



## Stemming epigenetics in marine stramenopiles

Florian Maumus, Pablo Rabinowicz, Chris Bowler, Máximo Rivarola

### ► To cite this version:

Florian Maumus, Pablo Rabinowicz, Chris Bowler, Máximo Rivarola. Stemming epigenetics in marine stramenopiles. *Current Genomics*, 2011, 12 (5), pp.357-370. hal-02644011

**HAL Id: hal-02644011**

**<https://hal.inrae.fr/hal-02644011>**

Submitted on 28 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Stemming Epigenetics in Marine Stramenopiles

Florian Maumus<sup>\*1</sup>, Pablo Rabinowicz<sup>3</sup>, Chris Bowler<sup>2</sup> and Maximo Rivarola<sup>\*4</sup>

<sup>1</sup>Unité de Recherche en Génomique-Info, UR 1164, INRA Centre de Versailles-Grignon, Versailles, France

<sup>2</sup>Institut de Biologie de l'Ecole Normale Supérieure, Section de Génomique Environnementale et Evolutive, CNRS UMR 8197 INSERM U1021, Paris, France

<sup>3</sup>Institute for Genome Sciences, Department of Biochemistry and Molecular Biology, University of Maryland, School of Medicine, BioPark Building II, 801 West Baltimore Street, Baltimore, MD 21201, USA

<sup>4</sup>Instituto Nacional de Tecnología Agropecuaria (INTA), Instituto de Biotecnología (CNIA), CC 25, Castelar (B1712WAA), Buenos Aires, Argentina

**Abstract:** Epigenetics include DNA methylation, the modification of histone tails that affect chromatin states, and small RNAs that are involved in the setting and maintenance of chromatin modifications. Marine stramenopiles (MAS), which are a diverse assemblage of algae that acquired photosynthesis from secondary endosymbiosis, include single-celled organisms such as diatoms as well as multicellular forms such as brown algae. The recent publication of two diatom genomes that diverged ~90 million years ago (mya), as well as the one of a brown algae that diverged from diatoms ~250 Mya, provide a great system of related, yet diverged set of organisms to compare epigenetic marks and their relationships. For example, putative DNA methyltransferase homologues were found in diatoms while none could be identified in the brown algal genome. On the other hand, no canonical DICER-like protein was found in diatoms in contrast to what is observed in brown algae. A key interest relies in understanding the adaptive nature of epigenetics and its inheritability. In contrast to yeast that lack DNA methylation, homogeneous cultures of diatoms constitute an attractive system to study epigenetic changes in response to environmental conditions such as nutrient-rich to nutrient-poor transitions which is especially relevant because of their ecological importance. *P. tricornutum* is also of outstanding interest because it is observed as three different morphotypes and thus constitutes a simple and promising model for the study of the epigenetic phenomena that accompany cellular differentiation. In this review we focus on the insights obtained from MAS comparative genomics and epigenomic analyses.

Received on: May 05, 2011 - Revised on: May 24, 2011 - Accepted on: June 17, 2011

**Keywords:** Marine stramenopiles, epigenomics, DNA methylation, chromatin, diatom, genomics, Small RNA, Brown algae, Transposable elements.

## INTRODUCTION

Epigenetics has different biological meanings for different people. It was historically defined by Conrad H. Waddington as how the “canalization” of genotypes can give rise to inheritable phenotypes through “genetic assimilation”. This concept, which involves evolutionary feedback through natural selection of acquired traits was therefore re-introducing the Lamarckian concept into adaptive evolution. Decades later, Arthur Riggs and collaborators defined epigenetics as “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence”. Rather than pointing to one defined biological phenomena, the contemporary use of the word “epigenetics” embraces different basic biological

mechanisms that were found to participate in setting and maintaining heritable changes that do not rely on changes in DNA sequence. These encompass various phenomena that modify chromatin including the modification of histone marks, DNA methylation, RNA interference (RNAi), and changes in higher-order structure of the chromosomes in the nucleus. These processes play key roles in generating and maintaining specific chromatin states that impact its compaction and accessibility and thus affect the readout of the underlying DNA sequence. In some cases, such modifications can be transmitted to subsequent generations. For instance, it was found that the phenotype of a peloric reversible mutant of *Linaria vulgaris* was due to extensive DNA methylation and transcriptional silencing at the *Lcyc* gene, which controls floral dorsoventral asymmetry [1]. Such inherited epigenetic marks are often referred to as “epimutations”. By contrast, some chromatin marks can be highly dynamic. For example, the phosphorylation of the variant histone H2AX after double-strand break is only transient [2]. As a consequence, an incorrect but widespread use of the word epigenetics refers to the study of these chromatin-related processes, whether their outcomes are inherited or not. Therefore, the temporary chromatin modifications accompanying processes such as

\*Address correspondence to this author at the Unité de Recherche en Génomique-Info, UR 1164, INRA Centre de Versailles-Grignon, Versailles, France; Tel: (0033) 1 30 83 31 74; Fax: (0033) 1 30 83 38 99; E-mail: fmaumus@gmail.com

Instituto Nacional de Tecnología Agropecuaria (INTA), Instituto de Biotecnología (CNIA), CC 25, Castelar (B1712WAA), Buenos Aires, Argentina; Tel: (0054-11) 4621-1278 extension: 173; E-mail: rivabros2@gmail.com

DNA repair, DNA replication, alternative splicing [3], and reprogramming [4] are also sometimes dubbed “epigenetic” phenomena [5]. One reason for this is that in many organisms, including those that are the subject of this review, there are very few studies of the heritability of such phenomena.

In eukaryotes, cytosine methylation is a common epigenetic mark that can impact gene expression. It is essential for various biological processes, including transposon silencing [6], imprinting [7], and X chromosome inactivation [8, 9]. The addition of a methyl group on cytosine occurs specifically in the symmetric CG and CHG contexts (in which H = A, T or C), and the asymmetric CHH context [10, 11]. “*De novo*” methylation acts on un-methylated residues in a small RNA-guided fashion known as RNA-dependent DNA methylation (RdDm) [12, 13]. RdDm was first discovered in plants and acts through small double-stranded RNA (dsRNA) that guide DNA methylation at specific loci and in most cases represses gene expression [14]. *De novo* DNA methylation also occurs through spreading from existing methylated loci [15]. Upon DNA replication, methylated symmetric DNA sites become hemi-methylated and are recognized by the maintenance DNA methylation machinery that methylates the corresponding cytosine in the newly synthesized strand [16].

Histone tails can undergo various post-translational modifications such as methylation, acetylation, phosphorylation, and ubiquitylation. Each histone (H2A, H2B, H3, and H4) and histone variant in a nucleosome can carry different modifications. Additional layers of complexity come for example from the fact that a histone can be modified on various residues and that different lysines can be mono-, di-, or trimethylated. As a result, over sixty different modified histones have been reported from different organisms [17], leading to a highly variable pool of modified nucleosomes. Each type of nucleosome can be organized into clusters along the chromosomes and index chromatin with reference to the associated DNA sequence to transmit specific information [18], positively or negatively impacting the activity of the transcriptional machinery either by direct interaction and/or by triggering chromatin remodeling. This vast array of modifications also provides an enormous potential for tuning transcriptional levels in response to signaling conditions within a cell or in different cell types. Through the S phase, at the replication fork, the parental nucleosomes are disrupted into two H2A–H2B dimers and an (H3–H4)<sub>2</sub> tetramer or two H3–H4 dimers. Their transfer or recycling onto the newly synthesized daughter strands provides a first source of histones and the nucleosomal density is restored by assembly of new histones on the duplicated material. Parental histone post translational modifications potentially preserved during transfer can be used as a template to reproduce marks on newly incorporated histones [19].

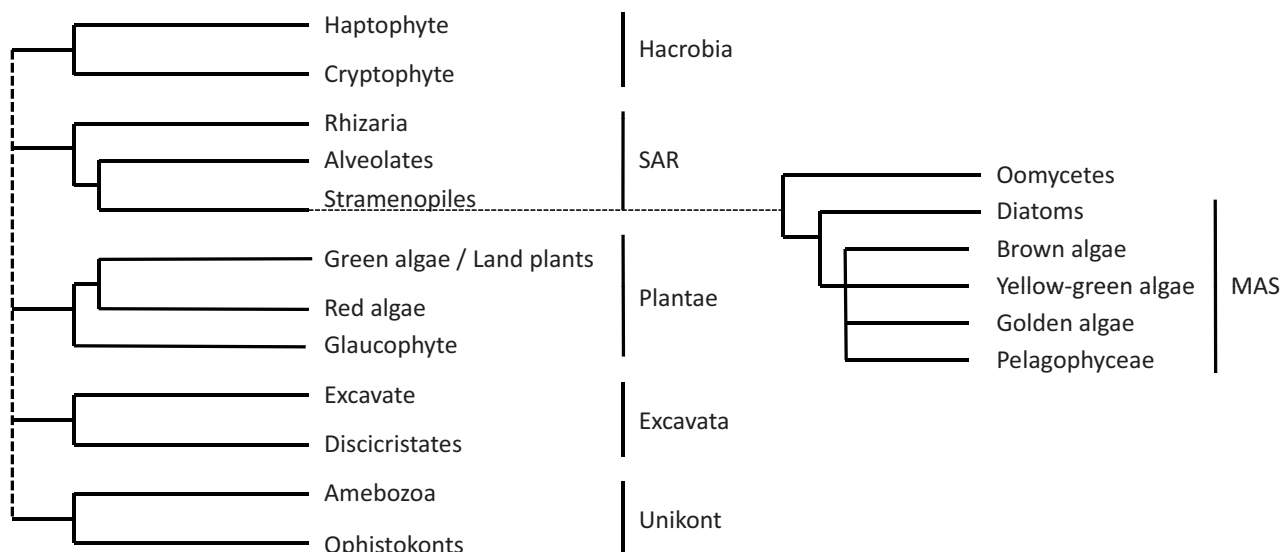
The RNAi machinery in eukaryotes is involved in transcriptional and translational regulation as well as in the primary defense against viruses and transposable elements. This regulatory system relies on the production and action of various classes of small (~20-30 nucleotides) RNA with specific functions. In addition to guiding DNA methylation, other classes of small RNAs including micro RNAs (miRNA) are involved in post-transcriptional gene regulation by directing the cleavage or translational inhibition of tran-

scripts through complete or partial complementarity of their sequence to the target sequence [20, 21].

These processes are the subject of intense studies in model organisms including plants, fungi, and animals where they were found to accomplish key functions. The availability of complete genome sequences and new technologies such as whole-genome tiling arrays and new generation sequencing (NGS) have recently bolstered genome-wide analysis of epigenetic marks by enabling the realization of protocols such as ChIP on chip (a technique which combines chromatin immunoprecipitation to isolate DNA associated with specific proteins such as modified histones, and its quantification using microarray technology), ChIP-seq (similar to ChIP on chip but with the use of new generation sequencing to quantify DNA), RNA-seq (also called “Whole Transcriptome Shotgun Sequencing”, uses new sequencing technologies to quantify a library of cDNA), small RNA-seq (similar to RNA-seq but starts with a library of only small RNAs), BS-seq (also referred to as “Whole genome bisulfite sequencing”, is a high-throughput genome-wide analysis of DNA methylation), and so on [22-24]. Although huge efforts were recently contributed to deciphering the DNA methylome of various species [11, 10, 25, 26] at single-base resolution, little is known about the occurrence and functions of epigenetic regulation in eukaryotes beyond the animal, plant and fungal lineages.

