Histone modifications during the life cycle of the brown alga *Ectocarpus*

Simon Bourdareau[1], Leila Tirichine[2], Bérangère Lombard[3], Damarys Loew[3], Delphine Scornet[1], Susana M. Coelho[1,*], J. Mark Cock[1,*]

[1]CNRS, Sorbonne Université, UPMC University Paris 06, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff, France, [2]Université de Nantes, CNRS, UFIP, UMR 6286, F-44000 Nantes, France, [3]Institut Curie, PSL Research University, Centre de Recherche, Laboratoire de Spectrométrie de Masse Protéomique, 26 rue d'Ulm, 75248 Cedex 05 Paris, France.

*Authors for correspondence (cock@sb-roscoff.fr; coelho@sb-roscoff.fr)

Abstract

Background: Brown algae evolved complex multicellularity independently of the animal and land plant lineages and are the third most developmentally complex phylogenetic group on the planet. An understanding of developmental processes in this group is expect to provide important insights into the evolutionary events necessary for the emergence of complex multicellularity. Here we have focused on mechanisms of epigenetic regulation involving post-translational modifications (PTMs) of histone proteins.

Results: A total of 45 histone PTMs were identified, including two novel marks, but *Ectocarpus* lacks both H3K27me3 and the major polycomb complexes. ChIP-seq identified PTMs associated with transcription start sites (TSSs) and gene bodies of active genes, and with transposons. H3K79me2 exhibited an unusual pattern, often marking large genomic regions spanning several genes. TSSs of closely spaced divergently transcribed gene pairs shared a common nucleosome depleted region and exhibited shared histone PTM peaks. Overall, patterns of histone PTMs were stable through the life cycle. Analysis of histone PTMs at generation-biased genes provided insights into their functions during gene induction.

Conclusions: The overview of the nature and functions of histone PTMs in the brown algae presented here will provide a foundation for future studies aimed at understanding epigenetic processes in the brown algae.

Keywords: brown algae, ChIP-seq, chromatin, *Ectocarpus*, gametophyte, histone modification, life cycle, multicellularity, polycomb complex, sporophyte

Introduction

Very few eukaryotic lineages have independently evolved complex multicellularity and the brown algae are of particular interest as the third most developmentally complex lineage after animals and land plants. The size and complexity attained by some brown algae is illustrated by the giant kelp *Macrocystis,* which can reach up to 50 metres in length.

The deployment of developmental programs and the establishment of different cell fates in complex multicellular organisms implies the acquisition of different epigenetic states in different cell types. A number of mechanisms underlie the establishment and maintenance of epigenetic states. These include, for example, cell signalling and regulation of gene transcription but another important factor is chemical modification of chromatin. In particular, histones have been shown to undergo a broad range of different post-translational modifications (PTMs) involving the addition of various chemical moieties to multiple amino acid residues, particularly within the unstructured amino-terminal "tails" of these proteins [1]. These modifications affect chromatin function either by directly modifying interactions between the different components of the nucleosome or via the action of proteins that bind to specific histone modifications and effect specific tasks. In this manner, histone PTMs act as a "histone code", mediating the establishment and maintenance of different epigenetic states across the genome.

Epigenetic processes not only allow cells that carry the same genomic information to assume different functions in different parts of a multicellular organism but they also allow temporal changes, i.e. establishment of different cellular functions at different stages of the life history. This latter aspect is particularly important in organisms that have complex life cycles. For example, many brown algae have haploid-diploid life cycles involving an alternation between two different organisms, the sporophyte and the gametophyte, often with very distinct bodyplans [2]. These life cycles imply complex regulation of epigenetic states but, at present, very little is known about the role of epigenetic processes in the regulation of developmental and life cycle processes in the brown algae.

The filamentous alga *Ectocarpus* is being used as a model system to study brown algal developmental biology [3–7]. *Ectocarpus* has a haploid-diploid life cycle involving an alternation between a gametophyte, which is usually haploid, and a sporophyte, which is usually diploid. However, there is clear evidence that the identity of each life cycle generation is not determined by its ploidy because haploid sporophytes (partheno-sporophytes) can be produced by parthenogenetic development of haploid gametes [8,9] and diploid gametophytes can be

constructed using mutants that are unable to deploy the sporophyte developmental pathway [3,7]. These observations indicate that epigenetic processes play an important role during the *Ectocarpus* life cycle. Recent work has shown that the deployment of the sporophyte program requires two different three amino acid loop extension homeodomain transcription factors (TALE HD TFs), OUROBOROS (ORO) and SAMSARA (SAM) [7]. Remarkably, TALE HD TFs appear to have been recruited convergently to regulate sporophyte development in both the brown algal and the land plant lineages [7]. In land plants, the PRC2 polycomb complex has also been implicated in life cycle regulation [10–17], indicating that chromatin modification processes play in important role in life cycle regulation in that lineage. Chromatin modification has been proposed to play a similar role in the brown algae [2] but this hypothesis has not been investigated experimentally.

In this study, we have carried out a broad census of histone PTMs in *Ectocarpus* chromatin and have developed a method to evaluate the genome-wide distribution of specific histone PTMs in this species. Application of this method allowed the identification of histone PTMs associated with transcriptional start sites (H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac) and gene bodies (H3K36me3) of actively transcribed genes and a histone PTM associated with transposons and repetitive sequences (H4K20me3). We also show that H3K79me2 often marks extensive regions spanning several genes and suggest some possible functions for this PTM. Overall, genome-wide histone PTM patterns were found to remain stable following transition between the sporophyte and gametophyte generations of the life cycle, consistent with the observation that only 4% of genes exhibited a significant level of generation-biased expression. Analysis of generation-biased genes allowed changes in chromatin state (combinations of histone PTMs) to be related to changes in gene expression.

Results

Ectocarpus histones and histone modifier enzymes

Analysis of the *Ectocarpus* genome sequence [18] identified 34 core histone and nine H1 histone genes (Additional file 5: Table S1). Four of the core histone genes are predicted to encode variant forms, including probable CenH3, H2A.X and H2A.Z proteins. All nine H1 histone genes appear to encode *bona fide* H1 proteins and are not members of related families such as the plant GH1-HMGA or GH1-Myb families [19]. All but eight of the histone genes were located in five gene clusters on chromosomes 4, 7 and 26 and on an unmapped scaffold (sctg_442). The organisation of the clusters suggests multiple duplication, rearrangement and

fragmentation of an ancestral cluster with the organisation H4, H1, H3, H2B, H2A (Additional file 2: Figure S1).

A search for genes encoding histone modifying enzymes identified both putative histone acetyltransferases and methyltransferases, and predicted deacetylase and demethylase enzymes (Additional file 6: Table S2). Most of the acetyltransferases were sufficiently similar to well-characterised homologues to allow prediction of their target residues but the methyltransferases tended to be less conserved at the sequence level and, in many cases, had novel domain structures. Direct functional information, for example based on mutant analysis, will therefore be necessary to investigate the specificity of the *Ectocarpus* methyltransferases.

Identification of histone PTMs in *Ectocarpus*

Histone PTMs were detected using mass spectrometry of enzyme-digested histone preparations. In addition, a broad range of commercially available antibodies were tested against *Ectocarpus* histone preparations on western blots to further confirm the presence of a subset of the PTMs identified by mass spectrometry. A total of 45 PTMs of core and variant histones were detected in *Ectocarpus* (Fig. 1a, Additional file 3: Figure S2). Six of these marks were only detected by western blotting and should therefore be treated with caution (marked with an asterisk in Fig. 1a, Additional file 7: Table S3). Note that two of these PTMs, H3K9me2 and H3K9me3, were also not detectable using mass spectrometry in *Arabidopsis* but have been detected using immunoblotting [20,21]. Most of the histone PTMs detected have been reported previously in species from one or more of the land plant, animal or fungal lineages, either at exactly the same position or at an equivalent position in the corresponding protein (Fig. 1b, Additional file 7: Table S3). However, a number of marks have only been described so far in stramenopiles. Of these, some PTMs such as acetylation of lysines nine and 15 of H2A.Z were detected in both *Ectocarpus* and the diatom *P. tricornutum*, whereas others were detected only in the diatoms (e.g. acetylation of lysine 59 on H4) or only in *Ectocarpus* (e.g. acetylation and methylation of lysine 20 and arginine 38, respectively, on H2A.Z; Fig. 1b).

Genome-wide distribution of selected histone PTMs

To investigate the functions of histone PTMs in *Ectocarpus*, we analysed the distribution of eight selected marks across the genome using chromatin immunoprecipitation and sequencing (ChIP-seq). The PTMs analysed were H3K4me2, H3K4me3, H3K9ac, H3K14ac, H3K27ac, H3K36me3, H3K79me2 and H4K20me3 (Fig. 2). Five of these PTMs, H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac, were preferentially associated with the transcription start

sites (TSSs) of genes but tended to be depleted from gene bodies (Fig. 2a,b,c). Genome-wide patterns of these five PTMs exhibited a range of levels of positive correlation (Pearson's r 0.13 to 0.86; Fig. 2d), consistent with these marks having overlapping functions. Similar PTM patterns were detected at protein-coding and lncRNA genes (Fig. 2c), suggesting that histone PTMs may be used in a similar manner to regulate these two types of gene. For each of the five TSS-associated PTMs a peak was detected within 500 bp of the TSS for between 77% and 83% of the genes in the genome. For H3K4me2, a pair of adjacent peaks was detected, one on each side of the TSS (Fig. 2b,c). Genome-wide, peaks of H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac marks were associated with between 13, 416 and 14,423 of the 17,406 genes. We noted a strong positive correlation between gene expression level, measured in transcripts per million (TPM) and the strength of mapping of each mark to TSSs (Fig. 3). Taken together, these observations indicate a strong association of H3K14ac, H3K4me3, H3K9ac, H3K14ac and H3K27ac with transcriptionally active genes in *Ectocarpus*.

The three remaining PTMs, H3K36me3, H3K79me2 and H4K20me3, were depleted from TSSs (Fig. 2). H3K36me3, which was detected at 12,863 genes (74% of the genome), was most strongly associated with gene bodies and was depleted from both TSSs and transcription end sites (TESs). As observed for the TSS-localised PTMs discussed above, presence of H3K36me3 on gene bodies was positively correlated with expression, indicating that this mark is also associated with gene activation (Fig. 3).

Almost all (94.6%) of the H4K20me3 peaks were associated with repeated sequences or transposons, predominantly in intergenic regions but also in introns. Note that introns make up approximately 40% of the *Ectocarpus* genome [22], and therefore represent a significant proportion of the non-coding fraction of the genome. The distribution of H4K20me3 was consistent with it playing a role in the regulation of inserted transposons, possibly in the silencing of these elements. Genes that were strongly marked with H4K20me3 tended to have very low expression levels (Fig. 3). This correlation could be interpreted as indicating a role for H4K20me3 in gene regulation but, given the co-localisation of this PTM with transposon sequences, a more likely explanation is that the observed effect on gene expression was an indirect effect of silencing of intronic transposon sequences.

