

Dicer-dependent heterochromatic small RNAs in the model diatom species *Phaeodactylum tricornutum*

Emilia Grypioti^{1,2,3,4*} , Hugues Richard^{4,5*} , Nikoleta Kryovrysanaki^{1,2}, Marianne Jaubert^{4,6}, Angela Falciaiore^{4,6}, Frédéric Verret^{1,2,3†}  and Kriton Kalantidis^{1,2†}

¹Department of Biology, University of Crete, PO Box 2208, 70013, Heraklion, Crete, Greece; ²Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, 70013, Heraklion, Crete, Greece; ³Institute of Marine Biology and Aquaculture, Hellenic Center for Marine Research, 71500, Gournes, Crete, Greece; ⁴Institut de Biologie Paris-Seine, Laboratory of Computational and Quantitative Biology, UMR 7238 Sorbonne Université, 75005, Paris, France; ⁵Bioinformatics Unit, Genome Competence Center (MF1), Robert Koch Institute, 13353, Berlin, Germany; ⁶Institut de Biologie Physico-Chimique, Laboratory of Chloroplast Biology and Light Sensing in Microalgae, UMR7141 Centre National de la Recherche Scientifique (CNRS), Sorbonne Université, 75005, Paris, France

Summary

Authors for correspondence:

Frédéric Verret

Email: fverret@hcmr.fr

Kriton Kalantidis

Email: kriton@imbb.forth.gr

Received: 6 September 2023

Accepted: 17 October 2023

New Phytologist (2024) 241: 811–826

doi: 10.1111/nph.19429

Key words: Dicer, epigenetic, nitrate, *Phaeodactylum tricornutum*, RNA interference, small RNAs, transposon.

- Diatoms are eukaryotic microalgae responsible for nearly half of the marine productivity. RNA interference (RNAi) is a mechanism of regulation of gene expression mediated by small RNAs (sRNAs) processed by the endoribonuclease Dicer (DCR). To date, the mechanism and physiological role of RNAi in diatoms are unknown.
- We mined diatom genomes and transcriptomes for key RNAi effectors and retraced their phylogenetic history. We generated DCR knockout lines in the model diatom species *Phaeodactylum tricornutum* and analyzed their mRNA and sRNA populations, repression-associated histone marks, and acclimatory response to nitrogen starvation.
- Diatoms presented a diversification of key RNAi effectors whose distribution across species suggests the presence of distinct RNAi pathways. *P. tricornutum* DCR was found to process 26–31-nt-long double-stranded sRNAs originating mostly from transposons covered by repression-associated epigenetic marks. In parallel, *P. tricornutum* DCR was necessary for the maintenance of the repression-associated histone marks H3K9me2/3 and H3K27me3. Finally, PtDCR-KO lines presented a compromised recovery post nitrogen starvation suggesting a role for *P. tricornutum* DCR in the acclimation to nutrient stress.
- Our study characterized the molecular function of the single DCR homolog of *P. tricornutum* suggesting an association between RNAi and heterochromatin maintenance in this model diatom species.

Introduction

RNA interference (RNAi) is a conserved mechanism of regulation of gene expression mediated by small RNAs (sRNAs; Farazi *et al.*, 2008; Ghildiyal & Zamore, 2009). Intracellular double-stranded RNAs (ds-RNA) are cleaved by RNase III-like ribonuclease enzymes of the Dicer (DCR) family into small (*c.* 20–35 nt) ds-sRNA, which are subsequently incorporated into an ARGONAUTE (AGO)-containing complex known as the RNA-induced silencing complex (RISC). After strand separation, the remaining single-stranded sRNAs (ss-sRNA) guide the sequence-specific targeting of RISC to complementary RNAs. In some organisms, RNA-dependent RNA (RDR) polymerases synthesize dsRNAs from ss-sRNA to initiate and/or amplify RNAi. One of the most conserved functions of RNAi across eukaryotes is the

repression of transposable elements (TEs; Ketting *et al.*, 1999; Tabara *et al.*, 1999; Schramke & Allshire, 2003; Zilberman *et al.*, 2003; Shi *et al.*, 2004; Slotkin *et al.*, 2005). RNA-induced silencing complex (RISC) can direct TE repression post-transcriptionally in the cytosol via degradation of TE mRNA and transcriptionally in the nucleus via epigenetic silencing. Epigenetic silencing involves the formation of heterochromatin at targeted loci via methylation of DNA at cytosine bases by DNA methyltransferases (DNMTs) and/or methylation at lysine of histone 3 by histone methyltransferases. Heterochromatin formation entails the tight compaction of nucleosomes, which restricts the accessibility of DNA-dependent RNA polymerase to DNA and thereby represses gene transcription. To date, mechanisms of sRNA-directed epigenetic silencing have been described in a few established model organisms such as *Arabidopsis thaliana* and *Saccharomyces pombe* (Matzke & Mosher, 2014; Holoch & Moazed, 2015; Martienssen & Moazed, 2015). Knowledge on the role of RNAi in suppressing gene expression in organisms outside the eukaryotic supergroup of Opisthokonta (animals and

*These authors contributed equally to this work and shared first authorship.

†These authors have jointly supervised this work and shared senior authorship.

fungi) and Archaeplastida (land plants and red/green algae), however, is limited.

Diatoms are diverse and prominent eukaryotic unicellular algae contributing up to 20% of the global primary productivity playing a pivotal role in the marine food web and biogeochemical cycles of carbon and silicates (Falkowski *et al.*, 1998; Field *et al.*, 1998; Malviya *et al.*, 2016). Diatoms are Stramenopile organisms whose common ancestor has been proposed to be derived from an endosymbiotic event between a heterotrophic eukaryotic host, a red alga and possibly a green alga (Cavalier-Smith, 2003; Moustafa *et al.*, 2009). Two major groups of diatoms can be defined: the centric diatoms with radial symmetry and the pennate diatoms with elongated bilateral symmetry (Medlin *et al.*, 1996). It is currently estimated that ancestral centric diatoms arose *c.* 190 Ma and diverged into centric and pennate species nearly 150 Ma. The pennates are further divided in nonmotile araphid species and the more recent (*c.* 129 Ma) motile raphid species (Falciatore *et al.*, 2020). Whole genome sequencing of the model diatom species *Thalassiosira pseudonana* and *Phaeodactylum tricorutum* has revealed diatoms peculiar genomic makeup and fast evolution rate (Armbrust *et al.*, 2004; Bowler *et al.*, 2008). Recent developments in transgenesis and targeted mutagenesis in *P. tricorutum* have empowered reverse genetic approaches addressing diatom gene repertoires and metabolic pathways (Huang & Daboussi, 2017; Falciatore *et al.*, 2020). In a seminal study in *P. tricorutum*, De Riso *et al.* (2009) reported transcriptional silencing and methylation of a GUS transgene following expression of antisense and hairpin RNA constructs of sequences homologous to the GUS target. Later, sRNA transcriptome and whole genome epigenetic analysis in *P. tricorutum* reported the presence of *c.* 26–31-nt ds-sRNA mapped to transcriptionally repressed TE and non-TE genes harboring cytosine methylation and repression-associated histone marks (Veluchamy *et al.*, 2013, 2015; Rogato *et al.*, 2014). Combinatorial analysis of epigenetic marks and gene expression in *P. tricorutum* have suggested a role for epigenetic silencing in the control of TE and non-TE gene expression in standard condition and in the response to nitrogen starvation (Maumus *et al.*, 2009; Veluchamy *et al.*, 2013, 2015; Huguin *et al.*, 2021). Previous *in silico* analyses in a limited number of diatom species have reported the presence of a single homolog of DCR, AGO, and RDR in *P. tricorutum*, *T. pseudonana*, and *Fragilariopsis cylindrus* (De Riso *et al.*, 2009; Lopez-Gomollon *et al.*, 2014).

In this study, mining of a large number of recently sequenced diatom genomes and transcriptomes unveiled an unanticipated diversification of the DCR/AGO repertoire, suggesting the presence of distinct RNAi pathways in these organisms. *P. tricorutum* DCR knockout lines (DCR-KO) were generated by CRISPR-Cas9 mutagenesis and their mRNA and sRNA transcriptomes were characterized. DCR-KOs presented a drastic reduction of *c.* 26–31 ds-sRNA mapped mostly to TE genes covered by repression-associated epigenetic marks with a strong bias toward autonomous retrotransposons (TEs carrying the genes indispensable for transposition). DCR-KOs showed

transcriptional reactivation of autonomous retrotransposons, a global depletion of the repression-associated histone marks H3K9me2/3 and H3K27me3, and an impaired recovery from nitrogen starvation. Collectively, our results demonstrate the presence of DCR-dependent heterochromatic sRNA in *P. tricorutum* suggesting an association between RNAi and epigenetic silencing in this model diatom species.

Materials and Methods

Diatom cultures

P. tricorutum Bohlin 1897 (CCAP 1055/1) was obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton. Axenic cultures of wild-type (WT) and DCR-KO mutant lines were grown at 18°C in *f/2* medium (Guillard, 1975) without silica addition and subjected to a 12 h : 12 h light : dark photoperiod using white fluorescence lights, at 80 $\mu\text{mol m}^{-2} \text{s}^{-1}$.

Growth experiments under fluctuating nitrogen availability

The growth experiments under fluctuating nitrogen availability were carried out in three successive phases each one starting with 5×10^4 cells ml^{-1} . Media were artificial seawater (ASW) based *f/2* containing either 1 mM nitrate (nitrogen-replete condition) or 50 μM nitrate (nitrogen-deplete condition). *P. tricorutum* cells were first grown under nitrogen-replete condition, then washed twice with plain ASW and grown in nitrogen-deplete condition, and finally transferred to nitrogen-replete condition. A diagram depicting the growth experiment is shown in Supporting Information Fig. S1. Statistical analysis of growth parameters is described in Methods S1.

Phylogenetic analyses

DCR, AGO, and RDR homolog sequences were searched in NCBI database (www.ncbi.nlm.nih.gov), DOE JGI database (genome.jgi-psf.org), Ensemble database, and the Marine Microbial Eukaryote Transcriptome Sequencing Project, (MMETSP; Keeling *et al.*, 2014). Protein domains were identified by prediction tools as PFAM (EMBL-EBI, pfam.xfam.org) and CDD (Conserved Domain Database, NCBI, www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). Alignments were performed using the software MUSCLE from MEGA v.7 (Kumar *et al.*, 2016) and MAFFT (www.ebi.ac.uk/Tools/msa/mafft; Katoh & Standley, 2013) using default parameters. For DCR, AGO, and RDR proteins concatenated RNase IIIa and RNase IIIb, PIWI, and RDRP domains were aligned, respectively. Alignments were analyzed by Bayesian phylogenetic approach using MrBayes-3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). Phylogenetic trees of diatom DCR, AGO, and RDR were created with maximum likelihood (ML) analysis.

Details of the identified genes, their alignments, and phylogenetic analysis are provided in Methods S2 and Dataset S1.

Cloning of *PtDCR*, *PtAGO*, and *PtRDR* genes

Genomic and complementary DNA (cDNA) sequences of *PtDCR* (Phatr3_J48138), *PtAGO* (Phatr3_J52260), and *PtRDR* (Phatr3_J45417) genes were cloned. Genomic DNA was isolated from *P. tricornutum* following a modified CTAB protocol (Doyle, 1991). Genomic sequences were PCR amplified from genomic DNA using High-Fidelity DNA Polymerase Kapa Hi-Fi (Kapa Biosystems, Wilmington, MA, USA), gel purified, and ligated in pSpark[®] I vector (CANVAX, Boecillo, Valladolid, Spain). Total RNA was extracted using **guanidinium thiocyanate-phenol-chloroform extraction** method (Chomczynski & Sacchi, 1987). cDNAs were synthesized using the reverse transcriptase PrimeScript[™] RT-PCR kit (TaKaRa, Kusatsu, Shiga, Japan). cDNA sequences were PCR amplified using Phusion[™] High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA), gel purified, and ligated pUC19 vector. The cloned genes were transformed into competent *E. coli* cells and their sequences validated by Sanger sequencing. Primers used are indicated in Table S1.

