LEGENDS FOR SUPPLEMENTAL FIGURES

Figure S1. Novel sequence-specific cytosine methylations and SNF2-family proteins, related to Figure 1. (A) High levels of methylation are apparent at specific cytosines. Inspection reveals two distinct palindromic sequence specificities: 5′-gCnnGc-3′ (blue) and 5′-gGCc-3′ (green; the capitalized base pairs contain the cytosines methylated). All other cytosines (gray) show no evidence of methylation. This type of methylation resembles that typically found in bacteria (e.g. restriction modification systems) and is unlike any previously reported in eukaryotes. The nuclear *A. anophagefferens* genome encodes predicted DNA methyltransferases without homologs known in other eukaryotes ("orphan"; see also Figure 1A), but with similarity to bacterial cytosine methyltransferases, making them excellent candidates for these sequence-specific activities. Scale bar is 10 kilobases (kb). (B) Predicted SNF2-family regions (from the ATPase to helicase domains) from eukaryotes were aligned and a maximum likelihood tree was inferred. Only Dnmt5 proteins also have N-terminal methyltransferase domains (Figure 1A,E) in addition to the SNF2-family regions. The C-terminal SNF2 regions of Dnmt5 are united with strong support (blue branches, 89% of bootstraps). The grouping of Dnmt5 SNF2 regions with the other RING finger-containing SNF2-family proteins, which include Rad5, Rad16, HLTF and SHPRH, is also strongly supported (green oval, 98% of bootstraps). SNF2-family proteins are DNA-dependent ATPases and multiple groups can remodel nucleosomes (i.e. Snf2-like, Swr1-like, Mot1-like and Rad54-like groups; pink oval) (Flaus et al., 2006). The nucleosome remodelers do not have RING fingers within their SNF2 regions and conversely, the RING finger-containing SNF2-family proteins are not known to remodel nucleosomes, but the RING finger provides E3 ubiquitin ligase activity in addition to the DNA-dependent ATPase of the SNF2 region (Unk et al., 2010). The RING finger proteins may also remodel chromatin proteins through their DNA-dependent ATPase activity, as has been shown for HLTF (Achar et al., 2011). *A RING finger is not detectable in a few members of the "RING finger" proteins, because the RING finger has either degenerated beyond recognition or been lost.

Figure S2. Symmetrical methylation, in transposable elements in some species, related to Figure 2. (A) Example genome snapshots at high resolution. These data are for small regions of Figures 2A (top panel; wild-type *C. neoformans*) and 4A (bottom panel; *M. pusilla*). The fraction of CG cytosines methylated is blue. Cytosines in other

sequence contexts are shown, but have background levels of methylation. CG positions (indicated in blue in the sequences) have similar levels of cytosine methylation on both strands (symmetrical), being either nearly 100% methylated or nearly 100% unmethylated in the populations of cells. The lower plots show the base-10 logarithms of cytosine coverage informative for methylation (black), indicating that all cytosine positions are confidently quantified across these regions. (B) Distributions of fractional methylation for individual CG sites are shown as boxplots, with the medians, first and third quartiles indicated by boxes, and the ranges indicated by whiskers at left. For each organism all sites in the whole genome are gray (upper) and only those found within transposable elements (TEs) are black (lower). For the panels at center right and far right, we assessed transcription by RNA-seq of total RNA. The barplot (center right) shows the fraction of predicted genes (gray) or TEs (black) that overlap a detectable transcript by at least 1 base. The boxplot (far right) shows the expression levels in FPKM (fragments per kilobase of transcript per million mapped reads) of detectable transcripts from the barplot at center right that overlap genes (gray) or TEs (black). Boxes indicate the medians, first and third quartiles with whiskers indicating the most extreme values up to 1.5 times the interquartile ranges away from the boxes. RNA-seq data for *A. anophagefferens*, *E. huxleyi*, *O. lucimarinus* and *M. pusilla*, for which we could not confidently identify classical TEs, are summarized in Figure S7.

Figure S3. Periodic methylation and methylation in gene bodies, related to Figure 3. (A) Mean fractional CG methylation is displayed as in Figure 3B, but instead it is for all genes, because we did not assess RNA levels in *B. prasinos*. (B) The autocorrelation function estimates for CG methylation are shown for each lag (offset) across the largest scaffold/chromosome of each organism. There is no apparent periodicity for the organisms at the top. For *C. neoformans* (see Figure 2A), we included wild-type strains (WT, blue and WT47, black) as well as the *dnmt5Δ* strain (gray). *P. tricornutum*, *F. cylindrus*, *T. pseudonana* and WT *C. neoformans* have large regions of either unmethylated or CG-methylated DNA, the latter found in and near transposable elements (Figures 2A and S2B). Apparent periodicities for periodically methylated organisms (left side) are indicated in base pairs (bp). (C) Histograms (bin size = 0.01) show mean fractions of CG methylation in each gene body. Almost every gene in *A. anophagefferens*, *E. huxleyi*, *B. prasinos*, *O. lucimarinus* and *M. pusilla* has CG methylation within it.

