



Epigenetic Control of Diatom Genomes: An Overview from In Silico Characterization to Functional Studies

Xue Zhao, Antoine Huguin, Timothée Chaumier, and Leila Tirichine

Abstract

Epigenetics and its role in genome regulation is one of the most exciting areas of modern science. After a brief history of epigenetics and an introduction to the molecular basics of this discipline of science, this chapter describes the current knowledge of epigenetic components in diatoms, namely writers and erasers of DNA methylation and histone modifications. With a particular focus on the model pennate diatom *Phaeodactylum tricoratum*, we describe our current understanding of the contribution of few epigenetic factors to diatoms biology. Further, short regulatory non-coding RNAs (ncRNAs) as well as long ncRNAs are described in light of recent research. We highlight future studies and directions with a focus on epigenomic editing and environmental epigenetics.

Keywords

Adaptation · Diatoms · DNA methylation · Evolution · Non-coding RNAs · *Phaeodactylum tricoratum* · Post-translational modifications of histones

Xue Zhao and Antoine Huguin contributed equally with all other contributors.

X. Zhao · T. Chaumier · L. Tirichine (✉)
Nantes Université, CNRS, US2B, UMR 6286, Nantes, France
e-mail: tirichine-l@univ-nantes.fr

A. Huguin
Nantes Université, CNRS, US2B, UMR 6286, Nantes, France

Département de biologie, École normale supérieure, Institut de Biologie de l'ENS (IBENS), CNRS, INSERM, Université PSL, Paris, France

Abbreviations

5mC	5-methylcytosine
ALKBH	alpha-ketoglutarate-dependent hydroxylase
<i>COLDAIR</i>	COLD ASSISTED INTRONIC NONCODING RNA
DCL	Dicer endonuclease
DEXDc/Helicase	Dead like helicase domain
DME	DEMETER
DNA	Deoxyribonucleic acid
DNMTs	DNA methyltransferases
DUF	Domain of unknown function
ENDO3c	Endonuclease IIIc
FLC	Flowering Locus C
HDACs	Histone deacetylases
HDMs	Histone demethylases
LncRNAs	Long non-coding RNAs
miRNA	micro RNAs
MMETSP	Marine Microbial Eukaryote Transcriptome Project
PAZ	Piwi-Argonaute-Zwille
PTMs	Post-translational modifications of histones
RDRs	RNA-dependent RNA polymerases
RID/DMTA	RIP deficient/DNA methyltransferase activity
RIP	Repeat-induced point mutation
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SAM	S-adenosyl-methionine
siRNA	small interfering RNAs
SNF	Sucrose non-fermentable
TEs	Transposable elements

1 Introduction

The history of epigenetics demonstrates the fast progress and the dramatic increase of knowledge gained in the last 70 years of this young discipline. This is illustrated by the definition of epigenetics which evolved from a science reflecting phenomena that cannot be explained to a more precise field known as the study of mitotically and/or meiotically heritable changes in gene regulation that are not due or cannot be explained by changes in DNA sequence. These changes are represented by biochemical modifications of DNA, post-translational modifications of histones (PTMs), the

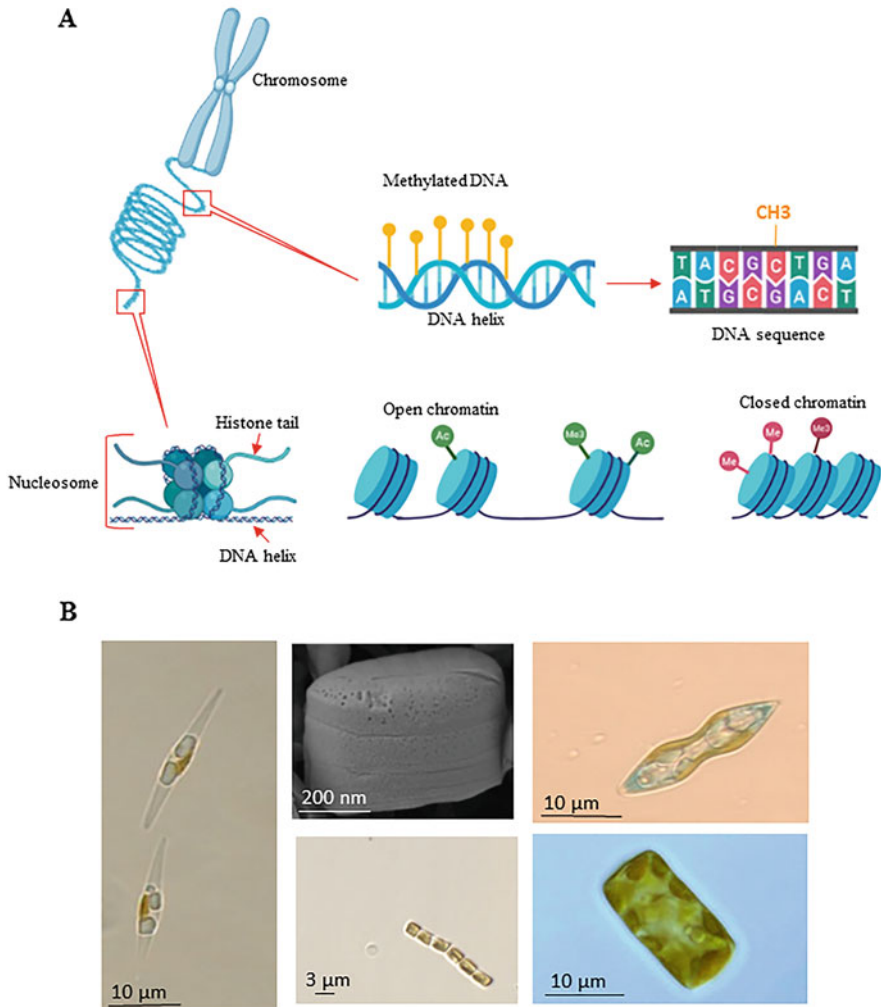


Fig. 1 Main targets of epigenetics regulation and model diatoms used for epigenetic studies. **(a)** A simplified scheme of epigenetic factors associated with activation (e.g., open chromatin, active histone marks) or repression (e.g., closed chromatin, DNA methylation) of gene transcription. The scheme was drawn using BioRender tool. **(b)** Photos of diatom species used as models of epigenetic studies, *P. tricornutum* (left), a scanning electron microscopy image of *F. cylindrus* (top middle, courtesy of N. Joli), a chain of *T. pseudonana* (bottom middle), *Haslea ostreraria* (top right) and *Cyclotella cryptica* (bottom right, courtesy of NCMA)

proteins that coat DNA, nucleosomes positioning, chromatin remodeling (Fig. 1a) and non-coding RNA-associated gene silencing (Waddington 1942).

The first reference to epigenetics goes back to 1942 by Conrad Waddington, an embryologist who referred to epigenetics as the whole complex of developmental processes that take place between genotype and phenotype (Waddington 1957). His

research led to the famous model of epigenetic landscape illustrating the different fates or developmental pathways a cell might take during differentiation with branches in the landscape structured by underlying genes. This concept quickly evolved in modern science, which extended epigenetics studies to several model organisms including bacteria, mammals, plants, insects, fungi and microalgae. Nowadays, the concept of epigenetics also includes changes that are not necessarily inherited in gene regulation, without modifying the underlying DNA sequence.

Epigenetics permitted different discoveries with the diversification of model organisms including the yeast species *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* used to elucidate chromatin structure and telomere silencing (Huang 2002). Long before, the genetic fly model *Drosophila melanogaster* was used to study the effect of the position of genes on the phenotype (position effect variegation), which led to the discovery of heterochromatin, chromatin remodeling and histone modifying proteins (Elgin and Reuter 2013). In plants, *Arabidopsis thaliana* emerged naturally as a model for epigenetic studies revealing widely conserved mechanisms as well as some specificities compared to animals. As an example, the seasonal regulation of the flowering locus C (FLC) involved in the vernalization process is associated with dynamics in deposition and removal of histone PTMs (Hepworth and Dean 2015). In mammals, the mouse is so far the greatest model to learn about epigenetic regulation mechanisms in humans in particular in stem cell research and environmental studies. It was demonstrated that histone PTMs are implicated in the transcriptional regulation of the homeobox-containing ‘Hox’ genes during the establishment of the antero-posterior axis (Deschamps and van Nes 2005). More recently, three unicellular species from the red, green and brown lineages of microalgae contributed to advance our fundamental knowledge in epigenetic research, *Cyanidioschyzon merolae*, *Chlamydomonas reinhardtii* and the diatom *Phaeodactylum tricorutum* respectively. Using these species is an interesting opportunity to address the questions of epigenetic regulation mechanisms in an evolutionary context. Sequencing of the *P. tricorutum* genome (Bowler et al. 2008) revealed a conserved epigenetic machinery including writers, erasers and readers of its different components (Rastogi et al. 2015; Tirichine et al. 2017) which were investigated in recent studies (Veluchamy et al. 2013a, 2015; Zhao et al. 2019). Few other diatoms were used to address the role of epigenetics in genome regulation (Fig. 1b) and evolution of epigenetic factors, such as DNA methylation which was investigated in *Thalassiosira pseudonana*, *Fragilariopsis cylindrus* (Tirichine et al. 2014; Huff and Zilberman 2014) (Joli et al., unpublished), *Cyclotella cryptica* (Traller et al. 2016) and *Haslea ostreraria* (Jean Luc Mouget, personal communication). In this chapter, we summarize our current knowledge about DNA methylation, PTMs of histones and non-coding RNAs in diatoms with a particular focus on in silico studies drawing a snapshot of the progress made at a time where epigenetics is coming to the forefront of diatoms biology.

