Mobilization of a diatom *mutator-like element* (*MULE*) transposon inactivates the uridine monophosphate synthase (UMPS) locus in *Phaeodactylum tricornutum*

Raffaela M. Abbriano[1,†,‡] (iD), Jestin George[1,†,§,¶] (iD), Tim Kahlke[1,†,#] (iD), Audrey S. Commault[1,**] (iD) and Michele Fabris[1,2,*,††] (iD)

[1]Climate Change Cluster, University of Technology, 15 Broadway, Ultimo, NSW 2007, Australia,
[2]CSIRO Synthetic Biology Future Science Platform, GPO Box 2583, Brisbane, QLD 4001, Australia

*For correspondence (e-mail mif@igt.sdu.dk).
[†]These authors contributed equally to this work.
[‡]Present address: Leidos Biological Innovations Center, San Diego, CA 92121, USA
[§]Present address: Department of Molecular Sciences, Macquarie University, Sydney, NSW 2109, Australia
[¶]Present address: ARC Centre of Excellence in Synthetic Biology, Macquarie University, Sydney, NSW 2109, Australia
[#]Present address: Blue Mountains Analytics, 129 Bee Farm Rd, Springwood, NSW 2777, Australia
[**]Present address: Fermentalg, 4 Rue Rivière, 33500, Libourne, France
[††]Present address: SDU Biotechnology, Department of Green Technology, University of Southern Denmark, Campusvej 55 5230, Odense M, Denmark

SUMMARY

Diatoms are photosynthetic unicellular microalgae that drive global ecological phenomena in the biosphere and are emerging as sustainable feedstock for an increasing number of industrial applications. Diatoms exhibit enormous taxonomic and genetic diversity, which often results in peculiar biochemical and biological traits. Transposable elements (TEs) represent a substantial portion of diatom genomes and have been hypothesized to exert a relevant role in enriching genetic diversity and making a core contribution to genome evolution. Here, through long-read whole-genome sequencing, we identified a mutator-like element (MULE) in the model diatom *Phaeodactylum tricornutum*, and we report the direct observation of its mobilization within the course of a single laboratory experiment. Under selective conditions, this TE inactivated the *uridine monophosphate synthase* (*UMPS*) gene of *P. tricornutum*, one of the few endogenous genetic loci currently targeted for selectable auxotrophy for functional genetics and genome-editing applications. We report the observation of a recently mobilized transposon in diatoms with unique features. These include the combined presence of a MULE transposase with zinc-finger SWIM-type domains and a diatom-specific E3 ubiquitin ligase of the zinc-finger UBR type, which are suggestive of a mobilization mechanism. Our findings provide new elements for the understanding of the role of TEs in diatom genome evolution and in the enrichment of intraspecific genetic variability.

Keywords: diatom, transposable element, genome editing, genome stability, selectable marker, *Phaeodactylum tricornutum*.

INTRODUCTION

Diatoms are unicellular photosynthetic microalgae with core ecological roles and are the focus of increasing biotechnological interest. They are widespread among the planet's diverse aquatic and oceanic environments because of their unique and efficient metabolic traits. As such, diatoms are among the most relevant primary producers and main drivers of oceanic geochemical cycles, in addition to being a major component of aquatic food chains (Armbrust, 2009). A number of diatom species are also found in industrial and commercial applications related to their biochemical traits, including aquaculture and the photosynthetic production of high-value bioproducts, such as polyunsaturated fatty acids, pigments and proteins (Butler et al., 2020; Fabris, Abbriano, et al., 2020). In particular, the genetically tractable *Phaeodactylum tricornutum* is a model organism for advanced bioengineering applications, including metabolic engineering and synthetic biology approaches, that could expand the biosynthetic capacity of diatoms beyond their natural traits (D'Adamo et al., 2018; Fabris, George, et al., 2020; Hempel, Bozarth, et al., 2011; Hempel, Lau, et al., 2011; Pampuch et al., 2022; Slattery et al., 2022; Windhagauer et al., 2022).

The unique metabolic capacities of diatoms have originated through a peculiar evolutionary path, which included at least two independent endosymbiotic events and numerous horizontal gene transfers (HGTs) (Vancaester et al., 2020). Through these, diatoms have acquired genes from all domains of life (Bowler et al., 2008). These genetic features have then been reshuffled and reshaped in a wealth of configurations, resulting in enormous genetic and taxonomic diversity, by mechanisms that are still largely unknown.

A major driver of genetic diversity in diatoms, as in other organisms, is the presence and action of transposable elements (TEs) (Maumus et al., 2009). In several eukaryotes, TEs have been shown to be widespread and to undergo frequent mobilization and rearrangement, particularly during stress, and have been hypothesized to play central roles in genome evolution (Sun et al., 2020). Recent analyses highlighted that the portion of *P. tricornutum* genome composed by TEs is much larger than previously thought and suggested that the effect and influence of TEs in diatoms is a key evolutionary factor in the wild (Filloramo et al., 2021).

Even in controlled laboratory conditions diatoms can accumulate high haplotype diversity, as well as within the same clonal cultures, where genotype diversity is enhanced through the frequent inter-homologous recombination events that are favoured under stress conditions

(Bulankova et al., 2021). Aspects of genome stability in diatoms are important to understand their evolution and role in ecosystems, but also in assessing the robustness of genetically engineered strains. In this latter scenario, many applications are based on the artificial perturbation of specific small portions of their genomes, such as in the case of the maintenance of functional selectable markers or transgenes, or both.

The enzyme uridine monophosphate synthase (UMPS) is involved in *de novo* pyrimidine biosynthesis and its genetic locus is a relevant selectable marker in diatoms (Pampuch et al., 2022; Sakaguchi et al., 2011; Serif et al., 2018). UMPS knockout mutants are tolerant to 5-fluoroorotic acid (5-FOA) and are uracil auxotrophic. 5-FOA is a chemical analogue of oroate, which is converted by UMPS to form orotidine-5′-phosphate (EC 2.4.2.10), which is then converted by UMPS again to form uridine-5′-monophosphate (EC 4.1.1.23), a precursor of uracil (Figure 1b). In the presence of 5-FOA and uracil, UMPS catabolizes the 5-FOA analogue compound to generate 5-fluorouracil (5-FU), a toxic molecule that causes cell death (Figure 1a). When UMPS is inactivated, the resultant mutant requires uracil supplementation in the medium, but is also unable to catabolize 5-FOA into toxic 5-FU, making this a useful selectable marker gene for screening microbes on 5-FOA selection plates (Figure 1a). Although auxotrophic strains for histidine and tryptophan in *P. tricornutum* have
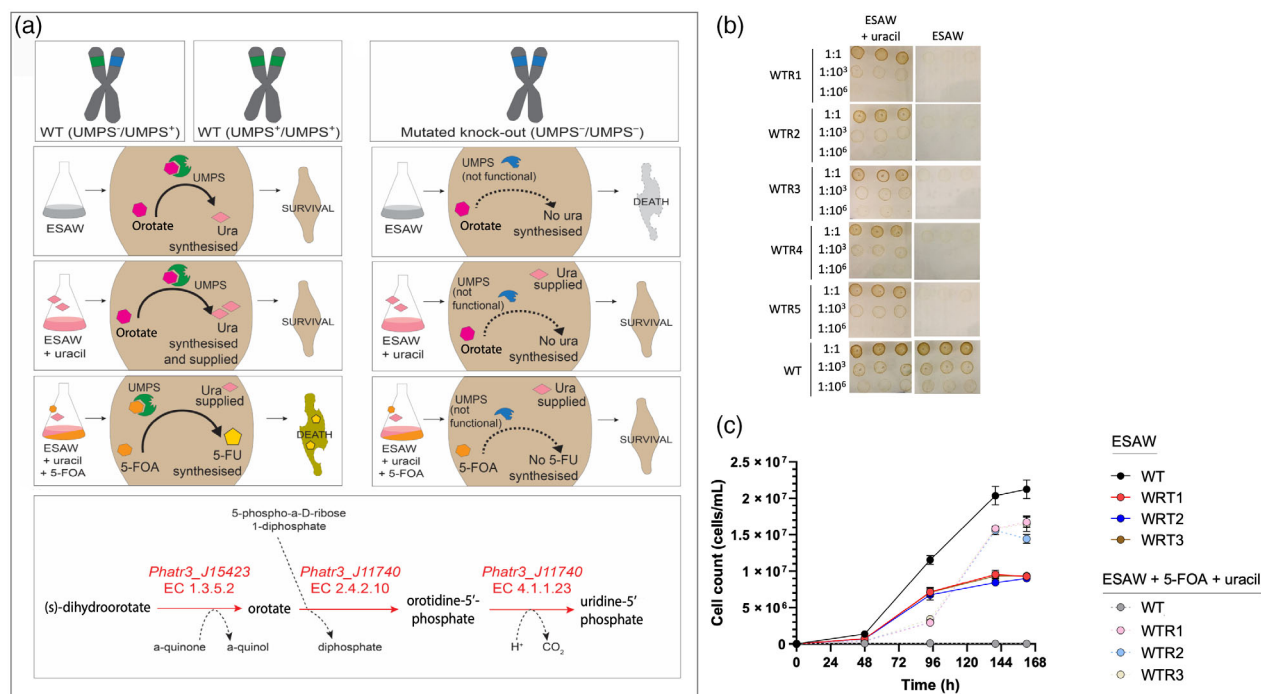