Stramenopiles (also called heterokonts) are a major line of eukaryotes that comprises mostly autotrophic algae such as diatoms and brown algae, but also heterotrophs such as oomycetes, including plant parasites such as *Phytophthora* (Fig. 1). They derive from a secondary endosymbiotic event in which an extant heterotrophic eukaryote either engulfed or was invaded by a red alga [27]. As of 2011, stramenopiles are supposed to belong to a monophyletic eukaryotic supergroup dubbed SAR after the major groups that constitute it: Stramenopiles, Alveolates, and Rhizaria [28] (Fig. 1). It has also been postulated, as the ‘Chromalveolata’ supergroup hypothesis [29], that a single secondary endosymbiosis event, about 1,300 Mya [30], originated the presence of red algal plastids in members of both the stramenopiles and the alveolates and in more distantly related groups such as haptophytes and cryptophytes but not in Rhizaria. However, the analysis of recent sequencing data suggests that chromalveolates constitute a paraphyletic group [31-33]. The origins, types, and acquisition times of secondary plastids in diverse eukaryotic groups including stramenopiles are still the subject of open debate [33-35].

Diatoms (*Bacillariophyceae*) are among the most successful and diversified groups of photosynthetic eukaryotes, with possibly over 100,000 extant species [36]. The contribution of diatom photosynthesis to marine primary productivity has been estimated to be around 40% [37-40]. They are traditionally divided into two orders: the centric diatoms which are radially symmetrical and are thought to have arisen around 180 Million years ago (Mya), followed by the pennate diatoms around 90 Mya which are bilaterally symmetrical [41]. A less ubiquitous group, the brown algae (*Phaeophyceae*) such as giant kelps, are mostly restricted to near-shore lines and rocky areas although some genus such as *Sargassum* are also found free-floating. Brown algae are



**Fig. (1). Stramenopiles are a major lineage of eukaryotes.** Consensus cladogram of selected eukaryotes highlighting five major eukaryotic supergroups with support in recent phylogenetic studies [28, 31, 32]. The dotted polytomy indicates uncertainty regarding the order of early branching events. The inset is a consensus cladogram of selected stramenopiles.

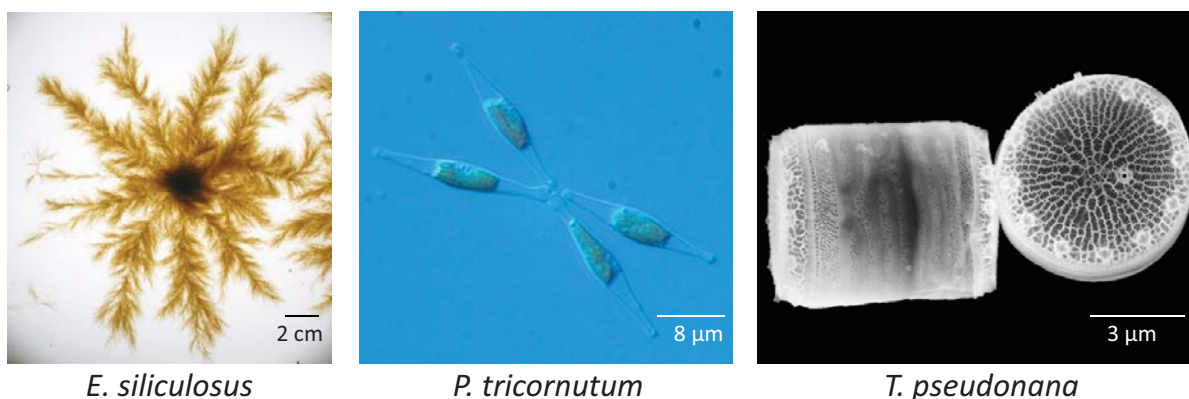
an essential source of ecological habitats and food for zooplankton, crustaceans, and fish. Interestingly, brown algae represent one of the few eukaryotic lineages that evolved multicellularity. Due to their evolutionary, biological and ecological interest, the genome sequences of three marine stramenopiles (MAS) (two highly divergent diatoms and a brown algae) have recently been published. *Thalassiosira pseudonana* (32 Mb) was the first diatom and eukaryotic marine phytoplankton genome to be sequenced [42] (Fig. 2; Table 1). *T. pseudonana* is a centric diatom present throughout the world's oceans and for which sexual reproduction has never been reported. The second diatom genome sequenced belongs to the pennate diatom *Phaeodactylum tricornutum* (27 Mb) [43]. *P. tricornutum* is mostly confined to coastal habitats and also appears to lack a sexual life cycle (Fig. 2; Table 1). Both genomes are relatively small respect to other diatoms, which can reach up to 50 Gb in *Coscinodiscus* sp. [44]. The brown algal genome is from the filamentous seaweed *Ectocarpus siliculosus* (214 Mb) [45], which is now considered as a model for the brown algae (Fig. 2; Table 1). Its genome sequence provides new opportunities for

the application of genomic and epigenomic approaches to study the brown algal biology. More recently, a sequence-tagged genetic map substantially improved the assembly of the *E. siliculosus* genome sequence [46]. These three genomes constitute levers enabling comprehensive studies of epigenetic regulation in stramenopiles.

In this review, we describe the tools available for genetic and epigenetic studies of MAS, highlight some of the most important findings coming from genomic and comparative genomic analyses in these species, and provide an overview and perspectives on our current knowledge about epigenetic mechanisms in sequenced MAS species.

## THE MAS TOOLBOX

On the route towards turning MAS species into model systems, the research community has developed various means to carry out functional genomic studies in MAS organisms with sequenced genomes. Extensive cDNA sequence databases containing over 200,000 *P. tricornutum* and *T. pseudonana* expressed sequence tags (ESTs) from



**Fig. (2). Sequenced MAS species.** Images from *Ectocarpus siliculosus* (Delphine Scornet, CNRS, Roscoff, France), *Phaeodactylum tricornutum* (Alessandra De Martino, CNRS, Paris, France), and *Thalassiosira pseudonana* (Nils Kröger, Georgia Institute of Technology, Atlanta, USA).



**Table 1.** Major Features of the *P. tricornutum*, *T. pseudonana* and *E. siliculosus* Genomes

	<i>P. tricornutum</i>	<i>T. pseudonana</i>	<i>E. siliculosus</i>
Sequencing center	JGI	JGI	VIB/Ugent
Genome size	27.4 Mb	32.4 Mb	195.8 Mb
Predicted genes	10,402	11,776	16,256
Introns per gene	0.79	1.52	6.98
Bacterial genes*	587	> 328	NA
Known TEs	6.4%	1.9%	12.5%
Other repeats*	NA	NA	9,9%

\*For *T. pseudonana*, the number indicates that although species-specific estimation is not available, 56% of the *P. tricornutum* bacterial gene pool was also found in *T. pseudonana*. The number of bacterial genes was not yet estimated in *E. siliculosus*. The amount of unknown repeats in diatoms has not yet been estimated.

cells grown under 16 and 7 different conditions, respectively [47] (<http://www.diatomics.biologie.ens.fr/EST3/>) constitute an important resource for gene expression analysis. Many of those culture conditions reflect ecologically significant stresses that commonly occur in contemporary oceans, such as nitrate and iron limitations [48]. In addition, over 90,000 *E. siliculosus* ESTs have been obtained from 4 different life stages and one stress condition [49]. EST data from *E. siliculosus* were used to design a microarray containing features representing some 17,000 genes, and which has been used to measure gene expression levels during acclimation to three different abiotic stress conditions (hyposaline, hypersaline, and oxidative stress) [49] (<http://www.sb-roscoff.fr/UMR7139/ectocarpus/transcriptomics/>). Whole genome tiling microarrays have also been designed for *T. pseudonana* [50], *P. tricornutum* (Maumus *et al.*, unpublished) and *E. siliculosus* [45].

Functional analyses in diatoms are further bolstered by a molecular toolbox providing protocols and materials for stable transformation of the nuclear genome by biolistics [51, 52], gene cloning using diatom-optimized gateway vectors for protein subcellular localization and immunodetection [53], and targeted gene silencing strategies based on RNA interference mechanisms [54]. Classical genetic techniques, genetic markers and mutagenesis protocols are also available for *E. siliculosus* [55].

## INSIGHTS FROM MAS GENOMICS

Comparative analyses of diatom genomes revealed that *P. tricornutum* has fewer genes than *T. pseudonana* (10,402 opposed to 11,776) and no major synteny could be detected between the two genomes [43]. *T. pseudonana* genes show an average of ~1.52 introns per gene as opposed to 0.79 in *P. tricornutum*, suggesting recent widespread intron gain in the centric diatom [56]. Of greater significance, diatom genomes have provided a wealth of information about the processes of gene transfer (EGT) from endosymbiotic organelles to the host nuclear genome that occurred during stramenopile evolution. For example, the *T. pseudonana* genome encodes six proteins, which are most closely related to genes from the nucleomorph (remnant nucleus of algal endosymbiont [57]) genome of the cryptophyte *Guillardia theta*. Four of these genes were also found in red algal plastid genomes, thus

demonstrating successive EGT from red algal plastid to red algal nucleus (nucleomorph) to stramenopile host nucleus [42]. Moreover, Moustafa and collaborators [35] provided evidence for a green algal-like endosymbiont in the common ancestor of chromalveolates as supported by the fact the 70% of diatom genes of Plantae origin are of green algae lineage provenance and that such genes are also found in the genome of other stramenopiles. Therefore, the authors postulate that chromalveolates were “green” prior to becoming “red” and are the result of serial secondary endosymbioses that conserved the genomic footprints of the previous endosymbioses, but finally displaced the plastid of green algal origin. However, the dataset used for comparison was biased against sequences of red algal origin. Indeed, the only complete or nearly complete red algal genomes available to date come from the acidothermophiles *Cyanidioshizon merolae* [58] and *Galdieria sulphuraria* [59] which contain reduced gene sets (only 5,331 genes in *C. merolae*). Therefore, such analysis is likely to gain refinement by taking into account the forthcoming sequences of larger red algal genomes such as those from *Chondrus crispus* and *Porphyra umbilicalis* that are currently being assembled at Genoscope, France, and at the Joint Genome Institute (JGI), USA, respectively (Boyen C., personal communication; [60]).

Independently of an endosymbiotic event, horizontal gene transfer (HGT) between organisms has a major role in species evolution. Although commonly observed among bacteria, archaea, and between bacteria and archaea, it is much less frequent in eukaryotes [61,62]. Unexpectedly, analysis of the *P. tricornutum* proteome revealed that as much as 5% of its predicted genes were most closely related to bacterial genes, greatly surpassing the level of HGT previously described in eukaryotes. About half of those genes of putative bacterial origin were also found in the *T. pseudonana* genome, attesting their ancient incorporation in the diatom lineage [43]. Hence, stramenopile proteomes appear to consist of a melting pot of genes of various origins.

