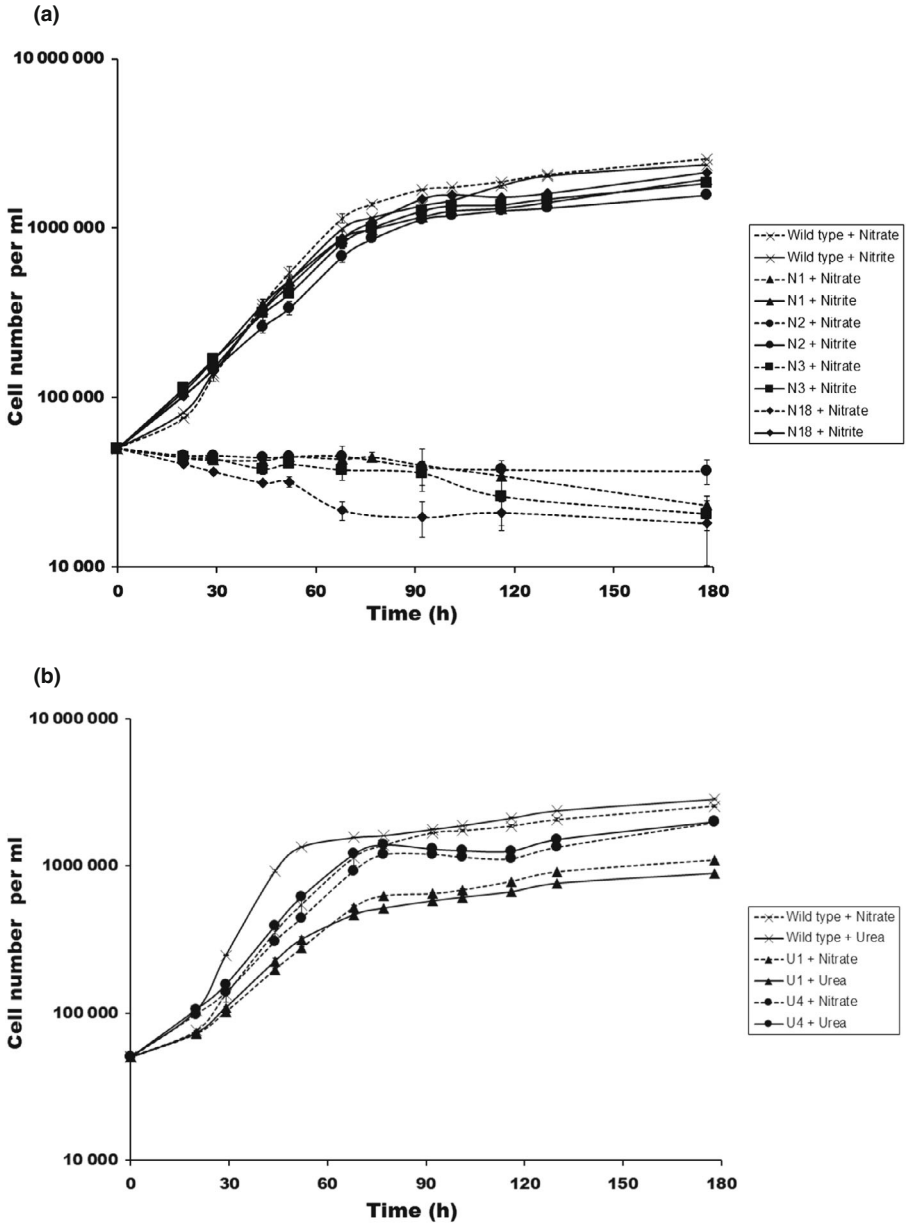

Fig. 8 Effect of homologous recombination-mediated gene deletion and Cas9-induced editing on *Thalassiosira pseudonana* cell growth of (a) *NitR* transformants and (b) urease transformants on nitrate- or urea-containing medium, compared with the wild-type (WT). Data represent mean ± SEM ($n = 3$).