Figure 1. Phenotyping of 5-FOA-resistant wild-type (WT) *P. tricornutum* strains (WTR). (a) Schematic representation of WT and UMPS-deficient phenotypes in the presence or absence of 5-FOA and uracil. (b) Uracil auxotrophy assessed by spotting either WTR or WT diatoms on ESAW–agar supplemented with 50 μg ml^{-1} uracil or ESAW–agar. (c) Growth of three WTR strains compared with WT controls cultivated in liquid ESAW medium in the presence or absence of 5-FOA (300 μg ml^{-1}) and uracil 50 μg ml^{-1} ($n = 3$, error bars represent standard error).

been developed (Slattery et al., 2020), UMPS and adenine phosphoribosyl transferase (APT) are currently the only endogenous counterselectable markers available for ribonucleoprotein (RNP)-based CRISPR applications in diatoms (Serif et al., 2018). For the UMPS locus to be a reliable genetic marker, it needs to be particularly stable. However, numerous genome-editing experiments carried out in our laboratory targeting *PtUMPS* suggested that this locus in *P. tricornutum* (*Phatr3_J11740*) is particularly prone to produce false positives, as we consistently observed the emergence of diatom colonies able to grow on 5-FOA without the *PtUMPS* locus being targeted by nuclease. To date, there are no reports of 5-FOA-induced mutation in diatoms. However, it has previously been reported that wild-type *P. tricornutum* could become reliant on uracil and resistant to 5-FOA (RURF) following chemical mutagenesis using *N*-ethyl-*N*-nitrosourea (ENU) and selection with 100–300 μg ml^{-1} of 5-FOA (Sakaguchi et al., 2011).

Here, we isolated and characterized *P. tricornutum* cell lines that spontaneously emerged as RURF-like in the absence of exogenous genetic manipulations and known mutagens. We provide phenotypic and genotypic evidence of the systematic and reproducible disruption of the *PtUMPS* locus through the activation of a TE mobilized during 5-FOA selection.

RESULTS AND DISCUSSION

Wild-type *P. tricornutum* exposed to 5-FOA produces colonies with a stable RURF-like phenotype

To investigate the occurrence of the 5-FOA-resistant diatom phenotype, we subjected wild-type (WT) *P. tricornutum* cultures (1.5×10^8 cells) to a range of concentrations of 5-FOA (0–1000 μg ml^{-1}), including concentrations typically used in CRISPR-Cas9 RNP-mediated experiments (100–300 μg ml^{-1}; Serif et al., 2018), and supplemented with 50 μg ml^{-1} uracil to rescue the possible inactivation of *PtUMPS*. Hundreds of colonies emerged from selection with 5-FOA concentrations ranging from 50 to 150 μg ml^{-1}, 16 colonies (with an average frequency of $3.56 \times 10^{-8} \pm 2.78 \times 10^{-8}$; $n = 3$) were found on plates with 5-FOA 300 μg ml^{-1} (Figure S1) and no diatoms survived higher concentrations (600–1000 μg ml^{-1}; Figure S1). We hypothesized that the wild type resistant (WTR) strains generated in this study had the same RURF phenotype described by Sakaguchi et al. (2011).

To characterize the WTR phenotype, we selected five random colonies that originated from five different plates containing 300 μg ml^{-1} 5-FOA. To confirm the auxotrophy of WTR cell lines for uracil, we grew them in enriched seawater, artificial water (ESAW) without uracil for 1 week to deplete their endogenous cellular uracil pools and then dilution-plated them onto ESAW or ESAW supplemented with uracil. Whereas the WT lines were able to grow on ESAW plates without uracil supplementation, none of the

WTR cell lines were able to survive. However, the WTR phenotype could be rescued by the presence of uracil in the medium (Figure 1b). To further confirm the phenotype, three WTR primary colonies and one untreated WT colony were grown in the absence or presence of 300 μg ml^{-1} 5-FOA and 50 μg ml^{-1} uracil over 7 days (Figure 1c). As expected, untreated WT diatoms grew normally in ESAW medium, indicating the presence of a functional PtUMPS enzyme and endogenous uracil biosynthesis, whereas when cultured in the presence of 5-FOA and uracil, no growth was observed because of the functional PtUMPS enzyme metabolizing 5-FOA into toxic 5-FU. In the absence of 5-FOA and uracil, WTR diatoms grew at a reduced rate in the early exponential phase, from 48 to 96 h, most likely because of the declining presence of an intracellular uracil pool available to the cells for RNA biosynthesis. Cell growth drastically slowed after 96 h and was eventually arrested at approximately 144 h, suggesting that the uracil reserves were depleted. In contrast, these strains were all able to grow in the presence of 5-FOA and uracil, confirming that they were unable to metabolize 5-FOA into toxic 5-FU but were able to use supplemented uracil for RNA biosynthesis (Figure 1c).

Together, these results confirmed that 5-FOA exposure results in an RURF-like phenotype (Sakaguchi et al., 2011) in WTR diatoms and strongly suggests that the PtUMPS enzyme might have been inactivated in these cell lines.