The *E. siliculosus* genome sequence includes 16,256 protein-coding genes that on average contain 7 introns per gene. Genome analysis has revealed the presence of an extensive set of light-harvesting and pigment biosynthesis genes that may explain brown algal fitness in tidal environments. Comparative genomics also revealed the emergence of a rich as-

sortment of signal transduction genes including a group of membrane-spanning receptor kinases in brown algae since divergence from diatoms [45]. Interestingly, such receptor kinases play key roles in developmental processes including cellular patterning [63] in the multicellular animal and green plant lineages in which they evolved independently [64, 65]. The evolution of membrane-spanning receptor kinase in at least three of the five groups that have evolved complex multicellularity may therefore reflect a convergent evolution towards developmental sophistication [45].

Regulation of gene expression also seems to have evolved independently in each stramenopile lineage. A recent comparative study of transcription factors (TFs) in sequenced stramenopile genomes found that the Heat Shock Factor (HSF) family of TFs has massively expanded in the diatom lineage, where HSF genes account for about 35% of the TFs in both centric and pennate diatoms. The second and third largest families in diatoms are Myb and C2H2-Zn finger families of TFs, which are the most abundant TF classes in *E. siliculosus* and other stramenopiles [66].

## THE MAS JUNK

Transposable elements (TEs) are mobile genetic sequences that represent a variable fraction of most eukaryotic genomes such as 3% in baker's yeast, 45% in our own, and over 80% in maize. TEs are thought to be important contributors to genome evolution by inserting in genes or genetic regulatory elements, thereby disrupting gene function or altering expression levels, triggering chromosomal rearrangements, and influencing the physical genome size. In order to moderate the impact of TEs on populations, genomes have evolved specific epigenetic defense strategies that limit their activity. Indeed, in most eukaryotic genomes, TEs and other repeat-like sequences show characteristic epigenetic features that commonly involve RNAi-mediated DNA methylation and specific histone modifications such as di-methylation of H3(K9) ([67]; Rivarola *et al.*, unpublished results). Therefore, the identification and mapping of repeat sequences is a crucial step towards understanding the epigenetic landscape of a genome.

There are two main approaches to identify TEs in a genome: sequence similarity-based methods that scan genomic sequences for known elements or TE-characteristic structural features, and *de novo* methods that identify overrepresented sequences in a genome.

Several sequence similarity-based approaches have recently been applied to diatom genomes with the aid of programs that detect sequence repeats (i.e. RepeatMasker program [A.F.A. Smit, R. Hubley & P. Green RepeatMasker at <http://repeatmasker.org>]) in order to map the matching sequences on diatom genomes and to assess the abundance of each class of TE [43, 68]. It appears from these analyses that the TEs identified cover 6.4 and 1.9% of the *P. tricornutum* and *T. pseudonana* genomes, respectively, being long terminal repeat-retrotransposons (LTR-RT) the most abundant ones, accounting for 90% and 58% of the respective TE sequences. Diatom-specific *Copia*-type LTR-retrotransposons (LTR-RT) called CoDis have been significantly amplified in *P. tricornutum* with respect to *T. pseudonana*, constituting 5.8 and 1% of the respective genomes [43, 68]. Some copies

of Gypsy-like LTR-RT were also found in *T. pseudonana* but not in *P. tricornutum*. However, Gypsy-like LTR-RT are not thought to be absent from all pennate genomes as corresponding sequences were found in cDNA libraries prepared from the pennate species *Pseudonitzschia multistriata* and *Pseudonitzschia multiseriata*. Two thirds of the remaining TEs from *T. pseudonana* are composed by Harbinger-related DNA transposons that were not found in *P. tricornutum*.

The potential impact of CoDis on genome dynamics has been studied in more depth. Interestingly, transcription of two CoDi elements from *P. tricornutum* called *Blackbeard* and *Surcouf* was activated in response to nitrate starvation and to treatment with diatom-produced aldehydes, respectively, which constitute two ecologically relevant conditions [68]. The recent activity of CoDis in *P. tricornutum* species was demonstrated by revealing insertion polymorphism across several ecotypes. Furthermore, genomic features suggested the presence of TE-mediated recombination events that may have played a role in the evolution of gene families [66, 68]. Few additional repeats including potential TEs were identified using *de novo* approaches, suggesting that yet unclassified repeats constitute a limited fraction of diatom repeat content (Maumus *et al.*, unpublished). Repetitive element analysis in larger diatom genomes may reveal the presence of other types of TEs that were present in a common ancestor and may have been lost in *P. tricornutum* and *T. pseudonana* as a consequence of genome reduction.

In the case of the *E. siliculosus* genome, a pipeline using a combination of multiple *de novo* and similarity-based programs [69, 70] and aiming at identifying and annotating repeated sequences in whole genome sequences was applied [45]. Manual curation confirmed the presence of different types of LTR-RT including *Ty1/copia*, *Ty3/gypsy*, and *DIRS/Ngaro*-like elements, as well as non-LTR retrotransposons (LINEs), and non-autonomous retrotransposable elements (TRIMs/LARDs). In addition, the presence of a novel type of LTR-RT dubbed large GAG-related elements (LGA) was reported [45]. LGAs present typical LTR-RT structural features and a ~400 aa open reading frame with a predicted C-terminal zinc finger that shares weak similarity with those from retroviral proteins. It remains to be determined if these proteins enable LGA elements to transpose autonomously. DNA transposable elements were also found including *Harbinger*, *JERKY*, *POGO*-like, and *Helitrons*. Altogether, the classified TE sequences were found to cover ~12.5 % of the *E. siliculosus* genome comprising ~9.4% from retrotransposable elements. In addition many unclassified repeats were identified and found to compose ~9.9% of the genome. Interestingly, one of those, denoted as “*Sower*”, was found to be the most abundant repeated sequence covering as much as 1.2% of the genome (~2.5 Mb) [45].

## DNA METHYLATION IN MAS

DNA (cytosine-5) methyltransferases (C5-MTases) are the enzymes that catalyze the transfer of a methyl group from S-adenosyl-L-methionine to the C5 position of cytosine residues in DNA. Functionally, C5-MTases are classified following two types of activities, namely “maintenance” and “*de novo*” methylation [71, 72]. Maintenance methylation occurs after DNA replication on hemimethylated symmetric

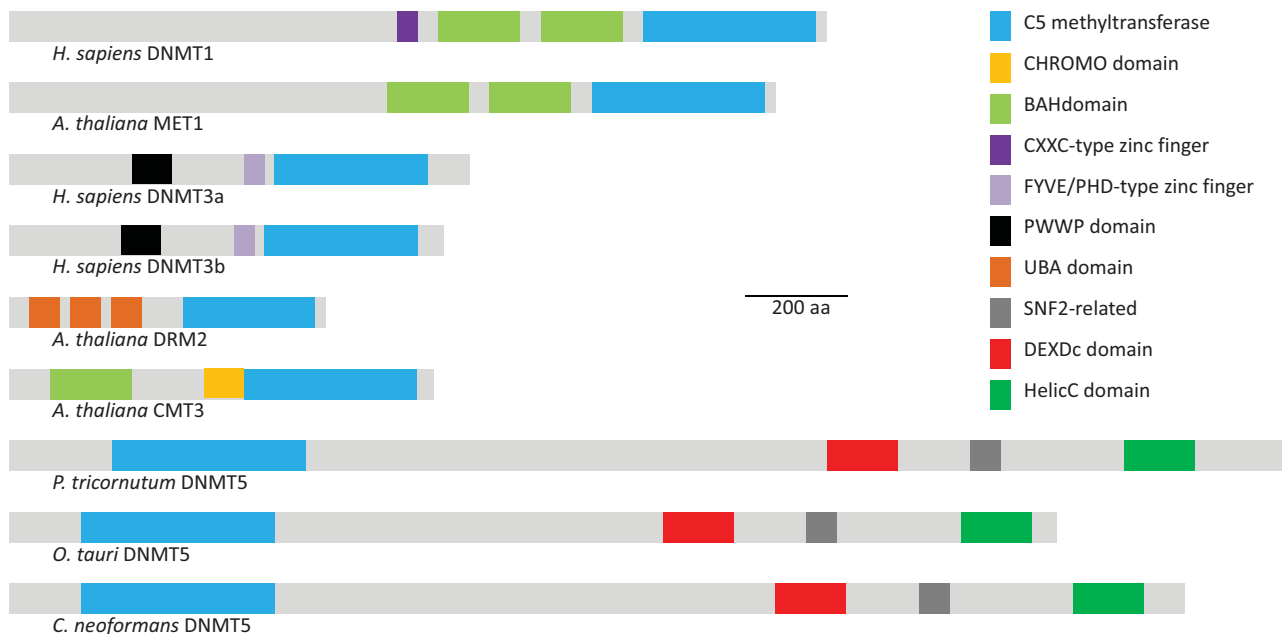
motifs (CpG and CpNpG), whereas *de novo* methylation occurs at previously unmethylated cytosines (in symmetrical or asymmetrical sites). Eukaryotic C5-MTases can be grouped into several classes based on sequence similarity within their C-terminal catalytic domains and are mostly represented by the mammalian DNMT1, DNMT2, and DNMT3a/3b, and the chromomethylases (CMTs) which are specific to flowering plants [73]. In addition to the catalytic domain that contains the highly conserved C5-MTase motifs, these enzymes can contain extra domains which are believed to define the structural basis for the differences in biological functions (Fig. 3). On the other side, DNMT2 homologs, which are present in most eukaryotes, were shown to methylate tRNA instead of DNA [74]. In addition, other types of C5-MTases such as DNMT4, DNMT5, and DNMT6 have been described in specific eukaryotic groups [75].

Both the *T. pseudonana* and the *P. tricornutum* genomes contain four gene models with C5-MTase domain (Maumus *et al.*, unpublished). Predicted proteins showing sequence similarity to members of the DNMT2 (Thaps3|22139 and Phatr2|16674) and DNMT3 (Thaps3|9575 and Phatr2|46156) gene families were detected in each genome (Fig. 4). Also, as in most eukaryotes apart from plants and metazoans [75], DNMT1 homologs were not found in diatom genomes. Interestingly, both diatom genomes contain predicted proteins similar to members of the DNMT5 family (Thaps3|3158 and Phatr2|45071-45072), commonly found in ascomycete and basidiomycete fungi [75], as well as in the Pelagophyceae *Aureococcus anophagefferens* and the Prasinophyceae (green algae) *Ostreococcus tauri*, *Ostreococcus lucimarinus*, and *Micromonas pusilla* (Maumus *et al.*, unpublished) (Fig. 4). In contrast to other classes of eukaryotic DNMTs that present an extension in their N-terminal region, DNMT5 proteins contain a SNF2-type DEXDc/HELICc helicase domain at the C-terminus (Fig. 3). Phylogenetic reconstruction and taxonomic distribution suggest that extant DNMT5-type pro-