The genome is partitioned into H3K79me2-marked and H3K79me2-depleted regions

H3K79me2 exhibited an unusual distribution pattern. This PTM was detected throughout the genome, often in large, discrete, continuously marked regions of several kilobases separated by H3K79me2-depleted regions (Fig. 4a). Together the H3K79me2 regions covered 74.5 Mbp,

37.2% of the genome. About a third (36.6%) of the H3K79me2 regions were longer than 5 kbp and many of these regions spanned more than one gene (on average, H3K79me2 regions of >5 kbp contained 1.85 genes). The borders of the long H3K79me2 regions tended to be localised near a TSS (when H3K79me2 regions of >5 kbp were considered, 42.2% of the borders were located within a kilobase of a TSS). Comparison of ChIP-seq data for the sporophyte and gametophyte generations of the life cycle indicated that the H3K79me2 regions were stably maintained throughout the life cycle. Only 0.3% or 1.2% of H3K79me2 regions longer than 5 kbp were detected uniquely in the sporophyte or the gametophyte generation, respectively. Genes within H3K79me2 zones were significantly shorter, possessed fewer introns and their transcripts were less abundant on average than genes that were outside the H3K79me2 regions (Fig. 4b-d). The TSSs of genes within the H3K79me2 regions were also significantly less strongly marked by H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac (Fig. 4e). Consistent with these observations, there was a negative correlation between gene expression level (measured in TPM) and the quantitative presence of H3K79me2 over genes (Fig. 3). However, based on a similar argument to that proposed for H4K20me3 above, the organisation of deposition of H3K79me2 across blocks of contiguous genes (Fig. 4) does not appear to be consistent with a direct role in gene regulation and we therefore favour the hypothesis that the effect of H3K79me2 on gene regulation is an indirect consequence of the involvement of this PTM in another, currently unknown, role in chromatin homeostasis.

Chromosome six contains an integrated copy of a large DNA virus (spanning 0.3 Mbp from approximately 4,244,200 to 4,547,200) that is closely related to the *Ectocarpus* phaeovirus EsV-1. The inserted virus has been shown to be transcriptionally silent [18,22]. A large H3K79me2 region of about 0.42 Mbp was detected that spanned the entire inserted viral genome (Additional file 4: Fig. S3). Chromatin within this H3K79me2 region was also marked with H4K20me3 but the other histone PTMs assayed were depleted from the inserted viral genome (Additional file 4: Figure S3).

Overlapping TSS regions of divergently transcribed gene pairs

One unusual feature of the *Ectocarpus* genome, compared to genomes of similar size, is that there is a relatively strong tendency for adjacent genes to be transcribed from opposite strands of the DNA helix [22]. Consequently, the genome contains many divergently transcribed gene pairs. To investigate the effect of this pattern of gene organisation on the chromatin characteristics of TSSs, we searched for pairs of adjacent genes located on the same sequence scaffold. This search identified 10,399 TSS-adjacent intergenic regions and 61.7% of the genes

flanking these regions were part of a divergently transcribed gene pair. The intergenic regions of divergently transcribed gene pairs were significantly shorter than those of tandem gene pairs (median 409 and 2293, respectively, Wilcoxon test, *p*-value<$2e^{-16}$; Fig. 5a). When the intergenic regions of divergent gene pairs were shorter than about 600 bp, the two TSS chromatin regions overlapped and shared the same nucleosome depleted region (NDR) based on micrococcal nuclease digestion data (Fig. 5a). This overlap correlated with the presence of double peaks for the PTMs H3K4me3, H3K9ac, H3K14ac and H3K27ac, one on each side of the TSS (Fig. 5a,b). For H3K4me2, which was detected as a double peak at most individual TSSs (Fig. 2b), the two peaks were further apart at the overlapping TSSs of divergent gene pairs than they were at the single intergenic TSSs of tandem gene pairs (Fig. 5a,b). Therefore, the pattern of PTMs also indicated that the TSSs of proximate divergent genes are located within a shared chromatin domain. Signals for all five of the above PTMs were significantly stronger for divergent than for tandem gene pairs (Fig. 5c, measured at the TSS of the second gene of the gene pair, see schema in Fig. 5a). This difference appeared to be due to the overlap of the two divergent TSS regions as the signal markedly increased for intergenics of divergent gene pairs that were shorter than 1 kbp (Fig. 5c). Consistent with this observation, no significant difference was detected between the median expression levels of divergent and tandem genes (Fig. 5d). Note that transcript abundances were not correlated for the two genes of a divergent gene pair (Pearson coefficients were 0.042 for all the divergent pairs and 0.054 for divergent pairs closer than 600 bp; Additional file 4: Figure S4), despite the presence of a shared chromatin domain at the TSSs and correlated histone PTM signals immediately downstream of each TSS when divergent genes were closer than 600 bp (Additional file 4: Figure S4).

Epigenetic reprogramming during the *Ectocarpus* life cycle

To relate patterns of histone modification to epigenetic reprogramming of gene expression, ChIP-seq analysis was used to compare the distributions of histone PTMs during the sporophyte and gametophyte generations of the *Ectocarpus* life cycle. Overall, the distribution of PTMs was stable between the two life cycle generations. For example, the six PTMs associated with actively expressed genes, H3K4me2, H3K4me3, H3K9ac, H3K14ac, H3K27ac and H3K36me3, were either stably present or stably absent for between 82.7% and 97.1% of the 17,406 genes genome-wide, depending on the PTM. This analysis indicated that there were not any major, genome-wide changes in patterns of histone PTMs associated with alternation between life cycle generations. To focus more specifically on changes in patterns of histone PTMs associated with life-cycle-related changes in gene expression, we analysed epigenetic

marks at genes that were differentially regulated between the two life cycle generations (Fig. 6a).

A comparison of gene expression patterns in the sporophyte and gametophyte, based on RNA-seq data generated using the same biological samples as were used for the ChIP-seq analysis, identified 774 genes that exhibited a generation-biased pattern of expression (padj<0.05, fold change>2, TPM>1; Fig. 6a, Additional file 5: Table S4). We will refer to these 774 genes as generation-biased genes (GBGs). Interestingly, this set of differentially expressed genes was significantly enriched in long non-coding RNAs (lncRNAs; Table S5; χ^2 test, $1.52e^{-15}$) and 19 of the 72 differentially expressed lncRNAs were adjacent to a differentially expressed protein-coding GBG (including four lncRNA/protein-coding gene pairs with overlapping transcripts; Table S5). In most cases (84%), the lncRNA and its adjacent protein-coding gene were co-ordinately upregulated during the same generation of the life cycle. These observations suggest a possible role for lncRNAs in the regulation of adjacent genes in *Ectocarpus*.

To analyse changes in histone PTMs at GBGs between life cycle generations, 16 chromatin states were defined based on different combinations of four TSS-localised PTMs that exhibited clear patterns of presence and absence: H3K4me2, H3K4me3, H3K9ac and H3K27ac (Fig. 6b). In most cases when GBGs changed chromatin state during the transition between the two generations, increased transcript abundance was correlated with acquisition of additional activation-associated PTMs (76.1% of the GBGs that changed chromatin state). However, the majority of GBGs (61.1%) did not change chromatin state between generations, although we noted that most of these genes (81.4%) were marked with all four activation-associated PTMs during both generations. Taken together, these observations are consistent with a correlation between the presence of the four TSS-localised marks and gene activation but indicate that additional mechanisms may be necessary to induce differential gene expression. Therefore, the histone PTMs may act to facilitate or potentiate transcription rather than directly activating gene transcription.

We detected more sporophyte-biased (503) than gametophyte-biased (271) genes (χ^2 test, $7.49e^{-17}$) and the former were significantly more strongly upregulated than the latter (4.2 and 2.8 mean log2TPM fold changes, respectively; Kruskal-Wallis rank sum test *p*-value = $4.0e^{-11}$). Overall, sporophyte-biased and gametophyte-biased genes showed similar patterns of transitions to new chromatin states (Spearman's rank correlation rho 0.61, *p*-value = 0.011). However, the sporophyte-biased genes exhibited a stronger tendency to switch to chromatin state 1 (presence of all four activation-associated PTMs) than the gametophyte-biased genes (74.8% compared with 26.4% of chromatin state transitions, respectively; z score 7.7857,

$p<0.00001$). This observation is consistent with the larger fold changes in transcript abundance (TPM) observed for sporophyte-biased genes as transitions to chromatin state 1 were associated with significantly larger fold changes than transitions to other states (Wilcoxon test, p-value<0.0078 for all GBGs). We also noted that the GBGs that exhibited the highest fold changes also exhibited a stronger tendency to change chromatin state between generations (Fig. 6c). This was true for both sporophyte-biased and gametophyte-biased genes.

Ectocarpus lacks a polycomb repressive complex 2 and H3K27me3

Polycomb repressive complexes have been shown to play important roles in the regulation of a broad range of developmental processes in both animals and land plants [23,24] and a key role has been proposed for Polycomb repressive complex 2 (PRC2) in the regulation of life cycle transitions in both *Arabidopsis* [10,12–15] and the moss *Physcomitrella patens* [16,17]. However, in contrast to the conservation and functional importance of these complexes in the land plant and animal lineages, a homology search failed to identify the core proteins of PRC2 in the *Ectocarpus* genome, and a similar result was obtained for PRC1 ([2]; Fig. 7, Additional file 5: Table S6). This analysis did identify an orthologue of the WD domain protein RbAp48 but this protein is also known to be a key component of other regulatory complexes in the cell, such as CAF-1, NuRD and NURF [25]. We cannot rule out the possibility that the brown algae possess highly diverged versions of the PRC2 and PRC1 complexes but this would seem unlikely as the evolutionary time separating brown algae from plants and animals is similar to that separating the latter two lineages. Moreover, orthologues of all the PRC2 and PRC1 core components were found in the genome of the diatom *Phaeodactylum tricornutum*, which also belongs to the stramenopile lineage ([26]; Additional file 5: Table. S6). Analysis of the limited number of complete genomes available for the stramenopiles indicated that PRC1 was lost after divergence from the diatoms and PRC2 at a later stage after divergence from the Pelagophyceae (Fig. 7b). Absence of PRC2 in *Ectocarpus* was supported by the fact that the mass spectrometry analysis did not detect any evidence of tri-methylation of H3K27. Histone western blots with an antibody raised against H3K27me3 (Additional file 5: Table S7) weakly detected a protein of the expected size but when this antibody was employed in a ChIP-seq experiment we did not detect any specific chromatin immunoprecipitation (data not shown). Therefore, taken together, these analyses indicated that *Ectocarpus* lacks the PRC2 (and probably also the PRC1) polycomb complex and tri-methylation of H3K27.

Discussion

Epigenetic regulation in a multicellular brown alga

Brown algae are the third most developmentally complex phylogenetic group on the planet and include members whose bodyplans rival those of land plants in their cellular and developmental complexity. However, compared to animals and land plants, many aspects of brown algal developmental biology are still very poorly understood, including epigenetic regulatory mechanisms. The objective of this study was to provide a comprehensive overview of histone PTMs in the model brown alga *Ectocarpus* and to investigate the relationship between patterns of key PTMs across the genome and the developmental programs that mediate alternation between the sporophyte and gametophyte generations during this seaweed's life cycle.