et al., 2016), the current work provides the first evidence for highly efficient CRISPR/Cas9-mediated HR in *T. pseudonana*. This result was achieved by designing exogenous donor matrices to substitute the cell-wall gene silacidin, the nitrate reductase gene and the urease gene with a resistance cassette encoding the NAT gene using Golden Gate cloning. Similar experiments with the diploid diatom *P. tricornutum*, which were based on TALEN-mediated HR, resulted in a maximum HR efficiency of only 27% when taking into account the number of colonies, which tested positive for HR on selective plates (Daboussi *et al.*, 2014; Table 3). In the present study, however, CRISPR/Cas9-mediated HR was detected in up to 85% of obtained colonies, making it the most efficient incidence of HR reported for any photosynthetic eukaryote with a predominantly diploid life-cycle phase. Other photosynthetic eukaryotes with a similar HR efficiency were either haploids such as *Nannochloropsis* sp. (Kilian *et al.*, 2011; Naduthodi *et al.*, 2019) or at least had a predominantly haploid life phase such as *P. patens* (Schaefer & Zryd, 1997; Kamisugi *et al.*, 2005). To date, the highest HR efficiency mediated by any nuclease among all photosynthetic eukaryotes was observed in *Nannochloropsis oceanica* (Naduthodi *et al.*, 2019). Thus, the results of our study represent an unexpectedly high rate of HR events overall and compared with non-homologous end joining in particular. Usually, nonhomologous end joining is expected to be the primary repair mechanism if Cas9, or any other exogenous nuclease, induces breaks within the target DNA sequence. Nonhomologous end joining was reported to be preferred over HR in human cells and diatoms (Mao *et al.*, 2008; Daboussi *et al.*, 2014). In our study, evidence for nonhomologous end joining was detected in the *NitR* and urease genes, where significant deletions were observed. However, based on our sequencing results, the truncated forms of the silacidin

Table 3 Overview of homologous recombination (HR) occurrence and efficiency in photosynthetic eukaryotes.

Species	Ploidy	Nuclease(s)	Homologous flanking sequence size	Efficiency[1]	Reference
Thalassiosira pseudonana	2	SpCas9	0.63–0.87 kb	85%	Present study
Phaeodactylum tricornutum	2	TALEN	1 kb	24%	Daboussi *et al*. (2014);
		TALEN, meganucleases	0.75 kb	27%	Weyman *et al*. (2015)
P. tricornutum	2	SpCas9	1 kb	17%	Moosburner *et al*. (2020)
Nannochloropsis sp.	1	–	1–1.4 kb	11–94%	Kilian *et al*. (2011)
Nannochloropsis oceanica IMET1	1	FnCas12a	1 kb	93%	Naduthodi *et al*. (2019)
Cyanidioschyzon merolae 10D	1	–	–[2]	+	Minoda *et al*. (2004)
Chlamydomonas reinhardtii	1 (2)	Zinc-finger	0.1–0.5 kb	8%	Greiner *et al*. (2017)
		Sp/SaCas9[3]		9%	
C. reinhardtii	1 (2)	LbCas12a (CpfI)	45–49 nt (ssODN)	18%	Ferenczi *et al*. (2017)
C. reinhardtii	1 (2)	SpCas9	2 kb	5%	Angstenberger *et al*. (2020)
Ostreococcus tauri	1	–	0.25–2.5 kb	1–4%	Lozano *et al*. (2014)
Physcomitrella patens	1 (2)	–	2.3–3.6 kb	66–100%	Schaefer & Zryd (1997);
		–	0.5–1 kb	70–85%	Kamisugi *et al*. (2005)
Nicotiana benthamiana	4	SpCas9	0.11–0.53 kb	10.7%	Li *et al*. (2013)
Arabidopsis thaliana	2	SpCas9	0.67 kb	0.14%	Schiml *et al*. (2014)
A. thaliana	2	SpCas9	?	9.1%	Miki *et al*. (2018)
Oryza sativa	2	FnCpf1, LbCpf1 nucleases	1 kb	8%	Begemann *et al*. (2017)
O. sativa	2	Cas9	75 nt (RNA)	16%	Butt *et al*. (2017)
Zea mays	2	Cas9	0.41–0.42 kb	18%	Barone *et al*. (2020)

Fn, *Fransicella novicida*; Lb, Lachnospiraceae bacterium; Sa, *Staphylococcus aureus*; Sp, *Streptococcus pyogenes*. Heterokonta, brown; Rhodophyta, red; Plantae, green; ?, flanking size not provided. Bold value is this study.
[1]Out of number of obtained transformant colonies targeting genomic wild-type (WT) loci.
[2]Entire sequence was exchanged, HR between WT sequence and sequence containing point mutation in resistance gene.
[3]Transient expression (heat-shock promoters).

gene seemed to be caused by HR between the direct repeats of the silacidin gene leading to 'looping out' and deletion of one or more of the six repeats.