CRISPR/Cas9-mediated generation of *P. tricornutum* DCR KnockOuts mutants

PtDCR gene was targeted by two single-guide RNAs (sgRNAs) in order to produce a *c.* 500-bp deletion. sgRNA sequences were designed using the PhytoCRISP-Ex (Rastogi *et al.*, 2016) and checked with CRISPOR (<http://crispor.tefor.net/>) and CHOP-CHOP (chopchop.cbu.uib.no) applications. Annealed complementary oligonucleotides corresponding to the sgRNA sequences were cloned into the pKS_diaCas9 vector (Nymark *et al.*, 2016) generating the vectors pKS_diaCas9_sgRNA1 and pKS_diaCas9_sgRNA2. Nucleotide sequences of the sgRNAs used are indicated in Table S2. The absence of genomic polymorphisms inside the targeted region of DCR was validated by PCR amplification and Sanger sequencing. The vector pNAT (Zaslavskaja *et al.*, 2000) carrying the nourseothricin resistance gene was used for mutant selection.

Vectors pKS_diaCas9_sgRNA1, pKS_diaCas9_sgRNA2, and pNAT were co-delivered in *P. tricornutum* cells by microparticle bombardment, as described in Apt *et al.* (1996), using tungsten beads M17, and the Biolistic PDS-1000/He Particle Delivery System (Bio-Rad) fitted with 1350 psi rupture disks. Cell lysates from nourseothricin-resistant single colonies (primary transformants) were used as templates in PCR to screen the presence of deletions in the *PtDCR* gene and conduct Sanger sequencing. Nucleotide and deduced protein sequences of DCR in the DCR-KO lines m4, m8, and m9 are provided in Dataset S2. Primers used are indicated in Table S1.

RNA extraction and transcriptome sequencing

Axenic cultures of *P. tricornutum* were grown to exponential phase ($1.8\text{--}2.4 \times 10^6$ cells ml⁻¹). Pelleted cells were ground in liquid nitrogen, and total RNAs were extracted using the Total RNA Purification kit from Norgen Biotek (Thorold, ON, Canada). RNA

quality was controlled by migration on 1.2% MOPS/formaldehyde gel and bioanalyser (Agilent, Santa Clara, CA, USA).

In a first step, sRNA library construction and sequencing were carried out in the WT and DCR-KO m4 lines. sRNA libraries were prepared in our laboratory using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs, Ipswich, MA, USA). Libraries were sequenced at the IMBB-FORTH sequencing facility on an Illumina Nextseq 500 platform using single-end 75-bp run format. In a second step, sRNA library construction and sequencing were carried out in the WT and DCR-KO m8 and m9 lines by the service provider Omega Biotek (Norcross, GA, USA). sRNA libraries were prepared using the Bioo NEXTflex[™] Small RNA-Seq Kit v3 (PerkinElmer, Waltham, MA, USA), which employed a randomized adaptor protocol reported to reduce sequence bias during sRNA ligation. Libraries were sequenced on a HiSeq 2500 platform using single-end 50-bp run format.

mRNA cDNA library construction and sequencing were carried out by the service provider Genewiz (Leipzig, Germany). Libraries were prepared from 2 µg of total RNA following Poly(A) selection and were sequenced on an Illumina NovaSeq platform using strand-specific RNA-seq 2 × 150-bp format.

The number of reads obtained in sRNA and mRNA library sequencing is listed in Dataset S3.

Chromatin immunoprecipitation and sequencing (ChIP-seq)

Chromatin immunoprecipitation (ChIP) was conducted as described in Lin *et al.* (2012) with few modifications. ChIP was carried out using the ChIP grade monoclonal antibodies Anti-Histone H3 (dimethyl-K9) antibody (#ab1220; Abcam, Cambridge, UK), Tri-Methyl-Histone H3 (Lys27; C36B11) Rabbit mAb (#9733; Cell Signaling Technology, Danvers, MA, USA), and Tri-Methyl-Histone H3 (Lys9; D4W1U) Rabbit mAb (#13969; Cell Signaling Technology).

DNA libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit (E7645) and sequenced at the IMBB-FORTH sequencing facility on an Illumina Nextseq 500 platform. The number of reads obtained is listed in Dataset S3. The material and methods used in the ChIP-seq experiment are detailed in Methods S3.

Sequencing data analysis

sRNA analysis The fragments from the sRNA libraries were aligned using BWA MEM v.0.7.17 (Li & Durbin, 2009) onto *P. tricornutum* genome v.2 with command line parameters: -k 12 -t 7 -A 1 -L 2 -T 13. The minimum aligned length was 14 bp, and reads with a minimum mapping quality of 5 after being soft-clipped on their end were kept for further analysis and defined the sRNA fragments. For each replicate, the number of aligned reads was used to normalize fragment counts to 5 million mapped fragments. Processing of alignment was done using SAMTOOLS (Li *et al.*, 2009). Genomic interval algebra was performed using BEDTOOLS v.2.29.2 (Quinlan & Hall, 2010).

Differential expression (DE) analysis was performed using the DESeq2 package (Love *et al.*, 2014), and an adjusted *P*-value threshold of 0.1 was used to detect DE when comparing each KO strain to the WT.

Definition of sRNA-associated loci and DCR-dependent sRNA clusters (DDSCs) For each WT strain and each library preparation protocol, we first classified genomic clusters covered by sRNA reads and then annotated DCR-dependent sRNA clusters and genes based on fragment properties (see Methods S4; Fig. S2).

We defined a locus as DCR-dependent sRNA cluster (DDSC) if its average strandedness value was between 0.2 and 0.8, its average peak of fragment length value was between 26 and 31 nt, and it was significantly DE in each mutant when compared to WT with a \log_2 fold change below -0.8 (Fig. S2c). A total of 387 DDSCs were defined.

Definition of DCR-sRNA-associated and independent genes (DAGs and DIGs) Most of the DDSC overlapped genes, including TE-associated genes. In order to avoid reporting too many short sRNA loci corresponding to the same gene region, we further aggregated DDSCs in a gene-centered way and reapplied our filters on fragment composition and sRNA regulation in KO (Fig. S2d). Thus, we defined DDSC-associated genes (DAGs) by selecting only genes whose set of reads (sRNAs) had properties as defined previously for DDSCs (Fig. S2c) and also presented a significant drop in fragment coverage in at least one of the KOs (adjusted *P*-value < 0.1 , \log_2 fold change KO/WT < -0.8). This resulted in a set of 306 DAGs. As a control, we also define DDSC-independent genes (DIGs), which correspond to genes either mapped with DCR-independent sRNAs or not mapped with any sRNA.

Gene expression analysis RNA-seq reads were aligned using STAR (Dobin *et al.*, 2013) with default parameters. Read counts were then aggregated at the gene level using FEATURECOUNTS (Liao *et al.*, 2014). Gene expression analysis was performed with the DESeq2 package (Love *et al.*, 2014), using an adjusted *P*-value threshold of 0.1 to call a gene differentially expressed (DE). For each KO, we define four categories of regulated genes: (1)

not-expressed 'NotExp', (2) upregulated 'Up', (3) downregulated 'Down', and (4) not significantly affected 'Not significant'. Parameters used to define each category for DAGs and DIGs in merged or individual mutants are described in Methods S5.

ChIP-sequencing analysis The marks were generated in duplicate libraries for all marks in both the WT and the DCR lines, with each library being constructed with chromatin extracted from two independent cultures. Reads were aligned to the genome using bwa mem (with parameters $-k\ 15\ -t\ 10\ -A\ 1\ -L\ 2\ -T\ 20$). Replicates showed a good correlation (PCC > 0.95), except one replicate in H3K9me3-m8 mutant (library 2134) that had a very low mapping rate so we put this library aside. The details of mapping results of all sequenced libraries are listed in Dataset S3. Histone marks were characterized in and compared between the WT and DCR-KO lines using a combination of the SICER2 (Zang *et al.*, 2009) and DESeq2 (Love *et al.*, 2014) tools (see Methods S6).

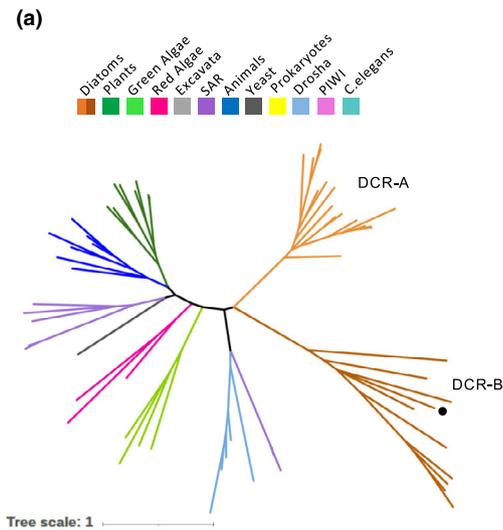
Results

Phylogenetic analysis of diatom DCR, AGO, and RDR

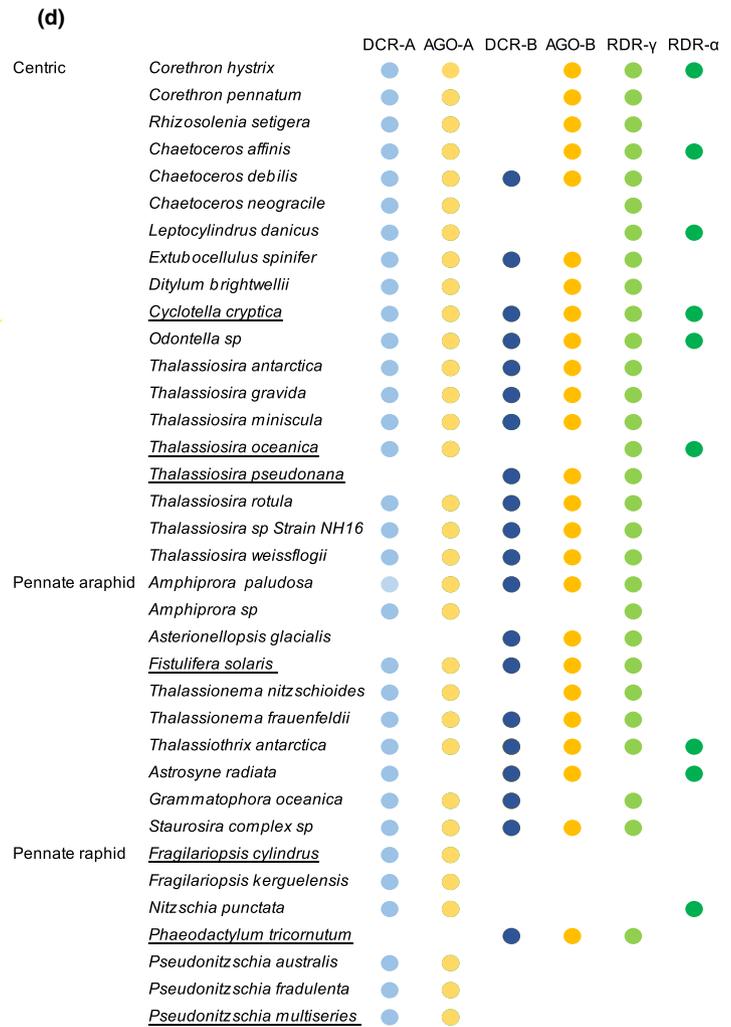
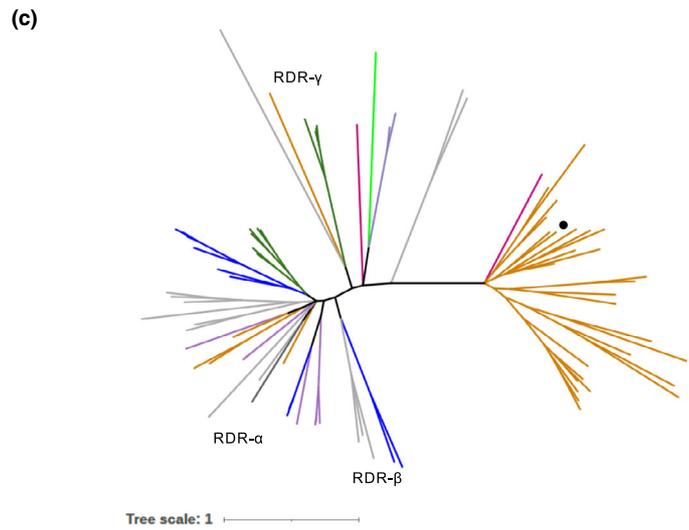
We identified a single homolog of DCR, AGO, and RDR encoding genes in the recently reannotated version of *P. tricornutum* genome Phatr3 (<https://www.diatomicsbase.bio.ens.psl.eu/>). *PtDCR* (Phatr3_J48138), *PtAGO* (Phatr3_J52260), and *PtRDR* (Phatr3_J45417) gene models were confirmed in RT-PCR amplification, cloning, and Sanger sequencing.