WTR diatom cell lines harbour large mutations in the *PtUMPS* locus

To determine whether the inactivation of PtUMPS in WTR lines resulted from genetic disruption of the *PtUMPS* gene, we interrogated this locus using PCR and Sanger sequencing in the five selected WTR lines and in one WT cell line (Figure 1b). Although we were able to amplify the full *PtUMPS* locus from genomic DNA of untreated WT diatoms, we were not able to amplify it from any of the WTR strains except from WTR4 (Figure 2a). Next, we used different primer combinations that spanned all three exons of *PtUMPS* (Figure 2b; Table S1) to try to detect disruptions at the origin of the *PtUMPS* inactivation. We obtained correctly amplified PCR products from untreated WT samples using all four primer combinations (Figure 2a). However, in the WTR samples we could only amplify portions of *PtUMPS* when primers were designed to anneal the first two exons. We were unable to amplify the terminal region in the WTR samples, with the exception of WTR4, suggesting the disruption of the *PtUMPS* gene in the remaining WTR cell lines. The PCR results indicate the presence of a large insertion or deletion (INDEL) in all WTR cell lines except for WTR4 (Figure 2c), which nonetheless clearly showed a phenotype associated with a non-functional *PtUMPS*. Sanger sequencing of the full *PtUMPS* gene from
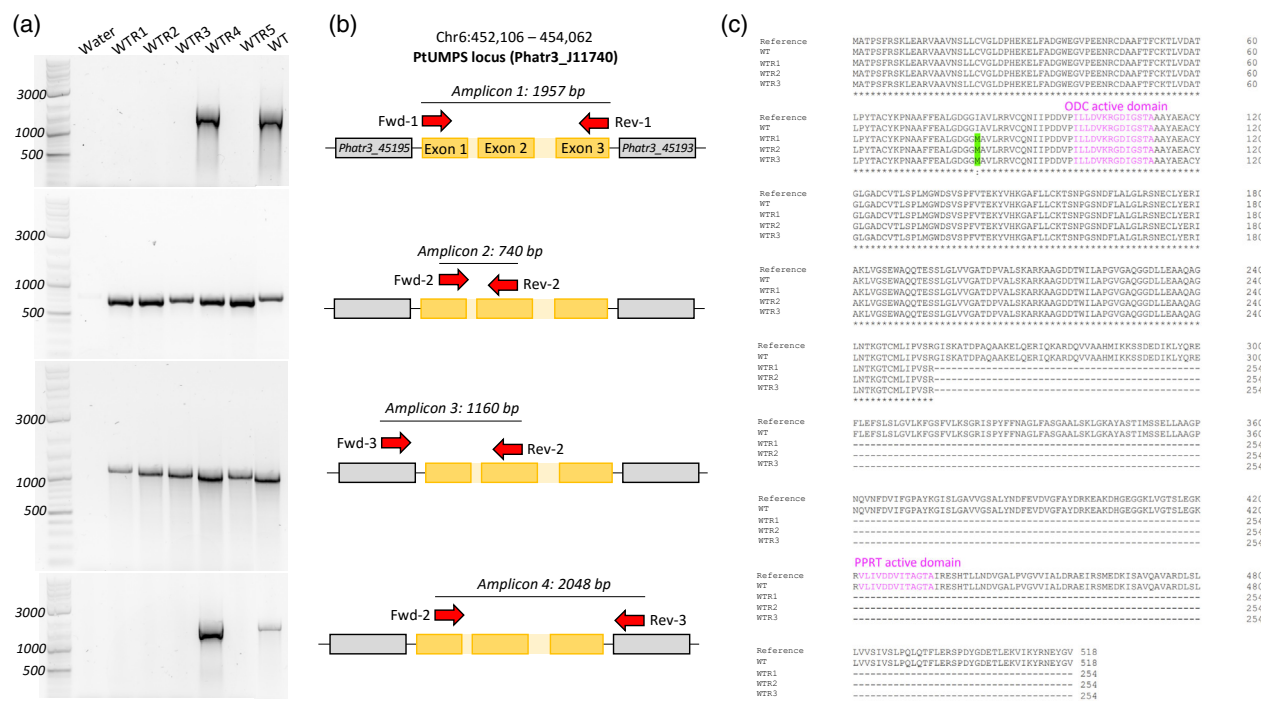
Figure 2. Genotyping of 5-FOA-resistant wild-type (WT) *P. tricornutum* strains (WTR). (a) PCR amplification of the *PtUMPS* locus in WTR and WT strains, with different primer combinations. (b) Schematic representation of the amplified genomic regions and expected amplicon size with an intact *PtUMPS* locus (in reversed configuration, for ease of interpretation). (c) Multiple sequence alignment of PtUMPS protein (UniProt accession C6L824) with translated nucleotide sequences obtained from *PtUMPS* amplicon 3 of WT, WTR1, WTR2 and WTR3 obtained in this study. All three WTR strains demonstrate a single point mutation (green) causing an amino acid change from isoleucine to methionine. The inability to amplify sequence from primers designed in exon 3 suggest a stunted coding region lacking the PPRT active domain (pink).

the untreated WT sample revealed a heterozygous genotype consisting of the same 16 single-nucleotide polymorphisms (SNPs) identified by Sakaguchi et al. (2011, SNP 1–SNP 16), as well as an additional three SNPs (SNP A–SNP C) (Figures S2 and S3). These were identified by the presence of peak doublets in the sequencing chromatogram (Figure S2), and suggested that untreated WT *P. tricornutum* harboured one functional *PtUMPS* allele (allele A) and a second non-functional allele (allele B) (Figure 1a). Therefore, we hypothesized that WTR mutants contained only the non-functional allele B. Given that we were only able to amplify the 5′ end of *UMPS* in the WTR mutants, we Sanger-sequenced amplicon 3 from WTR1, WTR2 and WTR3 (Figure 2c). Interestingly, the sequences obtained for amplicon 3 matched functional allele A (Figure S2). The WTR genotype obtained differed from the predicted *PtUMPS* WT sequence and putatively encoded a 254-amino-acid-long protein (instead of the 518 amino acids of the full-length protein) without the purine/pyrimidine phosphoribosyl transferase (PPRT) active domain, and with a single point mutation in the coding sequence causing the substitution of a methionine by an isoleucine (Figure 2c).

Considering these results, we rejected our original hypothesis that the RURF phenotype in *P. tricornutum* WTR strains was caused by point mutations in the

functional allele, as reported by Sakaguchi et al. (2011), but we hypothesized instead that the phenotype could originate from a larger chromosomal mutation that caused the severe truncation of the *PtUMPS* gene and the loss of the PPRT active domain. Such large chromosomal mutations have indeed been reported in other species, such as the loss of entire chromosomes following 5-FOA exposure in *Candida albicans* (Wang et al., 2004; Wellington et al., 2006; Wellington & Rustchenko, 2005).

A *mutator-like element* (*MULE*) transposon inactivated the *PtUMPS* locus in WTR diatoms

To test for the presence of large rearrangements affecting the *PtUMPS* locus in WTR diatom strains, we employed Oxford Nanopore long-read sequencing and assembled the genomes of all WTR cell lines as well as that of the WT control (for genome sequencing and assembly statistics, see Tables S2 and S3, respectively). Investigation of the *PtUMPS* locus using the INTEGRATED GENOMICS VIEWER (Robinson et al., 2011) revealed mutations of the *PtUMPS* gene in all WTR cell lines, whereas the corresponding WT locus was found to be intact. Interestingly, all cell lines except for WTR4 presented a single large insertion ranging from 2.3 to 3.7 kb in the *PtUMPS* coding sequences. These were confirmed by analysing the assembly with the structural

variant caller SNIFFLES (Sedlazeck et al., 2018) using the raw sequencing reads mapped against the published *P. tricornutum* genome. In contrast to the other WTR samples, WTR4 only presented a small in-frame deletion (of 15 bp) in exon 3, which occurs within the active site of the *PtUMPS* phosphoribosyl transferase domain (Figure 3). All mutations in the *PtUMPS* locus of WTR cell lines resulted in a disrupted gene sequence, corroborating the hypothesis that the observed phenotypes could be attributed to *PtUMPS* inactivation by larger-scale chromosomal modifications, as suggested by the PCR-based genotyping (Figure 2).