Mass spectrometry analysis of *Ectocarpus* histone preparations demonstrated that brown algal histones are subject to a broad range of PTMs. Most of the modifications detected had been reported previously for histones of organisms from other eukaryotic supergroups, such as the land plants within the Archaeplastida or animals and fungi from the Opisthokonta. However, the analysis also confirmed some stramenopile-specific histone PTMs such as H2A.ZK9ac and H2A.ZK15ac and detected two previously unreported (and therefore possibly brown-algal-specific) modifications, H2A.ZK20ac and H2A.ZR38me1. Overall, however, this study confirmed the conclusion, based on an analysis of diatom histone PTMs [27], that a large number of histone PTMs detected in other eukaryotic supergroups are conserved in the stramenopiles. This conclusion is consistent with the observation that histone molecules have been highly conserved during eukaryotic evolution [28,29] and with evidence that many of the proteins involved in post-translational modification of histones can be traced back to the last eukaryotic common ancestor [30]. Indeed, genes encoding histone modifying enzymes with highly conserved domain structures and domain sequences, such as the histone acetyltransferases, were detected in the *Ectocarpus* genome. For other families, it was difficult to assign predicted proteins to specific functions due to sequence divergence. Experimental analysis of protein function will therefore be necessary to identify the specific writers and readers of many histone PTMs in brown algae.

The *Ectocarpus* genome does not appear to encode the core components of the polycomb complexes PRC1 and PRC2 and mass spectrometry analysis of *Ectocarpus* histones did not detect any evidence for the H3K27me3 PTM in *Ectocarpus* chromatin. Absence of polycomb complexes in a complex multicellular organism may seem surprising from the perspective of animal and land plant model systems but the components of PRC2 are known to

exhibit a patchy distribution across the diverse eukaryotic supergroups [31], indicating that they have been repeatedly lost over evolutionary time. In this respect, the absence of polycomb complexes in brown algae represents a striking example of how comparative analysis of groups that have evolved complex multicellularity independently of plants and animals can be used to test the degree to which specific genetic systems (in this case polycomb regulation) have been essential for the evolution of complex multicellularity.

Analysis of ChIP-seq data allowed inferences to be made about the functional roles of eight histone PTMs in brown algae. H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac were all detected predominantly at TSSs and the degree to which genes were marked with these PTMs was proportional to their level of expression (transcript abundance), indicating a role in promoting or facilitating gene transcription. Further evidence supporting a role in promoting gene expression was provided by a comparative analysis of chromatin states at GBGs between life cycle generations, which indicated a broad correlation between accumulation of H3K4me2, H3K4me3, H3K9ac and H3K27ac at the TSS and increased expression of GBGs during the life cycle. However, patterns of histone PTMs at TSSs were not strictly correlated with gene expression indicating that these marks facilitate transcription but are not sufficient alone to induce gene expression. This conclusion was supported by a genome-wide analysis of closely spaced (<600 bp) pairs of divergently transcribed genes, which showed that the TSS regions of gene pairs overlapped but the two genes were nonetheless independently regulated. At present, it is not clear what processes are necessary, in addition to the deposition of histone PTMs at the TSS, to induce gene expression in *Ectocarpus*, but these could include, for example, the presence of specific transcription factors and the transcription machinery or the presence of additional histone PTMs, including alternative acyl marks such as propionylation, butyrylation and crotonylation which have been shown to stimulate transcription in cell-free assays [32].

Peaks of H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac are detected near the TSSs of active genes in both animals and land plants and all of these marks have been associated with active gene expression in both lineages ([33–44]; with the possible exception of H3K4me2, which may represent a repressive mark on plants; [44,45]). Overall, therefore, these five histone PTMs appear to have similar roles associated with gene activation at promoters in animals, land plants and brown algae, indicating that at least part of the "histone code" may be universal and therefore have a very deep evolutionary origin.

H3K36me3 marked gene bodies and was absent from TSSs and TESs but its presence was also positively correlated with the level of gene expression. H3K36me3 also marks the bodies of active genes in both animals and land plants, although its distribution along the gene

body differs in the two lineages. In *Arabidopsis* and rice a strong peak was observed at the 5' end of the gene [40,43,46], whereas strong signals have been reported at the 3' ends [47] and over exons [48] of animal genes. H3K36me3 is catalysed by SETD2, which is associated with Pol II and is therefore directly linked to transcript elongation. H3K36me3 has also been shown to be involved in alternative splicing in animals and plants [49,50]. The position of H3K36me3 within *Ectocarpus* genes and the positive correlation with transcript abundance are consistent with an evolutionarily conserved role associated with transcript elongation.

H4K20me3, was associated with transposons and repeat sequences, both between and within genes. The presence of this mark at intronic transposons was inversely correlated with levels of gene expression but its effect on gene expression may be indirect, for example as a consequence of silencing intronic transposons. The distribution of H4K20me3 suggests a similar role to that observed in animals, where H4K20me3 is enriched at transposons and in heterochromatin [51–53] and where it has been shown to repress transposons [53]. Note that H4K20me3 appears to have a different role in land plants as it localises to euchromatin and is associated with transcriptional activation [54].

H3K79me2 exhibited an unusual distribution pattern, marking discrete regions of the genome that often spanned multiple genes. The function of the H3K79me2 is not clear, although H3K79me2-marked regions commonly included transposons (88.6% of H3K79me2 regions of >5 kbp included a ≥400 bp repeated element) and the inhibitory effect of H3K79me2 on gene expression (reduced transcript abundance) would be consistent with a role in the silencing of repeated elements (based on a similar argument to that proposed above for H4K20me3). If H3K79me2 does have a role in the silencing of transposons, it is not clear why the H3K79me2 regions often cover much larger regions than the transposon(s) itself but one possible explanation may be that the regions extend to neighbouring TSSs in order to regulate transcriptional activity in a genomic region containing a transposon. We also noted that H3K79me2 tended to mark convergent gene pairs, even in regions that contained little repeated sequence. A possible additional role of H3K79me2 at these loci may be to regulate transcription in some manner to limit or respond to the formation of double-stranded RNA due to transcript overlap. Such double-stranded RNA could be problematic within the cell if it activates small RNA pathways but note that an earlier analysis did not find any evidence that convergent gene pairs were a preferential source of sRNAs in *Ectocarpus* [55].

The distribution pattern observed for H3K79me2 in *Ectocarpus* is very different from the pattern observed in animals, where H3K79me2 marks the 5' part of the gene body as part of a mutually exclusive pattern with H3K36me3, which marks the 3' part [48,56,57]. The

H3K79me2/H3K36me3 boundary corresponds to the first internal exon [48]. In animals, both H3K79me2 and H3K36me3 are considered to be associated with active transcription [56–58], and more specifically transcript elongation, although conflicting evidence has been reported [47]. In addition, a role has been proposed for H3K79me2 in the regulation of alternative splicing [59], although again there is evidence to the contrary [48]. H3K79 methylation has not been detected in land plants [21]. Based on the different distribution patterns, H3K79me2 would appear to have different roles in animals and in brown algae.

We also observed that a region of chromosome six, containing a large, inserted DNA virus genome, was extensively marked with both H3K79me2 and H4K20me3. The viral genes at this locus have been shown to be transcriptionally silent in a range of tissues and life cycle stages [18], suggesting possible roles for H3K79me2 and H4K20me3 in defending the alga against lysogenic viral infections by silencing inserted viral genomes.

Epigenetic regulation during the *Ectocarpus* life cycle

To our knowledge this is the first study to have compared patterns of histone PTMs across the two generations of a haploid-diploid life cycle (although previous studies have looked at PTMs associated with the repression of germline genes during the sporophyte generation; [60,61]). Overall, genome-wide patterns of histone PTM were found to be stable between the *Ectocarpus* sporophyte and gametophyte generations, with the marks deposited at most genes remaining unchanged. This stability was consistent with the observation that only about 4% of genes, genome-wide, were significantly differentially expressed between the two generations. The *Ectocarpus* sporophyte and gametophyte are very similar morphologically (near-isomorphy) and it is possible that more marked modifications of the epigenetic landscape occur during the life cycles of brown algal species (such as kelps) that exhibit greater differences between the sporophyte and gametophyte generations. However, note that there does not appear to be a strong correlation between the degree of morphological heteromorphy between life cycle generations and the proportion of GBGs in the genome [62]. Analysis of the epigenetic status of GBGs during the life cycle indicated that multiple TSS-localised histone PTMs are involved in facilitating or potentiating gene upregulation.

Conclusion

This study provides a first overview of the nature and functions of histone PTMs in the brown algae, a group that has evolved complex multicellularity independently of the animal and land plant lineages. The general, emerging picture is that epigenetic regulatory mechanisms are

broadly conserved with those of other eukaryotic supergroups with some key differences that will potentially provide important insights into the epigenetic regulation of developmental processes, including novel histone modifications, unusual patterns of deposition of known PTMs such as H3K79me2 and the absence of polycomb complexes. The data and resources generated by this study will provide a foundation for future studies aimed at understanding epigenetic processes in the brown algae.

Material and Methods

Strains and growth conditions

The *Ectocarpus* sp. strain used in this study was Ec32 [63]. *Ectocarpus* was cultivated as described previously [64]. Cultures were grown at 13°C with a 12h/12h day/night cycle and 20 µmol photons.m^{-2}.s^{-1} irradiance.

Detection of histone PTMs using mass spectrometry

Ectocarpus histone proteins were isolated using the method described by Tirichine *et al*. [65]. Briefly, histones were extracted from isolated nuclei in 1 M $CaCl_2$, 20 mM Tris HCl pH 7.4 in the presence of a cocktail of protease inhibitors. After precipitation of the acid-insoluble fraction in 0.3 M HCl, the histones were precipitated by dropwise addition of TCA, centrifuged and the pellet washed with 20% TCA and 0.2% HCl.

Gel purification and digestion of histones and mass spectrometry analysis were carried out essentially as described by Veluchamy *et al*. [27]. Briefly, histone proteins, excised from a 14% SDS-polyacryamide gel, were digested overnight with endoproteinase (12.5 ng/µl), trypsin (Promega), chymotrypsin (12.5 ng/µl, Promega), ArgC (12.5 ng/µl, Promega) or elastase (20 ng/µl Sigma-Aldrich). Peptides were analysed using a Q Exactive HF-X (Thermo Scientific) or a TripleTOF™ 6600 (ABSciex) mass spectrometer coupled to a RSLCnano system (Ultimate 3000, Thermo Scientific, Bremen, Germany). Spectra were generated using Xcalibur (version 2.0.7) and analysed with Mascot™ (version 1.4, Thermo Scientific) using an in-house database consisting of the complete *Ectocarpus* proteome available at Orcae [66]. The mass spectrometry proteomics data have been deposited in the PRIDE database [67] via ProteomeXchange with identifier PXD013535. Details of the liquid chromatography-tandem mass spectrometery parameters are provided in the Supplementary Methods.

Detection of histone PTMs using western blots

Commercially available antibodies (Additional file 5: Table S7) for a broad range of histone PTMs were tested against *Ectocarpus* chromatin preparations using western blots as described previously [68].