A number of factors are thought to influence the efficiency of gene replacement by HR including the efficiency of Cas9 digestion, the size of the homologous flanking regions and the distance between the gRNA target site and the homologous flanking regions. The efficiency of Cas9 digestion of each target was not evaluated in this study, but it should be expected that using two sgRNAs to target Cas9 to the same region would lead to the introduction of a greater number of DSBs than using only one sgRNA. An increase in the observed efficiency of HR at the *NitR* gene from 37% to 50% with two sgRNAs compared with one provides support for this; however, this requires further study. The sizes of the homologous flanking regions used here were similar for each of the three genes and comparable to those used to target urease in the diatom *P. tricornutum* (Weyman *et al*., 2015). It has been suggested that the homologous flanking sequences should be at least 400 bp for efficient HR; however, recent studies have observed efficient HR in other organisms with flanking regions as small as 50 bp (Picariello *et al*., 2020).

Interestingly, phenotyping the WT, complete biallelic HR and HR colonies with different truncated versions of the silacidin gene revealed a significant negative correlation (P-value = 0.007) between the level of mutations caused by HR (genotype) and the cell size expressed as the intensity of light scattering (phenotype) (Fig. 7a). This correlation suggests that the truncated versions of the silacidin gene (shorter bands in Fig. 3c(iii) lanes 1, 6, 7 and 10) were still functional in the HR cell lines. The largest cell size (Table 2) was only observed when the silacidin gene was completely replaced by the FCP:*NAT* cassette as observed for transformant S12 (Fig. 3c). This is supported by results from Kirkham *et al*. (2017) in which valve size increased after silencing silacidin in *T. pseudonana*. This paper also showed a correlation in size linked to downregulation of the silacidin gene following experimental evolution under low temperatures. The repetitive nature of the silacidin gene in addition to known post-translational modifications of the encoded protein might explain the significant link between genotype and phenotype. The truncated sequences displayed several different deletions, which nevertheless were still in frame and therefore probably translated into shorter versions of the repetitive WT silacidin protein sequence (Fig. 6). Repetitive motifs that undergo extensive post-translational modification such as phosphorylation ((SS)SEDSXDSXPSDESEESEDSVSSED(ED)) were shown to be the functional units of silacidin proteins responsible for silica precipitation in the cell wall (Fig. 6; Wenzl *et al*., 2008; Richthammer *et al*., 2011). Even beyond this exact peptide motif, the overall number of serines and acidic amino acids is most likely contributing to the functionality of the truncated versions. These data represent the first example of a direct link between genotype and phenotype with respect to how mutations in a single gene encoding a cell-wall protein impact the cell size of a diatom.

The inactivation of nitrate reductase activity in transformants N3 (truncated form of nitrate reductase with a 32-bp deletion and a frameshift mutation that leads to amino acid changes from position 342 to 349 followed by a premature stop codon) and N18 (183-bp deletion leading to the in-frame deletion of amino

acids 350–410) indicates that deletion of 61 amino acids (350–410) from the *T. pseudonana* nitrate reductase enzyme, which occurs in both transformants, induces a loss of activity. This suggests that this region of the protein is critical either for enzyme activity or its loss may impact the correct folding of the remaining polypeptide.