Figure S4. Endonuclease assay for periodically methylated genomic DNA, related to Figure 4. MspJI (top) is an endonuclease that makes a double-stranded cleavage (arrows) downstream of C5-methylated cytosine (mC) followed by a purine (R is A or G; Y is T or C on the other strand) (Cohen-Karni et al., 2011). It can cleave near methylated CG and CHG sites. The diatoms *P. tricornutum*, *F. cylindrus* and *T. pseudonana* do not have periodic methylation (Figure S3B) and are not extensively digested by MspJI. *E. huxleyi*, *B. prasinos*, *O. lucimarinus* and *M. pusilla* have abundant periodic methylation (Figures 3, 4 and S3) and are cleaved by MspJI to yield fragments between 100 and 200 bp long and fragments with lengths that are multiples thereof. All of the species with apparent periodically cleaved fragments are labeled in bold and include *Isochrysis galbana* and *Imantonia rotunda*. Notably, all the periodically methylated organisms identified by MspJI digestion belong to either the haptophyte class Prymnesiophyceae (brown) or the Chlorophyta green algal class Mamiellophyceae (green) (Marin and Melkonian, 2010). The divergence time of the lineage leading to *B. prasinos* has not been estimated and its relationship is thus presented as a polytomy, although it is actually more closely related to *O. lucimarinus* than *M. pusilla* (Moreau et al., 2012). We included a more distant Chlorophyta relative, *Pyramimonas parkeae*, which shows extensive digestion by MspJI, but no evidence of periodically cleaved fragments. †We were unable to digest *A. anophagefferens* DNA with MspJI despite abundant periodic methylation detectable by bisulfite sequencing (Figures 3A,B and S3B,C). The marker is 50 ng per lane O'GeneRuler 100 bp DNA ladder (Fermentas). *Small activating DNA (New England Biolabs) added to all reactions.

Figure S5. Alignment of DNA methyltransferase domains of Dnmt5 proteins, related to Figure 5. As a first step toward elucidating the mechanism by which periodic methylation occurs, we identified amino acids (5 positions; underlined) that are conserved in the DNA methyltransferase domain of at least one Dnmt5 protein in each species with periodic methylation and are absent in all species with non-periodic methylation. All positions with a gap in more than 2 Dnmt5 sequences were removed. We did not identify any amino acids that are conserved in species without periodic methylation and are absent in all species with periodic methylation. Functional amino acids known from structural studies of M.HhaI and mouse Dnmt1 (Song et al., 2011, 2012) are indicated below their positions:

*****catalytic cysteine

[a]contacts the methyl donor molecule, S-adenosyl methionine

[b]contacts the cytosine that is methylated, which is flipped out of the DNA duplex in active DNA methyltransferase structures

[c]contacts the phosphate backbone and deoxyribose next to the cytosine that is methylated

Figure S6. CG enrichment despite methylation, both inside and outside of methylated regions, related to Figure 6. (A) The $\log_2$(observed-to-expected ratios of CG content) versus mean percent CG methylation of individual genomes are plotted. The colors are as in Figure 6A. Log_2(observed-to-expected ratios of CG content) values for genomes above 0 are enriched for CG sites relative to G+C contents and those below 0 are depleted of CG sites. Mean methylation was calculated from both experiments we performed and published results (Feng et al., 2010; Zemach et al., 2010). Genomes known to lack methylation were plotted with a mean methylation of 0. (B) Substitution rates were estimated using a general reversible dinucleotide model, R2 (Siepel and Haussler, 2004), from aligned chromosomes of *Ostreococcus* species (Table S2). The non-coding sequence used in Figure 6B was separated into methylated clusters in *O. lucimarinus* (y-axis) with the remaining sequence constituting unmethylated regions (x-axis). Transversions and transitions are colored gray and black, respectively. Substitutions that lose and gain CG sites are labeled with circles and diamonds, respectively.

Figure S7. Genomic transcription, related to Figure 7. We assessed transcription by RNA-seq of total RNA. The barplot (left) shows the fraction of predicted genes that overlap a detectable transcript by at least 1 base. The boxplot (right) shows the expression levels in FPKM (fragments per kilobase of transcript per million mapped reads) of detectable transcripts from the barplot at left that overlap genes. Boxes indicate the medians, first and third quartiles with whiskers indicating the most extreme values up to 1.5 times the interquartile ranges away from the boxes. Classical TEs could not be identified with confidence in these G+C- and CG-rich genomes, perhaps because algorithms for identifying TEs are not optimized for such genomes, because classical

TEs have become unrecognizable in these genomes, and/or because the genomes have unusual families of TEs.

		Strand coverage	^{m}CG	^{m}CHG	^{m}CHH
Aureococcus anophagefferens					
	all	34×	24.20	0.19	0.18
	mitochondria[1]	457×	0.54	0.64	0.53
	chloroplasts	589×	0.14	0.11	0.11
Bathycoccus prasinos[2]					
	all	99×	20.90	0.21	0.19
	mitochondria	1,225×	0.15	0.11	0.12
	chloroplasts	783×	0.14	0.11	0.12
Cryptococcus neoformans[2] **WT**					
	all	293×	1.64	0.40	0.44
	mitochondria	18,843×	0.39	0.41	0.45
Cryptococcus neoformans[2] **WT47 (*sxi1Δ*)**					
	all	119×	1.58	0.39	0.41
	mitochondria	4,966×	0.38	0.40	0.42
Cryptococcus neoformans[2] ***dnmt5Δ***					
	all	85×	0.18	0.14	0.14
	mitochondria	4,611×	0.16	0.13	0.12
Cyanidioschyzon merolae[2,3]					
	all	98×	0.20	0.18	0.19
	mitochondria	829×	0.19	0.17	0.20
	chloroplasts	1,990×	0.20	0.19	0.20
Emiliania huxleyi					
	all	14×	16.36	5.63	0.38
	mitochondria	414×	0.58	0.56	0.59
	chloroplasts	683×	0.23	0.23	0.24
Fragilariopsis cylindrus					
	all	31×	8.63	0.13	0.13
	mitochondria	182×	0.12	0.14	0.12
	chloroplasts	401×	0.14	0.11	0.10
Leishmania major[2,4]					
	all	139×	1.28	1.35	1.71
	maxicircles	879×	1.15	1.33	1.57
Micromonas pusilla **CCMP1545**[2]					
	all	98×	17.70	0.14	0.13
	mitochondria	704×	0.13	0.11	0.11
	chloroplasts	2,484×[†] (2,579×)	1.99[†] (0.13)	0.11[†] (0.11)	0.10[†] (0.10)
Ostreococcus lucimarinus[2]					
	all	55×	22.17	0.13	0.13
Phaeodactylum tricornutum[2]					
	all	61×	1.03	0.19	0.20
	mitochondria	152×	0.20	0.18	0.19
	chloroplasts	555×	0.20	0.19	0.20
Thalassiosira pseudonana[2]					
	all	83×	2.57	0.13	0.12
	mitochondria	450×	0.13	0.10	0.11
	chloroplasts	707×	0.13	0.11	0.11