Further investigation of the insertion sequences at the *PtUMPS* locus of cell lines WTR1, WTR2, WTR3 and WTR5 revealed the presence of terminal inverted repeats (TIRs) of approximately 200 bp at the beginning and end of the insertions as well as a high degree of overall sequence similarity (Figures 4a and S4), indicating the insertion of a TE into the *PtUMPS* region of these cell lines. Sequence analysis around the insertion sites also revealed short (9 bp) target site duplications (TSDs) that externally flanked the TIRs (Figures 4a and S4). TSDs delimit the TE sequence as molecular 'scars' of the TE insertion (Figures 4a and S4), and are distinctive features of MULEs (Wicker et al., 2008). The similar nature and low GC content of the TSDs (Table S4) suggests that TE insertion may be specific to TA-rich regions.

Gene prediction and functional annotation of the TE sequences using INTERPROSCAN (Jones et al., 2014) revealed the presence of an open reading frame (ORF), putatively encoding a protein with a MULE transposase domain and a zinc-finger SWIM-type domain in WTR1 and WTR3 (Figure 4). Analysis of the TE sequences of WTR2 and WTR5 showed multiple mutations and frame shifts, possibly disrupting the predicted transposase ORF in WTR5 as well as a large deletion in the WTR2 TE sequence, resulting in the deletion of most of the transposase ORF (Figures 4b; Table S5). We identified multiple orthologs of the putative MULE transposase in the TE sequence of *P. tricornutum* (*Phatr3_EG01202*, *Phatr3_EG00036*, *Phatr3_EG00117*, *Phatr 3_EG02053*, *Phatr3_J33421*, *Phatr3_EG01875*, *Phatr3_ EG00351* and *Phatr3_J48581*), which show similarities to other predicted diatom proteins that contain MULE transposase domains and, to a lesser extent, to hypothetical proteins in eudicots that do not contain these domains. Interestingly, INTERPROSCAN analyses of the inserted TE sequences revealed the presence of the *Phatr3_J50367* gene, either in full or partial configuration. Searches on the US National Center for Biotechnology Information (NCBI) BLASTp and Plaza Diatoms databases (Vandepoele et al., 2013) revealed that this gene possesses three homologues with high similarity in *P. tricornutum* (*Phatr3_ EG02053*, *Phatr3_J48581* and *Phatr3_EG02300*). *Phatr3_ J50367* belongs to a gene family specific to raphid
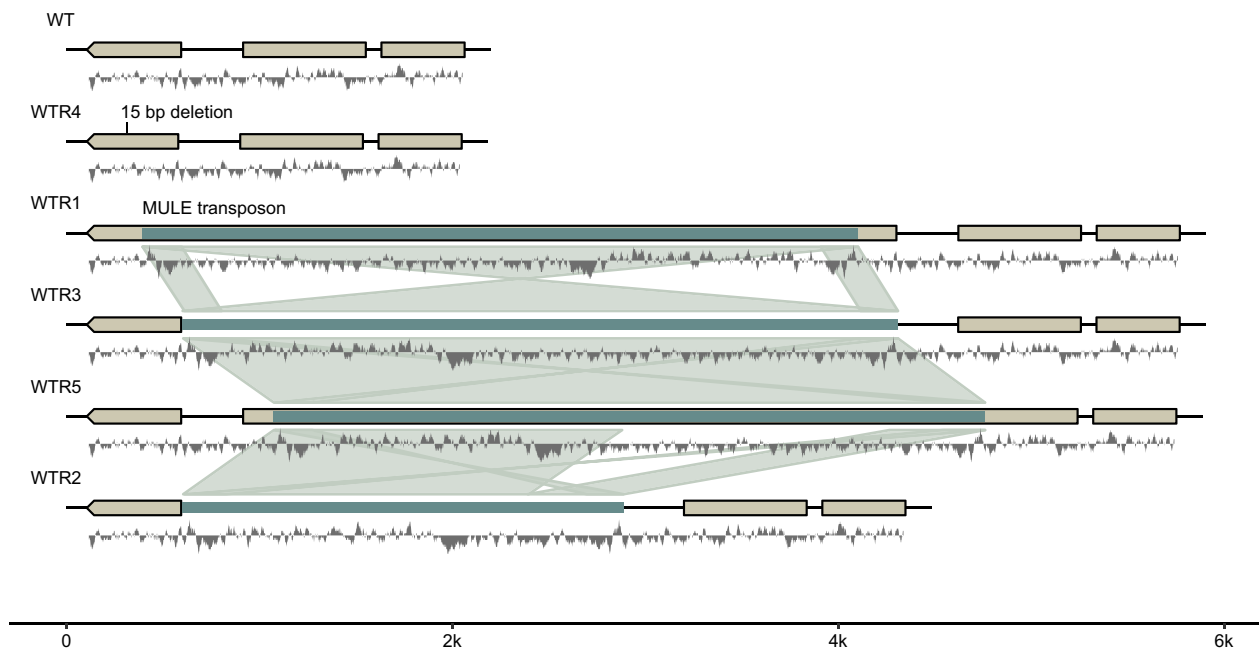

Figure 3. Schematic representation of genetic rearrangements in the *PtUMPS* locus. Assembled genome sequences at the *PtUMPS* locus, in wild-type (WT) and 5-FOA-resistant wild-type (WTR) lines. The coding sequence of the *PtUMPS* gene is shown in beige (in reverse orientation). Teal blocks represent insertion sequences annotated as MULE transposons, and light-teal ribbons represent synteny links between insertions at a cut-off value of >80% nucleotide sequence similarity. A GC track calculated with a 20-bp sliding window with a midline of 50% GC content is shown in grey beneath each sequence.
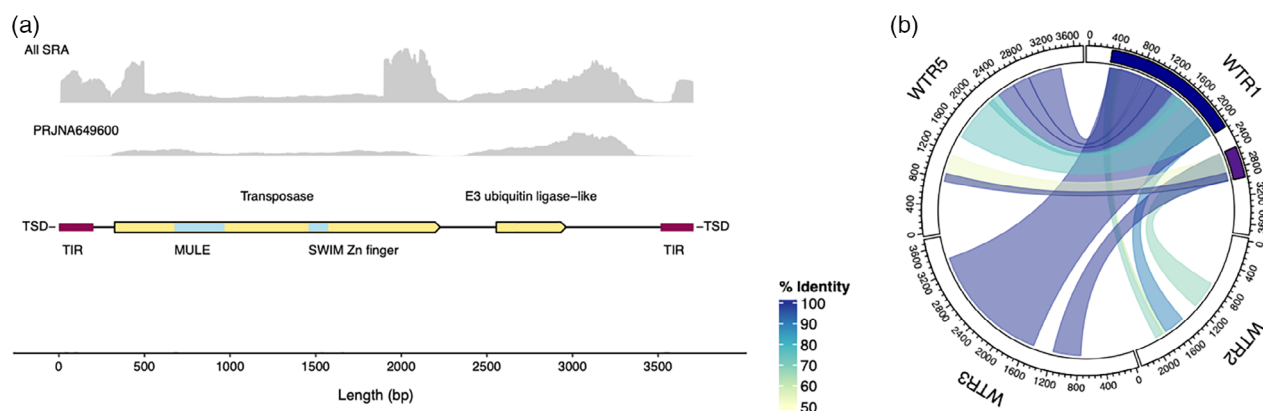
Figure 4. Structure and features of the transposable element (TE) that disrupted *PtUMPS* in WTR1. (a) Full length (3707 bp) of the entire transposon sequence in WTR1. Open reading frames (ORFs) corresponding to the transposase and the E3 ubiquitin ligase-like protein are depicted in yellow, with predicted MULE and SWIM zinc-finger domains in blue. The terminal inverted repeat (TIR) sequences are represented in maroon and flanked by 9-bp target site duplication (TSD) sequences. The RNA-seq data are from the available reads in the Sequence Read Archive (NCBI) (top) and from PRJNA649600. (b) Conservation of structure and features of the TE from WTR1 (dark blue) in WTR2, WTR3 and WTR4.

diatoms, with no significant similarities outside this group, and putatively encodes an E3 ubiquitin ligase of the zinc-finger UBR type (Figure 4a), according to EGGNOG predictions (Huerta-Cepas et al., 2019) reported in PLAZA diatoms 1.0. Similar to the MULE transposase, the complete sequence of this ORF was only found in some of the inserts. We detected the full sequence of *Phatr3_J50367* in WTR1 and WTR3, but only partial sequences in the other samples (Figure 4b; Table S2).