2 DNA Methylation

DNA methylation is a major epigenetic mark in eukaryotes and prokaryotes. In plants and animals, methylation at the fifth carbon of cytosines, 5-methylcytosine (5mC) (Fig. 2a), is an epigenetic mark involved in the repression of transposable elements in many species and the establishment and maintenance of genomic imprinting (Ideraabdullah et al. 2008; Kohler et al. 2012; Galagan and Selker 2004). 5mC patterns are very diverse within the eukaryotic tree of life which reflects a fine-tuning of lineage-specific regulatory networks (de Mendoza et al. 2019; Schmitz et al. 2019). Hence, 5mC can be found at cytosines in different contexts. In vertebrates and invertebrates, DNA methylation mainly occurs at cytosines found at CG dinucleotides, while in fungi and plants, methylation at non-CG sites is more widely observed (Zemach et al. 2010; Feng et al. 2010; Stroud et al. 2014).

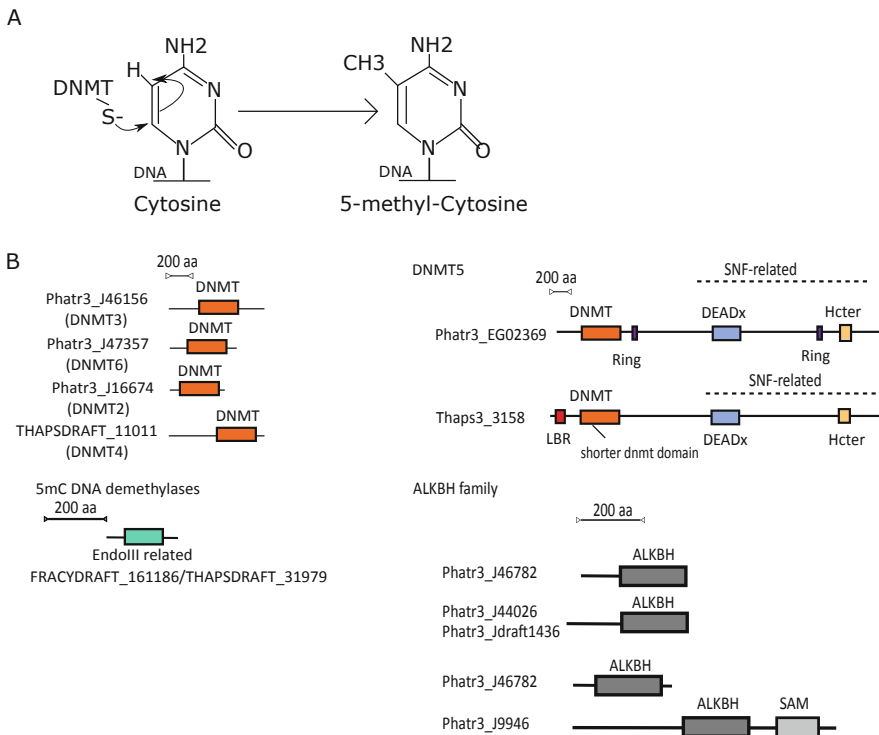


Fig. 2 DNA methylation mediated silencing. (a) Schematic representation of the DNA methyltransferase activity of 5mC DNA methyltransferases. (b) Structure of representative enzymes putatively involved in DNA methylation and de-methylation as well as DNA repair proteins

3 5mC Patterns and Functions in Diatoms

To date, 5mC have been reported in four diatoms, namely *P. tricornutum*, *T. pseudonana* and *F. cylindrus* (Veluchamy et al. 2013a; Huff and Zilberman 2014) as well as *Cyclotella cryptica* (Traller et al. 2016). In diatoms, 5mC is mainly found in a CG context over repeats and transposable elements usually (but not exclusively) concentrated in telomeric regions (Veluchamy et al. 2013a; Huff and Zilberman 2014). Although scarce, non-CG methylation is also detected. This is opposite to what is observed in the closely related multicellular brown alga *Saccharina japonica*, in which genes are mainly marked by non-CG methylations and transposable elements (TEs) are devoid of DNA methylation (Fan et al. 2020). In addition, in *P. tricornutum*, *T. pseudonana* and *F. cylindrus*, total levels of DNA methylation are low and range from 8% to as low as 1% of cytosines in the CG context compared to other species such as rice where 18% of total Cs are methylated (Huff and Zilberman 2014). Protein coding gene methylation is also sparsely observed. Among all diatoms with known methylome, *C. cryptica* shows the highest level of DNA methylation, which correlates with a higher amount of TEs found in its genome. In addition, DNA methylated regions can span up to 30 kb, a pattern not found in the other diatoms. Furthermore, no CG rich (CpG islands) and no promoter methylation patterns have been clearly described so far (Veluchamy et al. 2013a; De Riso et al. 2009). Overall, diatoms methylation pattern strongly contrasts with the patterns observed in animals in which nearly all CG dinucleotides are heavily methylated including within exons (Lister et al. 2009). This also contrasts with the pattern observed in the dinoflagellates *Symbiodinium minutum* and *S. kawaguttii* with ‘hypermethylated’ genomes exceeding 70% of CG methylation over genes and TEs (de Mendoza et al. 2018).

In diatoms, methylated TEs often have low expression (Veluchamy et al. 2013a; Huff and Zilberman 2014; Traller et al. 2016). This is very consistent with the repressive role of DNA methylation in other eukaryotes and further traces back 5mC mediated control of TE expression to the last eukaryotic common ancestor. In *P. tricornutum*, while DNA methylation over TEs correlates with low expression (Veluchamy et al. 2013a; Rastogi et al. 2018), its repressive role on genes depends on its pattern. Extensive methylation of genes correlates with low expression, while partial methylation correlates with moderate to high levels of expression (Veluchamy et al. 2013a). Nitrogen depletion triggers the concomitant loss of DNA methylation and over-expression of the transposable element ‘Blackbeard’ in *P. tricornutum*, suggesting that 5mC negatively regulates TEs expression under specific environmental triggers (Veluchamy et al. 2013a; Maumus et al. 2009). However, these observations are not consistently reported. In *C. cryptica*, 5mC patterns in any context are stable in response to silica depletion, including at transposable elements (Traller et al. 2016). Hence, within diatoms, DNA methylation, while likely involved in TE repression, might have evolved lineage-specific dynamics and functions.

4 Diatoms Have a Peculiar Set of DNA Methyltransferases

Eukaryotes possess diverse mechanisms to set, propagate and remove methylated cytosines. The deposition of 5mC is performed by DNA methyltransferases (DNMTs). Based on similarity with prokaryotic enzymes involved in the restriction-methylation system (Bestor 1990), six main eukaryotic DNMT families have been described: the DNMT1, DNMT2, DNMT3, DNMT4, DNMT5 and DNMT6 (Huff and Zilberman 2014; Ponger and Li 2005). All DNMTs contain a conserved protein domain with S-adenosyl-L-methionine binding and methyltransferase activity (PF00145 domain) referred to as DNMT domain. A summary of the DNMTs found in the three model diatoms as well as their associated protein domains are listed in Table 1 and Fig. 2.

Diatoms encode a unique set of DNMT2, DNMT3, DNMT4, DNMT5 and DNMT6 enzymes. The DNMT2 enzymes are responsible of RNA methylation at diverse cytosine positions that occur during the maturation of t-RNAs (Jeltsch et al. 2017). DNMT2 enzymes are highly conserved DNA methyltransferases that evolved RNA modifying functions and are found in animals, plants and micro-algae (Huff and Zilberman 2014). The DNMT3 proteins are widespread de novo DNA methylases in eukaryotes (Huff and Zilberman 2014). Typical DNMT3 proteins in metazoans therefore include chromatin domains that connect the DNA methylation pathways and the histone post-translational code deposited during development and meiosis (Laisne et al. 2018). Diatoms DNMT3, however, are short proteins lacking all known chromatin-associated domains (Fig. 2).