Genome-wide detection of histone PTMs

The ChIP-seq experiments were carried out in two batches, the first for H3K4me3, H3K9ac, and H3K27ac, the second for H3K4me2, H3K14ac, H3K36me3, H4K20me3, H3K79me2 and H3K27me3. RNA-seq data (see below) was generated for both batches to ensure that the histone PTM and gene expression data were fully compatible. For ChIP-seq, *Ectocarpus* tissue was fixed for five minutes in seawater containing 1% formaldehyde and the formaldehyde eliminated by rapid filtering followed by incubation in PBS containing 400 mM glycine. Nuclei were isolated by grinding in liquid nitrogen and in a Tenbroeck Potter in nuclei isolation buffer (0.1% triton X-100, 125 mM sorbitol, 20 mM potassium citrate, 30 mM MgCl2, 5 mM EDTA, 5 mM β-mercaptoethanol, 55 mM HEPES at pH 7.5 with cOmplete ULTRA protease inhibitors), filtering through Miracloth and then washing the precipitated nuclei in nuclei isolation buffer with, and then without, triton X-100. Chromatin was fragmented by sonicating the purified nuclei in nuclei lysis buffer (10 mM EDTA, 1% SDS, 50 mM Tris-HCl at pH 8 with cOmplete ULTRA protease inhibitors) in a Covaris M220 Focused-ultrasonicator (duty 25%, peak power 75, cycles/burst 200, duration 900 sec at 6°C). The chromatin was incubated with an anti-histone PTM antibody (Additional file 5: Table S7) overnight at 4°C and the immunoprecipitation carried out using Dynabeads protein A and Dynabeads protein G. Following immunoprecipitation and washing, a reverse cross-linking step was carried out by incubating for at least six hours at 65°C in 5 M NaCl and the samples were then digested with Proteinase K and RNAse A. Purified DNA was analysed on an Illumina HiSeq 4000 platform with a single-end sequencing primer over 50 cycles. At least 20 million reads were generated for each immunoprecipitation (Additional file 5: Table S8).

Nuclei were prepared for MNase digestion using the same procedure as that used for ChIP samples except that the isolation buffer did not contain EDTA. Each sample of nuclei was digested with 400 agarose gel units of MNase for 10 min at 37°C in MNase reaction buffer (5 mM $CaCl_2$, 60 mM KCl, 0.5 mM DTT, 15 mM NaCl, 125 mM sorbitol, 50 mM Tris-HCl pH 7.5). Following lysis of the nuclei, reverse cross-linking and digestion with Proteinase K and RNAse A, 120-210 bp fragments were excised from a 2% agarose gel and sequenced on a HiSeq 4000 platform (Additional file 5: Table S8).

The ChIP-seq and MNase digestion datasets have been deposited in the NCBI Gene Expression Omnibus database, along with the associated RNA-seq data, under the accession number GSE146369.

Quality control of the sequence data was carried out using FASTQC [69]. Poor quality sequence was removed and the high quality sequence was trimmed with Flexcat [70]. Illumina reads were mapped onto the *Ectocarpus* genome ([18]; available at Orcae; [66]) using Bowtie [71]. Peaks corresponding to regions enriched in PTMs were identified using the MACS2 (version 2.1.1) callpeak module (minimum FDR of 0.01) and refined with the MACS2 bdgpeakcall and bdgbroadcall modules [72]. Co-localised peaks corresponding to regions enriched in several PTMs were detected using MACS2 outputs in BedTools multiinter [73]. Overlaps of PTM peaks with other genomic features were analysed using BedTools intercept [73]. Heatmaps, average tag graphs and coverage tracks were plotted using EaSeq software [74] or pyGenomeTracks (https://pypi.org/project/pyGenomeTracks/). Circos graphs were generated using Circos software [75]. These analyses were carried out for two biological replicates for each PTM during both the sporophyte and gametophyte generations of the life cycle. Pearson correlation analysis between replicates was performed with DeepTools 2.5.4. Replicate samples were strongly correlated (Pearson correlations of 0.87 to 0.99; Additional file 4: Figure S5). Normalisation was carried out using the input chromatin data (Additional file 5: Table S8). Genome-wide analyses (e.g. Fig. 2) used data from sporophytes unless stated otherwise.

Comparisons of sporophyte and gametophyte transcriptomes using RNA-seq

RNA for transcriptome analysis was extracted from the same duplicate sporophyte and gametophyte cultures that were analysed for each of the two ChIP-seq experiments using the Qiagen RNeasy plant mini kit with an on-column Deoxyribonuclease I treatment. RNA quality and concentration was then analysed on a Qubit® 2.0 Fluorometer using Qubit RNA BR assay kit (Invitrogen, Life Technologies, Carlsbad, CA, USA). Between 49 and 107 million sequence reads were generated for each sample on an Illumina HiSeq 4000 platform with a single-end sequencing primer over 150 cycles (Additional file 5: Table S8). Quality trimming of raw reads was carried out with Flexcat [70] and reads of less than 50 nucleotides after trimming were removed. Tophat2 [76] was used to map reads to the *Ectocarpus* genome [18] and the mapped sequencing data was processed with HTSeq [77] to obtain counts for sequencing reads mapped to exons. Expression values were represented as TPM. Differential expression was detected using the DESeq2 package (Bioconductor; [78]) using an adjusted *p*-value cut-off of 0.05 and

a minimal fold-change of 2. The set of GBGs corresponded to genes that were significantly differentially regulated between life cycle generation in both of the ChIP-seq experiments.

Searches for histone and histone modifying enzyme encoding genes in *Ectocarpus*

Histone and histone modifier genes were detected in the *Ectocarpus* genome [18] using Blast [79] and manually reannotated when necessary, using GenomeView for structural reannotation [80].

Supplementary information

Additional file 1: Supplementary methods
Additional file 2: Figure S1. Histone gene clusters in the *Ectocarpus* genome.
Additional file 3: Figure S2. Representative MS/MS spectra showing the identification of histone modifications.
Additional file 4: Figure S3. H3K79me2, H4K20me3, H3K36me3, H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac signals for a region of chromosome 6 spanning an inserted viral genome. **Figure S5.** Correlations of transcript abundances and histone PTM signals for divergently transcribed pairs of genes. **Figure S4.** Pearson correlation scores for comparisons of the genomic distributions of ChIP-seq signal peaks for duplicate assays of the eight histone PTMs during both the sporophyte and gametophyte generations.
Additional file 5: Table S1. Histone proteins in *Ectocarpus*. **Table S4.** Genes with significantly different transcript abundances during the gametophyte and sporophyte generations. **Table S5.** Analysis of long non-coding RNA genes that were differentially expressed during the gametophyte and sporophyte generations. **Table S6.** Presence and absence of polycomb complex proteins in animals, land plants and stramenopiles. **Table S7**. Anti-histone-PTM antibodies used in this study. **Table S8.** Sequence data generated by this study.
Additional file 6: Table S2. Putative PTM writers and erasers in *Ectocarpus* sp.
Additional file 7: Table S3. Presence or absence of histone post-translational modifications in seven species from diverse eukaryotic supergroups.

Acknowledgments

We thank the Institut Français de Bioinformatique and the Roscoff Analysis and Bioinformatics for Marine Science platform ABiMS (http://abims.sb-roscoff.fr) for providing computing and data storage resources.

Funding

This work was supported by the Centre National de la Recherche Scientifique, the University Pierre et Marie Curie, the Agence Nationale de la Recherche projects Epicycle and Idealg (ANR-19-CE20-111111 and ANR-10-BTBR-04-01) and the European Research Council (grant

agreement 638240). DL acknowledges support from the Région Ile-de-France and the Fondation pour la Recherche Médicale.

Availability of data and materials

All sequence data used in this manuscript (Additional file 5: Table S8) have been deposited in the NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE146369. The *Ectocarpus* genome assembly is available through Orcae (https://bioinformatics.psb.ugent.be/orcae/overview/EctsiV2). The mass spectrometry proteomics data is available in the PRIDE database under accession number PXD013535.

Authors' contributions

SB prepared histones for mass spectrometry analysis and developed the ChIP-seq protocol with help from LT, SB carried out the ChIP-seq analysis, BL and DL carried out the mass spectrometry analysis of histone PTMs and analysed the data from these experiments, SB and JMC analysed the ChIP-seq data with input from LT and SMC, DS carried out additional experimental work, SB and JMC wrote the manuscript with input from the other authors. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Lawrence M, Daujat S, Schneider R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet. 2016;32:42–56.

2. Cock JM, Godfroy O, Macaisne N, Peters AF, Coelho SM. Evolution and regulation of complex life cycles: a brown algal perspective. Curr Opin Plant Biol. 2014;17:1–6.

3. Coelho SM, Godfroy O, Arun A, Le Corguillé G, Peters AF, Cock JM. *OUROBOROS* is a master regulator of the gametophyte to sporophyte life cycle transition in the brown alga *Ectocarpus*. Proc Natl Acad Sci USA. 2011;108:11518–23.

4. Cock JM, Coelho SM. Algal models in plant biology. J Exp Bot. 2011;62:2425–30.

5. Macaisne N, Liu F, Scornet D, Peters AF, Lipinska A, Perrineau M-M, et al. The *Ectocarpus IMMEDIATE UPRIGHT* gene encodes a member of a novel family of cysteine-rich proteins with an unusual distribution across the eukaryotes. Development. 2017;144:409–18.

6. Godfroy O, Uji T, Nagasato C, Lipinska AP, Scornet D, Peters AF, et al. DISTAG/TBCCd1 Is Required for Basal Cell Fate Determination in *Ectocarpus*. Plant Cell. 2017;29:3102–22.

7. Arun A, Coelho SM, Peters AF, Bourdareau S, Pérès L, Scornet D, et al. Convergent recruitment of TALE homeodomain life cycle regulators to direct sporophyte development in land plants and brown algae. eLife. 2019;8:e43101.

8. Müller DG. Generationswechsel, Kernphasenwechsel und Sexualität der Braunalge *Ectocarpus siliculosus* im Kulturversuch. Planta. 1967;75:39–54.

9. Bothwell JH, Marie D, Peters AF, Cock JM, Coelho SM. Role of endoreduplication and apomeiosis during parthenogenetic reproduction in the model brown alga *Ectocarpus*. New Phytol. 2010;188:111–21.

10. Ohad N, Margossian L, Hsu YC, Williams C, Repetti P, Fischer RL. A mutation that allows endosperm development without fertilization. Proc Natl Acad Sci U A. 1996;93:5319–24.

11. Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, et al. Mutations in *FIE*, a WD polycomb group gene, allow endosperm development without fertilization. Plant Cell. 1999;11:407–16.

12. Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ. Fertilization-independent seed development in *Arabidopsis thaliana*. Proc Natl Acad Sci U A. 1997;94:4223–8.

13. Köhler C, Hennig L, Bouveret R, Gheyselinck J, Grossniklaus U, Gruissem W. *Arabidopsis* MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. EMBO J. 2003;22:4804–14.

14. Guitton AE, Page DR, Chambrier P, Lionnet C, Faure JE, Grossniklaus U, et al. Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in *Arabidopsis thaliana*. Development. 2004;131:2971–81.

15. Guitton AE, Berger F. Loss of function of *MULTICOPY SUPPRESSOR OF IRA 1* produces nonviable parthenogenetic embryos in *Arabidopsis*. Curr Biol. 2005;15:750–4.

16. Mosquna A, Katz A, Decker EL, Rensing SA, Reski R, Ohad N. Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. Development. 2009;136:2433–44.

17. Okano Y, Aono N, Hiwatashi Y, Murata T, Nishiyama T, Ishikawa T, et al. A polycomb repressive complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution. Proc Natl Acad Sci U A. 2009;106:16321–6.