To assess whether both the ORFs encoding the transposases and the E3 ubiquitin ligases are actively transcribed in *P. tricornutum*, and not pseudogenes, we searched for evidence of their expression in publicly available transcriptomics data. Manual inspection of the mapped reads using the INTEGRATED GENOME VIEWER (Robinson et al., 2011) showed evidence of the expression of the complete transposase as well as large parts of the transposon (Figure 4a). Interestingly, the expression of the transposon was particularly pronounced in transgenic *P. tricornutum* cell lines overexpressing a *Thalassiosira pseudonana* chitin synthase (SRA project #PRJNA649600).

The presence of a MULE transposase in the TE sequence ruled out the possibility that the TE we identified falls into the subclass of Pack-MULE TEs, which instead are 'not autonomous' as they do not include a transposase and generally carry portions of several other host genes within them (Dupeyron et al., 2019). The presence of UBR-type zinc-finger E3 ubiquitin ligases in some of the TE sequences suggests that the mobilization of this MULE transposon mechanism might involve active post-translational regulation, based on protein degradation, which has never been described before. TE regulation mechanisms involving protein ubiquitination and histone binding in *Arabidopsis thaliana* have been reported (Kabelitz et al., 2016). Although negative regulatory roles of E3

ubiquitin ligases have been observed in LTR retrotransposons in mice (Maclennan et al., 2017), they have never been reported in class-2 DNA transposons of the MULE type. Given the lack of homologues of *Phatr3_J50367* outside the diatom taxonomic group, we hypothesize that this mechanism of DNA transposition might be diatom-specific. Inserts detected in WTR1, WTR3 and WTR5 exhibited the highest similarity, whereas the one found in WTR2, although still highly similar, lacks a large portion of the sequence present in the other samples (Figure 4b; Table S5). This suggests that there may be some variation in the insert length or modifications to the inserted sequence might occur during the insertion process.

In contrast to WTR1, WTR4 and WTR5, the presence of a small 15-bp deletion in the *PtUMPS* locus of WTR4 is likely to have emerged independently and subsequently was selected by the experimental conditions, in agreement with the uracil auxotrophy phenotype and the PCR amplification of the exons (Figure 2). This suggests that selection using 300 μg ml⁻¹ 5-FOA can induce presumably chemically-mediated mutations in *P. tricornutum* CCAP1055/1, resulting in RURF phenotypes.

The activation of the MULE transposon is favoured by 5-FOA selection

The fact that we observed interruption of the *PtUMPS* locus in all five WTR mutant cell lines suggests that selection with 300 μg ml⁻¹ 5-FOA clearly creates a strong bias towards the occurrence and selection of *PtUMPS* mutations. In four out of the five mutants analysed, the *PtUMPS* locus was disrupted by the same MULE transposon that was not present in the same location in the parental cell line before the experiment (Figure 3), suggesting that our selection conditions favoured TE-mediated mutations over TE-independent mutations (WTR4), or that TE-mediated

mutations are more frequent and/or more effective. In addition, the high degree of similarity between the TIR sequences, often used to estimate the age of the element (Hanada et al., 2009), indicates that the transposons were mobilized recently and that they originated from the same genomic source. Together, this provides evidence that in our experiments *PtUMPS* was disrupted by a TE, which was activated during the selection process, potentially triggered by sudden stress conditions represented by the accumulation of toxic 5-FU. Transposons and MULEs, in particular, are known to mobilize as a response to certain stress conditions in several organisms, either as an active coordinated response or through epigenetic deregulation (Negi et al., 2016).

We further explored whether the MULE-like TE was equally active in the genome of cell lines where it directly inactivated *PtUMPS* (WTR1–WTR3 and WTR5), in WTR4 where *PtUMPS* was disrupted by a TE-independent mutation and in WT where *PtUMPS* was left intact. Using the transposon that disrupted the *PtUMPS* locus in WTR1 as a query, we searched for highly similar regions with strict parameters (length >1.5 kb, lowest similarity of 80% and E-value of 0.0) throughout the genomic assemblies we generated (with genome coverages ranging 18× to 68×), as well as in the publicly available genome (Rastogi et al., 2018). In the untreated WT, this specific MULE transposon was found to occur in only five instances, which is comparable with the frequencies detected in publicly available genome assemblies. In contrast, the transposon was detected in much higher frequencies ranging from 20 to 131 copies (Table S2) in WTR cell lines subjected to 5-FOA selection. Interestingly, the WTR4 cell line that developed a TE-independent *PtUMPS* mutation exhibited a MULE TE mobilization frequency (25 copies) that is comparable with that of TE-dependent cell lines.

Although this does not provide evidence of the targeted mobilization of the MULE TE to the *PtUMPS* locus in direct response to 5-FOA selection, these results demonstrate that the observed MULE-mediated mutation occurred within the time frame of the experiment and suggest that MULEs may be activated during cellular stress, either as a regulated mechanism or as a consequence of derepression.

Implications of the use of 5-FOA as selection agent in diatoms

The mutant cell line WTR4 resulted from a TE-independent 15-bp deletion in the *PtUMPS* locus, suggesting that selection using 300 μg ml^{-1} 5-FOA can induce presumably chemically-mediated mutations in *P. tricornutum* CCAP1055/1 that result in RURF phenotypes. This phenotype has previously been obtained in *P. tricornutum* strain UTEX LB 642 following chemical mutagenesis using *N*-ethyl-*N*-nitrosourea and selection with 5-FOA (Sakaguchi

et al., 2011), and in *P. tricornutum* strain CCAP1055/1 following the targeted gene knockout of the *UMPS* gene via CRISPR-Cas9 gene editing (Serif et al., 2018; Slattery et al., 2020). In these works, diatoms were selected on 100 μg ml^{-1} 5-FOA. Our findings demonstrate that it is possible that 5-FOA can be mutagenic in *P. tricornutum* at concentrations above 100 μg ml^{-1}. These results also suggest that 5-FOA may produce false-positive 5-FOA-resistant mutants through the selection of transformants or genome-edited diatoms, as chemical and TE-induced mutagenesis cannot be excluded. 5-FOA has shown mixed mutagenic effects in different organisms, including *Saccharomyces cerevisiae* (Backhaus et al., 2017; Klein et al., 2014; Scott et al., 2018) and others, such as *Yarrowia lipolytica* (Lv et al., 2019), as well as *Escherichia coli* (Brandsen et al., 2018; Standage-Beier et al., 2015). However, 5-FOA is highly mutagenic in *C. albicans* (Wang et al., 2004; Wellington & Rustchenko, 2005). There is also a report of 5-FOA inducing mutation in *Sulfolobus acidocaldarius* (Kondo et al., 1991) and in *Saccharomyces cerevisiae* (Hao et al., 2016). Furthermore, 5-FOA has been used for generating spontaneous RURF phenotypes in the dinoflagellate *Symbiodinium* SSB01 following exposure to 200 μg ml^{-1} 5-FOA (Ishii et al., 2018), and in the red alga *Cyanidioschyzon merolae* 10D following treatment with 800 μg ml^{-1} 5-FOA (Minoda et al., 2004).