Genome and transcriptome mining, protein functional domain prediction, and phylogenetic analysis were conducted across a range of 38–68 different species to catalog the repertoire of DCR, AGO, and RDR in 36 diatom species (Dataset S1). Diatom DCRs clustered in a diatom-monophyletic group, including two clades (DCR-A and -B; Figs 1a, S3). DCR-A proteins had the typical domain architecture of animal and plant DCRs. They presented a Helicase-C (Hel), a Domain of unknown function (DUF), a dimerization (DMD), a PAZ, and two conserved RNase III domains (RNase IIIa and b; Fig. 1a). DCR-B proteins had a typical architecture of protist DCRs and animal DROSHA. DCR-B proteins are shorter than DCR-A proteins and presented only a PAZ

Fig. 1 Phylogenetic analysis of diatom DICER (DCR), ARGONAUTE (AGO), and RNA-dependent RNA (RDR) polymerase proteins. (a) Phylogenetic tree of DCR homologs based on the concatenated RNase IIIa and RNase IIIb domains among eukaryotes. RNase III domains of Drosha proteins are included. Protein domain architecture of diatom Clade A and B DCRs is indicated. Eukaryotic groups and protein families (Drosha and PIWI) are indicated by color code, with diatom Clade A and B DCRs with light and dark brown, respectively. A black dot is placed next to *Phaeodactylum tricornutum* DCR. The same color code is used in panels b and c. The complete list of species and posterior probabilities are given in Supporting Information Dataset S1 and Fig. S3. Helicase-C domain (Hel), Domain of unknown function (DUF), PIWI/Argonaute/Zwille (PAZ) domain, dimerization domain (DMD), RNase IIIa and RNase IIIb domains (RIIIa and RIIIIb), and Stramenopiles-Alveolates-Rhizaria (SAR). (b) Phylogenetic tree of AGO homologs based on PIWI domain among eukaryotes and a few representative prokaryotes. PIWI proteins as paralogs of AGOs are included. Protein domain architecture of diatom AGO Clade A and Clade B is indicated. The complete list of species and posterior probabilities are given in Dataset S1 and Fig. S8. N-terminal domain (N), PIWI/Argonaute/Zwille (PAZ) domain, middle (MID) domain, and the P-element-induced whimpy tested (PIWI) domain, *Caenorhabditis elegans* (*C. elegans*). (c) Phylogenetic tree of RNA-dependent RNA (RDR) polymerase homologs based on RDR domain among eukaryotes. Groups of type α , β , and γ RDR proteins are indicated. The complete list of species and posterior probabilities are given in Dataset S1 and Fig. S10. (d) Distribution of DCR, AGO, and RDR proteins, with their respective clades and types, among diatom taxa. Colored circles indicate the presence of each homolog in the respective species. Species with underlined name indicate that whole genome sequencing information is available.



DCR-A	Hel	DUF	DMD	PAZ	Rilla	Rlllb
DCR-B		PAZ	Rilla	Rlllb		



and two RNase III domains (Figs 1a, S4). In addition, RNase IIIb domains of DCR-B proteins were poorly conserved and most residues known to be critical for DCR dicing activity (Zhang *et al.*, 2004; Gan *et al.*, 2006) were absent (Figs S5–S7).

P. tricornutum and *T. pseudonana* single DCR homologs reported by De Riso *et al.*, 2009 clustered in DCR clade B.

Diatom AGOs clustered in two phylogenetic clades (AGO-A and -B), both containing conserved PAZ and PIWI domains

(Figs 1b, S8). AGO-A proteins clustered away from all known AGOs. AGO-B proteins, which include *P. tricornutum* single AGO homolog, were grouped with all known eukaryotic AGOs. The catalytic residues involved in PIWI slicing activity, however, were less conserved in AGO-B than in AGO-A proteins (Fig. S9).

Eukaryotic RDRs are generally classified into α , β , and γ types (Zong *et al.*, 2009). Most diatom RDRs, including *P. tricornutum* RDR, clustered with γ -type RDRs which include RDRs from plants, green/red algae, and chromalveolates (Figs 1c, S10). A few diatoms (e.g. *Leptocylindrus danicus* and *Thalassiosira oceanica*) also presented α -type RDRs, which included homologs found in animals, plants, and chromalveolates.

When looking at the distribution of DCR/AGO/RDR across diatoms, DCR and AGO but not RDR were found in all investigated species (Fig. 1d). Overall, we found a wide distribution of DCR and AGO of A/B clades in both centric and pennate species, suggesting that these forms were present in the diatom common ancestor. DCR- and AGO-B proteins, however, were generally absent in the evolutionary more recent raphid pennate species. When DCR of only one clade was present (either DCR-A or -B), the AGO found was again either only clade A or B, respectively. This observation suggests functional interaction between DCR- and AGO-A and between DCR- and AGO-B proteins. Taken together, our results indicate that DCR and AGO are conserved in diatoms and suggest interspecies differences in their number of RNAi pathways and possibly their function.

Identification of DCR-dependent sRNA in *P. tricornutum*

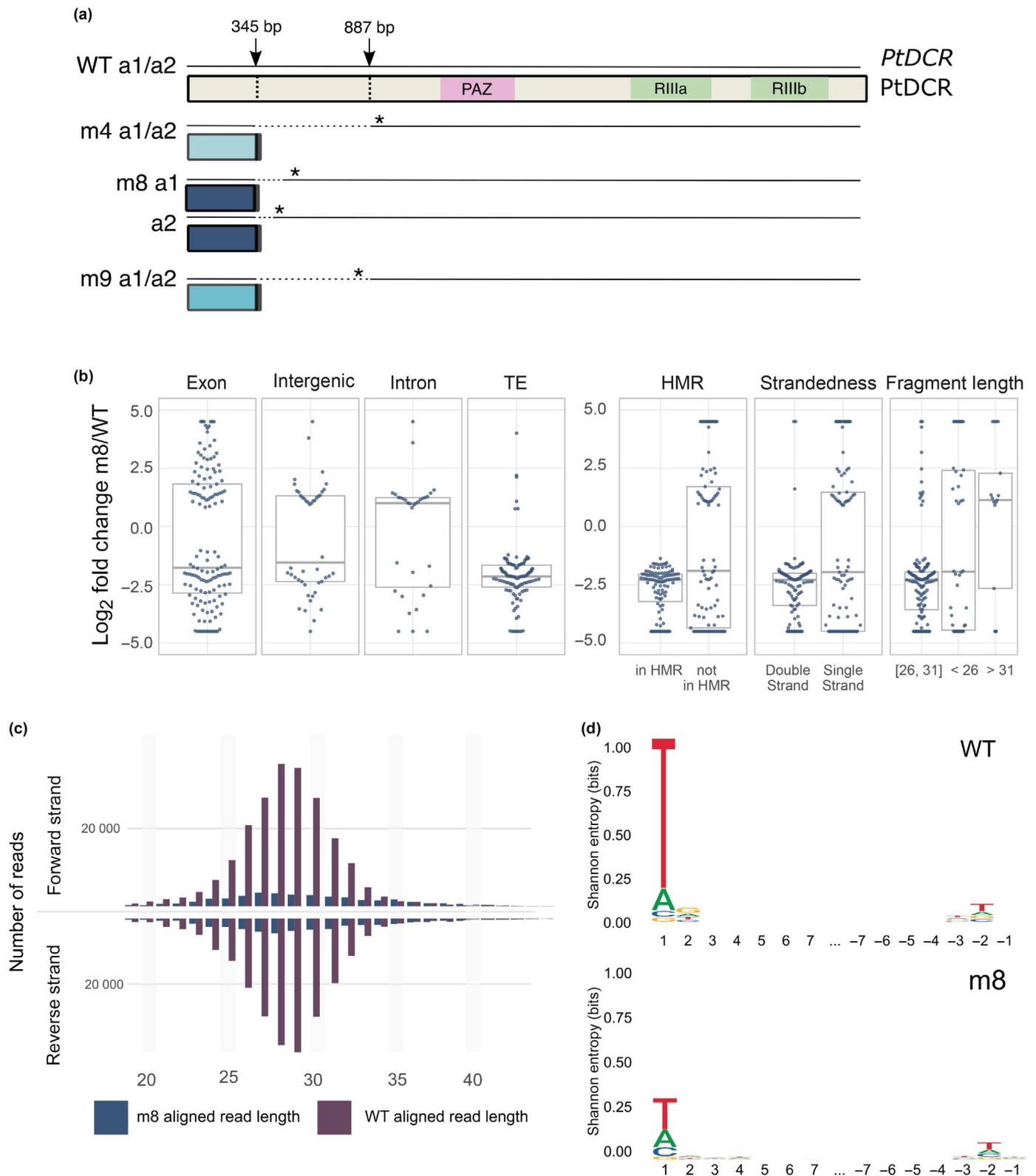
To gain insight into the biogenesis and functional role of DCR-dependent sRNA in diatoms, we focused our study on the model species *P. tricornutum*, which presents a single DCR homolog, publically available (epi)genomic information and validated genetic tools. We generated PtDCR knockout lines (DCR-KO) by CRISPR-Cas9 approach and conducted comparative analysis of WT vs DCR-KO sRNA transcriptomes. CRISPR-Cas9 mutagenesis was targeted at two genomic sites located upstream of the predicted PAZ and RNase III domains, which are known to be indispensable for DCR activity (Zhang *et al.*, 2004; Macrae *et al.*, 2006). Twenty-one mutants were found with an integrated Cas9 transgene (Table S3), and three biallelic DCR-KO lines that presented deletions which introduced a premature stop codon around the targeted sites were selected for further analysis

(m4, m8, and m9; Fig. 2a; Dataset S2). We carried out two independent sRNA transcriptome analyses using biological quadruplets of WT and m4 lines, and biological triplicates of WT, m8, and m9 lines grown in standard conditions. In these two analyses, we used different sRNA library preparation methods and next-generation sequencing platforms (see the **Materials and Methods**). A total of 143 million sRNA reads corresponding to 12 210 862 unique sequences with size ranging from 12 to 150 nt were sequenced (Dataset S3). Based on Phatr v3 genome and previous publications, we mapped sRNA and retrieved structural annotations and epigenetic marks (Veluchamy *et al.*, 2013, 2015) of the sRNA-mapped loci. Although the total sRNA coverage varied between the two methods, the overall patterns of sRNA alterations in DCR-KOs were consistent.