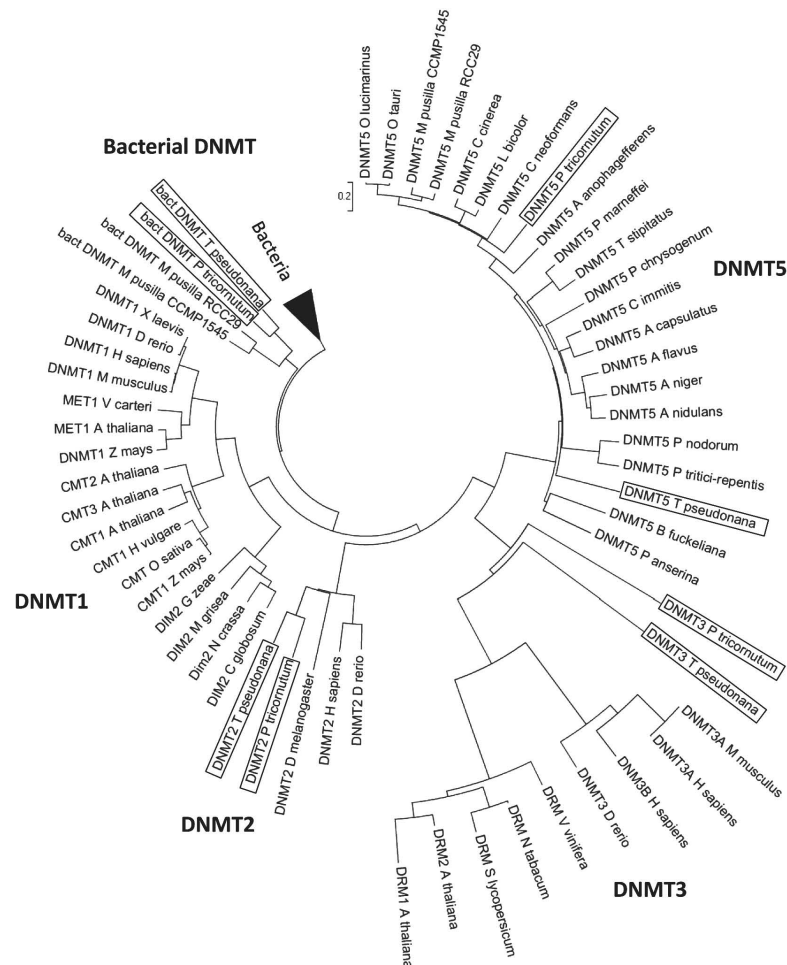
teins were present in a common eukaryotic ancestor. Interestingly, members of the SWI2/SNF2 family of ATP-dependent chromatin remodeling factors are necessary for proper cytosine methylation in plants and animals [76, 77] and they interact with DNMTs to repress transcription [78]. The direct involvement of certain SWI2/SNF2 helicases in DNA methylation processes and the conservation of their domain composition across eukaryotic groups suggest that DNMT5-like C5-MTases play a conserved and specific role in DNA methylation.

Phylogenetic analysis revealed that, in both genomes, a fourth predicted gene with C5-MTase domain appears to have evolved more recently from bacterial C5-MTases. Interestingly, these proteins are similar to sequences from *Micromonas* species (Fig. 4) (Maumus *et al.*, unpublished). Such evolutionary relationships suggest that these proteins have been transferred from bacteria to eukaryotes and then between eukaryotes. Their presence in both diatoms suggests a common ancestry and their absence in the remainder of the green lineage argues in favor of a transfer from extant diatoms towards extant prasinophytes.

The occurrence of DNA methylation in diatoms was first assessed by Jarvis and collaborators who used HPLC to analyze DNA nucleoside composition in several species of microalgae [79]. The authors described low levels of methylation in the DNA from two *P. tricornutum* ecotypes studied for which they reported a molar percentage for C5-methylcytosine of 0.11 and 0.16%, respectively. In the same study, DNA from other diatoms was shown to contain a higher 5mC molar percentage, up to 2.23% for the centric *Cyclotella cryptica* [79]. In contrast, preliminary genome-wide DNA methylation analysis of *P. tricornutum* using BS-seq showed that ~2.5% of all CpG motifs were methylated (Rivarola *et al.*, unpublished). It will be interesting to re-investigate global DNA methylation levels and patterns in the genomes of more microalgae using the BS-seq method.



**Fig. (3). DNMTs domains.** Schematic representation of the conserved domains found in various types of DNMTs including the DNMT5 family.



**Fig. (4). DNMTs phylogenetic reconstruction.** Bootstrap-supported phylogenetic tree constructed with the Neighbor-Joining method (1000 iterations) showing the relationships between the DNA methyltransferase domains of selected DNMT families. Bacterial DNMTs, indicated by a black dashed curve, come from the following species: *Bacillus subtilis*, *Clostridium cellulolyticum* H10, *Clostridium kluyveri* DSM 555, *Moraxella* sp., *Legionella pneumophila* str. Paris, *Legionella pneumophila* str. Lens, *Shewanella* sp. MR-4, *Roseobacter litoralis* Och 149, *Neisseria meningitidis* MC58, *Sulfurimonas denitrificans* DSM 1251, *Bacteroides fragilis* YCH46, *Microcystis aeruginosa* PCC 7806, *Microcystis aeruginosa* NIES-843, *Dactylococcopsis salina*, *Cyanobacteria* sp. ATCC 51142, *Collinsella stercoris* DSM 13279, and *Mycoplasma arthritidis* 158L3-1. Fungal DNMT5 proteins, indicated by a grey dashed curve, come from the following species: *Ajellomyces capsulatus* NAM1, *Aspergillus flavus* NRRL3357, *Aspergillus nidulans* FGSC A4, *Aspergillus niger*, *Botryotinia fuckeliana* B05.10, *Coccidioides immitis* RS, *Coprinopsis cinerea* okayama7 (#130), *Cryptococcus neoformans* var. *neoformans* B3501A, *Laccaria bicolor* S238NH82, *Penicillium chrysogenum* Wisconsin 541255, *Penicillium marneffeii* ATCC 18224, *Phaeosphaeria nodorum* SN15, *Podospora anserina*, *Pyrenophora tritici-repentis* Pt1CBFP, and *Talaromyces stipitatus* ATCC 10500. The tree was drawn using the SplitsTree4 software [107].

Indeed, although repetitive DNA is ubiquitously methylated in most eukaryotes, 5mC has also been observed in protein-coding genes in the DNA of various eukaryotic groups [25, 26], although its function and regulation remains insufficiently understood. In this line, preliminary analysis of results from an experiment aiming at unveiling genome-wide DNA methylation pattern in the *P. tricornutum* genome followed by experimental validation suggests that a few genes are densely methylated in the genome (Maurus *et al.* and Rivarola *et al.*, unpublished). Furthermore, recent use of the Mcr-PCR method to address the levels of DNA methylation in the *P. tricornutum* genome at specific loci also showed that most TEs were highly methylated in the *P. tricornutum* genome [68]. Unexpectedly, hypomethylation of the retrotransposon *Blackbeard* was detected and accompanied by its transcriptional activation in response to nitrate starvation

[68]. It remains to be determined whether this observation was due to a failure to maintain DNA methylation after subsequent cell divisions or to an active demethylation mechanism acting at this locus.

In strike contrast, no putative DNA C5-MTase could be detected in the *E. siliculosus* genome where only a DNMT2 homolog was found therefore suggesting that this organism may lack cytosine methylation in nuclear DNA completely. Indeed, HPLC analysis as well as Mcr-PCR experiments targeting TE loci failed to detect 5mC in this genome [45]. Therefore, TE silencing mechanisms in *E. siliculosus* are likely to be primarily sustained by other processes including histone modifications. It will be interesting to decipher whether the lack of 5mC is a brown algal-shared feature or if Ectocarpales abandoned this process later in evolution and what alternative strategies were evolved.



## HISTONE MODIFICATIONS IN MAS

There are over 60 different residues on histones where modifications have been detected either by specific antibodies or by mass spectrometry [17]. The characterization of the enzymes catalyzing histone modifications has been the focus of intense research over the last 10 years. The enzymes that are responsible for eight different types of histone modifications including acetylation, methylation, phosphorylation, and ubiquitylation have been identified.

In order to provide a view of the possible histone modifications occurring in MAS genomes, we have searched for homologs of histone modifiers among the respective sets of predicted proteins. A rich repertoire of proteins with significant similarities to known histone modifiers was found in the three sequenced MAS genomes (Table 2), such as putative homologs of the lysine acyltransferases (KAT) KAT1 and KAT2 [80, 81], CBP/p300 [82], MYST [83], and ELP3 (elongator complex protein 3) [84]. Putative class I, II, and III histone deacetylases (HDACs) such as RPD3 were also found in these genomes. Thus, dynamic acetylation/deacetylation of histones is likely to occur in these species. Preliminary data from ChIP-seq experiments on *P. tricornutum* utilizing specific antibodies to different modified histones under low and normal nitrogen conditions showed the expected dynamic nature of histone modifications (Rivarola *et al.*, unpublished results).

Various putative lysine methyltransferases were also identified in these species including candidates for ASH1 (absent, small, or homeotic 1), and SETD2 homologs, as well as MLL-like and SMYD-like proteins which appear to be more abundant in *P. tricornutum* and *E. siliculosus*, respectively. Other types of putative histone lysine methyltransferases were found to be species-specific. For example, putative homologs of SETD1, SETDB1, EHMT2 (Euchromatic Histone-lysine N-Methyltransferase 2) [85], and SETMAR proteins were only found in the *E. siliculosus* genome. In addition, the brown algal genome also encodes two Trithorax-like proteins as well as a group of three TRX-related proteins, whose C-terminus displays high similarity to the SET domain of the TRX candidates. In *P. tricornutum*, we also found a gene model encoding a protein containing a SET domain and a Jumonji C (JmjC) domain. The existence of transcripts encoding proteins containing both domains was confirmed by PCR on *P. tricornutum* cDNA (data not shown). The presence of such a multi-domain protein is of particular interest because it links two antagonistic activities: SET which is known to mediate lysine methylation, and JmjC that mediates lysine demethylation. This putative protein may have a dual activity, perhaps on different residues of the same histone. Intriguingly, such a SET-JmjC fusion is also found within the fungus *Neurospora crassa* but in the opposite configuration (JmjC-SET).

In the two diatoms, the histone lysine demethylase JmjC domain was also found in two combinations with other domains. One putative protein contains JmjC and MBT (Malignant Brain Tumor) domains. Interestingly, the MBT domain has been reported to bind mono-methylated H3 (K4) and different degrees of methylation on H4 (K20) depending on the MBT domain tested [86]. Furthermore, it was recently shown that the *Drosophila* L3MBTL1 protein with MBT

domain binds at least two nucleosomes, thus compacting nucleosomal arrays depending upon mono- and dimethylation of H4 (K20) [87]. Another domain found with JmjC in the same putative protein was the CHROMO domain. Along this line, CHROMO domain has been shown to interact with mono-, di-, and trimethylated H3 (K9) [88]. Furthermore, it was recently shown in fission yeast that the CHROMO domain-containing protein Swi6/HP1 interacts with the JmjC-domain-only protein Epe1 [89]. The function of MBT and CHROMO domains may be to address the histone demethylase JmjC activity to nucleosomes harboring specific histone marks. In addition, homologs of the JmjC-JmjN-containing protein JMJD2 were found in the three MAS genomes. Two other putative histone demethylase, LSD1 and FBXL homologs, were detected in the *P. tricornutum* and the *E. siliculosus* genomes. The presence of such a variety of histone methylases and demethylases in MAS genomes reinforces the dynamic histone methylation/demethylation that occurs on multiple lysine residues, as evident from our preliminary data from ChIP-seq experiments. In addition, we found in the three MAS genomes, genes encoding putative arginine methyltransferases, including various classes of PRMTs (protein arginine methyltransferase) that may act on different H3 and H4 arginine residues. Again, we also identified proteins with homology to the histone arginine demethylase JMJD6 [90] in all three genomes. The *in vivo* occurrence of modified histones was verified by Western blot and as mentioned above, by Chip-seq experiments using *P. tricornutum*. (Maumus *et al.* and Rivarola *et al.*, unpublished results).