Implications of the mobilization of TE-mediated, non-homologous genome rearrangements in diatom genome evolution and biotechnology

Recently, it has been reported that diatoms use an, as yet, elusive mechanism of mitotic inter-homologue recombination to increase phenotypic plasticity, which is particularly enhanced when diatoms face sudden environmental stress, and this phenomenon has been observed to occur specifically at mutated *PtUMPS* loci to restore the WT allele (Bulankova et al., 2021). Here, we provided evidence of the mobilization and integration of MULE-like transposons, another, possibly stress-activated, mechanism that can contribute to genetic and phenotypic diversity at this locus and across the *Phaeodactylum* genome, as previously hypothesized (Maumus et al., 2009). Our findings add relevant insights into the dynamics driving genome evolution in diatoms, suggesting that TEs can be rapidly mobilized and play a central role in enriching the genetic diversity of diatoms. Our findings raise intriguing questions on the mechanism of activation of these genetic elements, such as whether their mobilization is part of a regulated mechanism triggered by environmental stress or whether their activity is determined by the derepression of a silencing mechanism. In addition, it remains to be tested whether TE integration occurs randomly or if it is targeted towards some specific recombination hotspots in the genome, characterized by genetic or epigenetic factors. More broadly, as TEs compose a large part of all

sequenced diatom genomes (Maumus et al., 2009), our finding underscores the relevance that these interchromosomal non-homologous rearrangements have in the dynamic genome of diatoms. Not only are these involved in reshuffling endogenous genes, potentially they are also involved in incorporating external DNA during the mobilization process, possibly facilitating horizontal gene transfers, which are particularly frequent in diatoms (Bowler et al., 2008), as well as providing new micro-homology regions that could promote recombination events (Bourque et al., 2018).

Besides their central role in aquatic ecology, diatoms are also emerging microbial hosts for biotechnology and synthetic biology applications, and many genetic tools, including gene knockout, knockdown and overexpression, as well as genome editing, are now part of routine experimental practices in functional genetics to decipher diatom gene functions. In all these applications, a critical aspect is the genetic stability of the transformed cell lines. At the same time, all these applications involve a selection step, usually involving antibiotics or cytotoxic agents. Our findings show that these stressful environmental conditions caused by the accumulation of toxic compounds or by-products could mobilize TEs. Selection agents such as 5-FOA, zeocin and bleomycin, or others derived from the accumulation of heterologous products, or deriving from industrial cultivation settings, may affect loci that involve the introduced artificial genetic constructs, but also other regions of the genome, generating different genotypes in the resulting transformant lines that usually go unnoticed. As a direct consequence of the findings of this work, it is evident that the *PtUMPS* locus has important limitations as an auxotrophic selectable marker for genetic engineering and genome-editing approaches in diatoms, as it can be targeted by both TE-mediated and TE-independent mutations.

On the other hand, these results also provide inspiring avenues for strain engineering strategies. For example, if controllable through environmental conditions and coupled with suitable selection or reporter systems, the mobilization of MULE TEs could be exploited to enrich genetic variation in the process of improving diatom strains, as an inducible mutagenesis system.

CONCLUSION

By subjecting the *PtUMPS* genomic locus of *P. tricornutum* culture to strong selective pressure with 5-FOA at concentrations routinely employed for genetic manipulation experiments, within the time frame of a single experiment we observed the mobilization of a MULE transposon across four independent cell lines, which all conferred the same 5-FOA-resistant phenotype. In addition, we also observed a TE-independent deletion that occurred in a fifth cell line. Our findings report the observation of the

mobilization of a TE and suggest that this may be triggered by stressful environmental conditions. Moreover, our findings suggest that the TE that we identified might employ a mobilization mechanism involving protein degradation, facilitated by the presence of diatom-specific E3 ubiquitin ligases. Our findings have relevant implications in understanding the mechanisms underpinning the mobilization of TEs, in their role in the biology and evolution of diatoms, as well as in their biotechnological application.

EXPERIMENTAL PROCEDURES

Microbial strains and growth conditions

Axenic cultures of *P. tricornutum* CCAP1055/1 were grown in liquid ESAW medium (Berges et al., 2001). Where appropriate, *P. tricornutum* was cultured in ESAW containing 300 µg ml^{-1} 5-FOA (Sigma-Aldrich, Merck, https://www.sigmaaldrich.com) and 50 µg ml^{-1} uracil (Sigma-Aldrich), or ESAW supplemented with 50 µg ml^{-1} zeocin (Invitrogen, now ThermoFisher Scientific, https://www.thermofisher.com). Diatoms were cultured in fully controlled shaking incubators (Kuhner, https://kuhner.com) under 100 µE m^{-2} sec^{-1} light at 21°C shaking at 95 rpm.

5-FOA treatment and selection of RURF-like phenotypes

A population of 1.5×10^8 cells were plated on half-salts ESAW solid medium enriched with uracil 50 µg ml^{-1} and containing 5-FOA concentrations ranging from 0 to 1000 µg ml^{-1} (0, 10, 100, 150, 300, 600 and 1000 µg ml^{-1}), in triplicates (Figure S1). After 4–6 weeks, 5-FOA-resistant colonies (WTRs) started to emerge, which were large enough for visual inspection and transfer after 7 weeks. At this point, colonies obtained from 5-FOA 300 µg ml^{-1} were picked and inoculated in liquid ESAW medium enriched with uracil 50 µg ml^{-1} for use in experiments.

Flow cytometry

All cell counts were performed by flow cytometry, using a BD CytoFLEX S flow cytometer (BD Biosciences, https://www.bdbiosciences.com). The cells were analysed at medium speed until 20 000 events were counted. Fluorescein isothiocyanate (FITC) fluorescence was excited with a 488-nm laser and emission was acquired using a 525/40-nm optical filter.

PCR-based analyses

PCR amplification was performed using Q5 high fidelity polymerase (New England Biolabs, https://international.neb.com) and PCR screening was performed using GoTaq Flexi DNA polymerase (Promega, https://www.promega.com) according to the manufacturer's instructions. For high-throughput PCR screening, single colonies were grown in 200 µl of ESAW supplemented with relevant compounds for 10 days to increase biomass. A list of primers used can be found in Table S1.

High-molecular-weight genomic DNA (gDNA) extraction, sequencing and genome assembly

DNA was extracted and sequenced as described by George et al. (2020). Briefly, cell pellets containing 7×10^7 to 2×10^8 *P. tricornutum* cells were extracted in batches to obtain approximately 2 µg of genomic DNA following the protocol published at dx.doi.org/10.17504/protocols.io.qzudx6w. The extracted high-molecular-

weight DNA was resuspended in molecular-grade water at 22-24°C overnight. Sequencing libraries were prepared with the Ligation Sequencing Kit SQK-LSK109 and sequenced on R9.4.1 flow cells (Oxford Nanopore Technologies, https://nanoporetech.com). Individual sample sequencing runs were stopped after an estimated minimum of 11 Gb were sequenced, excepting samples WRT2 and WT. For sample WRT2, the coverage estimation during sequencing was inconsistent with the actual coverage after base calling, resulting in approximately half the estimated coverage. The sequencing run for the WT sample was performed last and only stopped once the flow cell was saturated, resulting in more than 2 Gb. After each sequencing run was stopped the flow cell was washed using wash kit EXP-WSH004 (Oxford Nanopore Technologies), according to manufacturer's specifications, and reloaded with a new sequencing library.