The DNMT4 family is a weakly supported DNMT1-related family, typified by fungi RIP deficient/DNA methyltransferase activity (RID/DMTA) and MASC1 (a member of the fungal-specific DMT-like family) proteins respectively involved in the Repeat Induced Point mutation (RIP) and methylation induced premeiotically (MIP) processes, which are two DNA methylation-dependent genomic defense mechanisms against TEs (Galagan and Selker 2004; Amselem et al. 2015; Gladyshev 2017). Earlier reports showed that DNMT4 is only conserved in fungi and diatoms which probably highlights a convergent evolutionary history of this gene family (Huff and Zilberman 2014; Ponger and Li 2005). In diatoms, DNMT4 is composed of a unique DNMT domain as no chromatin-associated protein features are found with this domain (Fig. 2). The role of DNMT4 in diatoms is unclear as it is unknown whether any RIP- or MIP-related process occurs in these species. Diatoms notoriously lack other DNMT1-related proteins that are the major 5mC maintenance enzyme in metazoans and plants. However, diatoms possess DNMT5 which was previously shown to maintain DNA methylation in a CG context in the parasitic yeast *Cryptococcus neoformans* (Huff and Zilberman 2014). This enzymes is likely responsible for CG methylation in other fungi, green algae, haptophytes and in the stramenopile *Aureococcus anophagefferens* (Huff and Zilberman 2014; Bewick et al. 2019). DNMT5 enzymes possess a long C-terminal region containing asp-glu-ala-asp box (DEADx) and Helicase domains (Fig. 2). This domain is related to Sucrose Non-Fermentable (SNF) chromatin remodeler with ATPase activity that is required for the DNA methylation function of the enzyme (Dumesic et al. 2020).

Table 1 Putative writers and erasers of C5 DNA methylation in three model diatoms

	<i>Phaeodactylum tricornutum</i>		<i>Fragilariopsis cylindrus</i>		<i>Thalassiosira pseudonana</i>	
DNA cytosine methyltransferases						
DNMT1	not found	not found	not found	not found	not found	not found
DNMT2	Phatr3_J16674	FRACYDRAFT_139371	FRACYDRAFT_139371	FRACYDRAFT_139371	THAPSDRAFT_22139	THAPSDRAFT_22139
DNMT3	Phatr3_J46156	not found	not found	not found	THAPSDRAFT_9575	THAPSDRAFT_9575
DNMT4	not found	FRACYDRAFT_250279, FRACYDRAFT_250634,FRACYDRAFT_242149*	FRACYDRAFT_250279, FRACYDRAFT_250634,FRACYDRAFT_242149*	FRACYDRAFT_250279, FRACYDRAFT_250634,FRACYDRAFT_242149*	THAPSDRAFT_11011	THAPSDRAFT_11011
DNMT5	Phatr3_EG02369	FRACYDRAFT_205902, FRACYDRAFT_239160,FRACYDRAFT_212855	FRACYDRAFT_205902, FRACYDRAFT_239160,FRACYDRAFT_212855	FRACYDRAFT_205902, FRACYDRAFT_239160,FRACYDRAFT_212855	THAPSDRAFT_3158	THAPSDRAFT_3158
DNMT6	Phatr3_J47357	FRACYDRAFT_148014	FRACYDRAFT_148014	FRACYDRAFT_148014	THAPSDRAFT_2094	THAPSDRAFT_2094
5mC DNA de-methylases						
ROS1	Phatr3_J46865, Phatr3_J12645	FRACYDRAFT_161186	FRACYDRAFT_161186	FRACYDRAFT_161186	THAPSDRAFT_31979	THAPSDRAFT_31979
DME	not found	FRACYDRAFT_161186	FRACYDRAFT_161186	FRACYDRAFT_161186	THAPSDRAFT_31979	THAPSDRAFT_31979
TET1/2/3	not found	not found	not found	not found	not found	not found
Alkhh family						
ALKBH1	Phatr3_J46782	FRACYDRAFT_196981,FRACYDRAFT_162731	FRACYDRAFT_196981,FRACYDRAFT_162731	FRACYDRAFT_196981,FRACYDRAFT_162731	THAPSDRAFT_6486, THAPSDRAFT_21832	THAPSDRAFT_6486, THAPSDRAFT_21832
ALKBH2	Phatr3_J44026	FRACYDRAFT_162731	FRACYDRAFT_162731	FRACYDRAFT_162731	THAPSDRAFT_6486	THAPSDRAFT_6486
ALKBH3	Phatr3_J44026, Phatr3_J449981	not found	not found	FRACYDRAFT_162731	not found	not found
ALKBH4	not found	not found	not found	not found	not found	not found
ALKBH5	not found	FRACYDRAFT_259236	FRACYDRAFT_259236	FRACYDRAFT_259236	THAPSDRAFT_1747	THAPSDRAFT_1747
ALKBH6	Phatr3_Jdraft1436	FRACYDRAFT_180788	FRACYDRAFT_180788	FRACYDRAFT_180788	THAPSDRAFT_31380	THAPSDRAFT_31380
ALKBH7	Phatr3_J50302	FRACYDRAFT_185395	FRACYDRAFT_185395	FRACYDRAFT_185395	THAPSDRAFT_908	THAPSDRAFT_908
ALKBH8	Phatr3_J9446	FRACYDRAFT_152506	FRACYDRAFT_152506	FRACYDRAFT_152506	THAPSDRAFT_42543	THAPSDRAFT_42543
THAPSDRAFT_42543-like clade	not found	FRACYDRAFT_152506	FRACYDRAFT_152506	FRACYDRAFT_152506	THAPSDRAFT_42543	THAPSDRAFT_42543

* This gene presents a Dicer double-stranded RNA-binding fold domain

Within diatoms, DNMT5 proteins are divergent. The DNMT domain of the DNMT5 enzyme of *Thalassiosira pseudonana* is shorter than the DNMT5 protein found in *P. tricornutum*. In addition its SNF-related domain lacks a RING finger domain. Whether these divergences lead to differences in the establishment or maintenance of the epigenetic landscape in both diatoms is unknown.

The DNMT6 family has been first described in the parasitic euglenozoa *Trypanosoma brucei* and *Leishmania major*, in the green alga *Micromonas pusilla* and dinoflagellates (Huff and Zilberman 2014; de Mendoza et al. 2018; Ponger and Li 2005). In *Leishmania major*, DNMT6 does not seem required for either de novo or maintenance of 5mC (Cuypers et al. 2020). In diatoms, DNMT6, whose function is unknown, is only composed of a highly conserved methyltransferase domain with no chromatin domains as observed for other diatoms DNMT3, DNMT4 and DNMT2 enzymes.

Our current understanding of the proteins involved in the regulation of DNA methylation in diatoms is in progress. Our investigation of the diversity of DNMTs found in unicellular eukaryotes of the Marine Microbial Eukaryote Transcriptome Project (MMETSP) data base (Keeling et al. 2014) using complementary in silico approaches and functional studies (Hoguin et al., unpublished) indicates that DNMT5 is an unappreciated diversified gene family in marine micro-eukaryotes. This is the only DNMT with chromatin-associated domains in diatoms. In addition, in *P. tricornutum*, DNMT5 knock-out associates with a loss of CG methylation and a transcriptional activation of otherwise silenced TEs revealing for the first time the mechanisms controlling TE expression in diatoms (Hoguin et al., unpublished). More questions nonetheless remain regarding the maintenance and establishment of 5mC in diatoms. Since no de novo DNA methylation activity has been found yet in diatoms, we may indeed ask whether diatom DNA methylation patterns are rather the results of strong maintenance activity as suggested in some fungal lineages. As mentioned, previous and current studies suggest that *P. tricornutum* methylome is responsive to environmental cues. It is therefore probable that DNA methylation evolved a condition-specific regulatory role in diatoms and hence might translate environmental changes into stable epigenetic inheritance of genes and or transposons regulation.

5 What Are the De-methylation Pathways in Diatoms?

DNA de-methylation machinery is not highly conserved in diatoms. There are no Ten-eleven translocation methylcytosine dioxygenase (TET) enzymes that are known DNA demethylases in animals (Choi et al. 2002; Agius et al. 2006; Gehring et al. 2006; Wu and Zhang 2017). Although with low similarity, two putative DNA demethylases, Phatr3_J46865, and Phatr3_J12645 with Endonuclease IIIc (ENDO3c) InterPro predicted domain were found in *P. tricornutum*. In *F. cylindrus* and *T. pseudonana*, orthologues of both Reactive Oxygen Species (ROS) and DEMETER (DME) proteins are detected by reciprocal BLAST analysis. Both enzymes are ENDO3c domain containing proteins but do not have additional

domains (Fig. 2). It is worth noting that ENDO3c domains are also associated with a wide range of evolutionary diverse DNA repair proteins (Kanchan et al. 2015) and the presence of this domain alone does not confer 5mC demethylation activity.