Despite screening a large number of transformants, only one transformant showed evidence of the biallelic replacement of urease and we were unable to passage this transformant further suggesting a loss of viability. This together with the observed significant impact on growth on both nitrate and urea as nitrogen source of the transformants retaining only one copy of the urease gene, which was further affected by the N-terminal truncation of the urease enzyme, raises the possibility that urease activity is essential for *T. pseudonana* cell survival. This is in contrast to the TALEN-mediated knockout of urease in the distantly related diatom *P. tricornutum*, which while unable to grow on urea, grew normally on nitrate (Weyman *et al.*, 2015). Our results also indicate that two copies of the urease gene are required for optimal growth of *T. pseudonana* on both nitrogen sources and also that the N-terminal truncation of the urease enzyme leads to at least some retention of activity. Our previous work targeting urease with CRISPR–Cas in *T. pseudonana* gave clear biallelic knockouts; however, biallelic mutants were still able to grow on urea but with a reduced growth rate. Investigation into the sequence showed that while the deletion, found within the gamma subunit, and frameshift were expected to result in a nonfunctional protein, alternative start sites would still allow the beta subunit and alpha subunit, containing the active site, to be expressed (Hopes *et al.*, 2016). This supports the hypothesis that urease is essential in *T. pseudonana*.

Our work presented here will contribute to firmly establishing *T. pseudonana* as a model not only for bionanotechnology, but also for questions addressing fundamental diatom biology including applications targeted at harnessing the metabolic potential of diatoms.

Acknowledgements

TM acknowledges support from the Gordon and Betty Moore Foundation (grant no. 4961) and support received from the Natural Environment Research Council (NERC) (NE/K013734/1). AH acknowledges funding from NERC for her PhD studentship.

Competing interests

None declared.

Author contributions

TM and AH conceived the project. NB and TM wrote the paper jointly with IG with contributions from AH and AG. NB and AH designed and conducted the laboratory and bioinformatics work with help from IG. LA conducted the cell-size measurements and produced the scanning electron micrographs.

ORCID

Assaf Gal https://orcid.org/0000-0003-1488-1227
Amanda Hopes https://orcid.org/0000-0001-5859-3951
Thomas Mock https://orcid.org/0000-0001-9604-0362

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Allen AE, Dupont CL, Oborník M, Horák A, Nunes-Nesi A, McCrow JP, Zheng H, Johnson DA, Hu H, Fernie AR *et al.* 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **473**: 203–207.

Angstenberger M, de Signori F, Vecchi V, Dall'Osto L, Bassi R. 2020. Cell synchronization enhances nuclear transformation and genome editing via Cas9 enabling homologous recombination in *Chlamydomonas reinhardtii*. *ACS Synthetic Biology* **9**: 2840–2850.

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M *et al.* 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**: 79–86.

Bae S, Park J, Kim JS. 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**: 1473–1475.

Barone P, Wu E, Lenderts B, Anand A, Gordon-Kamm W, Svitashev S, Kumar S. 2020. Efficient gene targeting in maize using inducible CRISPR-Cas9 and marker-free donor template. *Molecular Plant* **13**: 1219–1227.

Begemann MB, Gray BN, January E, Gordon GC, He Y, Liu H, Wu X, Brutnell TP, Mockler TC, Oufattole M. 2017. Precise insertion and guided editing of higher plant genomes using Cpf1 CRISPR nucleases. *Scientific Reports* **7**: 11606.

Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V. 2013. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* **9**: 39.

Bhattacharya D, Archibald JM, Weber APM, Reyes-Prieto A. 2007. How do endosymbionts become organelles? Understanding early events in plastid evolution. *BioEssays* **29**: 1239–1246.

Bozarth A, Maier UG, Zauner S. 2009. Diatoms in biotechnology: modern tools and applications. *Applied Microbiology and Biotechnology* **82**: 195–201.

Butt H, Eid A, Ali Z, Atia MAM, Mokhtar MM, Hassan N, Lee CM, Bao G, Mahfouz MM. 2017. Efficient CRISPR/Cas9-mediated genome editing using a chimeric single-guide RNA molecule. *Frontiers in Plant Science* **8**: 1441.

Daboussi F, Leduc S, Maréchal A, Dubois G, Guyot V, Perez-Michaut C, Amato A, Falciatore A, Juillerat A, Beurdeley M *et al.* 2014. Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology. *Nature Communications* **5**: 3831.

Doudna JA, Charpentier E. 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**: 1077–1086.

Falkowski PG, Raven JA. 2013. *Aquatic photosynthesis.* Princeton, NJ, USA: Princeton University Press.