Table S1 | Coverage and percent methylation from bisulfite sequencing, related to Figure 1

For each organism the median genomic strand coverage of cytosine positions obtained by bisulfite sequencing is shown in the first column. Columns two to four show the mean percent methylation of cytosines in the CG, CHG and CHH sequence contexts, respectively. For comparison, the values determined for cytosines in the mitochondrial (maxicircle for *Leishmania major*) and chloroplast genomes are shown, where available.

[1] *Aureococcus anophagefferens* mitochondria have specific methylations (see Figure S1A) not categorized in this table.

[2] The indicated species have genomes with assembled chromosomes.

[3] Among DNMT homologs, *Cyanidioschyzon merolae* encodes one Dnmt3.

[4] Among DNMT homologs, *Leishmania major* encodes one Dnmt6.

[†] One of the *Micromonas pusilla* CCMP1545 scaffolds annotated as "chloroplast" has considerable CG methylation, is sequenced to coverage depth similar to nuclear scaffolds and has sequence similarity to part of the second largest nuclear scaffold—all of these data suggest that this scaffold is actually part of the nuclear genome. Removal of this scaffold reveals methylation on the remaining chloroplast scaffolds (in parentheses) equivalent to background levels.

Ostreococcus lucimarinus	*Ostreococcus sp.* **RCC809**	*Ostreococcus tauri*
chr1	chr1 (RC)	chr1
chr2	chr2 (RC)	chr2 (RC)
chr3	chr3 (RC)	chr3
chr4	chr4 (RC)	chr4
chr5	chr5	chr5
chr6	chr6	chr6
chr7	chr7	chr7
chr8	chr8	chr8
chr9	chr9 (RC)	chr9
chr10	chr10 (RC)	chr10
chr11	chr11	chr11
chr12	chr12	chr12
chr13	chr13 (RC)	chr13
chr14	chr14	chr15
chr15	chr15 (RC)	chr16
chr16	chr16 (RC)	chr17
chr17	chr17	chr18
chr19	chr19 (RC)	chr20
chr20	chr20	chr14

Table S2 | Homologous chromosomes aligned from *Ostreococcus* species, related to Figure 6

Each row is a homologous set of chromosomes (each genome version is JGI v2.0) aligned for substitution rate analysis. Orientations relative to *O. lucimarinus* chromosomes were determined using the JGI synteny browser (Grigoriev et al., 2012). Where indicated, the reverse complement (RC) of the indicated chromosome sequence was used.

	References
1. *Homo sapiens* late erythroblast	Brown et al., 2006
2. *Homo sapiens* neutrophil granulocyte	Rosenbluth et al., 2006
3. *Homo sapiens* lymphocyte	Maul and Deaven, 1977
4. *Homo sapiens* intermediate erythroblast	Brown et al., 2006
5. *Homo sapiens* proerythroblast	Brown et al., 2006
6. *Gallus gallus* erythrocyte	García-Pérez et al., 1999
7. *Gallus gallus* neutrophil granulocyte	Grigoryev and Woodcock, 1998
8. *Gallus gallus* embryonic fibroblast	Habermann et al., 2001
9. *Emiliania huxleyi* diploid C cell	Bidle et al., 2007; Frada et al., 2008; van der Wal et al., 1983
10. *Drosophila melanogaster* interphase blastoderm	Chinta and Wasser, 2012
11. *Arabidopsis thaliania* shoot meristem cell	Berr and Schubert, 2007
12. *Arabidopsis thaliania* leaf cell	Jovtchev et al., 2006
13. *Arabidopsis thaliania* root meristem cell	Berr and Schubert, 2007
14. *Caenorhabditis elegans*, typical cell	Long et al., 2009
15. *Guillardia theta*, if diploid	Moore and Archibald, 2009
16. *Emiliania huxleyi* haploid S cell	Frada et al., 2008
17. *Aureococcus anophagefferens*, if diploid	Gastrich et al., 1998; Probyn et al., 2010
18. *Chlamydomonas reinhardtii*	Colón-Ramos et al., 2003
19. *Guillardia theta*, if haploid	Moore and Archibald, 2009
20. *Aureococcus anophagefferens*, if haploid	Gastrich et al., 1998; Probyn et al., 2010
21. *Chlorella variabilis*	Van Etten and Dunigan, 2012
22. *Schizosaccharomyces pombe* diploid	Neumann and Nurse, 2007
23. *Saccharomyces cerevisiae* diploid	Larson et al., 2011
24. *Micromonas pusilla*	Worden et al., 2009
25. *Cryptococcus neoformans*	Oliveira et al., 2009
26. *Cyanidioschyzon merolae*	Kuroiwa et al., 1994; Misumi et al., 2005
27. *Bathycoccus prasinos*	Moreau et al., 2012
28. *Ostreococcus lucimarinus*	Worden et al., 2004
29. *Ostreococcus tauri*	Henderson et al., 2007
30. *Schizosaccharomyces pombe* haploid spore	Neumann and Nurse, 2007
31. *Schizosaccharomyces pombe* haploid	Neumann and Nurse, 2007; Wu and Pollard, 2005
32. *Saccharomyces cerevisiae* haploid	Webster et al., 2010