The alpha-ketoglutarate-dependent hydroxylase (ALKBH) enzymes are diverse proteins known to regulate adenine methylations in vivo in mouse and *Caenorhabditis elegans* (Greer et al. 2015; Wu et al. 2016) and to produce oxidized 5-methylcytosines derivative in vitro (Bian et al. 2019), which can eventually lead to active DNA demethylation. ALKBH enzymes are also known to be involved in DNA repair of methylated DNA templates and they can modulate RNA methylation (Fu et al. 2010; Zdzalik et al. 2014; Iyer et al. 2016). BLAST analysis in diatoms revealed several putative ALKBH orthologues (Table 1). They all contain an alpha-ketoglutarate-dependent dioxygenase domain but lack RNA/DNA binding domains. The ALKBH8 orthologue in *P. tricornutum* possesses an S-adenosyl-methionine binding domain (SAM) highlighting a potential RNA modifying activity. It is important to note that the current phylogenetic assignment of the putative diatoms ALKBH enzymes must be further investigated. Nonetheless, these enzymes are potential new actors of the epigenetic regulation in diatoms.

6 Post-Translational Modifications of Histones and Their Enzymes in Diatoms

Histones are subject to a variety of post-translational modifications (PTMs) that alter gene expression and chromatin structure. *P. tricornutum* genome sequencing revealed a long list of histone modifying enzymes which were described previously (Rastogi et al. 2015; Veluchamy et al. 2015; Tirichine et al. 2014). Here we update the list of genes with predicted function in histone modifications in few diatom species (Table 2). Since the identification of PTMs using mass spectrometry in *P. tricornutum*, an epigenomic map of several histone marks known to be active or repressive was established using ChiP-Seq. Combined with previously published genome-wide DNA methylation data (Veluchamy et al. 2013b), comprehensive and combinatorial analyses revealed some conserved and specific epigenetic features in *P. tricornutum* extending the existence of the epigenetic code to Stramenopiles. One of the important findings is the co-occurrence of repressive histone marks and DNA methylation over genes and transposable elements (Veluchamy et al. 2015). These co-occurrence patterns define combinations of epigenetic marks unique to diatoms, suggesting a cooperation in repression and/or an interdependent recruitment mechanisms. This chapter section provides a general overview of predicted histone modifiers based on four fully sequenced diatom genomes. It is important to keep in mind that histone modifications are usually deposited by protein complexes which need to be taken into consideration in functional studies of these enzymes and the histone code in diatoms.

Table 2 Histone acetyltransferases and deacetylases in diatoms

Name	<i>H. sapiens</i>	<i>S. cerevisiae</i>	<i>P. tricornutum</i>	<i>T. pseudonana</i>	<i>F. cylindrus</i>	<i>P. multiseriis</i>
GNAT family	KAT1	HAT1	Hat1	Phatr3_J54343	223323	0082810
	KAT2A	Hgen5	Gen5	Phatr3_J2957	184437	0110000
	KAT2B	P/CAF		15161		
	KAT9	ELP3	Elp3	Phatr3_J50848	206186	0027300
MYST family	KAT5	TIP60	Esa1	Phatr3_J3062, Phatr3_J51406, P hatr3_J44463(MORF)	172955, 173464	0016210, 0055030
	KAT6A	MOZ	Sas3			
	KAT6B	MORF				
	KAT7	HBO1	-			
KAT8A	HMOF	Sas2				
KAT8B						
CBP/P300 family	KAT3A	CBP	-	Phatr3_J54505, Phatr3_J45703, Phatr3_J45764	168758, 207494, 181388	0024900, 004154 0, 0115290
	KAT3B	P300				
	KAT11	-	Rtt109			
transcription co-activators	KAT4	TAF1	Taf1	Phatr3_J42944	165291*	0067010
Deacetylase in diatoms (HDACs)						
Class I	HDAC1-3, HDAC8	Hos 1-3, Rpd 3		Phatr3_J49800, Phatr3_J43919, Phatr3_J51026, Phatr3_J4590, Phatr3_EC01943,	223488, 170265, 206710, 2 0681, 159918, 177910, 161 016(Hos3), 161020(Hos3) , 249643(Hos3)	0046900, 005846 0, 0116090, 00240 90(Hos3), 001868 0(Hos3), 0014140 (Hos3)
	HDAC4-7, HDAC9, 10	Hda 1		Phatr3_J35869, Phatr3_J50482, Phatr3_J8891, Phatr3_J45431, Phatr3_J4423, Phatr3_J45906		
Class III	SIRT1-7	Sir2, Hst1-4		Phatr3_J16859, Phatr3_J8827, Phatr3_J12305, Phatr3_J45850, Phatr3_J52135, Phatr3_J21543, Phatr3_J39523	16384, 16405, 35956, 264809, 2 69475	0067550, 009060 0, 0057370, 011165 00, 0121840,
	HDAC 11	Not found		Phatr3_J9278, Phatr3_J4821	206810, 159918	0086250, 004092 0

*: Low similarity

7 Histone Acetyltransferases and Deacetylases

Acetylation of ϵ -amino group in lysines leads to activation of transcription. This process is carried out by a group of proteins known as histone lysine acetyltransferase (HATs or KATs), which can be divided into 5 families: (1) Gcn5-related acetyltransferase (GNATs) family; (2) MYST family which includes MOZ-, Ybf2-, Sas2- and Tip60-related proteins; (3) p300/CBP family; (4) general transcription factor HATs including TFIIC90 and Taf1, and (5) the steroid receptor co-activators like SRC1, ACTR and CLOCK (Carrozza et al. 2003). Table 2 summarizes acetyl transferase families in diatoms except for steroid receptor co-activator family, since no homologs have been found. In GNATs family, three subgroups of KATs were found and listed in Table 2: KAT1, KAT2A/B and KAT9. Among those, KAT1 is the simplest which only contains a histone acetyltransferase HAT1 type domain. Interestingly, except the listed KAT1 homologs, there are more GNAT domain containing acetyltransferases in diatom genomes. Taking *P. tricornutum* as an example, there are around 48 genes with GNAT domain, but their function is unknown. Among them there are few unusual genes with another domain revealing a combination of protein domains, which have never been reported before. Such examples include Phatr3_J47498 which has a histidine phosphatase domain with GNATs, Phatr3_J46516 possesses two possible tRNA binding domains similar to bacterial acetyltransferase TmcA suggesting that these acetyltransferases might target non-histone proteins like ATAT1, α -tubulin K40 acetyltransferase (Akella et al. 2010; Shida et al. 2010). The expanded KAT1 subgroup of GNAT family in diatoms requires further investigations in future studies.

MYST domain containing proteins is another large family of KATs. In the yeast model species *S. cerevisiae*, only three genes were reported: Esa1, Sas2 and Sas3 (Osada et al. 2001). The common feature of MYST family acetyltransferase is a MYST-type histone acetyltransferase domain with a chromodomain shared in yeast (Esa1), humans (Tip60) and *Drosophila* (MOF). Similar domain features were also found in diatoms MYSTs, where an RNA binding activity-knot of a chromodomain (PF11717) can be found at the N-terminal. There are only three gene homologs that belong to MYST family in each of the model diatom species investigated (Table 2). In *P. tricornutum*, Phatr3_J44463 has a MORF-like acetyltransferase domain considered as a homolog of MORF which is responsible of H3K14 and H4K5/8/12/15 acetylation in vitro and H3K9 (Mishima et al. 2011; Kitabayashi et al. 2001) and H3K23 acetylation in vivo (Klein et al. 2019).

CBP/p300 family was initially identified in mammals, and they are a unique acetyltransferases group without any sequence similarity to GNATs (Goodman and Smolik 2000). In yeast there are no orthologs of human CBP and p300. However, a protein called Rtt109, which has a related 3-D protein structure of CBP/p300, was identified (Wang et al. 2008). It is a fungal-specific gene with no orthologs in uni- and multicellular organisms. Here we found fungal-like Rtt109 acetyltransferases in four diatom species (Table 2), three paralogs with similar domain structure in each species, suggesting an ancient origin of CBP/p300 family.

Histone lysine acetylation is a reversible process. Acetylated lysine can be removed by histone deacetylases (HDACs). Deacetylation of histones induces chromatin compaction leading to transcriptional repression. It is a constant balance between the antagonistic action of histone acetylases and deacetylases that contribute to transcriptional regulation of genes. Both enzymes were shown to have an important role in development and diseases (Haberland et al. 2009). Based on protein 3-D structure and domain feature, HDACs can be grouped into four classes (Seto and Yoshida 2014). In *T. pseudonana*, 7 genes were found in Class I, II and IV. In *F. cylindrus* and *Pseudo-nitzschia multiseriis*, 11 and 8 homologs were identified respectively as histone deacetylase genes (Table 2). In *P. tricornutum*, 13 homologs were identified as histone deacetylase similar to HDAC1-11 protein sequences from human and yeast Hos1-3, Rpd3 and Hda1 proteins. Of note, there are several deacetylase domain containing genes in *P. tricornutum* that do not fall into any of the classes suggesting a complex and diversified deacetylation mechanisms in diatoms.