Ferenczi A, Pyott DE, Xipnitou A, Molnar A, Merchant SS. 2017. Efficient targeted DNA editing and replacement in *Chlamydomonas reinhardtii* using Cpf1 ribonucleoproteins and single-stranded DNA. *Proceedings of the National Academy of Sciences, USA* **114**: 13567–13572.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**: 237–240.

Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zauner S, Maier UG, Grossman AR, Bhattacharya D, Lohr M. 2008. Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Molecular Biology and Evolution* **25**: 2653–2667.

Gaj T, Gersbach CA, Barbas CF. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology* **31**: 397–405.

Greiner A, Kelterborn S, Evers H, Kreimer G, Sizova I, Hegemann P. 2017. Targeting of photoreceptor genes in *Chlamydomonas reinhardtii* via zinc-finger nucleases and CRISPR/Cas9. *Plant Cell* **29**: 2498–2518.

Hopes A, Nekrasov V, Belshaw N, Grouneva I, Kamoun S, Mock T. 2017. Genome editing in diatoms using CRISPR-Cas to induce precise bi-allelic deletions. *Bio-Protocol* **7**: e2625.

Hopes A, Nekrasov V, Kamoun S, Mock T. 2016. Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant Methods* **12**: 49.

Ichimi K, Kawamura T, Yamamoto A, Tada K, Harrison PJ. 2012. Extremely high growth rate of the small diatom *Chaetoceros salsugineum* isolated from an estuary in the eastern Seto Inland Sea, Japan. *Journal of Phycology* **48**: 1284–1288.

Kamisugi Y, Cuming AC, Cove DJ. 2005. Parameters determining the efficiency of gene targeting in the moss *Physcomitrella patens*. *Nucleic Acids Research* **33**: 1–10.

Kamisugi Y, Schlink K, Rensing SA, Schween G, von Stackelberg M, Cuming AC, Reski R, Cove DJ. 2006. The mechanism of gene targeting in *Physcomitrella patens*: homologous recombination, concatenation and multiple integration. *Nucleic Acids Research* **34**: 6205–6214.

Kamisugi Y, Whitaker JW, Cuming AC. 2016. The transcriptional response to DNA double-strand breaks in *Physcomitrella patens*. *PLoS ONE* **11**: e0161204.

Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **365**: 729–748.

Kilian O, Benemann CSE, Niyogi KK, Vick B. 2011. High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proceedings of the National Academy of Sciences, USA* **108**: 21265–21269.

Kirkham AR, Richthammer P, Schmidt K, Wustmann M, Maeda Y, Hedrich R, Brunner E, Tanaka T, van Pée KH, Falciatore A *et al.* 2017. A role for the cell-wall protein silacidin in cell size of the diatom *Thalassiosira pseudonana*. *ISME Journal* **11**: 2452–2464.

Lander ES. 2016. The heroes of CRISPR. *Cell* **164**: 18–28.

Levitan O, Dinamarca J, Hochman G, Falkowski PG. 2014. Diatoms: a fossil fuel of the future. *Trends in Biotechnology* **32**: 117–124.

Li J-F, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J. 2013. Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nature Biotechnology* **31**: 688–691.

Lozano J-C, Schatt P, Botebol H, Vergé V, Lesuisse E, Blain S, Carré IA, Bouget F-Y. 2014. Efficient gene targeting and removal of foreign DNA by homologous recombination in the picoeukaryote *Ostreococcus*. *The Plant Journal* **78**: 1073–1083.

Mao Z, Bozzella M, Seluanov A, Gorbunova V. 2008. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair* **7**: 1765–1771.

Messina CD, Podlich D, Dong Z, Samples M, Cooper M. 2011. Yield-trait performance landscapes: from theory to application in breeding maize for drought tolerance. *Journal of Experimental Botany* **62**: 855–868.

Miki D, Zhang W, Zeng W, Feng Z, Zhu JK. 2018. CRISPR/Cas9-mediated gene targeting in Arabidopsis using sequential transformation. *Nature Communications* **9**: 1967.

Minoda A, Sakagami R, Yagisawa F, Kuroiwa T, Tanaka K. 2004. Improvement of culture conditions and evidence for nuclear transformation by homologous recombination in a red alga, *Cyanidioschyzon merolae* 10D. *Plant and Cell Physiology* **45**: 667–671.