Table S3 | References for nuclei sizes, related to Figure 7

EXTENDED EXPERIMENTAL PROCEDURES

Divergence time trees. For Figures 1B, 5A and S4, although the sistering of the haptophyte *Emiliania huxleyi* to the SAR supergroup (stramenopiles + alveolates + rhizarians; *Aureococcus anophagefferens* and the diatoms are stramenopiles) is tentative (Parfrey et al., 2011), it is supported by further analyses (Burki et al., 2012; Nozaki et al., 2012), and our conclusions would not be affected if a different phylogeny were instead correct. The divergence times of *Phaeodactylum tricornutum* from *Fragilariopsis cylindrus* and *Coprinopsis cinerea* from *Cryptococcus neoformans* are from TimeTree (Hedges et al., 2006) (http://www.timetree.org/). The age of the diversification of Chlorophyta green algae, other than Mamiellophyceae, was estimated previously (Herron et al., 2009). Trees were visualized with FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Phylogenetic inference on protein domains. Representative C5 cytosine methyltransferase domains were initially identified with SUPERFAMILY (Gough et al., 2001) (http://supfam.cs.bris.ac.uk/SUPERFAMILY/) and these were used in iteratively expanded BLAST (Altschul et al., 1997) searches of NCBI (http://blast.ncbi.nlm.nih.gov/) and JGI (Grigoriev et al., 2012) databases to identify C5 cytosine methyltransferases from a wide variety of organisms. The plant DRM methyltransferase domains were rearranged *in silico* to align with other Dnmt3 domains. To make alignments for the trees in Figures 1A and S1B, we removed many close orthologs that came from taxa with numerous sequenced genomes, generally leaving sequences from the species we studied here. Alignments were generated with default settings in MUSCLE v3.8.31 (Edgar, 2004). For SNF2-family proteins, we found homologs of the SNF2 regions of Dnmt5 sequences by BLAST searches, but only included close homologs and representatives of known nucleosome remodelers (Flaus et al., 2006) (Figure S1B). Trees were visualized with FigTree as for divergence time trees. Homologs of Uhrf1 (also known as Np95, VIM and ORTH; Figure 1C) were identified by reciprocal BLAST searches; a homolog is indicated where the predicted protein had reciprocal E-values against human and *Arabidopsis thaliana* sequences less than 10^{-10}.

Nucleic acid isolation. *Cyanidioschyzon merolae* cells were collected by centrifugation, and genomic DNA was isolated with a DNeasy Plant Mini Kit (QIAGEN). *C. neoformans var. grubii* H99 strains D632 and WT47 were grown overnight at 23 °C in the dark to

saturation in liquid Difco YM Broth (Becton, Dickinson and Company) and 10 µl was patched on autoclaved cellophane overlaid on Difco YM Agar (Becton, Dickinson and Company) and grown for 3 days at 23 °C in the dark. The cells and cellophane were placed in tubes with 0.5 mm Zirconia-Silica beads (BioSpec) and frozen in liquid nitrogen. The tubes were bead-beaten 4 × 1 min with replacement in liquid nitrogen in between, and pulverized material was stored at -80 °C. Pulverized material was partially thawed, and genomic DNA was extracted with a DNeasy Plant Mini Kit by starting at the step with addition of AP1 and RNase. Total RNA was extracted with TRIzol (Life Technologies) using Phase Lock Gel Heavy 1.5 ml tubes (5 PRIME).

Bisulfite sequencing. Genomic DNA was quantified with Qubit dsDNA Assay Kits (Life Technologies) and 50 to 250 ng were sonicated, end-repaired, and ligated to methylated adapters before bisulfite conversion (Ibarra et al., 2012). We performed 3 PCR reactions after bisulfite conversion and combined the products before sequencing. Single-end 100-base sequencing was performed with the Illumina HiSeq 2000 platform.

RNA sequencing. Total RNA was digested with RNase-free DNase I (QIAGEN) and concentrated with a RNeasy MinElute Cleanup Kit (QIAGEN). Purified total RNA was quantified by a Qubit RNA Assay Kit (Life Technologies) and between 67 and 125 ng was used to prepare a strand-specific library with the Encore Complete RNA-Seq Library System I (NuGEN). Single-end 100-base sequencing was performed with the Illumina HiSeq 2000 platform.

Micrococcal nuclease sequencing. We modified a protocol previously developed for in vivo nucleosome positions (Teves and Henikoff, 2012). Frozen *Ostreococcus lucimarinus* and *Micromonas pusilla* cells (collected from 5 L of culture) were thawed on ice and resuspended in 500 µl of cold 1× TM2 (10 mM Tris-HCl, pH 8, 2 mM $MgCl_2$, supplemented with 1:200 EDTA-free plant protease inhibitors (Sigma)) per 50 mg wet-pellet mass by pipetting with a 5-ml pipet up and down 100 times. The slurry was filtered through a 40 µm-pore cell strainer (BD Falcon) by gravity. Triplicate reactions from the same cells were performed. Five hundred µl of the cell suspension per reaction were collected by gentle, refrigerated centrifugation and washed once with 1 ml of 1× TM2. Cells were centrifuged again, resuspended in 200 µl of 1× TM2, and warmed to 37 °C for 5 min. Two hundred µl of 1× buffer (50 mM Tris-Hcl, pH 7.9, 5 mM $CaCl_2$), supplemented