8 Histone Methyltransferases and Demethylases

Methylation and demethylation of histones activates or represses genes depending on the amino acid that is methylated and how many methyl groups are attached to the residue. This activation or repression acts by loosening the attraction between histone tails and DNA allowing the transcriptional machinery and other regulatory proteins to access DNA or by compacting chromatin restricting the access to DNA respectively. Histone methylation is considered more stable than other modifications such as phosphorylation and acetylation, is involved in long-term maintenance of the expression status of regions of the genome and has been shown to play a role in virtually all biological processes (Greer et al. 2015).

Histone methyltransferases (HMTs) are one of the most well-studied histone modifiers. Unlike the broad range targeting strategy of HATs, HMTs are responsible for the methylation of specific residues. Almost all of the HMTs contain a Su(var)3-9, Enhancer-of-zeste and Trithorax (SET) domain except for DOT1 family. The DOT1 family is not structurally related to SET-domain proteins, but their members can methylate K79 of histone H3 (Feng et al. 2002; Ng et al. 2002). Interestingly, Dot1 homolog were detected in *P. tricornutum*, *F. cylindrus* and *P. multiseriis*, three pennate diatom species, but not in *T. pseudonana* (Table 3), a centric diatom. However, methylation of H3K79 was reported in *T. pseudonana* (Rastogi et al. 2015), it would be appealing to identify the putative methyltransferase of H3K79 in this diatom. SET-domain containing superfamily is a big group of HMTs that can be divided into seven families (Dillon et al. 2005). However three families were missing in diatom SET containing HMTs, namely SUV39, RIZ and SUV1–20 families (Table 3).

SUV39 family proteins are the most well-characterized HMTs. SUV39H1 was the first identified lysine methyltransferase which methylates lysine 9 of histone H3 (Tamaru and Selker 2001). In yeast model species, SUV39 is not found, but a

Table 3 Histone methyltransferases and demethylases in diatoms (KMTs)

Family	Common domain	<i>H. sapiens</i>	<i>S. cerevisiae</i>	<i>P. tricornutum</i>	<i>T. pseudonana</i>	<i>F. cylindrus</i>	<i>P. multiseriatus</i>
Methyltransferase (KMTs)							
SUV39 family	Pre-SET (9 Cys, 3 Zn), post-SET (CXCXC) SET domain	SUV39H1/2	Clr4 in <i>S.</i>	-	-	-	-
		ESET/CLL8	<i>pombe</i>	-	-	-	-
		G9a	-	-	-	-	-
Methyltransferase (KMTs)							
SET1 family	Post-SET (CXCXC), SET domain	MLL1-4	set1	Phatr3_J21456(SET1), Phatr3_J15937(NSD1), Phatr3_J6915(MLL1), Phatr3_J44935(SET2 family), Phatr3_J6093(SET2), Phatr3_EG022 11(SET1 family)	35182(SET1),35531(SET1),22757(SET2 family),264323(SET2)	182906(SET1),1924 71(SET2),152933(SET2 family), 0030240(SET2),0033940(SET2 family)	0078730(SET1 family),0068570(SET1),0047650(SET2 family), 0030240(SET2),0033940(SET2 family)
		STE1L	-	-	-	-	-
SET2 family	Pre-SET (7-9 Cys), post-SET (CXCXC), SET domain	NSD1-3	set2	-	-	-	-
		HIFI	-	-	-	-	-
		ASH1	-	-	-	-	-
RIZ family	C2H2 zinc finger, SET domain	RIZ	-	-	-	-	-
		BLIMP1	-	-	-	-	-
SMYD family	Post-SET (CXCXC), ZFYND, SET domain	SYMD1-3	-	Phatr3_J1647*Phatr3_J43708	23831,24988	231918, 248695	0076820,
EZ family	Pre-SET (~15 Cys), SET domain	EZH1-2	-	Phatr3_J6698	268872	181541	-
SUV1-20 family	Post-SET (CXCXC), SET domain	SUV4-20	-	-	-	-	-
Others		SET7-9	-	not found but MORN motif are found	-	-	-
DOT family		DOTL1	dot1	Phatr3_J50632	-	234840	0098260
	Demethylase (KDMs)						
KDM1	Amine oxidase	LSD1,	<i>S. pombe</i>	Phatr3_EG01090,Ph	-	138681,264053,2	0059180,0032800,0007610,010
		LSD2	wml/Sat1	atr3_J51708,	-	40044,264053,19	2820,0030440,0079500
			0	Phatr3_J44106, Phatr3_J48603	-	3836	-
KDM2	F-box + JmjC	FBXL10/11	Jhd1	Phatr3_J42595	14890,24345	184520	0025410,0061600
		JmjN + JmjC	Rph1	Phatr3_J48109	22122,	237169,	-
KDM5A-B	JmjN + JmjC + ZF-C5HC2 + ARID + ZF-PHD	JARID1A-C	Jhd2	Phatr3_J48747	2137	206463	0097490
		TPR + JmjC	UTX	-	not found but multiple TRP repeat proteins found	-	not found but multiple TRP repeat proteins found
KDM6B	JmjC	JMJD3	-	Too many, so not listed here	Too many, so not listed here	Too many, so not listed here	Too many, so not listed here

*: Low similarity

homolog named Clr4 was identified in *S. pombe* (Ivanova et al. 1998). Although H3K9me2 and me3 modifications were both found in *P. tricornutum*, no homologs of SUV39H1 in any of the four diatom species discussed here has been identified, suggesting the existence of a diatom-specific H3K9 methyltransferase. Another probability is that H3K9 methylation is deposited by other SET-domain HMTs such as enhancer of zeste which was shown recently to methylate lysine 9 of histone H3 in Paramecium (Frapporti et al. 2019). SUV1-20 family is also missing in diatom species and yeast, compared to their diversity in humans suggesting that these two families might have evolved with the emergence of multicellular species.

SET1 and SET2 families are two similar groups of proteins which both possess SET and Post-SET domains with sometimes another domain, a Pre-SET found in the SET2 family (Dillon et al. 2005). Although most of SET containing HMTs are similar to yeast Set1 or Set2, there are some unique diatom HMTs which show more similarity to human MLL1. For instance *P. tricornutum* Phatr3_J6915 and *P. multiseriis* 0078730 are homologs of human MLL1 with extra Bromo and PHD-finger domains which yeast Set1 does not have. *P. tricornutum* has homologs of not only human SET2 family genes such as NSD1-3 (Phatr3_J15937) but also yeast Set2 (Phatr3_6903). The redundancy of SET family in humans is believed to be related to multiple targets and functions of MLL/COMPASS complexes, but diatoms are unicellular species and the redundancy of SET-containing HMT is not clear. More genes with SET domain are found in diatoms but are not listed in Table 3 because of the lack of clarity on how to classify them with single SET domain. A SET domain containing enzyme is an enhancer of zeste E(z), which is known to methylate lysine 27 of histone H3. E(z) is the only methyltransferase characterized so far in diatoms, where it was shown to deposit H3K27me2/me3 in *P. tricornutum* (Zhao et al., doi: <https://doi.org/10.1101/2019.12.26.888800>).

Histone demethylases (HDMs) are enzymes responsible for the removal of methyl predominantly from lysine and arginine residues. HDMs can be categorized into two families with six classes, HDM1, HDM2, HDM4D, HDM5A/B, HDM6A and 6B. Except HDM1 which has an amine oxidase domain, the rest of the classes possess Jumonji C (JmjC) domain containing iron- and α -ketoglutarate (2OG)-dependent oxygenases (Klose et al. 2006). Each HDMs is a site-specific histone demethylase, including HDM2A which demethylates H3K36me1/2 (Tsukada et al. 2006) and HDM1 involved with demethylation of H3K4me1/2 (Rudolph et al. 2007). In *S. cerevisiae* there are only JmjC family histone demethylases, Jhd1, Jhd2 and Rhd2. However, in diatoms, additional domains were found including amine oxidase, multiple TRP repeat or SET domains suggesting that some of the diatom HDMs might have dual functions and/or specific recognition and demethylation mechanism.

9 Non-coding RNAs and the RNAi Machinery Components

A considerable portion of eukaryotic genomes can be transcribed to RNAs with no coding potential. According to their length, they can be classified into small and long non-coding RNAs (lncRNAs). They were shown to be involved in silencing, house-keeping functions, cell differentiation, development and stress response.