18. Cormier A, Avia K, Sterck L, Derrien T, Wucher V, Andres G, et al. Re-annotation, improved large-scale assembly and establishment of a catalogue of noncoding loci for the genome of the model brown alga *Ectocarpus*. New Phytol. 2017;214:219–32.

19. Kotliński M, Knizewski L, Muszewska A, Rutowicz K, Lirski M, Schmidt A, et al. Phylogeny-Based Systematization of *Arabidopsis* Proteins with Histone H1 Globular Domain. Plant Physiol. 2017;174:27–34.

20. Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, et al. Mass spectrometry analysis of *Arabidopsis* histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res. 2004;32:6511–8.

21. Zhang K, Sridhar VV, Zhu J, Kapoor A, Zhu J-K. Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*. PLoS One. 2007;2:e1210.

22. Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, et al. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature. 2010;465:617–21.

23. Mozgova I, Hennig L. The polycomb group protein regulatory network. Annu Rev Plant Biol. 2015;66:269–96.

24. Schuettengruber B, Bourbon H-M, Di Croce L, Cavalli G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell. 2017;171:34–57.

25. Philpott A, Krude T, Laskey RA. Nuclear chaperones. Semin Cell Dev Biol. 2000;11:7–14.

26. Zhao X, Rastogi A, Deton Cabanillas AF, Mohamed OA, Cantrel C, Lombard B, et al. H3K27me3 natural variation selectively marks genes predicted to be important for differentiation in unicellular algae. bioRxiv. 2019;2019.12.26.888800.

27. Veluchamy A, Rastogi A, Lin X, Lombard B, Murik O, Thomas Y, et al. An integrative analysis of post-translational histone modifications in the marine diatom *Phaeodactylum tricornutum*. Genome Biol. 2015;16:102.

28. Waterborg JH. Evolution of histone H3: emergence of variants and conservation of post-translational modification sites. Biochem Cell Biol. 2012;90:79–95.

29. Talbert PB, Ahmad K, Almouzni G, Ausió J, Berger F, Bhalla PL, et al. A unified phylogeny-based nomenclature for histone variants. Epigenetics Chromatin. 2012;5:7.

30. Iyer LM, Anantharaman V, Wolf MY, Aravind L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int J Parasitol. 2008;38:1–31.

31. Shaver S, Casas-Mollano JA, Cerny RL, Cerutti H. Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga *Chlamydomonas*. Epigenetics. 2010;5:301–12.

32. Barnes CE, English DM, Cowley SM. Acetylation & Co: an expanding repertoire of histone acylations regulates chromatin and transcription. Essays Biochem. 2019;63:97–107.

33. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NCT, et al. Active genes are tri-methylated at K4 of histone H3. Nature. 2002;419:407–11.

34. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell. 2007;130:77–88.

35. Akkers RC, van Heeringen SJ, Jacobi UG, Janssen-Megens EM, Françoijs K-J, Stunnenberg HG, et al. A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in *Xenopus* embryos. Dev Cell. 2009;17:425–34.

36. Charron J-BF, He H, Elling AA, Deng XW. Dynamic landscapes of four histone modifications during deetiolation in *Arabidopsis*. Plant Cell. 2009;21:3732–48.

37. Karlić R, Chung H-R, Lasserre J, Vlahovicek K, Vingron M. Histone modification levels are predictive for gene expression. Proc Natl Acad Sci U S A. 2010;107:2926–31.

38. Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, et al. Modeling gene expression using chromatin features in various cellular contexts. Genome Biol. 2012;13:R53.

39. Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics. 2012;13:424.

40. Yang H, Howard M, Dean C. Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at *Arabidopsis* FLC. Curr Biol CB. 2014;24:1793–7.

41. Zhang F, Qi B, Wang L, Zhao B, Rode S, Riggan ND, et al. EIN2-dependent regulation of acetylation of histone H3K14 and non-canonical histone H3K23 in ethylene signalling. Nat Commun. 2016;7:13018.

42. Jurkowska RZ, Qin S, Kungulovski G, Tempel W, Liu Y, Bashtrykov P, et al. H3K14ac is linked to methylation of H3K9 by the triple Tudor domain of SETDB1. Nat Commun. 2017;8:2057.

43. Wu Z, Tang J, Zhuo J, Tian Y, Zhao F, Li Z, et al. Chromatin Signature and Transcription Factor Binding Provide a Predictive Basis for Understanding Plant Gene Expression. Plant Cell Physiol. 2019;60:1471–86.

44. Liu Y, Liu K, Yin L, Yu Y, Qi J, Shen W-H, et al. H3K4me2 functions as a repressive epigenetic mark in plants. Epigenetics Chromatin. 2019;12:40.

45. Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. Genome Biol. 2009;10:R62.

46. Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, et al. Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. EMBO J. 2011;30:1928–38.

47. Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129:823–37.

48. Huff JT, Plocik AM, Guthrie C, Yamamoto KR. Reciprocal intronic and exonic histone modification regions in humans. Nat Struct Mol Biol. 2010;17:1495–9.

49. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. Science. 2010;327:996–1000.

50. Pajoro A, Severing E, Angenent GC, Immink RGH. Histone H3 lysine 36 methylation affects temperature-induced alternative splicing and flowering in plants. Genome Biol. 2017;18:102.

51. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448:553–60.

52. Balakrishnan L, Milavetz B. Decoding the histone H4 lysine 20 methylation mark. Crit Rev Biochem Mol Biol. 2010;45:440–52.

53. Xu J, Kidder BL. H4K20me3 co-localizes with activating histone modifications at transcriptionally dynamic regions in embryonic stem cells. BMC Genomics. 2018;19:514.

54. de la Paz Sanchez M, Gutierrez C. *Arabidopsis* ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. Proc Natl Acad Sci U S A. 2009;106:2065–70.

55. Tarver JE, Cormier A, Pinzón N, Taylor RS, Carré W, Strittmatter M, et al. microRNAs and the evolution of complex multicellularity: identification of a large, diverse complement of microRNAs in the brown alga *Ectocarpus*. Nucl Acids Res. 2015;43:6384–98.

56. Steger DJ, Lefterova MI, Ying L, Stonestrom AJ, Schupp M, Zhuo D, et al. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Mol Cell Biol. 2008;28:2825–39.

57. Bu J, Chen A, Yan X, He F, Dong Y, Zhou Y, et al. SETD2-mediated crosstalk between H3K36me3 and H3K79me2 in MLL-rearranged leukemia. Leukemia. 2018;32:890–9.

58. Schübeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, et al. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev. 2004;18:1263–71.

59. Li T, Liu Q, Garza N, Kornblau S, Jin VX. Integrative analysis reveals functional and regulatory roles of H3K79me2 in mediating alternative splicing. Genome Med. 2018;10:30.

60. Hoffmann RD, Palmgren MG. Epigenetic repression of male gametophyte-specific genes in the *Arabidopsis* sporophyte. Mol Plant. 2013;6:1176–86.

61. Jung CH, O'Brien M, Singh MB, Bhalla PL. Epigenetic landscape of germline specific genes in the sporophyte cells of *Arabidopsis thaliana*. Front Plant Sci. 2015;6:328.

62. Lipinska AP, Serrano-Serrano ML, Cormier A, Peters AF, Kogame K, Cock JM, et al. Rapid turnover of life-cycle-related genes in the brown algae. Genome Biol. 2019;20:35.

63. Peters AF, Scornet D, Ratin M, Charrier B, Monnier A, Merrien Y, et al. Life-cycle-generation-specific developmental processes are modified in the *immediate upright* mutant of the brown alga *Ectocarpus siliculosus*. Development. 2008;135:1503–12.

64. Coelho SM, Scornet D, Rousvoal S, Peters NT, Dartevelle L, Peters AF, et al. How to cultivate *Ectocarpus*. Cold Spring Harb Protoc. 2012;2012:258–61.

65. Tirichine L, Lin X, Thomas Y, Lombard B, Loew D, Bowler C. Histone extraction protocol from the two model diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. Mar Genomics. 2014;13:21–5.

66. Sterck L, Billiau K, Abeel T, Rouzé P, Van de Peer Y. ORCAE: online resource for community annotation of eukaryotes. Nat Methods. 2012;9:1041.

67. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 2019;47:D442–50.

68. Lin X, Tirichine L, Bowler C. Protocol: Chromatin immunoprecipitation (ChIP) methodology to investigate histone modifications in two model diatom species. Plant Methods. 2012;8:48.

69. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. 2016.

70. Hansen P, Hecht J, Ibn-Salem J, Menkuec BS, Roskosch S, Truss M, et al. Q-nexus: a comprehensive and efficient analysis pipeline designed for ChIP-nexus. BMC Genomics. 2016;17:873.

71. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.

72. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.

73. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma Oxf Engl. 2010;26:841–2.

74. Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. Nat Struct Mol Biol. 2016;23:349–57.

75. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45.

76. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.

77. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106.

78. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

79. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.

80. Abeel T, Van Parys T, Saeys Y, Galagan J, Van de Peer Y. GenomeView: a next-generation genome browser. Nucleic Acids Res. 2012;40:e12.

81. Brown JW, Sorhannus U. A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. PLoS One [Internet]. 2010;5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20862282