Base calling was performed using GUPPY 4.0.15 (Oxford Nanopore Technologies) with the high-accuracy model r9.4.1_450bps_hac. Base calling resulted in 678 874 Oxford Nanopore long reads ranging from 65 123 (WTR3) to 212 812 (WT control). Coverage of the samples range from approximately 18× for sample WTR3 to approximately 68× for the WT control (Table S5). Base-called reads were assembled using FLYE 2.8.1 (Kolmogorov et al., 2019) and error corrected with three rounds of RACON 1.4.17 (Vaser et al., 2017). Mapping of long reads for error correction was performed using MINIMAP 2.17 (Li, 2018, p. 2). Final assemblies ranged from 33 260 708 nt for sample WTR3 to 35 990 700 nt in the WT control (Table S5). We obtained 678 874 Oxford Nanopore long reads ranging from 65 123 (WTR3) to 212 812 (WT control), resulting in a per-sample sequencing coverage ranging from approximately 18× for sample WTR3 to approximately 68× for the WT control (Table S5). After quality control and error correction, the final FLYE 2.8.1 (Kolmogorov et al., 2019) assemblies ranged from 33 260 708 nt for sample WTR3 to 35 990 700 nt for the WT control (Table S5).

Comparative genomics and TE annotation

All final assemblies were mapped to the published *P. tricornutum* genome (https://protists.ensembl.org/Phaeodactylum_tricornutum) using MINIMAP 2.17. The resulting .sam files were sorted, converted to .bam format and indexed using SAMTOOLS 1.10 (Li et al., 2009). The *P. tricornutum* UPMS region was screened for insertions and deletions using the INTEGRATED GENOMICS VIEWER (IGV 2.10) (Robinson et al., 2011) and the indexed .bam files. Sequences of potential insertions at the UMPS locus were extracted using IGV and translated into their reverse complements online (https://www.bioinformatics.org/sms/rev_comp.html). Sequence alignments and visualization was performed using ALIVIEW 1.28 (Larsson, 2014) and the integrated aligner MUSCLE 3.8.31 (Edgar, 2004). The insertions/potential TEs in the *UMPS* region of the genome assemblies were confirmed as true insertions using SNIFFLES 1.0.12 (Sedlaczek et al., 2018). Read mapping for SNIFFLES was performed with MINIMAP 2.17 and the resulting .sam files were sorted, converted to .bam and indexed using SAMTOOLS 1.10.

Gene prediction and functional annotation of the coding sequences in the insertions were performed using INTERPROSCAN 5.54.87 (Jones et al., 2014).

For validation of the expression of predicted coding sequences of the insertion, publicly available *P. tricornutum* RNA-seq data sets were downloaded from the Sequence Read Archive (SRA) at NCBI and aligned with the complete transposon sequence of sample WTR1 using the splice-aware aligner STAR 2.7.10a (Dobin et al., 2013). To validate the expression of the transposase and other putative coding sequences of the transposon, publicly available *P. tricornutum* RNA-seq data sets were downloaded from the SRA (NCBI) and aligned with the complete transposon sequence of sample WTR1 using the splice-aware aligner STAR 2.7.10a (Dobin et al., 2013).

AUTHOR CONTRIBUTIONS

JG designed the study, conducted the experiments, analysed the data and wrote the article. TK designed the study, performed the DNA sequencing, analysed the data and contributed to writing the article. AC performed DNA extraction. RA analysed data and contributed to writing the article. MF designed the study, analysed the data and wrote the article. All authors read and edited the final version for publication.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest associated with this work.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Spontaneous formation of *Phaeodactylum tricornutum* colonies on solid ESAW medium and on solid ESAW medium with increasing concentrations of 5-FOA.

Figure S2. Analysis of *UMPS* genotype in untreated wild-type *Phaeodactylum tricornutum* strain CCAP1055/1.

Figure S3. Sequence alignment highlighting single-nucleotide polymorphisms (SNPs) within *Phaeodactylum tricornutum UMPS*, identified by Sakaguchi et al. (2011), and within wild-type strain CCAP1055/1, identified in this study.

Figure S4. Alignments of inverse terminal repeats of the TE sequences.

Table S1. Oligonucleotide primers used in this study.

Table S2. Oxford Nanopore sequencing statistics.

Table S3. Genome assembly summary statistics.

Table S4. Target site duplication (TSD) sequences flanking transposon insertions at the *PtUMPS* locus.

Table S5. BLAST analyses of the putative open reading frames containing MULE transposase and E3 ligase component domains in transposable element sequences.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at SRA link/bio project

(https://www.ncbi.nlm.nih.gov/sra/PRJNA873859) and Open Science Forum link (https://osf.io/7wtds/).

DATA AVAILABILITY STATEMENT

Base-called sequencing files for each sample in fastq format are available at the SRA under bioproject PRJNA873859. Error-corrected genome assemblies for all isolates, SNIFFLES variant calls in .vcf format, InterProScan annotations and fasta files of the complete TE sequences were deposited in an Open Science Forum repository at https://osf.io/7wtds/. Name mapping: WTR1 = WTR_1_1, WTR2 = WTR_2_1, WTR3 = WTR_2_2, WTR4 = WTR_1_3 and WTR5 = WTR_1_4.

REFERENCES

Armbrust, E.V. (2009) The life of diatoms in the world's oceans. *Nature*, 459, 185–192.

Backhaus, K., Ludwig-Radtke, L., Xie, X. & Li, S.M. (2017) Manipulation of the precursor supply in yeast significantly enhances the accumulation of prenylated β-Carbolines. *ACS Synthetic Biology*, 6, 1056–1064.

Berges, J.A., Franklin, D.J. & Harrison, P.J. (2001) Evolution of an artificial seawater medium: improvements in enriched seawater, artificial water over the last two decades. *Journal of Phycology*, 37, 1138–1145.

Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M. *et al.* (2018) Ten things you should know about transposable elements. *Genome Biology*, 19, 199.

Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A. *et al.* (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*, 456, 239–244.

Brandsen, B.M., Mattheisen, J.M., Noel, T. & Fields, S. (2018) A biosensor strategy for *E. coli* based on ligand-dependent stabilization. *ACS Synthetic Biology*, 7, 1990–1999.

Bulankova, P., Sekulić, M., Jallet, D., Nef, C., van Oosterhout, C., Delmont, T.O. *et al.* (2021) Mitotic recombination between homologous chromosomes drives genomic diversity in diatoms. *Current Biology*, 31, 3221–3232.e9.

Butler, T., Kapoore, R.V. & Vaidyanathan, S. (2020) *Phaeodactylum tricornutum*: a diatom cell factory. *Trends in Biotechnology*, 38, 606–622.

D'Adamo, S., Schiano di Visconte, G., Lowe, G., Szaub-Newton, J., Beacham, T., Landels, A. *et al.* (2018) Engineering the unicellular alga *Phaeodactylum tricornutum* for high-value plant triterpenoid production. *Plant Biotechnology Journal*, 17, 75–87.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S. *et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15–21.

Dupeyron, M., Singh, K.S., Bass, C. & Hayward, A. (2019) Evolution of mutator transposable elements across eukaryotic diversity. *Mobile DNA*, 10, 12.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797.

Fabris, M., Abbriano, R.M., Pernice, M., Sutherland, D.L., Commault, A.S., Hall, C.C. *et al.* (2020) Emerging technologies in algal biotechnology: toward the establishment of a sustainable, algae-based bioeconomy. *Frontiers in Plant Science*, 11, 279.

Fabris, M., George, J., Kuzhiumparambil, U., Lawson, C.A., Jaramillo Madrid, A.C., Abbriano, R.M. *et al.* (2020) Extrachromosomal genetic engineering of the marine diatom *Phaeodactylum tricornutum* enables the heterologous production of monoterpenoids. *ACS Synthetic Biology*, 9, 598–612.

Filloramo, G.V., Curtis, B.A., Blanche, E. & Archibald, J.M. (2021) Re-examination of two diatom reference genomes using long-read sequencing. *BMC Genomics*, 22, 379.