9.1 Small Non-coding RNAs

Different epigenetic mechanisms have evolved in eukaryotes to silence the expression of genes and mobility of transposable elements (TEs). They all require the cleavage of input double strand RNA into small RNAs (micro RNAs (miRNA) and small interfering RNAs (siRNA)) with a size between 19 to 31 nt in length) by an enzyme called dicer. The small RNAs are then bound by Argonaute proteins which are part of the RNA Induced Silencing Complex (RISC) with RNA-dependent RNA polymerases (RDRs) (Castel and Martienssen 2013). RISC uses the small RNAs as guides for sequence specific gene and TEs silencing via translational repression, mRNA degradation and heterochromatin formation by recruitment of histone and/or DNA methyltransferases to regulatory sequences of the target genes (Holoch and Moazed 2015). RNA mediated recruitment of DNA methyltransferase for silencing is known as RNA directed DNA methylation (RdDM), widely studied in *Arabidopsis thaliana*. In plants, canonical RdDM comprises 2 steps (1) biogenesis of 24 nt siRNAs mediated by Pol IV, the RNA-dependent RNA polymerase 2 (RDR2) and the dicer endonuclease 3 (DCL3) and (2) de novo methylation involving PolV, AGO4 and de novo dnmts (Law and Jacobsen 2010). Observation of systematic presence of siRNA over DNA-methylated TEs (Tirichine et al. 2017; Rogato et al. 2014) (Fig. 3) leading to their silencing in diatoms suggests that RdDM is not only restricted to plants and few other species but seems to have an evolutionary deep origin. RNA-mediated silencing was reported to occur in diatoms including both model species *P. tricornutum* (De Riso et al. 2009; Sakaguchi et al. 2011) and *T. pseudonana* (Shrestha and Hildebrand 2015). Small RNAs were characterized in both species as well as the polar diatom *F. cylindrus* (Rogato et al. 2014; Lopez-Gomollon et al. 2014; Norden-Krichmar et al. 2011; Huang et al. 2011).

Despite two reports of in silico prediction of miRNA (Norden-Krichmar et al. 2011; Huang et al. 2011), canonical miRNAs were not detected in any of the diatoms suggesting a diversified small RNA biogenesis pathway. Scanning of the MMETSP database reveals that diatoms encode all the components of the RNAi machinery (data not shown). Argonaute proteins which are highly conserved among species have multiple members in plants (10 in *Arabidopsis*, 19 in rice) (Kapoor et al. 2008; Baumberger and Baulcombe 2005), humans (8), *Drosophila melanogaster* (5) *C. elegans* (27), *Neurospora crassa* (2) but only one copy in the investigated diatoms except *Fragilariopsis cylindrus* and *Fustilifera solaris* which encode two copies. All the diatom Agos contain the Piwi-Argonaute-Zwille (PAZ), shared with

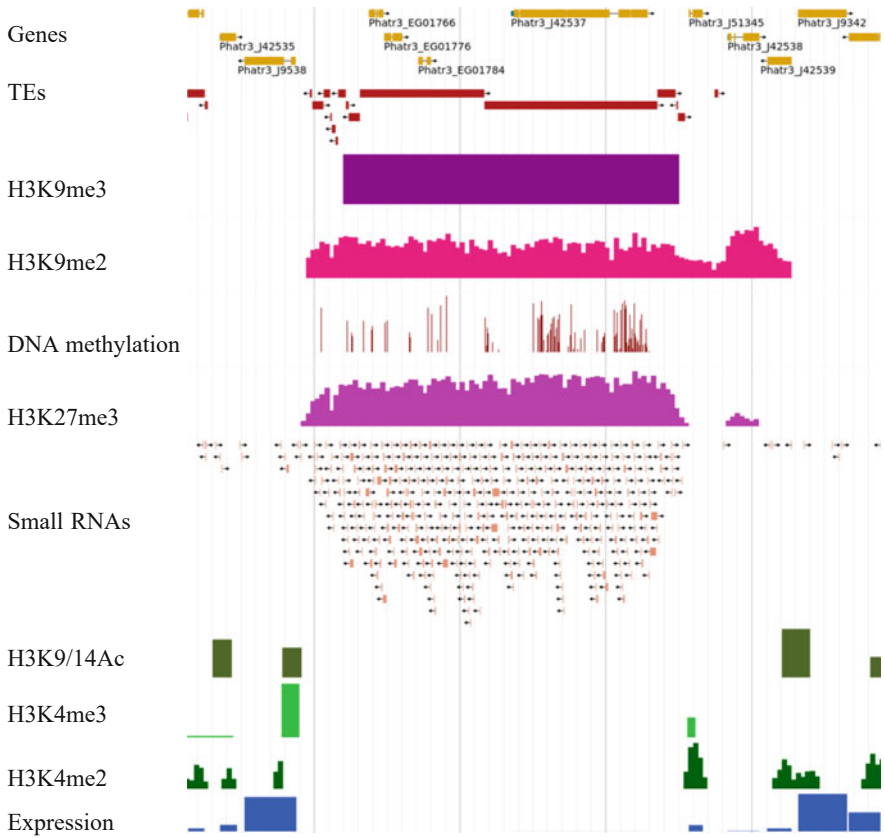


Fig. 3 Snapshot of the epigenome browser illustrating co-occurrence of repressive marks including DNA methylation and the presence of small RNAs over TEs suggesting RNA directed DNA methylation leading to silencing

Dicer and PIWI domains which are important for binding to small RNAs and cleavage.

Unlike Ago, Dicer shows a poor conservation in diatoms. Typical Dicer protein such as in humans shows an N-terminal Dead Like helicase domain (DEXDc/ Helicase), DUF283, a domain of unknown function, a PAZ, two RNaseIII domains (RNaseIIIa and RNaseIIIb) and a dsRNA binding domain. Diatom dicers have the two RNase III domains but miss the typical DEXDc/Helicase and for some of them the PAZ domain. We will therefore refer to diatom Dicer as Dicer like (DCL). *Giardia intestinalis* which is a unicellular parasite of the Excavates is the only species in which the crystal structure of Dicer was determined. Structural analysis has shown the importance of a conserved residue among all dicers (Proline at position 266) in dicer function (Macrae et al. 2006). This residue is in the platform domain between RNase and PAZ domains and was found conserved in the investigated diatoms of MMETSP except *Minutocellus polymorphus*.

G. intestinalis, which has a dicer protein similar to some diatoms with only tandem RNaseIII and PAZ domains was shown to be capable of dicing dsRNA in vitro and to support RNAi in vivo (Macrae et al. 2006) suggesting similar functionalities in diatoms. Most of the diatoms including *P. tricornutum* and *T. pseudonana* miss the PAZ domain and have only the RNase domains suggesting their importance in RNA-mediated silencing and diversified silencing pathways. Interestingly *F. cylindrus* DCL has unique features in that it is the only diatom with a DEXDc/ Helicase domain and an N terminal C5 DNA methylase domain similar to dnmt4 C5 methylase domain. This unique combination suggests an intimate interaction between DNA methylation and DCL domains to mediate silencing in an RNA-directed DNA methylation fashion.

9.2 Long Non-coding RNAs

LncRNAs are a class of transcripts with lengths superior to 200 nt and no coding potential. They can be intronic, intergenic or antisense transcripts. Although, not coding for proteins, they play an important role in gene regulation in combination with chromatin remodeling complexes and histone modifications (Fatica and Bozzoni 2014). A famous example is COLD ASSISTED INTRONIC NONCODING RNA (*COLDAIR*), which is a plant lncRNA encoding a flowering inhibitor protein Flowering Locus C (FLC) which regulates vernalization. Knockdown of FLC decreases its expression which causes late flowering after vernalization (Heo and Sung 2011). LncRNAs are poorly investigated in diatoms where the fraction of non-coding genome is estimated around 40% in both *P. tricornutum* and *T. pseudonana* (Rastogi et al. 2015). Few interesting studies reported the presence of lncRNAs in *P. tricornutum* under stress conditions or in natural variants of the species (Huang et al. 2018; Cruz de Carvalho et al. 2016; Rastogi 2016). The studies revealed the synthesis of intergenic lncRNAs under phosphate depletion and high CO₂ with some shared lncRNAs in stress-related studies suggesting an important and central regulatory role of lncRNAs in response to stresses. Validation of these lncRNAs using Phatr3 gene models (Rastogi et al. 2018) for those identified under phosphate depletion and quantitative RT PCR as well as functional studies is necessary to demonstrate the relevance of these lncRNAs to phosphate and high CO₂ metabolisms.

10 Conclusions and Future Directions

In recent years, some progress has been made in the characterization of epigenetic factors in few diatoms where still many questions remain to be addressed. Epigenetics in diatoms emerged only recently and it is experiencing classic gain in knowledge rate similar to previous disciplines with important progress expected to happen in the future. Diatoms and microalgae in general represent suitable species to address fundamental questions about epigenetic mechanisms involved in genome

regulation. Genome size, short life cycle, conservation of epigenetic components and lack of redundancy are all favourable factors in these microscopic living organisms to provide insightful findings about their epigenetic regulation.