Figures

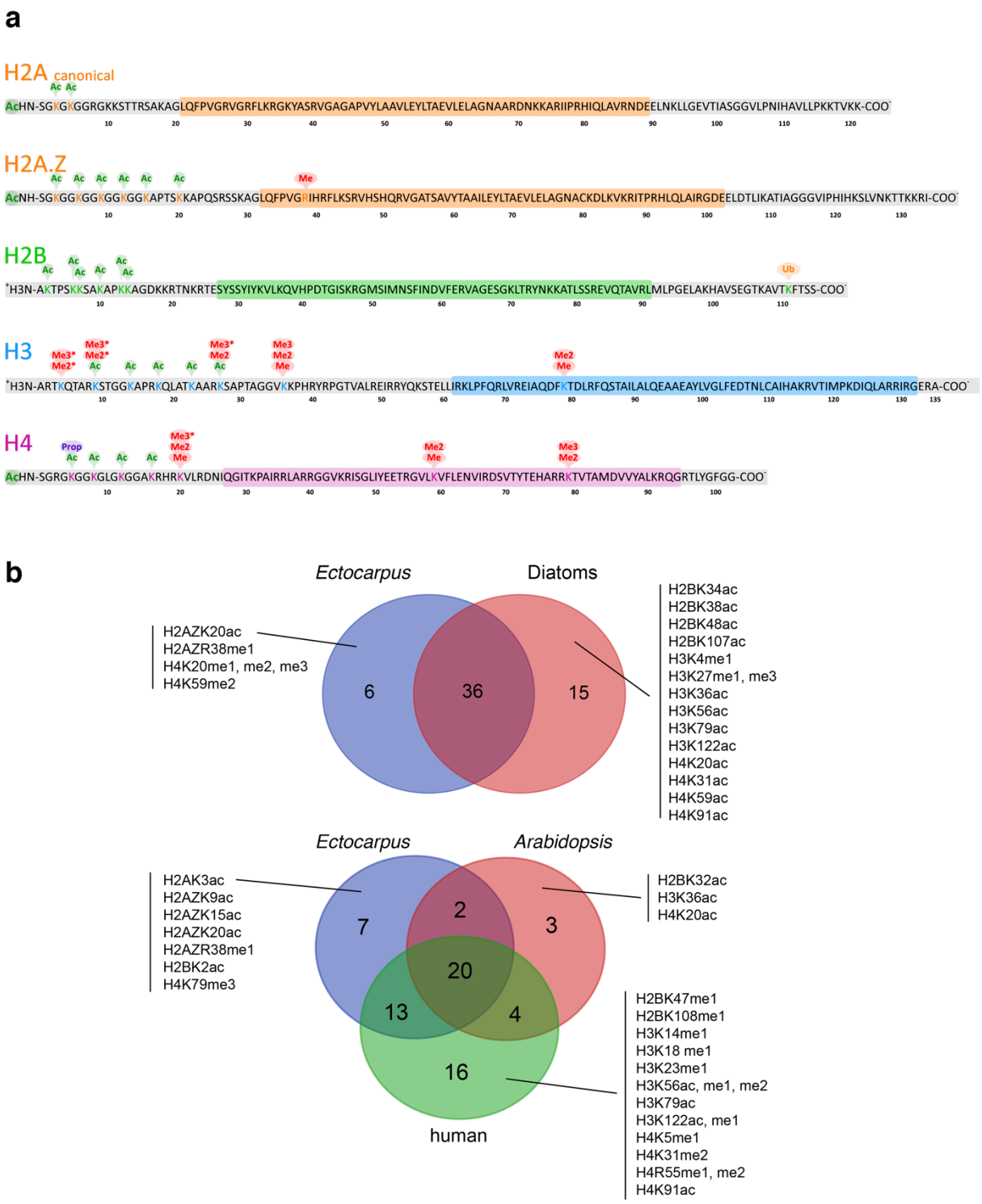


Fig. 1 a Post-translational modifications of *Ectocarpus* histones. Acetylation, methylation, ubiquitylation and propionylation modifications of core and variant histones identified in this study. Coloured boxes indicate globular core domains and grey boxes N- and C-terminal tails. Amino acid positions are indicated below the protein sequence. Asterisks indicate histone modifications that were only detected using antibodies. All other modifications were detected by mass spectrometry. **b** Comparison of the histone PTMs detected in *Ectocarpus* with PTMs reported for diatoms (combined data for *Phaeodactylum tricornutum* and *Thalassiosira*

pseudonana, upper panel) and for humans and the land plant *Arabidopsis* (lower panel). See Additional file 5: Table S3 for details.

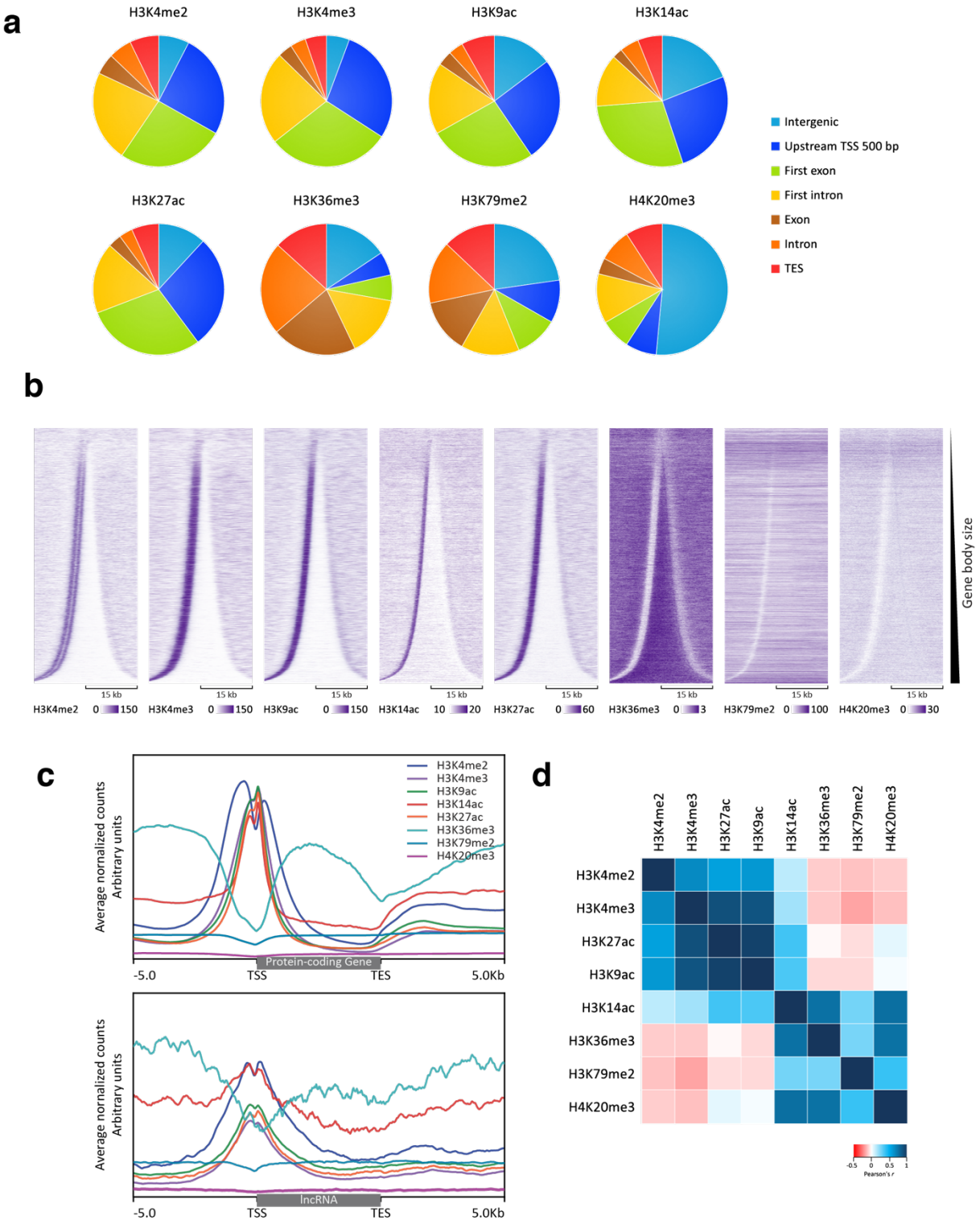


Fig. 2 Distribution of histone PTMs in the *Ectocarpus* genome. **a** Distribution of eight histone PTMs across seven genome feature classes. Intergenic, all intergenic regions except the TSS and TES regions; Upstream TSS 500 bp, 500 bp region 5' to the transcription start site; TES, 200 bp regions extending 100 bp on each side of the transcription end site; Exon, all exon sequence except first exons; Intron, all intron sequence except first introns. **b** Distribution of the eight histone PTMs across the complete set of *Ectocarpus* genes (17,447) sorted by gene body size and centred on the gene body. The black brackets correspond to 15 kilobases (kb). **c**

Enrichment profiles of the eight histone PTMs for all protein-coding (top panel) and all lncRNA (bottom panel) genes. Gene bodies are plotted as proportional lengths, upstream and downstream intergenic regions in kb. **d** Pearson correlation scores for comparisons of the genomic distributions of ChIP-seq signal peaks for the eight histone PTMs.

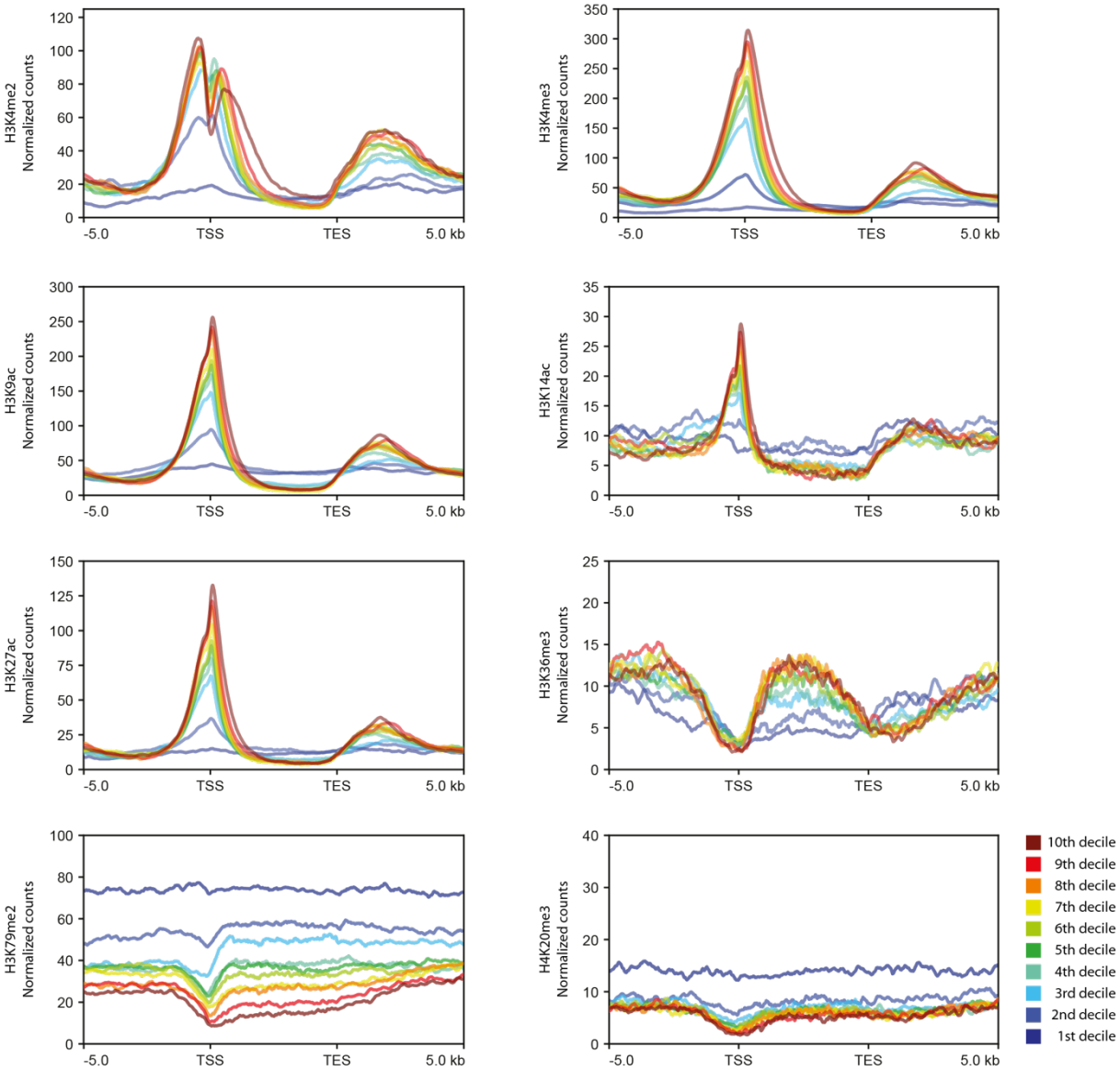


Fig. 3 Enrichment profiles for the histone PTMs H3K4me2, H3K4me3, H3K9ac, H3K14ac, H3K27ac, H3K36me3, H3K79me2 and H4K20me3 across genes expressed at different levels (TPM deciles). Gene bodies are plotted as proportional lengths, upstream and downstream intergenic regions in kilobases. TSS, transcription start site; TES, transcription end site, kb, kilobases.