## RNA INTERFERENCE IN MAS

Small (20-30 nt) RNAs are fundamental, sequence-specific regulatory elements in eukaryotes. They are processed from double-stranded RNA (dsRNA) with perfect or near-perfect complementarity. dsRNA can have a multitude of origins that can vary across organisms. For example in plants, dsRNA can come from the secondary structure of mRNA containing inverted repeats generating a hairpin or a stem-loop (miRNA) [91, 20], the annealing of complementary transcripts (siRNA) [92], or RNA-directed RNA polymerization (RdRP) using as template AGO-sliced mRNA including trans-acting siRNA transcripts [93]. Other sources of dsRNA are protein-coding transcripts (secondary siRNAs) [94], or RdRP-synthesized RNAs using RNA polymerase IV transcripts as templates (repeat-associated siRNA) [95]. Such dsRNA can be cleaved into 20-30 nt dsRNA by RNase III-like enzymes (typically Dicer-like (DCL) or Drosha-like) that can vary according to the type of dsRNA substrate. One strand (called the guide strand) is then loaded onto an Argonaute-like (AGO) protein that acts in an RNA-induced silencing complex (RISC) to ensure sequence-specific mRNA cleavage or translational inhibition, or in other complexes that act in the nucleus to direct sequence-specific cytosine methylation and histone modifications. In addition to the AGO/RNase III-associated pathways, another small RNA silencing system, called Piwi-associated RNAs (piRNAs), has been shown to prevent the amplification of selfish genetic elements in animals [94].

Although the transformation of diatom nuclear genomes is now routinely achieved [52], the lack of tools to generate

Table 2. Putative Histone Modifiers Found in MAS Genomes

Histone Modifiers	Residues Modified	Homologs in <i>P. tricornutum</i>	Homologs in <i>T. pseudonana</i>	Homologs in <i>E. siliculosus</i>
<b>Lysine Acetyltransferases (KATs)</b>				
HAT1 (KAT1)	H4 (K5, K12)	54343	1397, 22580	Esi0002_0151
GCN5 (KAT2)	H3 (K9, K14, K18, K23, K36)	46915	15161	Esi0090_0053
Nejire (KAT3); CBP/p300 (KAT3A/B)	H3 (K14, K18, K56) H4 (K5, K8); H2A (K5) H2B (K12, K15)	45703, 45764, 54505	24331, 269496, 263785	Esi0053_0101
MYST1 (KAT8)	H4 (K16)	24733, 24393	37928, 36275	Esi0090_0053, Esi0084_0032
ELP3 (KAT9)	H3	50848	9040	Esi0264_0014
<b>Lysine Deacetylases (HDACs)</b>				
RPD3 (Class I HDACS)	H2, H3, H4	51026, 49800	41025, 32098, 261393	Esi0147_0031, Esi0181_0051, Esi0092_0077
HDA1 (Class II HDACS)	H2, H3, H4	45906, 50482, 35869	268655, 269060, 3235, 15819	Esi0168_0016, Esi0157_0055, Esi0157_0057, Esi0040_0047
NAD <sup>+</sup> dependant (Class III HDACS)	H4 (K16)	52135, 45850, 24866, 45909, 52718, 21543, 39523	269475, 264809, 16405, 35693, 264494, 16384, 35956	Esi0054_0036, Esi0014_0023, Esi0026_0040
<b>Lysine Methyltransferases</b>				
MLL	H3 (K4)	40183, 54436, 42693, 47328, 49473, 49476, 44935	35182, 35531, 22757	Esi0069_0089
ASH1/WHSC1	H3 (K4)	43275	264323	Esi0016_0129, Esi0000_0259
SETD1	H3 (K36), H4 (K20)	not found	not found	Esi0070_0045, Esi0020_0054, Esi0043_0074
SETD2	H3 (K36)	50375	35510	Esi0028_0153
SETDB1	H3 (K9)	not found	not found	Esi0162_0064
SETMAR	H3 (K4, K36)	not found	not found	Esi0100_0024
SMYD	H3 (K4)	bd1647, 43708	23831, 24988	Esi0013_0081, Esi0189_0013, Esi0189_0015, Esi0286_0007, Esi0015_0158
TRX-related		not found	not found	Esi0115_0076, Esi0076_0098, Esi0043_0074, Esi0094_0082, Esi0455_0010
E(Z)	H3 (K9, K27)	32817	268872	not found
EHMT2	H3 (K9, K27)	not found	not found	Esi0453_0004
SET+JmjC	Unknown	bd1647	not found	not found
<b>Lysine Demethylases (KDM)</b>				
LSD1 (KDM1)	H3 (K4, K9)	51708, 44106, 48603	not found	Esi0073_0073, Esi0060_0020, Esi0014_0196
FBXL (KDM2)	H3 (K36)	42595	not found	Esi0025_0159
JMJD2 (KDM4)/JARID	H3 (K9, K36)	48747	2137	Esi0145_0049, Esi0014_0020, Esi0084_0076
JMJ-MBT	Unknown	48109	22122	not found
JMJ-CHROMO	Unknown	Pt1-40322	1863	not found

(Table 2). Contd.....

Histone Modifiers	Residues Modified	Homologs in <i>P. tricornutum</i>	Homologs in <i>T. pseudonana</i>	Homologs in <i>E. siliculosus</i>
<b>Arginine Methyltransferases</b>				
CARM1 (PRMT4)	H3 (R17)	17184	3690, 28185	Esi0137_0007, Esi0000_0236, Esi0165_0056
PRMT5	H3 (R8), H4 (R3)	49565	24429	Esi0006_0115
PRMT6	H3 (R2)	54710	bd1828	Esi0153_0004, Esi0000_0300
PRMT7	H4 (R3)	49245	25810	Esi0133_0084, Esi0122_0102
<b>Arginine demethylases</b>				
JMJD6	H3 (R2), H4 (R3)	3251, 35989, 46234	24024	Esi0055_0014, Esi0131_0061, Esi0189_0080, Esi0055_0061

Proteins with putative histone modifier function were identified in the sets of predicted proteins from the MAS genomes by reverse BLAST comparison with enzymes known to catalyze the modification of specific histone residues (as indicated) and by searching for corresponding catalytic InterPro domains in MAS annotation databases. Candidate proteins are indicated by their identification number.

specific knock-out or knock-down mutants has remained a major limitation for functional genomics studies in diatoms. The RNA interference pathway is often exploited to down-regulate target genes [96]. However, because published diatom genomes do not encode proteins with extensive similarities to Dicer-like proteins [97], it remained obscure whether such a machinery exists in diatoms until it was proven experimentally [54]. In this work, a *P. tricornutum* transgenic line that expressed the GUS reporter gene under a strong promoter was transformed with anti-sense or inverted repeat sequences of the GUS gene and resulted in a significant down-regulation of GUS in several clones for both types of constructs, thus demonstrating the existence of an RNAi-based silencing system in *P. tricornutum*. It is however unclear whether down-regulation of the GUS gene occurs through transcript cleavage, translational inhibition or RdDM processes. Interestingly, Mcr-PCR analysis of the resulting down-regulated lines revealed the presence of DNA methylation in the GUS gene. Furthermore, bisulfite-sequencing uncovered a high density of 5mC in the targeted region of the GUS gene as well as a discrete 5mc distribution up to the promoter [54]. This suggests that in this case, siRNA-mediated down-regulation of the GUS gene involves at least an RdDM process and that *de novo* DNA methylation is able to spread from 5mC-rich loci in *P. tricornutum* following a yet to be determined mechanism. Also, it remains to be clarified whether other siRNA-mediated processes are involved in the down-regulation of the GUS gene. Interestingly, this technique has also been proved to be efficient to down-regulate the expression of endogenous genes [54, 98] and thus constitutes a robust tool for reverse genetics in diatoms. Additional experiments will also be required to explore the existence of other silencing mechanisms in diatoms.

On the basis of sequence similarity with known core proteins of the eukaryotic RNAi machinery, we identified several predicted proteins in diatom genomes that may be involved in RNAi-based mechanisms. First, similar to what was observed by Cerruti & Casas-Mollano [97] for *T. pseudonana*, homologs of the canonical Dicer proteins could not be found in either of the published diatom genomes. However, although canonical DCLs are large proteins that typi-

cally contain an N-terminal DEADc/HELICASEc domain followed by a domain of unknown function (DUF283), a PAZ domain, two neighboring RNase III domains (RNase IIIa and RNase IIIb), and a double-stranded RNA-binding domain (dsRBD), in some eukaryotes and especially protists, less complex RNase IIIa-IIIb- containing proteins were found to be capable of slicer activity. For instance, DCL1 from the ciliate *Tetrahymena thermophila* (alveolate) contains only the two RNase III domains associated to a dsRBD domain [99]. In *Giardia intestinalis* (excavate) DCL1, the two RNase III domains associate with a PAZ and a DUF283 domain [100]. Also, DCL1 from the excavate *Trypanosoma brucei*, which consists of only two adjacent RNase III domains, is sufficient to drive the RNA interference pathway [101]. In fact, although they have evolved from a different branch of ancestral RNase III, RNase III-like proteins (RTLs) are also found in the genomes of “higher” eukaryotes including plants where they may also play a role in small RNA biogenesis in addition to the typical DCLs [102]. Such RTL proteins with putative *dicing* activity were identified in *P. tricornutum* and *T. pseudonana*: in the former, a gene model encodes a protein with dsRBD followed by two RNase III domains, and in the latter, a predicted protein with only two RNase III domains. For both proteins, the conservation of key amino acid residues supports their role as bona fide dsRNA dicers. Interestingly, *T. pseudonana* also encodes a protein with a DEADc/HELICASEc domain and a C-terminal dsRBD [54]. A protein with such domain composition could not be identified in other organisms and it is tempting to speculate a role in the RNAi machinery. In addition, one AGO-like protein with canonical domain composition has been identified in both published diatom genomes but Piwi-like proteins could not be found. Nevertheless, proteins with similarity to RdRPs were found: one in *P. tricornutum* and two in *T. pseudonana* [54]. However, although quite a similar set of putative core RNAi proteins were found in both the pennate and centric diatoms, the pennate diatom *Fragilariopsis cylindrus* appears to possess quite a diversified complement of such proteins (Mock *et al.*, unpublished).