Mock T, Otillar RP, Strauss J, McMullan M, Paajanen P, Schmutz J, Salamov A, Sanges R, Toseland A, Ward BJ *et al.* 2017. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541**: 536–540.

Moosburner MA, Gholami P, McCarthy JK, Tan M, Bielinski VA, Allen AE. 2020. Multiplexed knockouts in the model diatom *Phaeodactylum* by episomal delivery of a selectable Cas9. *Frontiers in Microbiology* **11**: 5.

Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* **324**: 1724–1726.

Naduthodi MIS, Mohanraju P, Südfeld C, D'Adamo S, Barbosa MJ, van der Oost J. 2019. CRISPR-Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae *Nannochloropsis oceanica* IMET1. *Biotechnology for Biofuels* **12**: 66.

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

Paajanen P, Strauss J, van Oosterhout C, McMullan M, Clark MD, Mock T. 2017. Data descriptor: building a locally diploid genome and transcriptome of the diatom *Fragilariopsis cylindrus*. *Scientific Data* **4**: 170149.

Picariello T, Hou Y, Kubo T, McNeill NA, Yanagisawa HA, Oda T, Witman GB. 2020. TIM, a targeted insertional mutagenesis method utilizing CRISPR/Cas9 in *Chlamydomonas reinhardtii*. *PLoS ONE* **15**: e0232594.

Poulsen N, Chesley PM, Kröger N. 2006. Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae). *Journal of Phycology* **42**: 1059–1065.

Price NM, Harrison GI, Hering JG, Hudson RJ, Nirel PM, Palenik B, Morel FM. 1989. Preparation and chemistry of the artificial algal culture medium Aquil. *Biological Oceanography* **6**: 443–461.

Richthammer P, Börmel M, Brunner E, van Pée KH. 2011. Biomineralization in diatoms: the role of silacidins. *Chembiochem* **12**: 1362–1366.

Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology* **32**: 347–355.

Schaefer DG, Zryd J-P. 1997. Efficient gene targeting in the moss *Physcomitrella patens*. *The Plant Journal* **11**: 1195–1206.

Schiml S, Fauser F, Puchta H. 2014. The CRISPR/Cas system can be used as nuclease for *in planta* gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. *The Plant Journal* **80**: 1139–1150.

Serif M, Lepetit B, Weißert K, Kroth PG, Rio Bartulos C. 2017. A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom *Phaeodactylum tricornutum*. *Algal Research* **23**: 186–195.

Tanaka T, Maeda Y, Veluchamy A, Tanaka M, Abida H, Maréchal E, Bowler C, Muto M, Sunaga Y, Tanaka M *et al.* 2015. Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. *Plant Cell* **27**: 162–176.

Tréguer PJ, De La Rocha CL. 2013. The world ocean silica cycle. *Annual Review of Marine Science* **5**: 477–501.

Wang JK, Seibert M. 2017. Prospects for commercial production of diatoms. *Biotechnology for Biofuels* **10**: 16.

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. 2011. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* **6**: e16765.

Wenzl S, Hett R, Richthammer P, Sumper M. 2008. Silacidins: highly acidic phosphopeptides from diatom shells assist *in silica* precipitation *in vitro*. *Angewandte Chemie International Edition* **47**: 1729–1732.

Weyman PD, Beeri K, Lefebvre SC, Rivera J, McCarthy JK, Heuberger AL, Peers G, Allen AE, Dupont CL. 2015. Inactivation of *Phaeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnology Journal* **13**: 460–470.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Alignment of DNA sequencing data of the *Thalassiosira pseudonana* silacidin amplicon in wild-type and transformants.

Fig. S2 Alignment of DNA sequencing data of the *Thalassiosira pseudonana NitR* amplicon in wild-type and transformants.

Fig. S3 Alignment of DNA sequencing data of the *Thalassiosira pseudonana* urease amplicon in wild-type and transformants.

Fig. S4 Alignment of DNA sequencing data of the inverse PCR amplicons from *Thalassiosira pseudonana* transformants.

Table S1 Plasmids used for cloning, assembly and transformation.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.