with 100 μg ml^{-1} BSA and 125 gel units ml^{-1} for *O. lucimarinus* and 250 gel units ml^{-1} for *M. pusilla* (~12.5 and 25 Kunitz units ml^{-1}, respectively, which we previously determined to yield mostly mononucleosomes) of micrococcal nuclease (New England Biolabs) were pre-heated to 37 °C and added to the 200 μl of resuspended cells, and the 400-μl mixture was further incubated at 37 °C for 10 min with occasional vortexing. Reactions were stopped with 5 mM final EGTA. Then 100 mM final NaCl, 0.625% final SDS, 20 μg DNase-free RNase A and 50 μg proteinase K were added, and the mixture was incubated at 75 °C for 10 min. Nucleosomal fragments were purified first by extraction with buffer-saturated phenol-chloroform-isoamyl alcohol using Phase Lock Gel Heavy 1.5 ml tubes, then with 2 volumes of Agencourt AMPure XP beads (Beckman Coulter), and 66.7 ng from each of the triplicate reactions was combined (200 ng total) and made into a library without PCR by using the Encore Rapid Library System (NuGEN). Paired-end 100-base sequencing of in vivo nucleosome fragments was performed with the Illumina HiSeq 2000 platform.

For in vitro nucleosome position analyses, we first generated an unmethylated equivalent of *O. lucimarinus* genomic DNA in vitro by isothermal multiple strand displacement amplification, using either Bst 2.0 (NEB) or an illustra Ready-To-Go GenomiPhi v3 Kit (GE Healthcare). For Bst 2.0, 15 replicate 50-μl reactions containing 10 ng *O. lucimarinus* genomic DNA each in 1× NEB Isothermal Amplification Buffer with 12.5 μM random hexamers, 400 μM dNTPs (with the deoxycytidine triphosphate unmethylated) and 8 units of Bst 2.0 DNA polymerase (New England Biolabs) were incubated at 50 °C for 8 hours. The DNA was amplified more than 20-fold in these reactions, so that approximately 95% of the amplified DNA molecules contain entirely unmethylated cytosines. The resulting amplified DNA and natively methylated genomic DNA were incubated separately at 75 °C for 10 min at a concentration of 4 μg ml^{-1} each with 0.62% SDS, 38 ng ml^{-1} DNase-free RNase A and 100 ng ml^{-1} proteinase K. DNA was purified by extracting twice with buffer-saturated phenol-chloroform-isoamyl alcohol using Phase Lock Gel Heavy 1.5 ml tubes. The DNA was further purified and concentrated using 0.7 volumes of Agencourt AMPure XP beads.

The purified natively methylated *O. lucimarinus* genomic DNA or unmethylated equivalent was assembled with purified recombinant human histones into nucleosomes by salt dilution using an EpiMark Nucleosome Assembly Kit (New England Biolabs) at a

1:2:1 molar ratio of DNA:histone H2A/H2B dimer:histone H3.1/H4 tetramer (assuming 1 nucleosome per 184 bp DNA for *O. lucimarinus*; Figure 4D). The assembly began at 640 nM of "184-bp" DNA in 2M NaCl and was completed when the final concentration of "184-bp" DNA was 80 nM and NaCl was 250 mM. The assembly was heated to 37 °C for 5 min and then mixed with an equal volume (80 µl) of pre-warmed 1× buffer (50 mM Tris-Hcl, pH 7.9, 5 mM $CaCl_2$), supplemented with 100 µg ml^{-1} BSA and 250 gel units ml^{-1} micrococcal nuclease (~25 Kunitz units ml^{-1}, which we previously determined to yield mostly mononucleosomes for assembly on either methylated or unmethylated DNA) and incubated at 37 °C for 10 min. Reactions were stopped with 5 mM final EGTA. Then 100 mM final NaCl, 0.625% final SDS and 20 µg proteinase K were added, and the mixture was incubated at 75 °C for 10 min. Nucleosomal fragments were purified first by extraction with buffer-saturated phenol-chloroform-isoamyl alcohol using Phase Lock Gel Heavy 1.5 ml tubes, then with 2 volumes of Agencourt AMPure XP beads, and 90 ng of each sample was made into a library without PCR by using the Encore Rapid Library System. Paired-end 70-base sequencing of in vitro nucleosome fragments was performed with the Illumina HiSeq 2500 platform.

To assess nucleosome assembly efficiencies (Figure 5B), we performed assembly reactions as above with or without histones, and then either mock-treated the DNA or digested it with 13.9, 41.7 or 125 gel units ml^{-1} micrococcal nuclease (final concentration) in 20-µl reactions at 37 °C for 10 min. Reactions were stopped with 5 mM EGTA and 0.625% SDS (final). Ten µg proteinase K were added, and the mixture was incubated at 75 °C for 10 min. We extracted the DNA with buffer-saturated phenol-chloroform-isoamyl alcohol using Phase Lock Gel Heavy 1.5 ml tubes to analyze by electrophoresis through a gel of 2% agarose in 1× TAE, containing 0.4× GelRed (Biotium) for visualization.

Digestion of naked purified natively methylated *O. lucimarinus* genomic DNA was performed similarly to the in vitro assembly of *O. lucimarinus* genomic DNA (*i.e.* with 125 gel units ml^{-1} micrococcal nuclease (final) at 37 °C), except without histones. The digestion was only run for 2.5 min—a full 10-min incubation would have eliminated most fragments of size comparable to nucleosomes (Figure 5B). Paired-end 70-base sequencing of naked micrococcal nuclease-digested fragments was performed with the Illumina HiSeq 2500 platform.