In contrast to diatoms, the *E. siliculosus* genome was found to encode a canonical Dicer-like protein as well as one AGO and two RdRP homologs [45]. The *E. siliculosus* ge-

nome project also performed deep sequencing of small RNAs in this species. When small RNA sequences were mapped on the genome it was found that they were distributed primarily on rRNA sequences (45%), followed by intergenic regions (26%), TEs and other repeated sequences (13%) and introns (9%). Size analysis revealed enrichment in 21 nt species for the small RNAs that map on intergenic regions, repeated sequences, introns and exons [45]. Interestingly, there was a statistically significant link between small RNAs and repeated sequences suggesting the existence of chromatin-level small RNA-directed histone modifications in this species that controls TE amplification. *E. siliculosus* small RNA data were successfully combined with secondary structure prediction algorithms to screen the genome for putative miRNA genes where 26 miRNA sequences defining 21 families were identified. Most of the mature miRNA sequences are 20 nt long and begin with a U which is a well known characteristic of miRNA loaded on the AGO1 protein in plants. Target prediction revealed 71 potential transcript targets for 12 out of the 26 miRNAs. Interestingly, most (75%) of the predicted targets contain leucine-rich repeat domains [45].

## DISCUSSION AND FUTURE PERSPECTIVES

Diatom genomes are well on their way. These organisms, long studied by marine biologists for their morphology, phylogeny, and ecological success, are now providing valuable information into epigenetic processes such as transposon behavior in genomes, and the evolutionary history of eukaryotes. The recent sequencing of the three MAS species (plus a few more soon to be finished) 'have' and 'are' providing a valuable resource from which to study different biological phenomena previously unreachable, opening up a whole new arena of possibilities. Scientists can now tackle fundamental questions of biology in a genome-wide manner with many 're-sequencing' techniques, that use the reference genome to aid in mapping of the reads. The ability to perform ChIP-seq assays with commercial antibodies that detect a vast array of different histone modifications and BS-seq experiments provide a clear quantitative landscape of the particular chromatin modification in study. On the other hand, if this is coupled with RNA-seq, which provides an approximate snapshot of gene expression levels, a better understanding of how epigenetics can alter gene expression can be obtained.

To date, several diatom genomes are close to being completed: *Pseudo-nitzschia multiseries* (~250 Mb genome size) and *Fragilariopsis cylindrus* (~81 Mb genome size) are being sequenced by the Joint Genome Institute (USA: <http://www.jgi.doe.gov/genome-projects>), and *Amphora sp.* CCMP2378 and *Atheya sp.* CCMP212 are being sequenced by J. Raymond, U. Nevada, USA (personal communication). Also, the genome of the pelagophyte *Aureococcus anophagefferens* sequenced at JGI has been published [103] meanwhile the resubmission of this manuscript and provides the first complete genomic data from the MAS clade *Pelagophyceae*.

This genomic revolution has provided solid ground to ask some exciting questions, some of which are being addressed in our labs. One such example is determining the gene expression and epigenetic landscapes in different MAS species

and the effect of environmental changes at a genome-wide level. Moreover, one can study the epigenetic modification stability after an environmental cue by examining the cells once they return to their normal condition (e.g. nitrogen-deficient medium, which is known to alter DNA methylation and transcription of certain transposons). Besides nutrient conditions, a very exciting area is to analyze the potential epigenetic remodeling that accompanies morphotype transition of *P. tricornutum* cells such as the switch in morphology from 'fusiform' to 'round' cell shape during biofilm formation [104]. The round cell morphotype is of particular interest because it may represent a cyst-like resting stage. Comparing the epigenomes of the three different *P. tricornutum* morphotypes may pinpoint different epigenetic profiles at specific genes that cause morphotype-specific expression levels and networks involved the regulation of morphological and metabolic processes [105]. These experiments will provide insight into how epigenetic changes influence gene expression and how environmental cues act upon chromatin. Moreover, the sequencing of *Pseudo-nitzschia multiseries*, another pennate diatom whose genome is 250 Mb (approx. 10 times larger than *P. tricornutum*) will provide a great system to compare the epigenomes of these two diatoms.

Another interesting area of biology that can be assessed using MAS species is DNA replication. How origins of DNA replication in eukaryotes are specified is a fundamental biological question that remains unresolved. No DNA consensus sequence that allows initiation of DNA replication has been identified in eukaryotes, with the exception of *Saccharomyces cerevisiae*. Therefore, epigenetic and/or functional elements may be responsible for origin specification. Several studies have correlated origins of replication with transcription or epigenetic marks. Interestingly, transposable elements have also been associated with the presence or timing of origins of replication. The unique characteristics of MAS species, with contrasting DNA methylation and histone modification profiles, represent unicellular experimental systems that can be synchronized to study DNA replication origins. The nature of centromeres and kinetochores also represents an exciting topic because of the unusual nature of mitosis in MAS organisms such as diatoms [106].

Moreover, with the upcoming 3rd generation of DNA sequencers such as from PacBio (<http://www.pacificbiosciences.com>), that will provide up to several Kb read lengths, assembling new diatom genomes will be faster, more accurate, and hopefully less expensive. A safe assumption is that in the near future tons of data from marine organisms will be produced and the challenge to analyze all that information will be daunting. The unique characteristics of the MAS species and all the 'omics' data that will come make it a great time to be in the marine world.

## ACKNOWLEDGEMENTS

We would like to thank Delphine Scornet, Nils Kröger, and Alessandra De Martino who kindly allowed us to illustrate the sequenced MAS species using pictures that they produced.

## REFERENCES

- [1] Cubas, P.; Vincent, C.; Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*, **1999**, *401*(6749), 157-161.



- [2] Rogakou, E. P.; Boon, C.; Redon, C.; Bonner, W. M. Megabase chromatin domains involved in DNA double-strand breaks *in vivo*. *J. Cell Biol.*, **1999**, 146(5), 905-916.
- [3] Luco, R. F.; Pan, Q.; Tominaga, K.; Blencowe, B. J.; Pereira-Smith, O. M.; Misteli, T. Regulation of alternative splicing by histone modifications. *Science*, **2010**, 327(5968), 996-1000.
- [4] Feng, S.; Jacobsen, S. E.; Reik, W. Epigenetic reprogramming in plant and animal development. *Science*, **2010**, 330(6004), 622-627.
- [5] Bird, A. Perceptions of epigenetics. *Nature*, **2007**, 447(7143), 396-398.
- [6] Teixeira, F. K.; Colot, V. Repeat elements and the Arabidopsis DNA methylation landscape. *Heredity*, **2010**, 105(1), 14-23.
- [7] Kohler, C.; Weinhofer-Molisch, I. Mechanisms and evolution of genomic imprinting in plants. *Heredity*, **2010**, 105(1), 57-63.
- [8] Barakat, T. S.; Jonkers, I.; Monkhorst, K.; Gribnau, J. X-changing information on X inactivation. *Exp. Cell Res.*, **2010**, 316(5), 679-687.
- [9] Wilkinson, L. S.; Davies, W.; Isles, A. R. Genomic imprinting effects on brain development and function. *Nat. Rev. Neurosci.*, **2007**, 8(11), 832-843.
- [10] Lister, R.; Pelizzola, M.; Dowen, R. H.; Hawkins, R. D.; Hon, G.; Tonti-Filippini, J.; Nery, J. R.; Lee, L.; Ye, Z.; Ngo, Q. M.; Edsall, L.; Antosiewicz-Bourget, J.; Stewart, R.; Ruotti, V.; Millar, A. H.; Thomson, J. A.; Ren, B.; Ecker, J. R. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **2009**, 462(7271), 315-322.
- [11] Cokus, S. J.; Feng, S.; Zhang, X.; Chen, Z.; Merriman, B.; Haudenschild, C. D.; Pradhan, S.; Nelson, S. F.; Pellegrini, M.; Jacobsen, S. E. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*, **2008**, 452(7184), 215-219.
- [12] Wassenegger, M.; Heimes, S.; Riedel, L.; Sanger, H. L. RNA-directed de novo methylation of genomic sequences in plants. *Cell*, **1994**, 76(3), 567-576.
- [13] Teixeira, F. K.; Heredia, F.; Sarazin, A.; Roudier, F.; Boccara, M.; Ciaudo, C.; Cruaud, C.; Poulain, J.; Berdasco, M.; Fraga, M. F.; Voinnet, O.; Wincker, P.; Esteller, M.; Colot, V. A role for RNAi in the selective correction of DNA methylation defects. *Science*, **2009**, 323(5921), 1600-1604.
- [14] Vance, V.; Vaucheret, H. RNA silencing in plants--defense and counterdefense. *Science*, **2001**, 292(5525), 2277-2280.
- [15] Ahmed, I.; Sarazin, A.; Bowler, C.; Colot, V.; Quesneville, H. Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. *Nucleic Acids Res.*, **2011**, [Epub ahead of print].
- [16] Pradhan, S.; Bacolla, A.; Wells, R. D.; Roberts, R. J. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J. Biol. Chem.*, **1999**, 274(46), 33002-33010.
- [17] Kouzarides, T. Chromatin modifications and their function. *Cell*, **2007**, 128(4), 693-705.
- [18] Roudier, F.; Teixeira, F. K.; Colot, V. Chromatin indexing in Arabidopsis: an epigenomic tale of tails and more. *Trends Genet.*, **2009**, 25(11), 511-517.
- [19] Corpet, A.; Almouzni, G. Making copies of chromatin: the challenge of nucleosomal organization and epigenetic information. *Trends Cell Biol.*, **2009**, 19(1), 29-41.
- [20] Voinnet, O. Origin, biogenesis, and activity of plant microRNAs. *Cell*, **2009**, 136(4), 669-687.
- [21] Carthew, R. W.; Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell*, **2009**, 136(4), 642-655.
- [22] Nowrousian, M. Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. *Eukaryot. Cell*, **2010**, 9(9), 1300-1310.
- [23] Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **2009**, 10(1), 57-63.
- [24] Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.*, **2009**, 10(10), 669-680.
- [25] Zemach, A.; McDaniel, I. E.; Silva, P.; Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, **2010**, 328(5980), 916-919.
- [26] Feng, S.; Cokus, S. J.; Zhang, X.; Chen, P. Y.; Bostick, M.; Goll, M. G.; Hetzel, J.; Jain, J.; Strauss, S. H.; Halpern, M. E.; Ukomadu, C.; Sadler, K. C.; Pradhan, S.; Pellegrini, M.; Jacobsen, S. E. Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. U S A*, **2010**, 107(19), 8689-8694.
- [27] Gibbs, S. P. The evolution of algal chloroplasts. In *Origins of plastids*, RA, L., Ed. Chapman and Hall, New York: **1993**; pp 107-121.
- [28] Burki, F.; Shalchian-Tabrizi, K.; Minge, M.; Skjaeveland, A.; Nikolaev, S. I.; Jakobsen, K. S.; Pawlowski, J. Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One*, **2007**, 2(8), e790.
- [29] Cavalier-Smith, T. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol.*, **1999**, 46(4), 347-366.
- [30] Yoon, H. S.; Hackett, J. D.; Ciniglia, C.; Pinto, G.; Bhattacharya, D. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.*, **2004**, 21(5), 809-818.
- [31] Hackett, J. D.; Yoon, H. S.; Li, S.; Reyes-Prieto, A.; Rummele, S. E.; Bhattacharya, D. Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Mol. Biol. Evol.*, **2007**, 24(8), 1702-1713.
- [32] Okamoto, N.; Chantangsai, C.; Horak, A.; Leander, B. S.; Keeling, P. J. Molecular phylogeny and description of the novel katablepharid *Roombia truncata* gen. et sp. nov., and establishment of the Hacrobia taxon nov. *PLoS One*, **2009**, 4(9), e7080.
- [33] Cavalier-Smith, T. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biol. Lett.*, **2010**, 6(3), 342-345.
- [34] Archibald, J. M. The puzzle of plastid evolution. *Curr. Biol.*, **2009**, 19(2), R81-8.
- [35] Moustafa, A.; Beszteri, B.; Maier, U. G.; Bowler, C.; Valentin, K.; Bhattacharya, D. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science*, **2009**, 324(5935), 1724-1726.
- [36] Round, F. E.; Crawford, R. M.; Mann, D. G. *The Diatoms: Biology and Morphology of the Genera*. Cambridge University Press: London, UK, **1990**.
- [37] Raven, J. A.; Waite, A. M. The evolution of silicification in diatoms: inescapable sinking and sinking as escape? *New Phytologist*, **2004**, 162, 45-61.
- [38] Nelson, D. M.; Treguer, P.; Brzezinski, M. A.; Leynaert, A.; Queguiner, B. Production and dissolution of biogenic silica in the ocean - revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochem. Cycles*, **1995**, 9, 359-372.
- [39] Falkowski, P. G.; Katz, M. E.; Knoll, A. H.; Quigg, A.; Raven, J. A.; Schofield, O.; Taylor, F. J. The evolution of modern eukaryotic phytoplankton. *Science*, **2004**, 305(5682), 354-360.
- [40] Bowler, C.; Vardi, A.; Allen, A. E. Oceanographic and biogeochemical insights from diatom genomes. *Ann. Rev. Mar. Sci.*, **2010**, 2, 333-365.
- [41] Kooistra, W. H.; Medlin, L. K. Evolution of the diatoms (Bacillariophyta). IV. A reconstruction of their age from small subunit rRNA coding regions and the fossil record. *Mol. Phylogenet. Evol.*, **1996**, 6(3), 391-407.
- [42] Armbrust, E. V.; Berges, J. A.; Bowler, C.; Green, B. R.; Martinez, D.; Putnam, N. H.; Zhou, S.; Allen, A. E.; Apt, K. E.; Bechner, M.; Brzezinski, M. A.; Chaal, B. K.; Chiovitti, A.; Davis, A. K.; Demarest, M. S.; Detter, J. C.; Glavina, T.; Goodstein, D.; Hadi, M. Z.; Hellsten, U.; Hildebrand, M.; Jenkins, B. D.; Jurka, J.; Kapitonov, V. V.; Kroger, N.; Lau, W. W.; Lane, T. W.; Larimer, F. W.; Lippmeier, J. C.; Lucas, S.; Medina, M.; Montsant, A.; Obornik, M.; Parker, M. S.; Palenik, B.; Pazour, G. J.; Richardson, P. M.; Rynearson, T. A.; Saito, M. A.; Schwartz, D. C.; Thamatrakoln, K.; Valentin, K.; Vardi, A.; Wilkerson, F. P.; Rokhsar, D. S. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*, **2004**, 306(5693), 79-86.
- [43] Bowler, C.; Allen, A. E.; Badger, J. H.; Grimwood, J.; Jabbari, K.; Kuo, A.; Maheswari, U.; Martens, C.; Maumus, F.; Otillar, R. P.; Rayko, E.; Salamov, A.; Vandepoele, K.; Beszteri, B.; Gruber, A.; Heijde, M.; Katinka, M.; Mock, T.; Valentin, K.; Verret, F.; Berges, J. A.; Brownlee, C.; Cadoret, J. P.; Chiovitti, A.; Choi, C. J.; Coesel, S.; De Martino, A.; Detter, J. C.; Durkin, C.; Falcitatore, A.; Fournet, J.; Haruta, M.; Huysman, M. J.; Jenkins, B. D.; Jiroutova, K.; Jorgensen, R. E.; Joubert, Y.; Kaplan, A.; Kroger, N.; Kroth, P. G.; La Roche, J.; Lindquist, E.; Lommer, M.; Martin-Jezequel, V.; Lopez, P. J.; Lucas, S.; Mangogna, M.; McGinnis, K.; Medlin, L. K.; Montsant, A.; Oudot-Le Secq, M. P.; Napoli, C.; Obornik, M.; Parker, M. S.; Petit, J. L.; Porcel, B. M.; Poulsen, N.; Robison, M.; Rychlewski, L.; Rynearson, T. A.; Schmutz, J.; Shapiro, H.; Siat, M.; Stanley, M.; Sussman, M. R.; Taylor, A. R.; Vardi, A.; von Dassow, P.; Vyverman, W.; Willis, A.; Wyrwicz, L.