In WT, we found abundant single-stranded sRNA of 12–20 nt mapped to exons of non-TE genes (27–17% of the average read counts per category), tRNA loci, and intergenic regions (6–13%), of 20–26 nt mapped to intergenic regions (0.5–6%), and of 60–150 nt mapped to introns and intergenic regions (18% Fig. S10a,b). In addition, we found double-stranded sRNA of 12–20 nt mapped to exon of non-TE genes (1.7–6.9%) and of 26–31 nt mapped to exons of non-TE genes (4.6–5.4%) and to TEs (16–24%). The 26–31-nt sRNA represented the bulk of the double-stranded sRNAs (70–86% of the double-stranded). The originating loci, strandedness, size range, and relative abundance of the sRNA populations we identified in *P. tricornutum* are in agreement with previous analysis (Rogato *et al.*, 2014; Dataset S3; Fig. S11).

Compared with WT, DCR-KOs presented a reduction of sRNA mapped to TEs (3- and 5.5-fold change reduction in normalized read count –NEB and NEXTFLEX, respectively) that extensively overlapped with highly methylated regions (HMRs) previously reported in Veluchamy *et al.*, 2013 (3- and 5.8-fold change reduction –NEB and NEXTFLEX, respectively; Figs 2b, S12, S13). When considering sRNA structural properties, double-stranded 26–31-nt sRNAs were markedly less abundant in DCR-KOs than in WT (3.3- and 8.2-fold change reduction –NEB and NEXTFLEX, respectively, Figs 2b, S12, S13). At HMRs specifically, double-stranded sRNA of 26–31-nt sRNAs (size-peak at 28–29 nt, see Figs 2c, S13) were near absent in DCR-KOs compared with WT (3.2- and 8.8-fold change reduction, respectively). These DCR-dependent sRNAs (DCR-sRNAs) presented a strong bias at their 5' end first nucleotide toward a uridine (Fig. 2d). The strandedness, discrete size range,

Fig. 2 Small RNAs (sRNAs) in *Phaeodactylum tricornutum* DCR KnockOuts (PtDCR-KOs). (a) Targeted PtDCR locus by CRISPR/Cas9-mediated genome editing and selected PtDCR-KOs m4, m8, and m9 used in this study. Dashed lines represent deleted loci, and asterisks indicate stop codons due to frame-shift at the a1/a2 alleles of diploid *P. tricornutum*. Predicted truncated proteins are illustrated with the mutant-specific color (m4 light blue, m8 gray, and m9 blue). (b) Change in small RNA (sRNA) coverage in the DCR m8 KO vs WT. Left: Log₂ fold change of sRNA coverage KO in m8 vs WT according to their overlap with different functional categories: exon, intergenic, intron, and transposable element (TE). Each point corresponds to an element in the corresponding functional category. Right: Log₂ fold change of sRNA coverage m8 vs WT according to their strandedness, their most represented fragment length, and their overlap either with highly methylated regions (HMRs) or not methylated loci. Each point corresponds to a sRNA cluster. Horizontal lines are, from bottom to top, first quartile, median, and third quartile. (c) Length of fragments overlapping HMRs in WT (brown) and in m8 KO (dark blue). For each aligned fragment length, the number of fragments aligning on the forward (positive values) and on the reverse (negative values) genomic strand are reported. (d) Nucleotide composition at the 5' end (positive values) and 3' end (negative values) of the fragments aligning to HMRs for the WT (top) and the m8 KO (bottom).



and 5' end nucleotide bias of the sRNA absent in DCR-KOs are in agreement with their processing by a RNase III enzyme of the DCR family. Collectively, our results demonstrate that *PtDCR* processes double-stranded 26–31-nt sRNA originating from methylated non-TE genes and TEs.

Characterization of DCR-dependent sRNA-associated genes

To characterize the repertoire of DCR-sRNA homologous loci, we first used reads strandedness, size range, and drop of sRNA

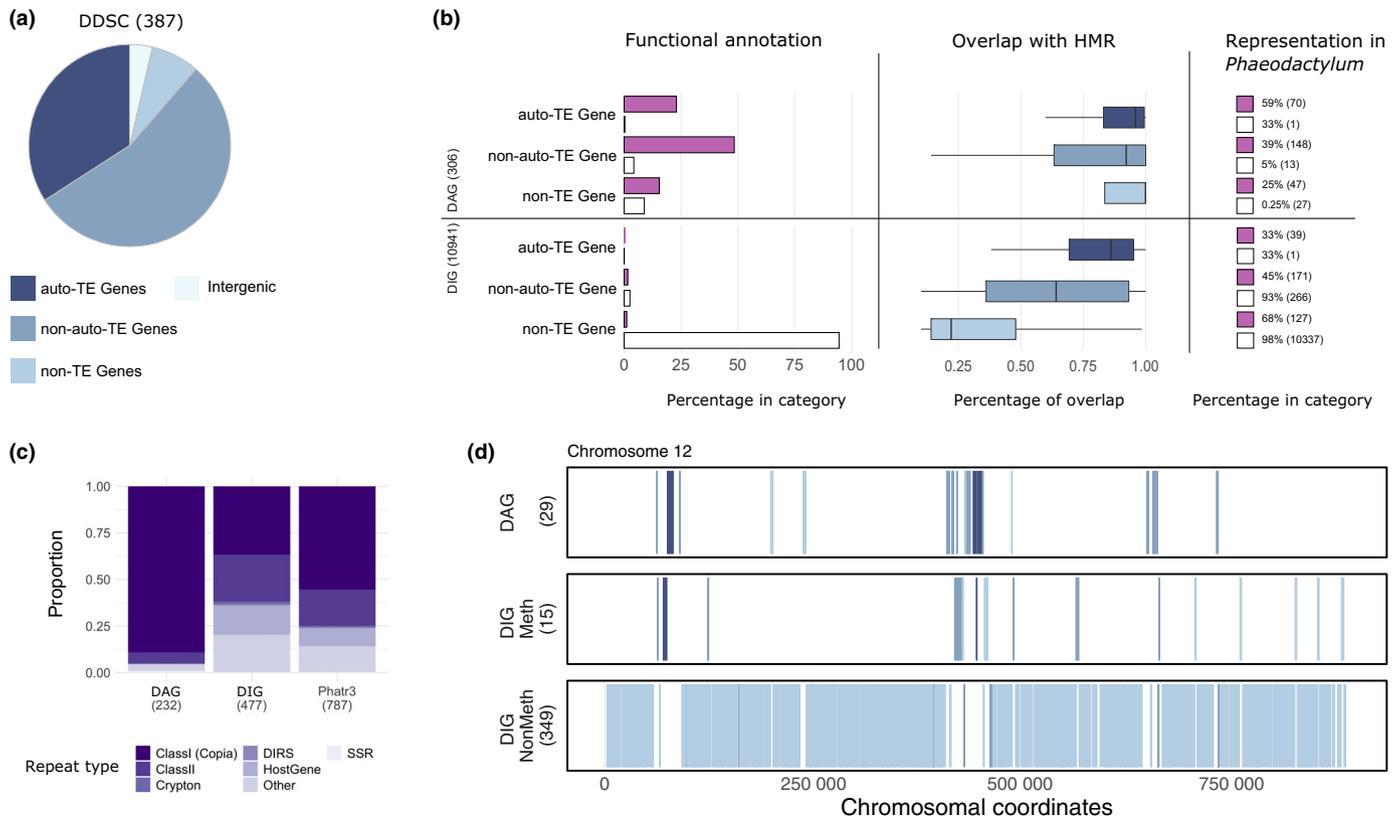


Fig. 3 Properties of the DCR-dependent sRNA clusters (DDSCs) and their associated genes. DDSCs are defined as loci exhibiting specific fragment properties (double-stranded and with a main fragment length *c.* 27–28 nt) and drop of whole sRNA coverage in any of the 3 KOs. (a) Annotation of DCR-dependent sRNA clusters (DDSC) according to their primary overlap: Autonomous TE (34%), not autonomous TE (54%), Non-TE Genes (8%), or intergenic regions (4%). (b) Left: Functional annotation of DAGs and DIGs (see the [Materials and Methods](#)) given that they are methylated (pink) or not (white). Middle: Proportion of the genes overlapping highly methylated regions (HMR) for methylated DAGs and methylated DIG, according to their functional category (colors as in a). Right: Representation of the DAGs and DIGs in each functional category of Phatr3 given their methylation status. Total numbers are given in parentheses. (c) Transposon family composition of transposable elements (TEs) within different groups. Transposon family for genes within TEs for different classes (Phatr3 annotation). From left to right: DAGs, DIGs, or all *Phaeodactylum tricornutum* genes. (d) Location of DAG and DIG on Chromosome 12. Color code is as in panel (a).

detected in each KO to determine DCR-dependent sRNA clusters (DDSC) at the genome level. More than three quarters (85%) of the DDSCs mapped to TEs, with nearly half of those being autonomous transposons (aTEs), that is TEs carrying the ORFs indispensable for transposition (see [Methods S7](#); [Fig. 3a](#)). This is significant since TEs cover only 6.4% of *Phaeodactylum* genome (Maumus *et al.*, 2009). Since most DDSCs overlapped genes ([Fig. 3a](#)), we aggregated DDSC in a gene-centric manner into DAGs for each KO. We also considered DIGs, corresponding to genes either covered with DCR-independent sRNAs or not covered by any sRNA at all (see [Methods S4](#); [Fig. S2](#)).

We retrieved a total of 306 unique DAGs (and 10 941 DIGs, see [Dataset S4](#)). DAGs were well shared between the DCR-KOs with almost three quarters (228) found in at least two mutants ([Fig. S14a](#)). DAGs corresponded mostly to TE genes (232, 76%), which were markedly less represented in DIGs (4%, $P < 2.2e^{-16}$ Fisher exact test; [Fig. 3b](#)). DCR-sRNA coverage on DAGs was the largest on autonomous TE genes (aTEGs), intermediate on nonautonomous TE genes (naTEGs), and the smallest on non-TE genes (aTEGs vs naTEGs $P = 0.0018$ one-sided *t*-test and naTEGs vs non-TE genes $P = 3.6e^{-10}$ one-sided *t*-test;

[Fig. S14b](#)). Consistent with a preferential processing of DCR-sRNA from methylated genes, the vast majority (87%, 265) of the DAGs were methylated. In contrast, DIGs were mostly composed of nonmethylated genes (94%). Overall, although DAGs accounted for only 2.6% of *P. tricornutum* gene complement, they cumulated near 39% of all methylated genes, concentrating a large fraction of the transposons ([Fig. 3b](#)). Furthermore, the proportion of aTEGs was 7.6 times more abundant in DAGs than in DIGs ($P < 2.2e^{-16}$ Fisher exact test), and more nonautonomous TE genes were methylated in DAGs ($P < 2.2e^{-16}$ Fisher exact test). Notably, methylated DAGs showed consistently higher methylation coverages than methylated DIGs, for TE genes (for aTEGs $P = 0.009$ *t*-test and for naTEGs $P = 4.9e^{-12}$ *t*-test) and especially for non-TE genes ($P < 2.2e^{-16}$ *t*-test; [Fig. 3b](#)), indicating a correlation between DCR-sRNAs and the level of DNA methylation. DAG TE repertoire consisted mainly (89%, 207) of class I retrotransposons of the long terminal repeat retrotransposon (LTR-RT) *Ty1/Copia* family, which were less represented in DIGs (37%, $P < 2.2e^{-16}$ Fisher exact test) or among all *Phaeodactylum* TEs (56%, $P < 2.2e^{-16}$ Fisher exact test; [Fig. 3c](#)). Global inspection of the chromosomal landscape of