Genomic analyses. Analyses were performed with Perl scripts from dzlab-tools (http://dzlab.pmb.berkeley.edu/tools/) and custom scripts written in AWK and R (http://www.r-project.org/). bs-sequel from dzlab-tools was used to map bisulfite sequencing reads with Bowtie v0.12.7 (Langmead et al., 2009) and quantify fractional methylation in CG, CHG, and CHH sequence contexts. Paired micrococcal nuclease sequencing reads were mapped with Bowtie 2 beta 5 (Langmead and Salzberg, 2012), and only nucleosomal (or similarly sized naked DNA) fragments of length 125 to 171 bp (inclusive) were analyzed. Genomic maps per base of nucleosomal fragment centers and coverage were generated with genomeCoverageBed from bedtools (http://bedtools.googlecode.com/). Nucleosome predictions for the *O. lucimarinus* and the *M. pusilla* genomes were generated using an existing algorithm (Kaplan et al., 2009). CG methylation data across *C. neoformans* chromosome 10 (Figure 2A) was smoothed with the default smooth() function of R, which is a concatenation of Tukey's running median smoothers that run until convergence. Autocorrelation functions were estimated using the acf() function of R, calculated using covariances about the sample means. The autocorrelation function is the correlation at each lag (offset) of the data to themselves for all of the CG or CHG sites that are available. For example, at position 182 bp for CG methylation, the correlation of the data to themselves offset by 182 bp is calculated. So, for *E. huxleyi* (Figure 3C) this results in strongly positive correlation, because the methylation of CG sites is correlated to that of other CG sites that happen to be offset 182 bp up- or downstream, due to the apparent periodicity of methylation (see Figure 3A). For Figure 5E, top in vivo nucleosome positions were defined as the $1/184^{th}$ of all genomic positions (allowing 1 per nucleosome on average in *O. tauri*; Figure S3B) with the highest numbers of nucleosome center reads from the in vivo dataset. RNA sequencing reads were mapped with Bowtie v0.12.7 / Tophat v1.4.1 (Trapnell et al., 2009) and FPKMs (fragments per kilobase transcript per million mapped reads) were calculated with Cufflinks v1.3.0 using bias correction (Roberts et al., 2011).

Transposable element (TE) annotation. TEs were predicted using RepeatModeler v1.7.0 with RMBlast v2.2.27 (http://www.repeatmasker.org/RepeatModeler.html). To reduce false positives, only those elements classified as similar to known eukaryotic TE types were used (*i.e.* transposable DNA and retrotransposable elements in Repbase (Jurka et al., 2005)).

Methylated regions/clusters. Methylated regions (which are typically <100 bp clusters in the periodically methylated species) were constructed by first defining contiguous runs of cytosines in a given sequence context with >0.25 fractional methylation in the population of cells. For alignment to methylated clusters, the individual regions were then fused if they were less than 50 bp apart. For the analysis of local methylation density in Figure 5A, we used individual regions that were between 3 and 50 base pairs (inclusive) in length in each species. We noticed that CG sites often lacked coverage on one strand or the other in previously published data (Zemach et al., 2010), so to ensure a fair comparison to the high-coverage data generated here (Table S1), we estimated missing methylation values at CG sites from the published data by copying *in silico* the value from the strand that did have sequencing coverage, which assumes that Dnmt1-mediated CG methylation is symmetrical (Law and Jacobsen, 2010). *For E. huxleyi*, we also combined CG and CHG methylation data and defined methylated regions similarly as for CG methylation alone (Figure 5A).

Methylation-dependent endonuclease digestions. Genomic DNA was treated to remove any possible contaminating chromatin components and digested with MspJI as previously described (Cohen-Karni et al., 2011). Briefly, several hundred ng of DNA were incubated in 200 µl of 100 mM NaCl, 0.625% SDS, containing 10 µg DNase-free RNase A and 25 µg proteinase K at 75 °C for 10 min. The DNA was then twice extracted with buffer-saturated phenol-chloroform-isoamyl alcohol using Phase Lock Gel Heavy 1.5 ml tubes. DNA was quantified with Qubit dsDNA Assay Kits and 50 ng DNA were digested in 10 µl of 1× NEB buffer 4 (50 mM KOAc, 20 mM Tris-OAc, 10 mM $Mg(OAc)_2$, 1 mM DTT, pH 7.9) with 100 µg ml^{-1} BSA, 1× activator solution, and 40 units ml^{-1} MspJI (New England Biolabs) at 37 °C for 16 hrs. Control samples were treated identically, except without MspJI. DNA was separated (Figure S4) by electrophoresis through a gel of 2% agarose in 1× TAE, containing 0.4× GelRed for visualization.

Dinucleotide substitution rates. Homologous *Ostreococcus* chromosomes (Table S2) were aligned with FSA v1.15.7 (Bradley et al., 2009), using MUMmer 3.23 (Kurtz et al., 2004) and exonerate v2.2.0 (http://www.ebi.ac.uk/~guy/exonerate/), along with soft-masking. The segments of the multiple sequence alignment overlapping *O. lucimarinus* non-coding DNA (*i.e.* intergenics, untranslated regions and introns) were used together to estimate dinucleotide substitution rates with RPHAST v1.3 (Hubisz et al., 2011), using

the EM algorithm with “high” precision. A second-order general time-reversible substitution model (R2; 63 parameters) and a second-order general unrestricted substitution model (U2; 96 parameters) were used to estimate each single-nucleotide substitution rate between dinucleotides (Siepel and Haussler, 2004). Substitution rates toward CG dinucleotides (gains) and those away from CG dinucleotides (losses) are indicated in Figures 6B and S6B. For Figure S6B, the non-coding DNA that does or does not overlap methylated clusters in *O. lucimarinus* (see above) was used.