- S.; Rokhsar, D. S.; Weissenbach, J.; Armbrust, E. V.; Green, B. R.; Van de Peer, Y.; Grigoriev, I. V. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature*, **2008**, 456(7219), 239-244.
- [44] Connolly, J. A.; Oliver, M. J.; Beaulieu, J. M.; Knight, C. A.; Tomaneck, L.; Moline, M. A. Correlated evolution of genome size and cell volume in diatoms (Bacillariophyceae). *J. Phycol.*, **2008**, 44(1), 124-131.
- [45] Cock, J. M.; Sterck, L.; Rouze, P.; Scornet, D.; Allen, A. E.; Amoutzias, G.; Anthouard, V.; Artiguenave, F.; Aury, J. M.; Badger, J. H.; Beszteri, B.; Billiau, K.; Bonnet, E.; Bothwell, J. H.; Bowler, C.; Boyen, C.; Brownlee, C.; Carrano, C. J.; Charrier, B.; Cho, G. Y.; Coelho, S. M.; Collen, J.; Corre, E.; Da Silva, C.; Delage, L.; Delarouge, N.; Dittami, S. M.; Doubeau, S.; Elias, M.; Farnham, G.; Gachon, C. M.; Gschloessl, B.; Heesch, S.; Jabbari, K.; Jubin, C.; Kawai, H.; Kimura, K.; Kloareg, B.; Kupper, F. C.; Lang, D.; Le Bail, A.; Leblanc, C.; Lerouge, P.; Lohr, M.; Lopez, P. J.; Martens, C.; Maumus, F.; Michel, G.; Miranda-Saavedra, D.; Morales, J.; Moreau, H.; Motomura, T.; Nagasato, C.; Napoli, C. A.; Nelson, D. R.; Nyvall-Collen, P.; Peters, A. F.; Pommier, C.; Potin, P.; Poulain, J.; Quesneville, H.; Read, B.; Rensing, S. A.; Ritter, A.; Rousvoal, S.; Samanta, M.; Samson, G.; Schroeder, D. C.; Segurens, B.; Strittmatter, M.; Tonon, T.; Tregear, J. W.; Valentin, K.; von Dassow, P.; Yamagishi, T.; Van de Peer, Y.; Wincker, P. The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature*, **2010**, 465(7298), 617-621.
- [46] Heesch, S.; Cho, G. Y.; Peters, A. F.; Le Corguille, G.; Falentin, C.; Boutet, G.; Coedel, S.; Jubin, C.; Samson, G.; Corre, E.; Coelho, S. M.; Cock, J. M. A sequence-tagged genetic map for the brown alga Ectocarpus siliculosus provides large-scale assembly of the genome sequence. *New Phytol.*, **2010**, 188(1), 42-51.
- [47] Maheswari, U.; Mock, T.; Armbrust, E. V.; Bowler, C. Update of the Diatom EST Database: a new tool for digital transcriptomics. *Nucleic Acids Res.*, **2009**, 37, (Database issue), D1001-5.
- [48] Allen, A. E.; Laroche, J.; Maheswari, U.; Lommer, M.; Schauer, N.; Lopez, P. J.; Finazzi, G.; Fernie, A. R.; Bowler, C. Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. *Proc. Natl. Acad. Sci. U S A*, **2008**, 105(30), 10438-10443.
- [49] Dittami, S. M.; Scornet, D.; Petit, J. L.; Segurens, B.; Da Silva, C.; Corre, E.; Dondrup, M.; Glatting, K. H.; Konig, R.; Sterck, L.; Rouze, P.; Van de Peer, Y.; Cock, J. M.; Boyen, C.; Tonon, T. Global expression analysis of the brown alga Ectocarpus siliculosus (Phaeophyceae) reveals large-scale reprogramming of the transcriptome in response to abiotic stress. *Genome Biol.*, **2009**, 10(6), R66.
- [50] Mock, T.; Samanta, M. P.; Iverson, V.; Berthiaume, C.; Robison, M.; Holtermann, K.; Durkin, C.; Bondurant, S. S.; Richmond, K.; Rodesch, M.; Kallas, T.; Huttlin, E. L.; Cerrina, F.; Sussman, M. R.; Armbrust, E. V. Whole-genome expression profiling of the marine diatom Thalassiosira pseudonana identifies genes involved in silicon bioprocesses. *Proc. Natl. Acad. Sci. U S A*, **2008**, 105(5), 1579-1584.
- [51] Apt, K. E.; Kroth-Pancic, P. G.; Grossman, A. R. Stable nuclear transformation of the diatom Phaeodactylum tricornutum. *Mol. Gen. Genet.*, **1996**, 252(5), 572-579.
- [52] Falcione, A.; Casotti, R.; Leblanc, C.; Abrescia, C.; Bowler, C. Transformation of Nonselectable Reporter Genes in Marine Diatoms. *Mar. Biotechnol. (NY)*, **1999**, 1(3), 239-251.
- [53] Siaut, M.; Heijde, M.; Mangogna, M.; Montsant, A.; Coesel, S.; Allen, A.; Manfredonia, A.; Falcione, A.; Bowler, C. Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. *Gene*, **2007**, 406(1-2), 23-35.
- [54] De Riso, V.; Raniello, R.; Maumus, F.; Rogato, A.; Bowler, C.; Falcione, A. Gene silencing in the marine diatom Phaeodactylum tricornutum. *Nucleic Acids Res.*, **2009**, 37(14), e96.
- [55] Peters, A. F.; Scornet, D.; Ratn, M.; Charrier, B.; Monnier, A.; Merrien, Y.; Corre, E.; Coelho, S. M.; Cock, J. M. Life-cycle-generation-specific developmental processes are modified in the immediate upright mutant of the brown alga Ectocarpus siliculosus. *Development*, **2008**, 135(8), 1503-1512.
- [56] Roy, S. W.; Penny, D. A very high fraction of unique intron positions in the intron-rich diatom Thalassiosira pseudonana indicates widespread intron gain. *Mol. Biol. Evol.*, **2007**, 24(7), 1447-1457.
- [57] Archibald, J. M. Nucleomorph genomes: structure, function, origin and evolution. *Bioessays*, **2007**, 29(4), 392-402.
- [58] Matsuzaki, M.; Misumi, O.; Shin, I. T.; Maruyama, S.; Takahara, M.; Miyagishima, S. Y.; Mori, T.; Nishida, K.; Yagisawa, F.; Yoshida, Y.; Nishimura, Y.; Nakao, S.; Kobayashi, T.; Momoyama, Y.; Higashiyama, T.; Minoda, A.; Sano, M.; Nomoto, H.; Oishi, K.; Hayashi, H.; Ohta, F.; Nishizaka, S.; Haga, S.; Miura, S.; Morishita, T.; Kabeya, Y.; Terasawa, K.; Suzuki, Y.; Ishii, Y.; Asakawa, S.; Takano, H.; Ohta, N.; Kuroiwa, H.; Tanaka, K.; Shimizu, N.; Sugano, S.; Sato, N.; Nozaki, H.; Ogasawara, N.; Kohara, Y.; Kuroiwa, T. Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature*, **2004**, 428(6983), 653-657.
- [59] Barbier, G.; Oesterheld, C.; Larson, M. D.; Halgren, R. G.; Wilkerson, C.; Garavito, R. M.; Benning, C.; Weber, A. P. Comparative genomics of two closely related unicellular thermo-acidophilic red algae, Galdieria sulphuraria and Cyanidioschyzon merolae, reveals the molecular basis of the metabolic flexibility of Galdieria sulphuraria and significant differences in carbohydrate metabolism of both algae. *Plant Physiol.*, **2005**, 137(2), 460-474.
- [60] Blouin, N. A.; Brodie, J. A.; Grossman, A. C.; Xu, P.; Brawley, S. H. Porphyra: a marine crop shaped by stress. *Trends Plant Sci.*, **2011**, 16(1), 29-37.
- [61] Gogarten, J. P.; Townsend, J. P. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.*, **2005**, 3(9), 679-687.
- [62] Keeling, P. J.; Palmer, J. D. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.*, **2008**, 9(8), 605-618.
- [63] De Smet, I.; Voss, U.; Jurgens, G.; Beeckman, T. Receptor-like kinases shape the plant. *Nat. Cell Biol.*, **2009**, 11(10), 1166-1173.
- [64] Cock, J. M.; Vanoosthuysen, V.; Gaude, T. Receptor kinase signalling in plants and animals: distinct molecular systems with mechanistic similarities. *Curr. Opin. Cell Biol.*, **2002**, 14(2), 230-6.
- [65] Shiu, S. H.; Bleeker, A. B. Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. U S A*, **2001**, 98(19), 10763-10768.
- [66] Rayko, E.; Maumus, F.; Maheswari, U.; Jabbari, K.; Bowler, C. Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol.*, **2010**, 188(1), 52-66.
- [67] Slotkin, R. K.; Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.*, **2007**, 8(4), 272-85.
- [68] Maumus, F.; Allen, A. E.; Mhiri, C.; Hu, H.; Jabbari, K.; Vardi, A.; Grandbastien, M. A.; Bowler, C. Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genomics*, **2009**, 10, 624.
- [69] Quesneville, H.; Bergman, C. M.; Andrieu, O.; Autard, D.; Nouaud, D.; Ashburner, M.; Anxolabehere, D. Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput. Biol.*, **2005**, 1(2), 166-175.
- [70] Flutre, T.; Duprat, E.; Feuillet, C.; Quesneville, H. Considering transposable element diversification in de novo annotation approaches. *PLoS One*, **2011**, 6(1), e16526.
- [71] Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.*, **1975**, 14(1), 9-25.
- [72] Holliday, R.; Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science*, **1975**, 187(4173), 226-232.
- [73] Goll, M. G.; Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **2005**, 74, 481-514.
- [74] Goll, M. G.; Kirpekar, F.; Maggert, K. A.; Yoder, J. A.; Hsieh, C. L.; Zhang, X.; Golic, K. G.; Jacobsen, S. E.; Bestor, T. H. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science*, **2006**, 311(5759), 395-398.
- [75] Ponger, L.; Li, W. H. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol. Biol. Evol.*, **2005**, 22(4), 1119-1128.
- [76] Jeddell, J. A.; Stokes, T. L.; Richards, E. J. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat. Genet.*, **1999**, 22(1), 94-97.
- [77] Miura, A.; Yonebayashi, S.; Watanabe, K.; Toyama, T.; Shimada, H.; Kakutani, T. Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. *Nature*, **2001**, 411(6834), 212-214.
- [78] Myant, K.; Stancheva, I. LSH cooperates with DNA methyltransferases to repress transcription. *Mol. Cell Biol.*, **2008**, 28(1), 215-226.