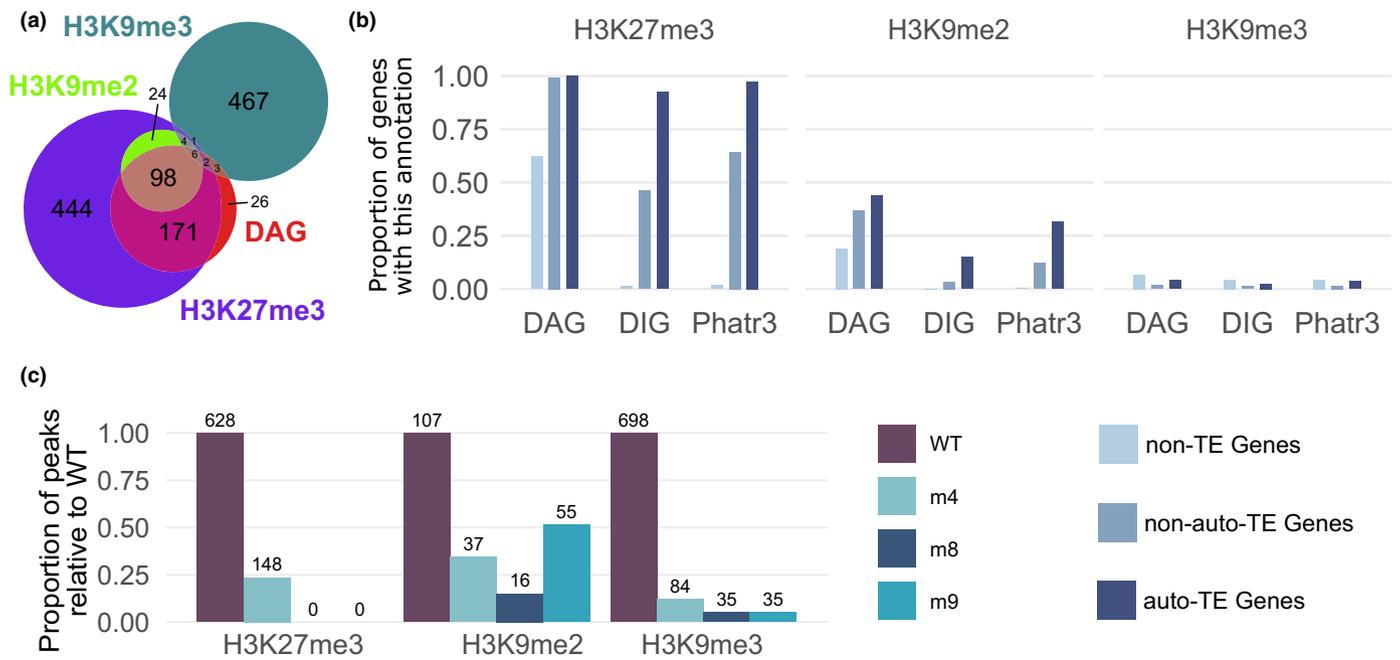


Fig. 4 DCR is involved in the maintenance of the repression-associated histone marks H3K9me2/3 and H3K27me3. (a) Overlap of the genes covered by H3K9me2/3 and H3K27me3, and of the DCR-sRNAs associated genes (DAGs). (b) Number of genes in each category possessing a given histone mark. (c) For each histone mark, proportion of islands relative to the WT in each of the DCR-KO. The corresponding number of islands is reported above each bar.

DAGs indicated that some DAGs were located next to methylated DIGs, while others were distant from the closest methylated DIGs by > 100 kb (Fig. 3d; Dataset S5). Taken together, our results indicate that DCR-sRNAs target a specific subset of methylated genes covering nearly half of the *P. tricornutum* repertoire of methylated TE genes with a strong bias toward autonomous LTR-RT elements.

Analysis of repression-associated histone marks in DCR-KO mutant lines

Because of the known association between RNAi and heterochromatin in various eukaryotes (Djupedal & Ekwall, 2009; Holoch & Moazed, 2015), we decided to conduct genome-wide ChIP-seq comparative analysis in WT vs DCR-KO lines of three most abundant repression-associated histone marks (H3K27me3, H3K9me2, and H3K9me3) previously reported in *P. tricornutum* (Veluchamy *et al.*, 2015; Zhao *et al.*, 2021).

In the WT line, we identified 628 regions (mean length: 3872 bp) covered by H3K27me3, 107 (mean length: 5415 bp) by H3K9me2, and 698 (mean length: 3224 bp) by H3K9me3. The number of genes covered by H3K27me3, H3K9me2, and H3K9me3 were 750, 134, and 483, respectively (Fig. 4a). The majority of the genes covered by H3K9me2 were also covered by H3K27me3 (98, 73% of the H3K9me2 covered genes). Conversely, the majority of the genes covered by H3K9me3 were not covered by the two other marks (467, 97% of the H3K9me3 covered genes). Almost all DAGs (90%) were covered by H3K27me3 and one third by H3K9me2 (Fig. 4a). Most of the H3K27me3 and H3K9me2 marks overlapped with (n)aTEGs (51% and 80%, respectively) (Fig. 4b).

We then assessed how the marks were affected in the DCR-KO mutants. In all three mutant lines, we observed a consistent effect of loss of most of the marks present in the WT (Fig. 4c). In two out of three mutants, 100% (H3K27me3), and at least 65% (H3K9me2) and 95% (H3K9me3) of the marks were lost. Conversely, no new marks were observed in the mutants. Altogether, these results indicate a correlation between DCR-sRNAs and the maintenance of the repression-associated epigenetic marks H3K27me3, H3K9me2, and H3K9me3.

Role of DCR-dependent sRNA in the control of gene expression

To gain insights into the role of DCR-sRNA in the control of gene expression, we conducted quantitative mRNA-seq analysis in *P. tricornutum* WT and DCR-KO strains. In a first step, we compared DIG and DAG expression levels in the WT strain (Figs 5a,b, S15). Normalized gene expression levels were used to sort genes in three groups: expressed, expressed at low levels, and not-expressed (see Methods S5). The proportion of genes not-expressed or expressed at low level was the largest in DAGs (*c.* 80%), intermediate in methylated DIGs (*c.* 60%) and the smallest in nonmethylated DIGs (*c.* 10%; Fig. 5a). When taking into consideration gene structural annotation together with their methylation characteristics, we found that the expression levels of methylated non-TE genes were markedly lower in DAGs than in DIGs (*t*-test *P*-value < 2.2e⁻¹⁶, Figs 5b, S15). Conversely, the expression levels of methylated (n)aTEGs were similar between DAGs and DIGs. These results indicate that DCR-sRNAs are associated with transcriptionally repressed genes.

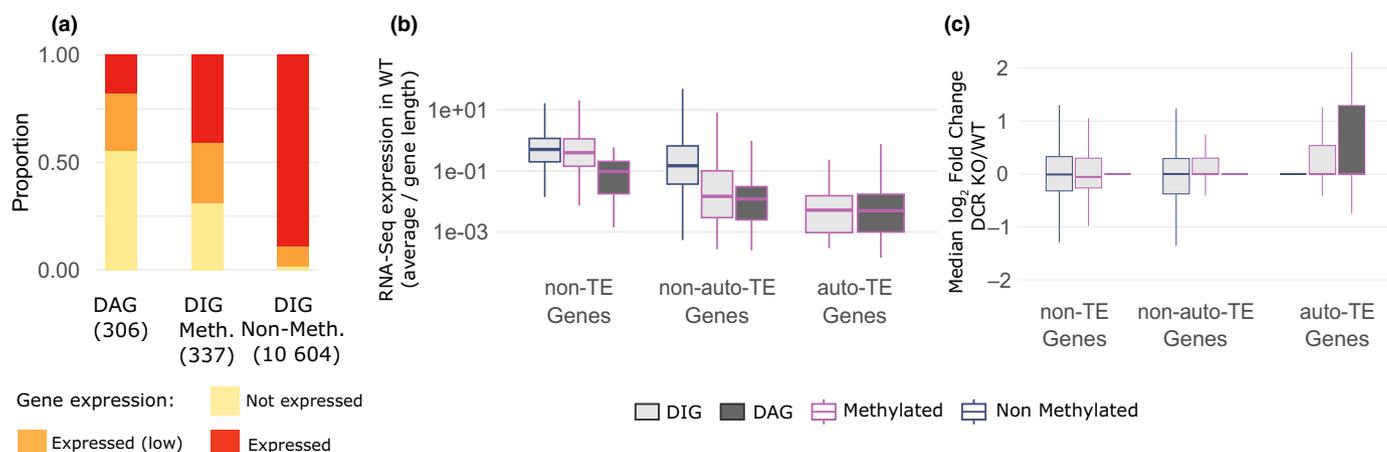


Fig. 5 Global gene expression analysis in *P. tricornutum* DCR-KOs vs wild-type (WT) lines. (a) Expression level of DCR-dependent sRNA clusters associated (DAGs) and independent (DIGs) genes in WT. Proportion of genes expressed (red), expressed in low levels (orange) or not-expressed (yellow) for all DAGs, methylated DIGs, and nonmethylated DIGs. (b) RNA-Seq gene expression in WT according to the functional category of the gene and whether it is DCR dependent or not: Nonmethylated DIG (dark blue line), methylated DIGs (pink line), or methylated DAGs (dark gray fill). Horizontal lines are, from bottom to top, first quartile, median, and third quartile. Vertical lines are 1.5 the interquartile range. (c) RNA-Seq fold change of KO/WT according to the functional category of the gene and whether it is DCR dependent or not (colors as in b). For each gene, the median of the \log_2 fold change of the three mutants was considered. Horizontal lines are, from bottom to top, first quartile, median, and third quartile. Vertical lines are 1.5 the interquartile range.

In a second step, we compared gene expression levels in the three DCR-KOs together vs the WT strain (Fig. 5c). Globally, DIG and DAG expression levels were stable. Some groups of DIGs and DAGs, however, presented distinct trends of transcriptional regulations. In DIGs, (methylated) non-TE genes and nonmethylated naTEGs tended to be differentially expressed toward up- and downregulation. Conversely, methylated aTEGs and naTEGs tended to be principally upregulated (Fig. 5c). In DAGs, while non-TE genes and naTEGs were stable, aTEGs tended to be principally upregulated and so with a larger amplitude than their DIG counterparts (Fig. 5c). These results indicate that while DCR-KO had no global effect on gene expression levels, some group of genes presented distinct trends of transcriptional regulation in function of their structural annotation and methylation characteristic.

In a third step, we compared gene expression levels in each DCR-KO individually vs the WT strain. Based on their regulation, genes were classified into four distinct transcriptional groups: upregulated (Up), downregulated (Down), not-significantly-affected expressed genes (Not significant), and not-expressed genes (NotExp; see the [Materials and Methods](#)). Between 1000 and 2000 differentially expressed genes (DEGs) were detected in each DCR-KO corresponding to 9–18% of the *P. tricornutum* gene repertoire (Table S4). DAGs accounted for only 3% of the DEGs divided in 43 up- and 43 downregulated genes in total (Fig. S16). Conversely, DIGs accounted for nearly 97% of the DEGs divided in 1561 up- and 1373 downregulated genes in total. The set of differentially expressed DAGs and DIGs was poorly shared across the DCR-KOs, indicating that their transcriptional regulation was mutant-specific (Fig. S16).