Veluchamy and co-authors have shown the importance of PTMs of histones and DNA methylation in the response of *P. tricornutum* to nitrate starvation which induced dramatic changes genome wide in the redistribution of H3K9me3, H3K9/14Ac and DNA methylation with a decrease or an increase in the expression of targeted regions upon loss or gain of one or more of these marks (Veluchamy et al. 2015). A recent study has established a link between PRC2 and its associated repressive mark H3K27me3 in cell differentiation in *P. tricornutum*. Knockout of the catalytic subunit of PRC2, enhancer of zeste in three morphotypes, fusiform, cruciform and triradiate, led to a change in the morphology and caused a genome wide depletion in H3K27me3 suggesting a role of the polycomb mark in cell differentiation (Zhao et al. 2021). As widely investigated in many model species, epigenetic factors are likely to regulate many biological processes in diatoms including but not limited to stress responses, cell cycle, differentiation, life cycle and reproduction.

The feasibility of gene editing in diatoms made these species attractive and boosted our knowledge of their gene function. The diversity of epigenetic factors and their peculiar domain combinations need to be addressed using the genetic tools that are now available in *P. tricornutum* and some other diatoms that emerge as additional and attractive models (Tirichine et al. 2017). The recent evolution of customizable epigenome engineering tools in mammals is a great inspiration for diatom biologists. Typical examples include a strategy that uses fusions of engineered transcription activator-like effector (TALE) repeat arrays and the TET1 hydroxylase catalytic domain for efficient targeted demethylation of specific CpGs in human cells (Maeder et al. 2013). This TALE system is effectively used in fusion with histone demethylase of LSD1 type to remove enhancer associated chromatin modifications from target loci (Mendenhall et al. 2013). Epigenetic factors, their writers, erasers and readers do not act in isolation, and multiple evidence point to complex interactions orchestrating epigenetic-mediated regulation of genomes. Typical examples include lncRNAs interactions with histone modifiers, shaping thus the outcomes of gene transcription and nuclear architecture. Likewise, DNA methylation and histone modifications maintain a close cross-talk that deserves further investigation.

Another epigenetic process that deserves attention is RNA editing, known as specific modifications to nucleotide within an RNA molecule after synthesis by the RNA polymerase. Such examples include pseudouridylation which is the isomerization of uridine residues and deamination which is the removal of an amine group from cytidine to give rise to uridine mostly known as C to U change (more RNA editing types exist). RNA editing can also be insertional or deletional, in which nucleotides are added to, and in some cases also removed from a transcript (Lin et al. 2008). In humans, Adenosine deaminase acting on RNA (ADAR) is the enzyme that converts adenosine (A) to inosine (I) by deamination. RNA editing takes place within the nucleus and the cytoplasm as well as in the mitochondria and the plastids. It is known to occur in animals, plants, trypanosomes, dinoflagellates (only later

diverging dinoflagellates) and even viruses. RNA editing was not detected in ciliates apicomplexans, basal lineages of dinoflagellates (Lin et al. 2008) and not yet documented in diatoms. However, in silico search detects homologues of ADAR enzyme in several diatom species including *P. tricornutum* (data not shown), which is likely going to be a great opportunity to investigate the role of RNA editing in generating protein diversity in microalgae.

Functional studies in diatoms provided important findings about adaptation to their environments, and because epigenetic variations are intimately connected to adaptation, it is important to investigate such connection over several generations asking, what is the role of DNA methylation, histone modifications and non-coding RNAs in the evolution of adaptive traits in response to specific changes in environmental factors. The inheritance of such modifications can be investigated in clonally propagating species such as *P. tricornutum* but also species reproducing sexually such as *H. ostrearia* where current studies in Mouget's lab are addressing such topics. Epigenetic studies in diatoms are undoubtedly going to provide exciting and important insights for still many years to come.

References

- Agius F, Kapoor A, Zhu JK (2006) Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc Natl Acad Sci U S A* 103:11796–11801
- Akella JS, Wloga D, Kim J, Starostina NG, Lyons-Abbott S, Morrissette NS, Dougan ST, Kipreos ET, Gaertig J (2010) MEC-17 is an α -tubulin acetyltransferase. *Nature* 467:218–222
- Amselem J, Lebrun MH, Quesneville H (2015) Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. *BMC Genomics* 16:141
- Baumberger N, Baulcombe DC (2005) Arabidopsis ARGONAUTE1 is an RNA slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A* 102:11928–11933
- Bestor TH (1990) DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philos Trans R Soc Lond Ser B Biol Sci* 326:179–187
- Bewick AJ, Hofmeister BT, Powers RA, Mondo SJ, Grigoriev IV, James TY, Stajich JE, Schmitz RJ (2019) Diversity of cytosine methylation across the fungal tree of life. *Nat Ecol Evol* 3:479–490
- Bian K, Lenz SAP, Tang Q, Chen F, Qi R, Jost M, Drennan CL, Essigmann JM, Wetmore SD, Li D (2019) DNA repair enzymes ALKBH2, ALKBH3, and AlkB oxidize 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine in vitro. *Nucleic Acids Res* 47:5522–5529
- 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
- Carrozza MJ, Utley RT, Workman JL, Cote J (2003) The diverse functions of histone acetyltransferase complexes. *Trends Genet* 19:321–329
- Castel SE, Martienssen RA (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet* 14:100–112
- Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell* 110:33–42

- Cruz de Carvalho MH, Sun HX, Bowler C, Chua NH (2016) Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. *New Phytol* 210:497–510
- Cuyppers B et al (2020) The absence of C-5 DNA methylation in *Leishmania donovani* allows DNA enrichment from complex samples. *Microorganisms*. 8(8):1252. <https://doi.org/10.3390/microorganisms8081252>
- de Mendoza A, Bonnet A, Vargas-Landin DB, Ji N, Li H, Yang F, Li L, Hori K, Pflueger J, Buckberry S et al (2018) Recurrent acquisition of cytosine methyltransferases into eukaryotic retrotransposons. *Nat Commun* 9:1341
- de Mendoza A, Lister R, Bogdanovic O (2019) Evolution of DNA Methylome diversity in eukaryotes. *J Mol Biol*
- De Riso V, Raniello R, Maumus F, Rogato A, Bowler C, Falciatore A (2009) Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res* 37:e96
- Deschamps J, van Nes J (2005) Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development* 132:2931–2942
- Dillon SC, Zhang X, Trievel RC, Cheng X (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol* 6:227
- Dumesic PA, Stoddard CI, Catania S, Narlikar GJ, Madhani HD (2020) ATP hydrolysis by the SNF2 domain of Dnmt5 is coupled to both specific recognition and modification of Hemimethylated DNA. *Mol Cell* 79(127–139):e124
- Elgin SC, Reuter G (2013) Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb Perspect Biol* 5:a017780
- Fan X, Han W, Teng L, Jiang P, Zhang X, Xu D, Li C, Pellegrini M, Wu C, Wang Y et al (2020) Single-base methylome profiling of the giant kelp *Saccharina japonica* reveals significant differences in DNA methylation to microalgae and plants. *New Phytol* 225:234–249
- Fatica A, Bozzoni I (2014) Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* 15:7–21
- Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y (2002) Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol* 12:1052–1058
- Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME et al (2010) Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A* 107:8689–8694
- Frapporti A, Miro Pina C, Arnaiz O, Holoch D, Kawaguchi T, Humbert A, Eleftheriou E, Lombard B, Loew D, Sperling L et al (2019) The Polycomb protein Ezh1 mediates H3K9 and H3K27 methylation to repress transposable elements in paramecium. *Nat Commun* 10:2710
- Fu D, Brophy JA, Chan CT, Atmore KA, Begley U, Paules RS, Dedon PC, Begley TJ, Samson LD (2010) Human AlkB homolog ABH8 is a tRNA methyltransferase required for wobble uridine modification and DNA damage survival. *Mol Cell Biol* 30:2449–2459
- Galagan JE, Selker EU (2004) RIP: the evolutionary cost of genome defense. *Trends Genet* 20:417–423
- Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* 124:495–506
- Gladyshev E (2017) Repeat-induced point mutation and other genome defense mechanisms in fungi. *Microbiol Spectr* 5
- Goodman RH, Smolik S (2000) CBP/p300 in cell growth, transformation, and development. *Genes Dev* 14:1553–1577
- Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizabal-Corrales D, Hsu CH, Aravind L, He C, Shi Y (2015) DNA methylation on N6-adenine in *C. elegans*. *Cell* 161:868–878
- Haberland M, Montgomery RL, Olson EN (2009) The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet* 10:32–42
- Heo JB, Sung S (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331:76–79