George, J., Kahlke, T., Abbriano, R.M., Kuzhiumparambil, U., Ralph, P.J. & Fabris, M. (2020) Metabolic engineering strategies in diatoms reveal unique phenotypes and genetic configurations with implications for algal genetics and synthetic biology. *Frontiers in Bioengineering and Biotechnology*, 8, 513.

Hanada, K., Vallejo, V., Nobuta, K., Slotkin, R.K., Lisch, D., Meyers, B.C. *et al.* (2009) The functional role of pack-MULEs in rice inferred from purifying selection and expression profile. *The Plant Cell*, 21, 25–38.

Hao, H., Wang, X., Jia, H., Yu, M., Zhang, X., Tang, H. *et al.* (2016) Large fragment deletion using a CRISPR/Cas9 system in *Saccharomyces cerevisiae*. *Analytical Biochemistry*, 509, 118–123.

Hempel, F., Bozarth, A.S., Lindenkamp, N., Klingl, A., Zauner, S., Linne, U. *et al.* (2011) Microalgae as bioreactors for bioplastic production. *Microbial Cell Factories*, 10, 81.

Hempel, F., Lau, J., Klingl, A. & Maier, U.G. (2011) Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PLoS One*, 6, e28424.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H. *et al.* (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47, D309–D314.

Ishii, Y., Maruyama, S., Fujimura-Kamada, K., Kutsuna, N., Takahashi, S., Kawata, M. *et al.* (2018) Isolation of uracil auxotroph mutants of coral symbiont alga for symbiosis studies. *Scientific Reports*, 8, 3237.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30, 1236–1240.

Kabelitz, T., Brzezinka, K., Friedrich, T., Górka, M., Graf, A., Kappel, C. *et al.* (2016) A JUMONJI protein with E3 ligase and histone H3 binding activities affects transposon silencing in *Arabidopsis*. *Plant Physiology*, 171, 344–358.

Klein, J., Heal, J.R., Hamilton, W.D.O., Boussemghoune, T., Tange, T.Ø., Delegrange, F. *et al.* (2014) Yeast synthetic biology platform generates novel chemical structures as scaffolds for drug discovery. *ACS Synthetic Biology*, 3, 314–323.

Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P.A. (2019) Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, 37, 540–546.

Kondo, T., Johnson, C.H. & Hastings, J.W. (1991) Action spectrum for resetting the circadian phototaxis rhythm in the CW15 strain of *Chlamydomonas*: I. Cells in darkness. *Plant Physiology*, 95, 197–205.

Larsson, A. (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30, 3276–3278.

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34, 3094–3100.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. *et al.* (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Lv, Y., Edwards, H., Zhou, J. & Xu, P. (2019) Combining 26s rDNA and the Cre-loxP system for iterative gene integration and efficient marker curation in *Yarrowia lipolytica*. *ACS Synthetic Biology*, 8, 568–576.

Maclennan, M., García-Cañadas, M., Reichmann, J., Khazina, E., Wagner, G., Playfoot, C.J. *et al.* (2017) Mobilization of LINE-1 retrotransposons is restricted by *Tex19.1* in mouse embryonic stem cells. *eLife*, 6, e26152.

Maumus, F., Allen, A.E., Mhiri, C., Hu, H., Jabbari, K., Vardi, A. *et al.* (2009) Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genomics*, 10, 624.

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.

Negi, P., Rai, A.N. & Suprasanna, P. (2016) Moving through the stressed genome: emerging regulatory roles for transposons in plant stress response. *Frontiers in Plant Science*, 7, 1448.

Pampuch, M., Walker, E.J.L. & Karas, B.J. (2022) Towards synthetic diatoms: the *Phaeodactylum tricornutum* Pt-syn 1.0 project. *Current Opinion in Green and Sustainable Chemistry*, 35, 100611.

Rastogi, A., Maheswari, U., Dorrell, R.G., Vieira, F.R.J., Maumus, F., Kustka, A. *et al.* (2018) Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Scientific Reports*, 8(1), 4834.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. *et al.* (2011) Integrative genomics viewer. *Nature Biotechnology*, 29, 24–26.

Sakaguchi, T., Nakajima, K. & Matsuda, Y. (2011) Identification of the UMP synthase gene by establishment of uracil auxotrophic mutants and the

phenotypic complementation system in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiology*, **156**, 78–89.

Scott, L.H., Mathews, J.C., Flematti, G.R., Filipovska, A. & Rackham, O. (2018) An artificial yeast genetic circuit enables deep mutational scanning of an antimicrobial resistance protein. *ACS Synthetic Biology*, **7**, 1907–1917.

Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A. et al. (2018) Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods*, **15**, 461–468.

Serif, M., Dubois, G., Finoux, A.-L., Teste, M.-A., Jallet, D. & Daboussi, F. (2018) One-step generation of multiple gene knock-outs in the diatom *Phaeodactylum tricornutum* by DNA-free genome editing. *Nature Communications*, **9**, 3924.

Slattery, S.S., Giguere, D.J., Stuckless, E.E., Shrestha, A., Briere, L.A.K., Galbraith, A. et al. (2022) Phosphate-regulated expression of the SARS-CoV-2 receptor-binding domain in the diatom *Phaeodactylum tricornutum* for pandemic diagnostics. *Scientific Reports*, **12**, 7010.

Slattery, S.S., Wang, H., Giguere, D.J., Kocsis, C., Urguhart, B.L., Karas, B. et al. (2020) Plasmid-based complementation of large deletion in *Phaeodactylum tricornutum* biosynthetic genes generated by Cas9 editing. *Scientific Reports*, **10**, 13879.

Standage-Beier, K., Zhang, Q. & Wang, X. (2015) Targeted large-scale deletion of bacterial genomes using CRISPR-nickases. *ACS Synthetic Biology*, **4**, 1217–1225.

Sun, L., Jing, Y., Liu, X., Li, Q., Xue, Z., Cheng, Z. et al. (2020) Heat stress-induced transposon activation correlates with 3D chromatin organization rearrangement in *Arabidopsis*. *Nature Communications*, **11**, 1886.

Vancaester, E., Depuydt, T., Osuna-cruz, C.M. & Vandepoele, K. (2020) Systematic and functional analysis of horizontal gene transfer events in diatoms. *Molecular Biology and Evolution*, **37**, 3243–3257.

Vandepoele, K., Van Bel, M., Richard, G., Van Landeghem, S., Verhelst, B., Moreau, H. et al. (2013) Pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. *Environmental Microbiology*, **15**, 2147–2153.

Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, **27**, 737–746.

Wang, Y.K., Das, B., Huber, D.H., Wellington, M., Kabir, M.A., Sherman, F. et al. (2004) Role of the 14-3-3 protein in carbon metabolism of the pathogenic yeast *Candida albicans*. *Yeast*, **21**, 685–702.

Wellington, M., Kabir, M.A. & Rustchenko, E. (2006) 5-Fluoro-orotic acid induces chromosome alterations in genetically manipulated strains of *Candida albicans*. *Mycologia*, **98**, 393–398.

Wellington, M. & Rustchenko, E. (2005) 5-Fluoro-orotic acid induces chromosome alterations in *Candida albicans*. *Yeast*, **22**, 57–70.

Wicker, T., Sabot, F., Hua-van, A., Bennetzen, J.L., Capy, P., Chalhoub, B. et al. (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature Reviews. Genetics*, **9**, 414. https://doi.org/10.1038/nrg2165-c2

Windhagauer, M., Abbriano, R.M., Pittrich, D.A. & Doblin, M.A. (2022) Phosphate-inducible poly-hydroxy butyrate production dynamics in CO_2-supplemented upscaled cultivation of engineered *Phaeodactylum tricornutum*. *Journal of Applied Phycology*, **34**, 2259–2270.