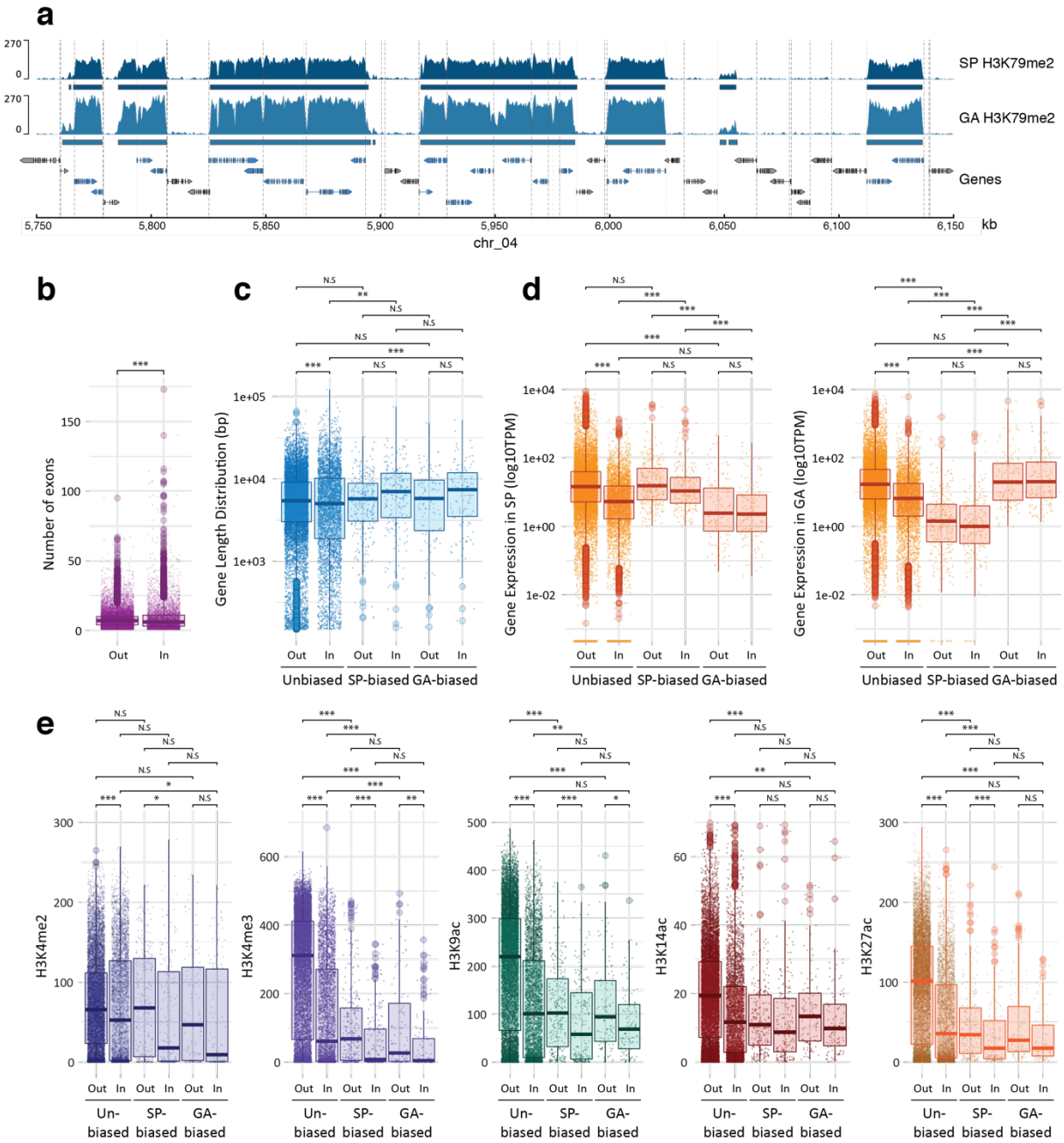


Fig. 4 Genomic distribution of the histone PTM H3K79me2. **a** Representative region of 0.4 Mbp from chromosome 4 (spanning the genes with LocusIDs Ec-04_005800 to Ec-04_006210) showing alternating regions marked or not with H3K79me2 during both the sporophyte and gametophyte generations of the life cycle. Blue bars indicate H3K79me2 regions, genes are shown as arrows with exons as boxes. Genes within a H3K79me2 region are in blue. Vertical dotted lines indicate the positions of TSSs. **b** Genes in H3K79me2 regions have fewer exons. **c** Lengths of genes within (In) or outside (Out) H3K79me2 regions. **d** Abundances of transcripts for genes within (In) or outside (Out) H3K79me2 regions. **e** Histone PTMs on genes within (In) or outside (Out) H3K79me2 regions. A gene was considered to be "In" if more than 50% of the gene body (TSS-TES) was inside a H3K79me2 region. SP, sporophyte; GA, gametophyte; kb, kilobase pairs. Asterisks indicate significant differences (Bonferonni-adjusted Wilcoxon tests; ***, $p < 0.0001$; **, $p < 0.001$; *, $p < 0.05$).

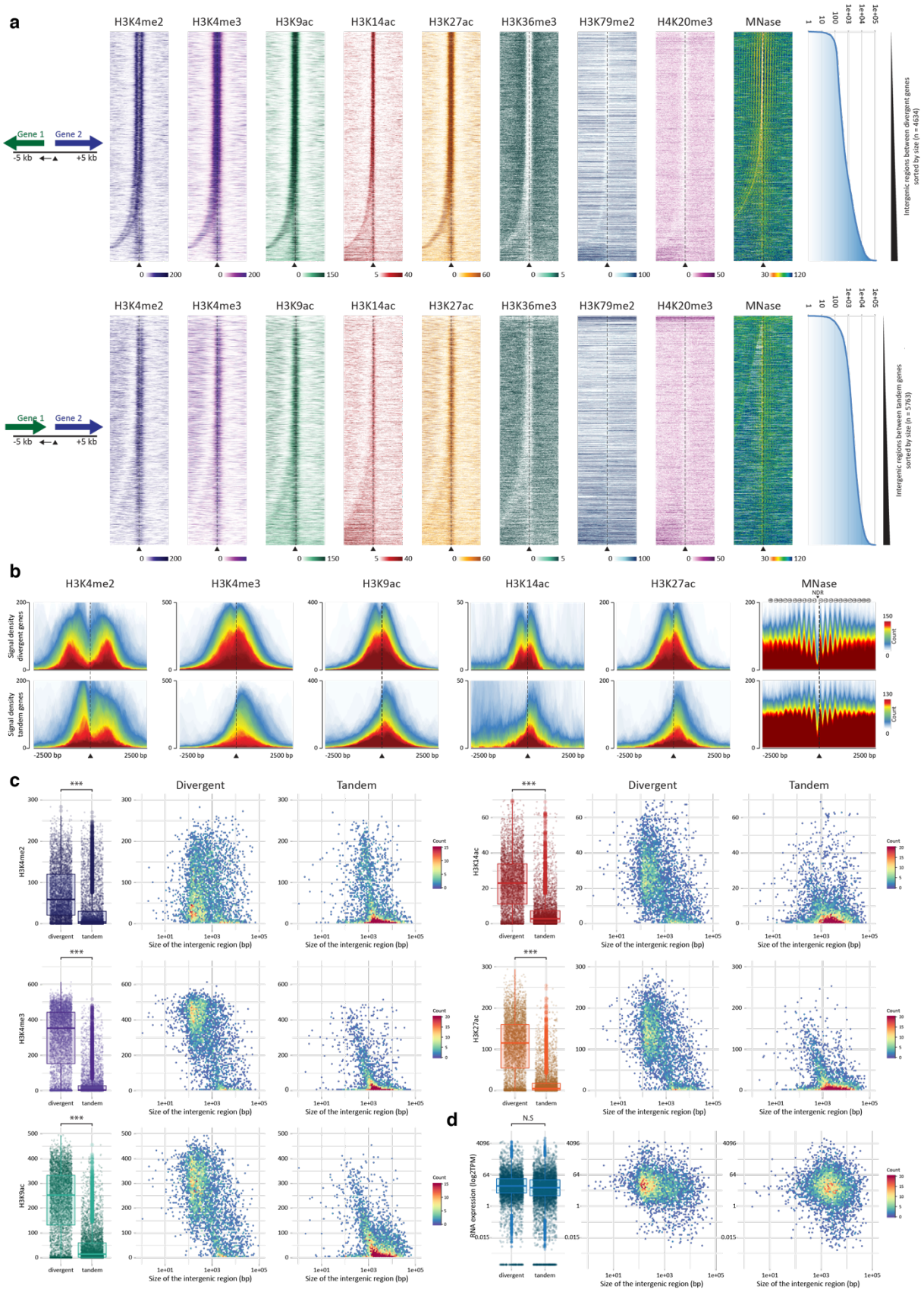
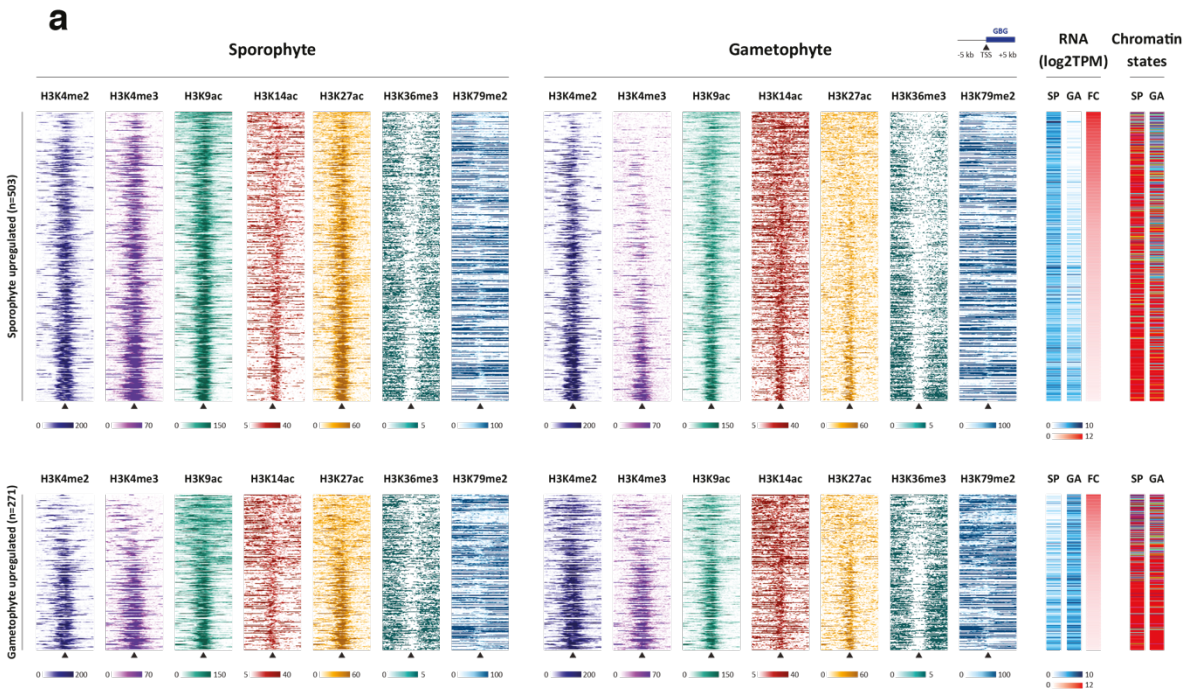
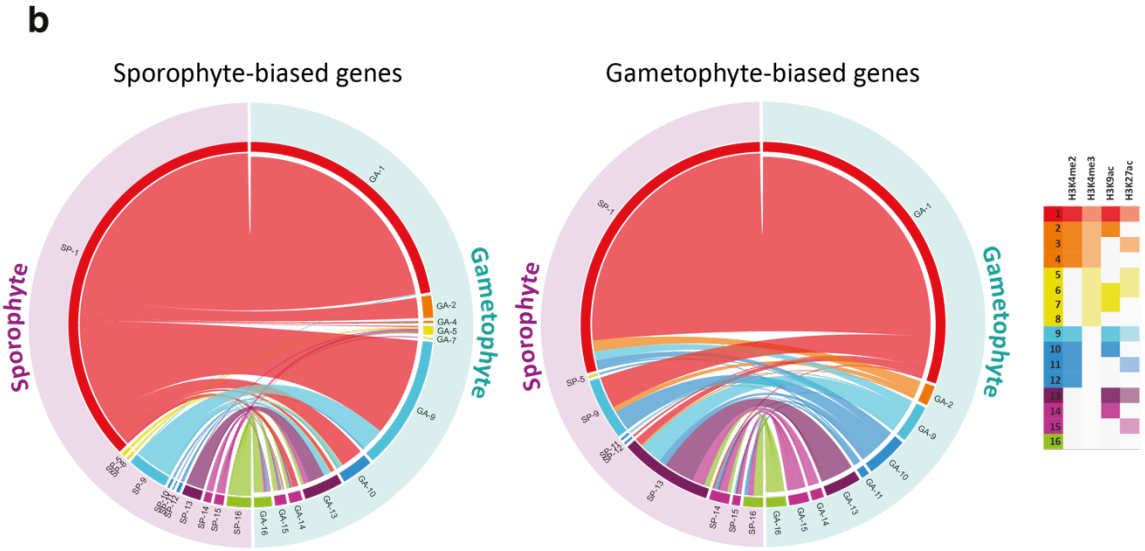