- Hepworth J, Dean C (2015) Flowering locus C's lessons: conserved chromatin switches underpinning developmental timing and adaptation. *Plant Physiol* 168:1237–1245
- Holoch D, Moazed D (2015) RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* 16:71–84
- Huang Y (2002) Transcriptional silencing in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Nucleic Acids Res* 30:1465–1482
- Huang A, He L, Wang G (2011) Identification and characterization of microRNAs from *Phaeodactylum tricornutum* by high-throughput sequencing and bioinformatics analysis. *BMC Genomics* 12:337
- Huang R, Ding J, Gao K, Cruz de Carvalho MH, Tirichine L, Bowler C, Lin X (2018) A potential role for epigenetic processes in the acclimation response to elevated pCO₂ in the model diatom *Phaeodactylum tricornutum*. *Front Microbiol* 9(3342)
- Huff JT, Zilberman D (2014) Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* 156:1286–1297
- Ideraabdullah FY, Vigneau S, Bartolomei MS (2008) Genomic imprinting mechanisms in mammals. *Mutat Res* 647:77–85
- Ivanova AV, Bonaduce MJ, Ivanov SV, Klar AJ (1998) The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast. *Nat Genet* 19:192–195
- Iyer LM, Zhang D, Aravind L (2016) Adenine methylation in eukaryotes: apprehending the complex evolutionary history and functional potential of an epigenetic modification. *BioEssays* 38:27–40
- Jeltsch A, Ehrenhofer-Murray A, Jurkowski TP, Lyko F, Reuter G, Anki S, Nellen W, Schaefer M, Helm M (2017) Mechanism and biological role of Dnmt2 in nucleic acid methylation. *RNA Biol* 14:1108–1123
- Kanchan S, Mehrotra R, Chowdhury S (2015) In silico analysis of the endonuclease III protein family identifies key residues and processes during evolution. *J Mol Evol* 81:54–67
- Kapoor M, Arora R, Lama T, Nijhawan A, Khurana JP, Tyagi AK, Kapoor S (2008) Genome-wide identification, organization and phylogenetic analysis of dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC Genomics* 9:451
- 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 Biol* 12:e1001889
- Kitabayashi I, Aikawa Y, Nguyen LA, Yokoyama A, Ohki M (2001) Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ–CBP fusion protein. *EMBO J* 20:7184–7196
- Klein BJ, Jang SM, Lachance C, Mi W, Lyu J, Sakuraba S, Krajewski K, Wang WW, Sidoli S, Liu J et al (2019) Histone H3K23-specific acetylation by MORF is coupled to H3K14 acylation. *Nat Commun* 10:4724
- Klose RJ, Kallin EM, Zhang Y (2006) JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* 7:715–727
- Kohler C, Wolff P, Spillane C (2012) Epigenetic mechanisms underlying genomic imprinting in plants. *Annu Rev Plant Biol* 63:331–352
- Laisne M, Gupta N, Kirsh O, Pradhan S, Defossez PA (2018) Mechanisms of DNA methyltransferase recruitment in mammals. *Genes (Basel)* 9
- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11:204–220
- Lin S, Zhang H, Gray MW (2008) RNA editing in dinoflagellates and its implications for the evolutionary history of the editing machinery. In: Smith HC (ed) *RNA and DNA editing: molecular mechanisms and their integration into biological systems*. Wiley, Boca Raton, FL, pp 280–309

- Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–322
- 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
- 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
- Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, Ho QH, Sander JD, Reyon D, Bernstein BE et al (2013) Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* 31:1137–1142
- 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
- Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK, Bernstein BE (2013) Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol* 31:1133–1136
- Mishima Y, Miyagi S, Saraya A, Negishi M, Endoh M, Endo TA, Toyoda T, Shinga J, Katsumoto T, Chiba T (2011) The Hbo1-Brd1/Brpf2 complex is responsible for global acetylation of H3K14 and required for fetal liver erythropoiesis. *Blood* 118:2443–2453
- Ng HH, Feng Q, Wang H, Erdjument-Bromage H, Tempst P, Zhang Y, Struhl K (2002) Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and sir protein association. *Genes Dev* 16:1518–1527
- Norden-Krichmar TM, Allen AE, Gaasterland T, Hildebrand M (2011) Characterization of the small RNA transcriptome of the diatom, *Thalassiosira pseudonana*. *PLoS One* 6:e22870
- Osada S, Sutton A, Muster N, Brown CE, Yates JR 3rd, Sternglanz R, Workman JL (2001) The yeast SAS (something about silencing) protein complex contains a MYST-type putative acetyltransferase and functions with chromatin assembly factor ASF1. *Genes Dev* 15:3155–3168
- Ponger L, Li WH (2005) Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol* 22:1119–1128
- Rastogi A (2016) *Phaeodactylum tricornutum* genome and epigenome: characterization of natural variants. PSL Research University, Paris
- 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, Tirichine L (2018) Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci Rep* 8:4834
- Rogato A, Richard H, Sarazin A, Voss B, Cheminant Navarro S, Champeimont R, Navarro L, Carbone A, Hess WR, Falciatore A (2014) The diversity of small non-coding RNAs in the diatom *Phaeodactylum tricornutum*. *BMC Genomics* 15:698
- Rudolph T, Yonezawa M, Lein S, Heidrich K, Kubicek S, Schäfer C, Phalke S, Walther M, Schmidt A, Jenuwein T (2007) Heterochromatin formation in *Drosophila* is initiated through active removal of H3K4 methylation by the LSD1 homolog SU (VAR) 3-3. *Mol Cell* 26:103–115
- 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 Physiol* 156:78–89
- Schmitz RJ, Lewis ZA, Goll MG (2019) DNA methylation: shared and divergent features across eukaryotes. *Trends Genet* 35:818–827
- Seto E, Yoshida M (2014) Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb Perspect Biol* 6:a018713

- Shida T, Cueva JG, Xu Z, Goodman MB, Nachury MV (2010) The major α -tubulin K40 acetyltransferase α TAT1 promotes rapid ciliogenesis and efficient mechanosensation. *Proc Natl Acad Sci* 107:21517–21522
- Shrestha RP, Hildebrand M (2015) Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. *Eukaryot Cell* 14:29–40
- Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* 21:64–72
- Tamaru H, Selker EU (2001) A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* 414:277–283
- Tirichine L, Lin X, Thomas Y, Lombard B, Loew D, Bowler C (2014) Histone extraction protocol from the two model diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. *Mar Genomics* 13:21–25
- Tirichine L, Rastogi A, Bowler C (2017) Recent progress in diatom genomics and epigenomics. *Curr Opin Plant Biol* 36:46–55
- Traller JC, Cokus SJ, Lopez DA, Gaidarenko O, Smith SR, McCrow JP, Gallaher SD, Podell S, Thompson M, Cook O et al (2016) Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. *Biotechnol Biofuels* 9:258
- Tsukada Y-i, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439:811–816
- Veluchamy A, Lin X, Maumus F, Rivarola M, Bhavsar J, Creasy T, O'Brien K, Sengamalay NA, Tallon LJ, Smith AD et al (2013a) Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. *Nat Commun* 4
- Veluchamy A, Lin X, Maumus F, Rivarola M, Bhavsar J, Creasy T, O'Brien K, Sengamalay NA, Tallon LJ, Smith AD et al (2013b) Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. *Nat Commun* 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 tricornutum*. *Genome Biol* 16:102
- Waddington CH (1942) The epigenotype. *Endeavour* 1:18–20
- Waddington CH (1957) The strategy of the genes. Macmillan, New York
- Wang L, Tang Y, Cole PA, Marmorstein R (2008) Structure and chemistry of the p300/CBP and Rtt109 histone acetyltransferases: implications for histone acetyltransferase evolution and function. *Curr Opin Struct Biol* 18:741–747
- Wu X, Zhang Y (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet* 18:517–534
- Wu F, Olson BG, Yao J (2016) DamID-seq: genome-wide mapping of protein-DNA interactions by high throughput sequencing of adenine-methylated DNA fragments. *J Vis Exp*. 107:e53620
- Zdzalik D, Vagbo CB, Kirpekar F, Davydova E, Puscian A, Maciejewska AM, Krokan HE, Klungland A, Tudek B, van den Born E, Falnes PO (2014) Protozoan ALKBH8 oxygenases display both DNA repair and tRNA modification activities. *PLoS One* 9:e98729
- Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328:916–919
- Zhao X, Rastogi A, Deton Cabanillas AF, Ait Mohamed O, Cantrel C, Lombard B, Murik O, Genovesio A, Bowler C, Bouyer D, Loew D, Lin X, Veluchamy A, Vieira FRJ, Tirichine L (2019) H3K27me3 natural variation selectively marks genes predicted to be important for differentiation in unicellular algae. *bioRxiv*. <https://doi.org/10.1101/2019.12.26.888800>
- 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 tricornutum*. *New Phytol* 229:3208–3220