Nuclear DNA density. Nuclear volumes were curated directly from the literature where available (references in Table S3). We used the mean volume where measurements from multiple nuclei were available. For those species/cell types with multiple source references, we then used the mean of volumes from the different sources. For most of the species/cell types, volumes have not been directly measured, so we first calculated nuclear area from micrographs using the provided scale bars (references in Table S3). From these areas we calculated the idealized circular radius, r_{circle}. If the cells were not prepared using dehydrating methods we then used this r_{circle} to calculate the idealized spherical volume. For preparations with dehydrating methods we used the following equation to estimate the idealized spherical volume, $Volume_{est}$, adjusting for the loss of water, which we assume was 70% of the nuclear volume:

$$Volume_{est} = \frac{\frac{4}{3}\pi r_{circle}^3}{30\%}$$

Although this may introduce bias, it will often be conservative in that it over-estimates the volume, because nuclei are typically smaller in volume than the idealized sphere of equivalent radius and it is not likely that the nuclei shrank by a full 70% in dehydrating preparations. We note that this method produces similar DNA concentration estimates for all of the periodically methylated species (Figure 7), suggesting that these estimates may be precise within a few fold. Also, the nuclear volumes we estimated for *O. lucimarinus* and *Bathycoccus prasinos* are close to the volume measured for the related *Ostreococcus tauri* using highly quantitative electron cryotomography on minimally processed cells (Henderson et al., 2007), indicating that at least in these cases our method is likely accurate.

SUPPLEMENTAL REFERENCES

Achar, Y.J., Balogh, D., and Haracska, L. (2011). Coordinated protein and DNA remodeling by human HLTF on stalled replication fork. Proc. Natl. Acad. Sci. U. S. A. *108*, 14073–14078.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Berr, A., and Schubert, I. (2007). Interphase chromosome arrangement in Arabidopsis thaliana is similar in differentiated and meristematic tissues and shows a transient mirror symmetry after nuclear division. Genetics *176*, 853–863.

Bidle, K.D., Haramaty, L., Barcelos E Ramos, J., and Falkowski, P. (2007). Viral activation and recruitment of metacaspases in the unicellular coccolithophore, Emiliania huxleyi. Proc. Natl. Acad. Sci. U. S. A. *104*, 6049–6054.

Bradley, R.K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I., and Pachter, L. (2009). Fast statistical alignment. PLoS Comput. Biol. *5*, e1000392.

Brown, J.M., Leach, J., Reittie, J.E., Atzberger, A., Lee-Prudhoe, J., Wood, W.G., Higgs, D.R., Iborra, F.J., and Buckle, V.J. (2006). Coregulated human globin genes are frequently in spatial proximity when active. J. Cell Biol. *172*, 177–187.

Burki, F., Okamoto, N., Pombert, J.-F., and Keeling, P.J. (2012). The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc. Biol. Sci.

Chinta, R., and Wasser, M. (2012). Three-dimensional segmentation of nuclei and mitotic chromosomes for the study of cell divisions in live Drosophila embryos. Cytom. Part J. Int. Soc. Anal. Cytol. *81*, 52–64.

Cohen-Karni, D., Xu, D., Apone, L., Fomenkov, A., Sun, Z., Davis, P.J., Kinney, S.R.M., Yamada-Mabuchi, M., Xu, S., Davis, T., et al. (2011). The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. Proc. Natl. Acad. Sci. U. S. A. *108*, 11040–11045.

Colón-Ramos, D.A., Salisbury, J.L., Sanders, M.A., Shenoy, S.M., Singer, R.H., and García-Blanco, M.A. (2003). Asymmetric distribution of nuclear pore complexes and the cytoplasmic localization of beta2-tubulin mRNA in Chlamydomonas reinhardtii. Dev. Cell *4*, 941–952.

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

Van Etten, J.L., and Dunigan, D.D. (2012). Chloroviruses: not your everyday plant virus. Trends Plant Sci. *17*, 1–8.

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., et al. (2010). Conservation and divergence of

methylation patterning in plants and animals. Proc. Natl. Acad. Sci. U. S. A. *107*, 8689–8694.

Flaus, A., Martin, D.M.A., Barton, G.J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res. *34*, 2887–2905.

Frada, M., Probert, I., Allen, M.J., Wilson, W.H., and de Vargas, C. (2008). The "Cheshire Cat" escape strategy of the coccolithophore Emiliania huxleyi in response to viral infection. Proc. Natl. Acad. Sci. U. S. A. *105*, 15944–15949.

García-Pérez, A.I., López-Beltrán, E.A., Klüner, P., Luque, J., Ballesteros, P., and Cerdán, S. (1999). Molecular crowding and viscosity as determinants of translational diffusion of metabolites in subcellular organelles. Arch. Biochem. Biophys. *362*, 329–338.

Gastrich, M.D., Anderson, O.R., Benmayor, S.S., and Cosper, E.M. (1998). Ultrastructural analysis of viral infection in the brown-tide alga, Aureococcus anophagefferens (Pelagophyceae). Phycologia *37*, 300–306.

Gough, J., Karplus, K., Hughey, R., and Chothia, C. (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J. Mol. Biol. *313*, 903–919.

Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R.A., et al. (2012). The genome portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res. *40*, D26–32.

Grigoryev, S.A., and Woodcock, C.L. (1998). Chromatin structure in granulocytes. A link between tight compaction and accumulation of a heterochromatin-associated protein (MENT). J. Biol. Chem. *273*, 3082–3089.

Habermann, F.A., Cremer, M., Walter, J., Kreth, G., von Hase, J., Bauer, K., Wienberg, J., Cremer, C., Cremer, T., and Solovei, I. (2001). Arrangements of macro- and microchromosomes in chicken cells. Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol. *9*, 569–584.

Hedges, S.B., Dudley, J., and Kumar, S. (2006). TimeTree: a public knowledge-base of divergence times among organisms. Bioinforma. Oxf. Engl. *22*, 2971–2972.

Henderson, G.P., Gan, L., and Jensen, G.J. (2007). 3-D ultrastructure of O. tauri: electron cryotomography of an entire eukaryotic cell. PloS One *2*, e749.