We next compared the different transcriptional groups of DAGs and DIGs based on their initial expression levels in the WT strain (Fig. S17a), their structural (Fig. S17b), and functional (gene ontology terms) annotations (Fig. S17c). DAGs and DIGs from

the three DCR-KOs were merged to pinpoint common trends across the mutants (see [Methods S5](#)). Up- and Down-DAGs were mostly (*c.* 60%) low expressed, while their DIGs counterparts were mostly (> 75%) expressed genes in the WT strain (Fig. S17a). Up- and Down-DIGs were mostly non-TE genes (> 90%). Conversely, Up-DAGs were mostly aTEGs and naTEGs (> 80%) while Down-DAGs were mostly naTEGs (> 50%; Fig. S17b). The different transcriptional groups of DIGs presented similar gene ontology (GO) profiles reflecting the overall GO profile of the *P. tricornutum* gene repertoire (Fig. S17c). The different transcriptional groups of DAGs differed from their DIG counterparts by a larger representation of the TE genes associated GO terms ‘Nucleic acid binding’ and ‘DNA integration’ concordant with the larger proportion of (n)aTEGs found in DAGs than in DIGs. Between DAGs, Up-DAGs presented the largest proportion of genes with the aforementioned GO terms concordant with the larger proportion of aTEGs in this group. These results indicate that the differentially expressed DAGs were principally (n)aTEGs initially low expressed in the WT strain. Conversely, the differentially expressed DIGs were principally non-TE genes without bias toward any particular molecular function and initially expressed in the WT strain.

Taken together, our mRNA-seq analysis indicates that while DCR-sRNAs are associated with transcriptionally repressed genes, they are apparently dispensable to maintain the repression of the majority of their targeted genes with the exception of few (n)aTEGs, which were derepressed in a stochastic manner.

Phenotypic analysis of DCR-KO during the acclimatory response to changes in nitrogen availability

Previous studies suggest that epigenetic mechanisms control LTR-RT reactivation and regulate the expression of non-TE genes involved in nitrogen metabolism in *P. tricornutum* (Maurus *et al.*, 2009; Veluchamy *et al.*, 2013, 2015). In the present

study, we hypothesized that DCR-KOs may be affected in their capacity to regulate TE and non-TE gene expression in response to fluctuating nitrogen availability, which may lead to reduced cell fitness. DCR-KOs and WT were sequentially grown in liquid batch cultures under nitrogen-replete condition, transferred to nitrogen-deplete condition, and returned to nitrogen-replete condition (Figs 6, S1; Method S1; Table S5; Dataset S6). We did not observe any consistent difference in cell morphology and length between the WT and the three DCR-KO lines (Dataset S6). In terms of growth kinetics, DCR-KO and WT grown under initial nitrogen-replete condition presented similar lag-phase durations (*c.* 3 d), growth rates at exponential phase, and maximum cell density at stationary phase. When transferred to nitrogen-deplete condition, both WT and DCR-KOs cultures were severely affected, with their maximum cell density reaching only 10–15% to that found in initial nitrate replete condition.

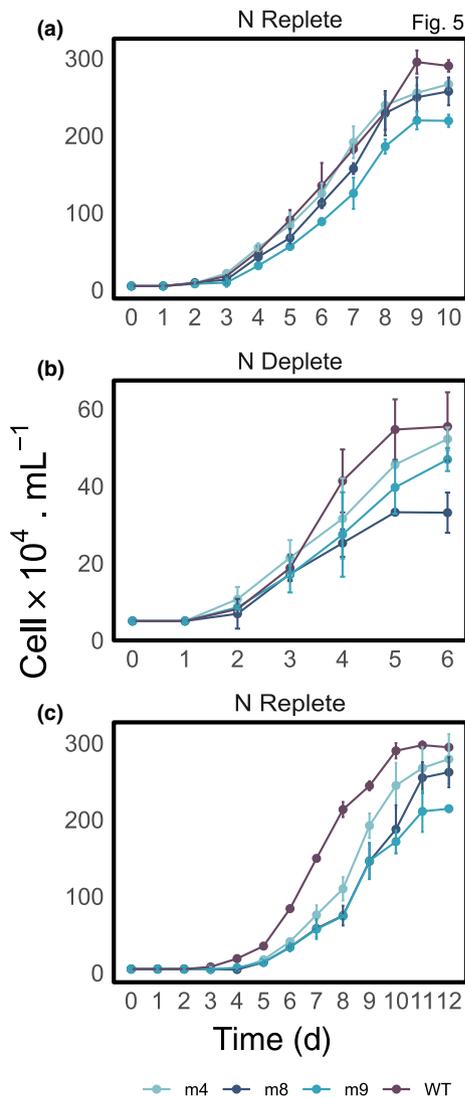


Fig. 6 Growth of *P. tricornutum* DCR-KO and wild-type (WT) lines under sequential changes in nitrate availability. Growth kinetics of WT and DCR-KO lines m4, m8, and m9 under (a) initial nitrate replete, (b) intermediate nitrate deplete, and (c) final nitrate replete conditions. Error bars indicate standard deviation (biological replicates $n = 3$).

Nitrogen limitation had the opposite effect on WT and DCR-KOs growth rate, that is increasing in WT while decreasing in DCR-KOs. Among the three DCR-KOs, however, only m8 presented a significantly ($P < 0.01$) lower growth rate compared with WT under nitrogen limitation. When returned to nitrogen-replete condition, both WT and DCR-KOs presented similar growth rates and maximum cell density values close to that found under initial nitrate replete condition. Lag-phase durations, however, were extended in DCR-KOs compared with WT by 1–2 d. Our results therefore indicate that DCR-KOs present a lower fitness when acclimating to rapidly changing nitrogen availability, suggesting a role for DCR in the regulation of gene expression under this fluctuating growth condition.

Discussion

Our study uncovered an unanticipated diversity of the key RNAi effectors in diatoms. It has been suggested that the last common ancestor of all modern eukaryotes possessed at least one DCR, AGO, and RDR homolog (Cerutti & Casas-Mollano, 2006; Shabalina & Koonin, 2008; Burroughs *et al.*, 2014). RNAi effectors have subsequently expanded through gene duplication events in some organisms (e.g. *A. thaliana*) or been lost altogether in others (e.g. *S. cerevisiae* and *Cyanidioschyzon merolae*). All diatom species investigated here possessed at least one DCR and AGO, indicating that RNAi may play an important role in diatoms (Fig. 1c). RDR was absent in some species, suggesting that RDR is dispensable for RNAi at least in these species. Some diatom species presented either clade A or B DCR/AGO, while others presented DCR/AGO of both clades. The distribution of DCR/AGO across diatoms supports a scenario in which the last common ancestor of diatoms presented DCR and AGO of both clades, with DCR-B and AGO-B members being lost in the evolutionarily more recent raphid pennate species. Interestingly, the two best studied model diatom species *P. tricornutum* and *T. pseudonana*, which present the two smallest diatom genome sizes sequenced to date, are exceptions among diatoms in presenting clade B DCR/AGO only. The distribution of clade A and B DCR/AGO across diatom species suggests functional interaction between DCR and AGO of the same clade (Fig. 1d). As clade A and B proteins, for both DCR and AGO, differ in their functional domain architectures and functional residues, this observation implies in turn the presence of distinct RNAi pathways in diatoms. Gene duplication events and gain/loss of functional domains have diversified DCR, AGO, and RDR isoforms in eukaryotes with different substrate and product specificities and subcellular localizations (Cerutti & Casas-Mollano, 2006; Shabalina & Koonin, 2008; Burroughs *et al.*, 2014). In turn, specialized RNAi pathways with distinct molecular mechanisms and physiological roles evolved through functional interaction between DCR, AGO, and RDR isoforms (Chapman & Carrington, 2007). Analysis of DCR-sRNAs and RNAi mechanisms in diatom species with contrasting repertoire of clade A and B DCR/AGO will be pivotal to comprehend the diversity of RNAi mechanism and role in diatoms.

Our results demonstrate an essential role for *P. tricornutum* DCR in processing sRNAs targeting preferentially LTR-RT of the

Ty1/Copia family covered by DNA and histone repression-associated epigenetic marks (Figs 3b,c, 4b). Interestingly, such pathway may present mixed characteristics with its plant, fungi and animal counterparts. The presence of these two types of repression-associated epigenetic marks coincides with what is found in plants and mammals but not in other animals such as *C. elegans* and *D. melanogaster* or the yeast *S. pombe*, which all lack DNA methylation. While sRNA-directed DNA methylation exists in mammals, it principally takes place in the germ line and is mediated by PIWI-interacting RNAs, which are DCR-independent (Girard *et al.*, 2006). In *Arabidopsis thaliana*, the plant-specific RNA polymerases IV and V (RNA Pol IV and V) together with RDR2 and DCL3 are involved in the processing of sRNA targeting epigenetically suppressed heterochromatic genes – a plant-specific pathway coined RNA-directed DNA methylation (RdDM; Matzke *et al.*, 2015). *P. tricornutum* lacks RNA Pol IV/V, suggesting that, as in *C. elegans* and *S. pombe*, DCR-sRNA might be processed from Pol II transcripts. LTR-RTs of the *Ty1/Copia* family, with some being specific to *P. tricornutum*, account for nearly half of the *P. tricornutum* TE complement and are known targets of DNA methylation (Maumus *et al.*, 2009; Veluchamy *et al.*, 2013; Rastogi *et al.*, 2018; Hoguein *et al.*, 2023). *In silico* whole genome analysis has revealed an evolutionarily recent burst of *Ty1/Copia*-like elements in *P. tricornutum* genome, suggesting that they may still be active and generate genetic diversity (Maumus *et al.*, 2009; Basu *et al.*, 2017; Filloramo *et al.*, 2021). The evolutionarily relatively recent expansion of LTR-RT *Ty1/Copia* elements in *P. tricornutum* may explain their active targeting by DCR-sRNAs.

Our results indicate that *P. tricornutum* DCR might play an important role in the genome-wide maintenance of the repression-associated marks H3K9me2/3 and H3K27me3 (Fig. 4c). A similar requirement has been evidenced in diverse eukaryotes, including *S. pombe* (Volpe *et al.*, 2002; Chen *et al.*, 2008; Kloc *et al.*, 2008; Zaratiegui *et al.*, 2011; Castel Stephane *et al.*, 2014), mouse (Kanellopoulou *et al.*, 2005; Gutbrod *et al.*, 2022), *T. thermophila* (Mochizuki & Gorovsky, 2005), *D. melanogaster* (Peng & Karpén, 2007), and *A. thaliana* (Parent *et al.*, 2021). Since H3K9me2, H3K27me3, and especially H3K9me3 were also found in DIGs, our results indicate that DCR abrogation imparts their maintenance at genes not directly targeted by DCR-sRNAs. The reason behind this observation is unclear. *P. tricornutum* genome-wide landscape of cytosine and histone methylation has been characterized (Veluchamy *et al.*, 2013, 2015; Zhao *et al.*, 2020). Current knowledge on the molecular effectors and mechanisms involved in the establishment and maintenance of these modifications, however, is limited. *P. tricornutum* genome encodes 5 and 13 putative homologs of DNA and histone methyltransferases, respectively (Rastogi *et al.*, 2015; Zhao *et al.*, 2020; Hoguein *et al.*, 2023). *P. tricornutum* homolog of the histone methyltransferase enhancer of zeste E(z), a component of the Polycomb Repressive Complex 2 (PRC2), has been shown to be essential for the maintenance of H3K27me2/3 and to play an important role in the control of cell morphology (Zhao *et al.*, 2021). In addition, *P. tricornutum* homolog of DNMT5 has been reported to be essential for the genome-wide maintenance of cytosine methylation (Hoguein *et al.*, 2023). Further studies pertained to these molecular effectors and their

homologs, together with their genetic interactions with DCR, are needed to better understand the role of DCR-sRNAs in the maintenance of repression-associated epigenetic marks.