- [79] Jarvis, E. E.; Dunahay, T. G.; Brown, L. M. DNA nucleoside composition and methylation in several species of microalgae. *J. Phycol.*, **1992**, 28(3), 356-362.
- [80] Verreault, A.; Kaufman, P. D.; Kobayashi, R.; Stillman, B. Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Curr. Biol.*, **1998**, 8(2), 96-108.
- [81] Wang, L.; Mizzen, C.; Ying, C.; Candau, R.; Barlev, N.; Brownell, J.; Allis, C. D.; Berger, S. L. Histone acetyltransferase activity is conserved between yeast and human GCN5 and is required for complementation of growth and transcriptional activation. *Mol. Cell Biol.*, **1997**, 17(1), 519-527.
- [82] Das, C.; Lucia, M. S.; Hansen, K. C.; Tyler, J. K. CBP/p300-mediated acetylation of histone H3 on lysine 56. *Nature*, **2009**, 459(7243), 113-117.
- [83] Taipale, M.; Rea, S.; Richter, K.; Vilar, A.; Lichter, P.; Imhof, A.; Akhtar, A. hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol. Cell Biol.*, **2005**, 25(15), 6798-6810.
- [84] Hawkes, N. A.; Otero, G.; Winkler, G. S.; Marshall, N.; Dahmus, M. E.; Krappmann, D.; Scheidereit, C.; Thomas, C. L.; Schiavo, G.; Erdjument-Bromage, H.; Tempst, P.; Svejstrup, J. Q. Purification and characterization of the human elongator complex. *J. Biol. Chem.*, **2002**, 277(4), 3047-3052.
- [85] Tachibana, M.; Sugimoto, K.; Fukushima, T.; Shinkai, Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J. Biol. Chem.*, **2001**, 276(27), 25309-25317.
- [86] Kim, J.; Daniel, J.; Espejo, A.; Lake, A.; Krishna, M.; Xia, L.; Zhang, Y.; Bedford, M. T. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep.*, **2006**, 7(4), 397-403.
- [87] Trojer, P.; Li, G.; Sims, R. J., 3rd; Vaquero, A.; Kalakonda, N.; Boccuni, P.; Lee, D.; Erdjument-Bromage, H.; Tempst, P.; Nimer, S. D.; Wang, Y. H.; Reinberg, D. L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell*, **2007**, 129(5), 915-928.
- [88] Flanagan, J. F.; Mi, L. Z.; Chruszcz, M.; Cymborowski, M.; Clines, K. L.; Kim, Y.; Minor, W.; Rastinejad, F.; Khorasanizadeh, S. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*, **2005**, 438(7071), 1181-1185.
- [89] Zofall, M.; Grewal, S. I. Swi6/HP1 recruits a JmjC domain protein to facilitate transcription of heterochromatic repeats. *Mol. Cell*, **2006**, 22(5), 681-692.
- [90] Chang, B.; Chen, Y.; Zhao, Y.; Bruick, R. K. JMJD6 is a histone arginine demethylase. *Science*, **2007**, 318(5849), 444-447.
- [91] Rajagopalan, R.; Vaucheret, H.; Trejo, J.; Bartel, D. P. A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev.*, **2006**, 20(24), 3407-3425.
- [92] Borsani, O.; Zhu, J.; Verslues, P. E.; Sunkar, R.; Zhu, J. K. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell*, **2005**, 123(7), 1279-1291.
- [93] Vaucheret, H. MicroRNA-dependent trans-acting siRNA production. *Sci. STKE* **2005**, 2005(300), pe43.
- [94] Cuperus, J. T.; Carbonell, A.; Fahlgren, N.; Garcia-Ruiz, H.; Burke, R. T.; Takeda, A.; Sullivan, C. M.; Gilbert, S. D.; Montgomery, T. A.; Carrington, J. C. Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat. Struct. Mol. Biol.*, **2010**, 17(8), 997-1003.
- [95] Herr, A. J.; Jensen, M. B.; Dalmay, T.; Baulcombe, D. C. RNA polymerase IV directs silencing of endogenous DNA. *Science*, **2005**, 308(5718), 118-120.
- [96] Voorhoeve, P. M.; Agami, R. Knockdown stands up. *Trends Biotechnol.*, **2003**, 21(1), 2-4.
- [97] Cerutti, H.; Casas-Mollano, J. A. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr. Genet.*, **2006**, 50(2), 81-99.
- [98] Bailleul, B.; Rogato, A.; de Martino, A.; Coesel, S.; Cardol, P.; Bowler, C.; Falcioratore, A.; Finazzi, G. An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. *Proc. Natl. Acad. Sci. U S A*, **2010**, 107(42), 18214-18219.
- [99] Mochizuki, K.; Gorovsky, M. A. A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes Dev.*, **2005**, 19(1), 77-89.
- [100] Macrae, I. J.; Zhou, K.; Li, F.; Repic, A.; Brooks, A. N.; Cande, W. Z.; Adams, P. D.; Doudna, J. A. Structural basis for double-stranded RNA processing by Dicer. *Science*, **2006**, 311(5758), 195-198.
- [101] Shi, H.; Tschudi, C.; Ullu, E. An unusual Dicer-like1 protein fuels the RNA interference pathway in Trypanosoma brucei. *RNA*, **2006**, 12(12), 2063-2072.
- [102] Kiyota, E.; Okada, R.; Kondo, N.; Hiraguri, A.; Moriyama, H.; Fukuhara, T. An Arabidopsis RNase III-like protein, AtRTL2, cleaves double-stranded RNA in vitro. *J. Plant Res.*
- [103] Gobler, C. J.; Berry, D. L.; Dyhrman, S. T.; Wilhelm, S. W.; Salamov, A.; Lobanov, A. V.; Zhang, Y.; Collier, J. L.; Wurch, L. L.; Kustka, A. B.; Dill, B. D.; Shah, M.; VerBerkmoes, N. C.; Kuo, A.; Terry, A.; Pangilinan, J.; Lindquist, E. A.; Lucas, S.; Paulsen, I. T.; Hattenrath-Lehmann, T. K.; Talmage, S. C.; Walker, E. A.; Koch, F.; Burson, A. M.; Marcoval, M. A.; Tang, Y. Z.; Lecleir, G. R.; Coyne, K. J.; Berg, G. M.; Bertrand, E. M.; Saito, M. A.; Gladyshev, V. N.; Grigoriev, I. V. Niche of harmful alga Aureococcus anophagefferens revealed through ecogenomics. *Proc. Natl. Acad. Sci. U S A*, **2011**, 108(11), 4352-4357.
- [104] Bowler, C.; De Martino, A.; Falcioratore, A. Diatom cell division in an environmental context. *Curr. Opin. Plant Biol.*, **2010**, 13(6), 623-630.
- [105] De Martino, A.; Bartual, A.; Willis, A.; Meichenin, A.; Villazan, B.; Maheswari, U.; Bowler, C. Physiological and Molecular Evidence that Environmental Changes Elicit Morphological Interconversion in the Model Diatom Phaeodactylum tricornutum. *Protist*, **2011**, 162(3), 462-481.
- [106] De Martino, A.; Amato, A.; Bowler, C. Mitosis in diatoms: rediscovering an old model for cell division. *Bioessays*, **2009**, 31(8), 874-884.
- [107] Huson, D. H.; Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.*, **2006**, 23(2), 254-267.