Herron, M.D., Hackett, J.D., Aylward, F.O., and Michod, R.E. (2009). Triassic origin and early radiation of multicellular volvocine algae. Proc. Natl. Acad. Sci. U. S. A. *106*, 3254–3258.

Hubisz, M.J., Pollard, K.S., and Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. Brief. Bioinform. *12*, 41–51.

Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.-F., Uzawa, R., Rodrigues, J.A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., et al. (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science *337*, 1360–1364.

Jovtchev, G., Schubert, V., Meister, A., Barow, M., and Schubert, I. (2006). Nuclear DNA content and nuclear and cell volume are positively correlated in angiosperms. Cytogenet. Genome Res. *114*, 77–82.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. *110*, 462–467.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362–366.

Kuroiwa, T., Kawazu, T., Takahashi, H., Suzuki, K., Ohta, N., and Kuroiwa, H. (1994). Comparison of ultrastructures between the ultra-small eukaryote Cyanidioschyzon merolae and Cyanidium caldarium. Cytologia (Tokyo) *59*, 149–158.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. *5*, R12.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A., and Singer, R.H. (2011). Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science *332*, 475–478.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220.

Long, F., Peng, H., Liu, X., Kim, S.K., and Myers, E. (2009). A 3D digital atlas of C. elegans and its application to single-cell analyses. Nat. Methods *6*, 667–672.

Marin, B., and Melkonian, M. (2010). Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. Protist *161*, 304–336.

Maul, G.G., and Deaven, L. (1977). Quantitative determination of nuclear pore complexes in cycling cells with differing DNA content. J. Cell Biol. *73*, 748–760.

Misumi, O., Matsuzaki, M., Nozaki, H., Miyagishima, S., Mori, T., Nishida, K., Yagisawa, F., Yoshida, Y., Kuroiwa, H., and Kuroiwa, T. (2005). Cyanidioschyzon merolae genome.

A tool for facilitating comparable studies on organelle biogenesis in photosynthetic eukaryotes. Plant Physiol. *137*, 567–585.

Moore, C.E., and Archibald, J.M. (2009). Nucleomorph genomes. Annu. Rev. Genet. *43*, 251–264.

Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., Van Bel, M., Poulain, J., Katinka, M., Hohmann-Marriott, M., et al. (2012). Gene functionalities and genome structure in Bathycoccus prasinos reflect cellular specializations at the base of the green lineage. Genome Biol. *13*, R74.

Neumann, F.R., and Nurse, P. (2007). Nuclear size control in fission yeast. J. Cell Biol. *179*, 593–600.

Nozaki, H., Yang, Y., Maruyama, S., and Suzaki, T. (2012). A case study for effects of operational taxonomic units from intracellular endoparasites and ciliates on the eukaryotic phylogeny: phylogenetic position of the haptophyta in analyses of multiple slowly evolving genes. PloS One *7*, e50827.

Oliveira, D.L., Nimrichter, L., Miranda, K., Frases, S., Faull, K.F., Casadevall, A., and Rodrigues, M.L. (2009). Cryptococcus neoformans cryoultramicrotomy and vesicle fractionation reveals an intimate association between membrane lipids and glucuronoxylomannan. Fungal Genet. Biol. FG B *46*, 956–963.

Parfrey, L.W., Lahr, D.J.G., Knoll, A.H., and Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc. Natl. Acad. Sci. U. S. A. *108*, 13624–13629.

Probyn, T.A., Bernard, S., Pitcher, G.C., and Pienaar, R.N. (2010). Ecophysiological studies on Aureococcus anophagefferens blooms in Saldanha Bay, South Africa. Harmful Algae *9*, 123–133.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., and Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. *12*, R22.

Rosenbluth, M.J., Lam, W.A., and Fletcher, D.A. (2006). Force microscopy of nonadherent cells: a comparison of leukemia cell deformability. Biophys. J. *90*, 2994–3003.

Siepel, A., and Haussler, D. (2004). Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. Mol. Biol. Evol. *21*, 468–488.

Song, J., Rechkoblit, O., Bestor, T.H., and Patel, D.J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science *331*, 1036–1040.

Song, J., Teplova, M., Ishibe-Murakami, S., and Patel, D.J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science *335*, 709–712.

Teves, S.S., and Henikoff, S. (2012). Salt fractionation of nucleosomes for genome-wide profiling. Methods Mol. Biol. Clifton NJ *833*, 421–432.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinforma. Oxf. Engl. *25*, 1105–1111.

Unk, I., Hajdú, I., Blastyák, A., and Haracska, L. (2010). Role of yeast Rad5 and its human orthologs, HLTF and SHPRH in DNA damage tolerance. DNA Repair *9*, 257–267.

Van der Wal, P., de Jong, E.W., Westbroek, P., de Bruijn, W.C., and Mulder-Stapel, A.A. (1983). Ultrastructural polysaccharide localization in calcifying and naked cells of the coccolithophorid Emiliania huxleyi. Protoplasma *118*, 157–168.

Webster, M.T., McCaffery, J.M., and Cohen-Fix, O. (2010). Vesicle trafficking maintains nuclear shape in Saccharomyces cerevisiae during membrane proliferation. J. Cell Biol. *191*, 1079–1088.

Worden, A.Z., Nolan, J.K., and Palenik, B. (2004). Assessing the dynamics and ecology of marine picophytoplankton: The importance of the eukaryotic component. Limnol. Oceanogr. *49*, 168–179.

Worden, A.Z., Lee, J.-H., Mock, T., Rouzé, P., Simmons, M.P., Aerts, A.L., Allen, A.E., Cuvelier, M.L., Derelle, E., Everett, M.V., et al. (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. Science *324*, 268–272.

Wu, J.-Q., and Pollard, T.D. (2005). Counting cytokinesis proteins globally and locally in fission yeast. Science *310*, 310–314.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916–919.