We determined that most DAGs were covered by repression-associated epigenetic marks (Figs 3b, 4a,b) and low/not-expressed (Fig. 5a,b) suggesting a correlation between DCR and epigenetic suppression. Yet, a significant proportion of low/not-expressed genes covered by repression-associated epigenetic marks corresponded to DIGs, suggesting that epigenetic suppression in *P. tricornutum* WT can take place in the absence of ongoing targeting by DCR-sRNAs. In addition, very few DAGs were found upregulated in *P. tricornutum* DCR-KOs (Figs 5c, S16). Hence, DAG transcriptional repression can apparently be maintained in the absence of DCR-sRNAs (and of H3K9me2/3 and H3K27me3). These observations are reminiscent to previous results obtained in *A. thaliana* (Lippman *et al.*, 2003; Zilberman *et al.*, 2003; Xie *et al.*, 2004; Woodhouse *et al.*, 2006). Genome-wide analysis of DNA methylation in DCR-KOs will be instrumental to characterize the dependence of these epigenetic marks with DCR-sRNAs and its possible contribution in the maintenance of DAG suppression in DCR-KOs. Besides a possible role of DNA methylation, transcriptional reactivation of DAGs might be conditioned by environmental stimuli such as stress (e.g. temperature and nitrogen starvation) as reported for some LTR-RTs in plants and diatoms (Maumus *et al.*, 2009; Ito *et al.*, 2011; Pargana *et al.*, 2019). In addition, antisense and hairpin-induced silencing of endogenous genes in *P. tricornutum* has been shown to take place at the transcriptional and post-transcriptional levels (De Riso *et al.*, 2009). Hence, analysis of DAG expression at the protein level in DCR-KOs will be necessary to comprehensively assess the role of DCR-sRNAs in the regulation of gene expression. Counterintuitively, the vast majority of the genes found upregulated in DCR-KOs were DIGs corresponding non-TE genes lacking repression-associated epigenetic marks and initially expressed in WT (Figs 5, S16, S17a,b). The sets of upregulated DIGs were DCR-KO specific (Fig. S16). We cannot rule out that the upregulation of DIGs resulted from the random genomic integration of DNA fragments during the generation of DCR-KOs by biolistic transformation. Alternatively, the upregulation of DIGs might be a consequence of reactivated DAGs, which might have affected DIG expression either in *cis* (e.g. transcription of derepressed DAGs acting on nearby DIGs) or in *trans* (e.g. DAGs acting as master regulators or transcription factors on distant DIGs). In disfavor of a regulation in *cis*, global analysis of DIGs and DAGs genomic location indicated that upregulated DIGs and DAGs did not have the tendency to be positioned next to each other (data not shown). In disfavor of a regulation in *trans*, the set of upregulated DIGs did not present an enrichment in any specific GO term (Fig. S17c), which could be otherwise expected in a set of genes co-regulated by a common master regulator or transcription factor. To better understand the possible cause and significance of the observed upregulation of DIGs, future studies should aim to generate and characterize DCR-KO lines using CRISPR-Cas9 constructs carried by a nonintegrative episomal vector and delivered by bacterial conjugation (Sharma *et al.*, 2018). In addition, functional complementation of our

DCR-KO lines with a CRISPR-Cas9-resistant version of the DCR gene driven by its own promoter will remain indispensable to establish and confirm the phenotype resulting specifically from the DCR-KO.

Nitrogen availability often controls diatom proliferation and nitrogen limitation is one of the most frequent and severe stresses encountered by diatoms in their natural environment (Armbrust, 2009; Bristow *et al.*, 2017). Nitrogen limitation in *P. tricornutum* laboratory cultures has been shown to trigger remodeling of epigenetic marks with concomitant transcriptional reactivation of LTR-RTs and DE of genes involved in nitrogen metabolism (Maumus *et al.*, 2009; Veluchamy *et al.*, 2013, 2015). In our study, *P. tricornutum* DCR-KOs and WT cultures grown under nitrogen-replete conditions had comparable growth kinetics, suggesting that DCR is not essential for diatom at least under our laboratory conditions (Fig. 6). WT and DCR-KOs were similarly impacted by nitrogen limitation, suggesting that DCR does not play a critical role in the acclimatory response to nitrogen starvation. Consistent with this observation, most of the nitrogen metabolism genes reported (Veluchamy *et al.*, 2013, 2015) as being differentially expressed under nitrogen starvation, including the two key genes carbamoyl phosphate synthase II (EG01947) and nitrite reductase (J12902), corresponded to DIGs (Dataset S4). When returned to nitrate replete condition, DCR-KOs presented an extended lag phase (i.e. by 1–2 d) compared with WT indicating an impaired capacity to quickly respond to nitrogen replenishment and/or to recover from nitrogen starvation stress. Although DCR-KOs eventually reached a maximum cell density close to WT when cultured in parallel, such delay in entering exponential growth phase may represent a severe disadvantage in the natural environment where various phytoplankton groups coexist and compete for resources. DCR-KO phenotype could be due to a possible role of DCR in re-establishing silencing of non-TE metabolic genes activated under nitrogen limitation. Moreover, DCR-KOs could be defective in keeping TE genes silenced during nitrogen limitation and/or in re-establishing TE genes silencing under nitrogen-replete conditions. Analysis of non-TE and TE gene expression and remodeling of epigenetic marks in DCR-KOs under fluctuating nitrogen availability was beyond the scope of our study and thus remains to be investigated.

Collectively, our results identified the function of the single DCR homolog of *P. tricornutum* establishing an association between RNAi and heterochromatin maintenance in this model diatom species. Future studies aiming to characterize the function of A and B clade DCR and AGO and their interaction with DNA- and histone-modifying enzymes will be necessary to better comprehend the role of RNAi in the evolution and structure of diatom genomes.

Acknowledgements

We thank Vasiliki Theodorou and the IMBB sequencing team for providing advices in sRNA library preparation. We thank Soizic Cheminant-Navarro for obtaining the cell morphology and cell length data. We are grateful to Leila Tirichine and Nathalie Jolie

for providing useful advices in chromatin extraction and immunoprecipitation. We thank Haroula Kontaki for valuable guidance and assistance in chromatin sonication. We thank Matthieu Lavigne for providing advices in the design and bioinformatic analysis of the ChIP-seq experiment. We also thank Antony James for editing in English. This work was funded by the project RADIO (RNA Silencing in Diatoms; EG and FV), which has received funding from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under grant agreement No 483 (FV). This publication benefited from the Horizon 2020 Research and Innovation Programme GHANA (The Genus Haslea, New marine resources for blue biotechnology and Aquaculture) under Grant Agreement No. 734708/GHANA/H2020-MSCA-RISE-2016 (FV). We acknowledge support (EG and FV) of this work by the project 'Centre for the study and sustainable exploitation of Marine Biological Resources (CMBR)' (MIS 5002670), which is implemented under the Action 'Reinforcement of the Research and Innovation Infrastructure', funded by the Operational Programme 'Competitiveness, Entrepreneurship and Innovation' (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund). EG's visit to AF laboratory was supported by an EMBO Short-Term Fellowship. Work taken place in AF laboratory was supported by the Fondation Bettencourt-Schueller (Coups d'élan pour la recherche française-2018) and the 'Initiative d'Excellence' program (Grant 'DYNAMO', ANR-11-LABX-0011-01). The publication of the article in OA mode was financially supported by HEAL-Link.

Competing interests

None declared.

Author contributions

FV conceived and designed the study. FV and EG designed the experiments. EG conducted sequence mining and phylogenetic analysis and all laboratory experiments except the ChIP-seq which was carried out by NK. HR analyzed the transcriptome and ChIP-sequencing data, conducted the bioinformatic analysis, and produced the respective figures. FV and KK supervised EG and coordinated the project. AF and MJ trained and supervised EG in generating the DCR-KO mutants in their laboratory and intellectually contributed throughout the project. EG drafted a first version of the manuscript. FV and EG wrote the manuscript with major intellectual inputs and contributions from KK, HR, AF and MJ. FV, KK and AF provided research funding. EG and HR contributed equally to this work and shared first authorship. FV and KK have jointly supervised this work and shared senior authorship. All authors contributed to the writing and revision of the manuscript.

ORCID

Emilia Grypioti  <https://orcid.org/0000-0002-0144-8160>
Hugues Richard  <https://orcid.org/0000-0001-7776-1340>
Frédéric Verret  <https://orcid.org/0000-0002-8175-7325>

Data availability

The sequence data have been submitted to the European Nucleotide Archive database (ENA) at EMBL-EBI under experiment accession no. PRJEB45526 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB45526>). Sequences of cloned *P. tricornutum* DCR, AGO, and RDR have been submitted as analysis (PRJEB45526) in the same ENA study under the accession nos. OU230856 (PtDCR), OU230857 (PtAGO), and OU230858 (PtRDR).

References

- Apt KE, Kroth-Pancic PG, Grossman AR. 1996. Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Molecular & General Genetics* 252: 572–579.
- Armbrust EV. 2009. The life of diatoms in the world's oceans. *Nature* 459: 185–192.
- 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.
- Basu S, Patil S, Mapleson D, Russo MT, Vitale L, Fevola C, Maumus F, Casotti R, Mock T, Caccamo M *et al.* 2017. Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytologist* 215: 140–156.
- Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, O'tillar RP *et al.* 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456: 239–244.
- Bristow LA, Mohr W, Ahmerkamp S, Kuypers MMM. 2017. Nutrients that limit growth in the ocean. *Current Biology* 27: R474–R478.
- Burroughs AM, Ando Y, Aravind L. 2014. New perspectives on the diversification of the RNA interference system: insights from comparative genomics and small RNA sequencing. *Wiley Interdisciplinary Reviews: RNA* 5: 141–181.
- Castel Stéphane E, Ren J, Bhattacharjee S, Chang A-Y, Sánchez M, Valbuena A, Antequera F, Martienssen RA. 2014. Dicer promotes transcription termination at sites of replication stress to maintain genome stability. *Cell* 159: 572–583.
- Cavalier-Smith T. 2003. Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 358: 109–133; discussion 133–104.
- Cerutti H, Casas-Mollano JA. 2006. On the origin and functions of RNA-mediated silencing: from protists to man. *Current Genetics* 50: 81–99.
- Chapman EJ, Carrington JC. 2007. Specialization and evolution of endogenous small RNA pathways. *Nature Reviews. Genetics* 8: 884–896.
- Chen ES, Zhang K, Nicolas E, Cam HP, Zofall M, Grewal SIS. 2008. Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature* 451: 734–737.
- Chomczynski P, Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* 162: 156–159.
- De Riso V, Raniello R, Maumus F, Rogato A, Bowler C, Falcatore A. 2009. Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Research* 37: e96.
- Djupedal I, Ekwall K. 2009. Epigenetics: heterochromatin meets RNAi. *Cell Research* 19: 282–295.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.
- Doyle J. 1991. DNA protocols for plants. In: Hewitt GM, Johnston AWB, Young JPW, eds. *Molecular techniques in taxonomy*. Berlin, Heidelberg, Germany: Springer Berlin Heidelberg, 283–293.
- Falcatore A, Jaubert M, Bouly J-P, Bailleul B, Mock T. 2020. Diatom molecular research comes of age: model species for studying phytoplankton biology and diversity. *Plant Cell* 32: 547–572.
- Falkowski PG, Barber RT, Smetacek VV. 1998. Biogeochemical controls and feedbacks on ocean primary production. *Science* 281: 200–207.
- Farazi TA, Juranek SA, Tuschl T. 2008. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* 135: 1201–1214.
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281: 237–240.
- Filloramo GV, Curtis BA, Blanche E, Archibald JM. 2021. Re-examination of two diatom reference genomes using long-read sequencing. *BMC Genomics* 22: 379.
- Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X. 2006. Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell* 124: 355–366.
- Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: an expanding universe. *Nature Reviews. Genetics* 10: 94–108.
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442: 199–202.
- Guillard RRL. 1975. Culture of phytoplankton for feeding marine invertebrates. In: Smith WL, Chanley MH, eds. *Culture of marine invertebrate animals*. Boston, MA, USA: Springer US, 29–60.
- Gutbrod MJ, Roche B, Steinberg JI, Lakhani AA, Chang K, Schorn AJ, Martienssen RA. 2022. Dicer promotes genome stability via the bromodomain transcriptional co-activator BRD4. *Nature Communications* 13: 1001.
- Hoguin A, Ait Mohamed O, Bowler C, Genovesio A, Jimenez Vieira FR, Tirichine L. 2021. Evolutionary analysis of DNA methyltransferases in microeukaryotes: insights from the model diatom *Phaeodactylum tricornutum*. *bioRxiv*. doi: 10.1101/2021.06.11.447926.
- Hoguin A, Yang F, Groisillier A, Bowler C, Genovesio A, Ait-Mohamed O, Vieira FRJ, Tirichine L. 2023. The model diatom *Phaeodactylum tricornutum* provides insights into the diversity and function of microeukaryotic DNA methyltransferases. *Communications Biology* 6: 253.
- Holoch D, Moazed D. 2015. RNA-mediated epigenetic regulation of gene expression. *Nature Reviews. Genetics* 16: 71–84.
- Huang W, Daboussi F. 2017. Genetic and metabolic engineering in diatoms. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 372: 20160411.
- Huelsbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.
- Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J. 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* 472: 115–119.
- Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K. 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes & Development* 19: 489–501.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software v.7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.
- Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ *et al.* 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biology* 12: e1001889.
- Ketting RF, Haverkamp THA, van Luenen HGAM, Plasterk RHA. 1999. mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 99: 133–141.
- Kloc A, Zaratiegui M, Nora E, Martienssen R. 2008. RNA interference guides histone modification during the S phase of chromosomal replication. *Current Biology* 18: 490–495.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis v.7.0 for bigger datasets. *Molecular Biology and Evolution* 33: 1870–1874.

- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The sequence alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Liao Y, Smyth GK, Shi W. 2014. FEATURECOUNTS: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923–930.
- Lin X, Tirichine L, Bowler C. 2012. Protocol: chromatin immunoprecipitation (ChIP) methodology to investigate histone modifications in two model diatom species. *Plant Methods* 8: 48.
- Lippman Z, May B, Yordan C, Singer T, Martienssen R. 2003. Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biology* 1: E67.
- Lopez-Gomollon S, Beckers M, Rathjen T, Moxon S, Maumus F, Mohorianu I, Moulton V, Dalmay T, Mock T. 2014. Global discovery and characterization of small non-coding RNAs in marine microalgae. *BMC Genomics* 15: 697.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 550.
- Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA. 2006. Structural basis for double-stranded RNA processing by Dicer. *Science* 311: 195–198.
- Malviya S, Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, Wincker P, Iudicone D, de Vargas C, Bittner L *et al.* 2016. Insights into global diatom distribution and diversity in the world's ocean. *Proceedings of the National Academy of Sciences, USA* 113: E1516–E1525.
- Martienssen R, Moazed D. 2015. RNAi and heterochromatin assembly. *Cold Spring Harbor Perspectives in Biology* 7: a019323.
- Matzke MA, Kanno T, Matzke AJM. 2015. RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. *Annual Review of Plant Biology* 66: 243–267.
- Matzke MA, Mosher RA. 2014. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews. Genetics* 15: 394–408.
- Maumus F, Allen AE, Mhiri C, Hu H, Jabbari K, Vardi A, Grandbastien MA, Bowler C. 2009. Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genomics* 10: 624.
- Medlin LK, Kooistra WH, Gersonde R, Wellbrock U. 1996. Evolution of the diatoms (Bacillariophyta). II. Nuclear-encoded small-subunit rRNA sequence comparisons confirm a paraphyletic origin for the centric diatoms. *Molecular Biology and Evolution* 13: 67–75.
- Mochizuki K, Gorovsky MA. 2005. A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes & Development* 19: 77–89.
- 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.
- 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.
- Parent JS, Cahn J, Herridge RP, Grimaneli D, Martienssen RA. 2021. Small RNAs guide histone methylation in Arabidopsis embryos. *Genes & Development* 35: 841–846.
- Pargana A, Musacchia F, Sanges R, Russo MT, Ferrante MI, Bowler C, Zingone A. 2019. Intraspecific diversity in the cold stress response of transposable elements in the diatom *Leptocylindrus aporus*. *Genes* 11: 9.
- Peng JC, Karpen GH. 2007. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nature Cell Biology* 9: 25–35.
- Quinlan AR, Hall IM. 2010. BEDTOOLS: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.
- Rastogi A, Lin X, Lombard B, Loew D, Tirichine L. 2015. Probing the evolutionary history of epigenetic mechanisms: what can we learn from marine diatoms. *AIMS Genetics* 2: 173–191.
- Rastogi A, Maheswari U, Dorrell RG, Vieira FRJ, Maumus F, Kustka A, McCarthy J, Allen AE, Kersey P, Bowler C *et al.* 2018. Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricorutum* genome and evolutionary origin of diatoms. *Scientific Reports* 8: 4834.
- Rastogi A, Murik O, Bowler C, Tirichine L. 2016. PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing. *BMC Bioinformatics* 17: 261.
- Rogato A, Richard H, Sarazin A, Voss B, Cheminant Navarro S, Champeimont R, Navarro L, Carbone A, Hess WR, Falcatore A. 2014. The diversity of small non-coding RNAs in the diatom *Phaeodactylum tricorutum*. *BMC Genomics* 15: 698.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
- Schramke V, Allshire R. 2003. Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science* 301: 1069–1074.
- Shabalina SA, Koonin EV. 2008. Origins and evolution of eukaryotic RNA interference. *Trends in Ecology & Evolution* 23: 578–587.
- Sharma AK, Nymark M, Sparstad T, Bones AM, Winge P. 2018. Transgene-free genome editing in marine algae by bacterial conjugation – comparison with biolistic CRISPR/Cas9 transformation. *Scientific Reports* 8: 14401.
- Shi H, Djikeng A, Tschudi C, Ullu E. 2004. Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance. *Molecular and Cellular Biology* 24: 420–427.
- Slotkin RK, Freeling M, Lisch D. 2005. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nature Genetics* 37: 641–644.
- Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, Fire A, Mello CC. 1999. The rde-1 Gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 99: 123–132.
- Veluchamy A, Lin X, Maumus F, Rivarola M, Bhavsar J, Creasy T, O'Brien K, Sengamalay NA, Tallon LJ, Smith AD *et al.* 2013. Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricorutum*. *Nature Communications* 4: 2091.
- Veluchamy A, Rastogi A, Lin X, Lombard B, Murik O, Thomas Y, Dingli F, Rivarola M, Ott S, Liu X *et al.* 2015. An integrative analysis of post-translational histone modifications in the marine diatom *Phaeodactylum tricorutum*. *Genome Biology* 16: 102.
- Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. 2002. Regulation of heterochromatin silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297: 1833–1837.
- Woodhouse MR, Freeling M, Lisch D. 2006. Initiation, establishment, and maintenance of heritable MuDR transposon silencing in maize are mediated by distinct factors. *PLoS Biology* 4: e339.
- Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC. 2004. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biology* 2: E104.
- Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* 25: 1952–1958.
- Zaratigui M, Castel SE, Irvine DV, Kloc A, Ren J, Li F, de Castro E, Marin L, Chang AY, Goto D *et al.* 2011. RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II. *Nature* 479: 135–138.
- Zaslavskaja LA, Lippmeier JC, Kroth PG, Grossman AR, Apt KE. 2000. Transformation of the diatom *Phaeodactylum tricorutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. *Journal of Phycology* 36: 379–386.
- Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W. 2004. Single processing center models for human Dicer and bacterial RNase III. *Cell* 118: 57–68.
- Zhao X, Deton Cabanillas AF, Veluchamy A, Bowler C, Vieira FRJ, Tirichine L. 2020. Probing the diversity of Polycomb and Trithorax proteins in cultured and environmentally sampled microalgae. *Frontiers in Marine Science* 7: 1–17.
- Zhao X, Rastogi A, Deton Cabanillas AF, Ait Mohamed O, Cantrel C, Lombard B, Murik O, Genovesio A, Bowler C, Bouyer D *et al.* 2021. Genome wide natural variation of H3K27me3 selectively marks genes predicted to be important for cell differentiation in *Phaeodactylum tricorutum*. *New Phytologist* 229: 3208–3220.

Zilberman D, Cao X, Jacobsen SE. 2003. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299: 716–719.

Zong J, Yao X, Yin J, Zhang D, Ma H. 2009. Evolution of the RNA-dependent RNA polymerase (RdRP) genes: duplications and possible losses before and after the divergence of major eukaryotic groups. *Gene* 447: 29–39.

Supporting Information

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

Dataset S1 Protein amino acid sequences used in phylogenetic analysis.

Dataset S2 Nucleotide sequences and deduced protein amino acid sequences of PtDCR in DCR-KO.

Dataset S3 SequencingStats.

Dataset S4 Gene expression Dicer categories.

Dataset S5 Chromosomal location of DAGs and DIGs.

Dataset S6 Cell size and morphology.

Fig. S1 Diagram describing the nitrogen starvation experiment.

Fig. S2 Summary of the steps used to define Dicer-dependent sRNA clusters (DDSCs) and DDSC-associated genes.

Fig. S3 Detailed phylogenetic tree of Dicer homologs based on alignment of their concatenated RNase IIIa and RNase IIIb domains (corresponding to main Fig. 1a).

Fig. S4 Alignment of Dicer PAZ domains.

Fig. S5 Alignment of Dicer RNase IIIa domains.

Fig. S6 Alignment of Dicer RNase IIIb domains.

Fig. S7 Phylogenetic tree of separate RNase IIIa/b domains from Diatom Dicer-A and -B showing the divergence of Diatom Dicer-B RNase IIIb (maximum likelihood method).

Fig. S8 Detailed phylogenetic tree of AGO homologs based on alignment of the PIWI domain (corresponding to main Fig. 1b).

Fig. S9 Alignment of AGO PIWI domains from diatoms, plants, animal AGO, Excavata, Stramenopiles, and animal PIWIs.

Fig. S10 Phylogenetic tree of RDR homologs based on alignment of the RDR domain (corresponding to main Fig. 1c).

Fig. S11 Properties of the sRNA-associated fragments in wild-type and agreement with previous publications using (A) the NEXTFLEX or (B) the NEB protocol.

Fig. S12 Fragment length distribution of the mapped small RNA fragments overlapping exons and transposable elements.

Fig. S13 Change in small RNA coverage in the Dicer-KOs m4 and m9.

Fig. S14 Shared DAGs between the different Dicer-KO strains and sRNA coverage in wild-type (NEXTFLEX) in function of their structural annotation.

Fig. S15 DAG and DIG expression levels in wild-type in function of their epigenetic marks, using RNAseq data from Rastogi *et al.* (2018) and from this study.

Fig. S16 Transcriptional regulation of DAGs and DIGs in each Dicer-KO vs wild-type and their sharing between Dicer-KOs.

Fig. S17 Initial expression levels in wild-type, structural annotation, and GO term enrichment analysis of the different transcriptional groups of DAGs and DIGs.

Methods S1 Statistical analysis of differences in growth rate and lag-phase duration between wild-type and Dicer-Kos grown under fluctuating nitrogen availability.

Methods S2 Genome and transcriptome mining and phylogenetic analyses.

Methods S3 Chromatin immunoprecipitation followed by sequencing.

Methods S4 Definition of sRNA-associated loci and annotation of Dicer-associated loci.

Methods S5 Gene expression analysis.

Methods S6 Characterization and comparative analysis of histone marks in the wild-type and Dicer-KO lines.

Methods S7 Definition of autonomous transposons.

Table S1 Primers used in this study.

Table S2 sgRNAs used in this study.

Table S3 Pt DCR-KO lines generated.

Table S4 Statistics of DAGs and DIGs.

Table S5 Culture growth rates under fluctuating nitrogen availability.

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.