Fig. 5 Histone PTM distributions in intergenic regions between divergent and tandemly organised pairs of genes. **a** Distribution of eight histone PTMs and MNase-sensitive sites in

regions of 10 kb surrounding the TSS of the right-hand gene for pairs of either divergently (upper panel) or tandemly (lower panel) organised genes. The lengths of the intergenic regions are plotted on the right in bp. **b** Signal distributions for five TSS-localised histone PTMs and MNase-sensitive sites in regions of 5 kb surrounding the TSS of the right-hand gene for pairs of either divergently (upper plots) or tandemly (lower plots) organised genes on chromosome 4. **c** Abundance of five histone PTMs at the TSS of Gene 2 (see schema in A) in the intergenic regions of divergently or tandemly organised gene pairs together with plots indicating abundance in relation to the size of the intergenic region. **d** Transcript abundances for Gene 2. For **c** and **d**, heat score (count) indicates the number of genes with both the same PTM or transcript abundance and the same size intergenic region. MNase, micrococcal nuclease; kb, kilobases; bp, base pairs.


a
Sporophyte
Gametophyte
RNA (log2TPM)
Chromatin states
H3K4me2
H3K4me3
H3K9ac
H3K14ac
H3K27ac
H3K36me3
H3K79me2
SP GA FC
SP GA
Sporophyte upregulated (n=503)
Gametophyte upregulated (n=271)


b
Sporophyte-biased genes
Gametophyte-biased genes
Sporophyte
Gametophyte
H3K4me2
H3K4me3
H3K9ac
H3K27ac

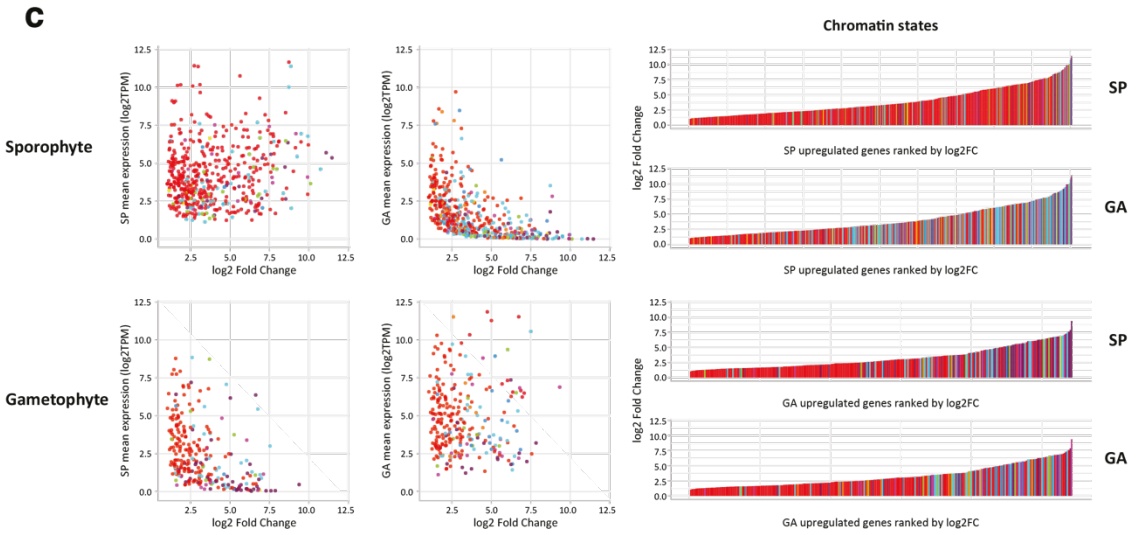

c
Chromatin states
Sporophyte
Gametophyte
SP mean expression (log2TPM)
GA mean expression (log2TPM)
log2 Fold Change
SP
GA
SP upregulated genes ranked by log2FC
GA upregulated genes ranked by log2FC

Fig. 6 Alterations of histone PTM patterns at generation-biased genes during the two generations of the life cycle. **a** Patterns of histone PTMs over 10 kb regions surrounding the TSSs of sporophyte-biased (upper panel) and gametophyte-biased (lower panel) genes during the sporophyte (left) and gametophyte (middle) generations. The heatmaps on the right show transcript abundance during the two generations and log2 fold-change between generations. **b** Circos plots comparing chromatin states (combinations of histone PTMs) at the transcription start sites (TSS) of sporophyte-biased (left) and gametophyte-biased (left) genes in chromatin from the sporophyte (mauve) and gametophyte (green) generations of the life cycle. Colours correspond to chromatin states one to 16 as indicated in the key. **c** Plots of fold change against transcript abundance (TPM) for sporophyte- and gametophyte-biased genes during the two generations (left panel) along with ranked plots of fold change (right panel). Colour coding corresponds to chromatin states.

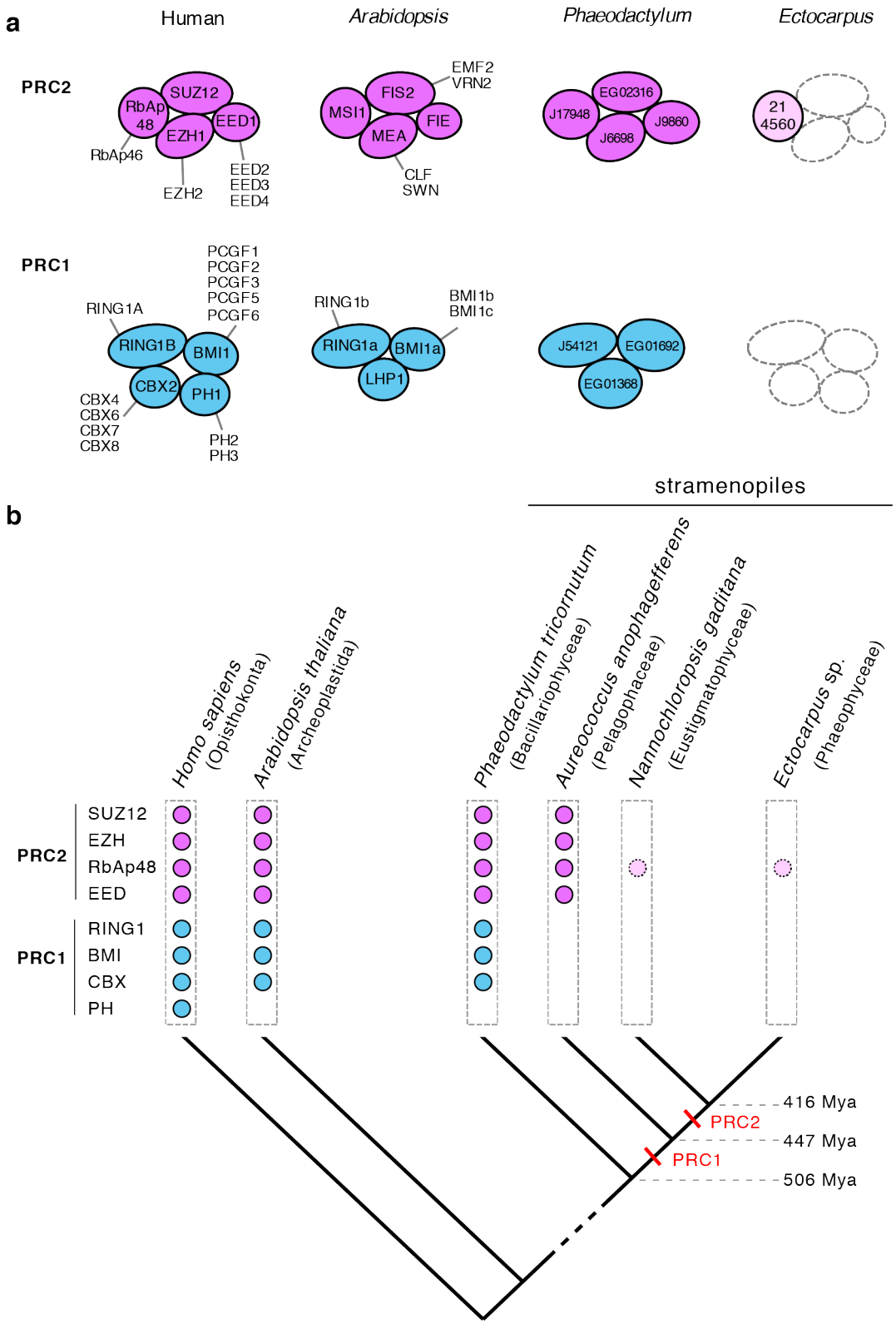


Fig. 7 Loss of polycomb complexes from brown algae. **a** Core components of the polycomb complexes PRC2 (mauve) and PRC1 (blue) in *Homo sapiens, Arabidopsis thaliana*, *Phaeodactylum tricornutum* and *Ectocarpus* sp.. Mauve or blue circles indicate the presence of a PRC2 or PRC1 gene, respectively. RbAp48 is indicated with a dotted grey line when it probably represents a component of other cellular complexes (see text for details). Alternative

proteins are shown for each component. Empty, dotted circles indicate components that are absent from *Ectocarpus*. All *P. tricornutum* and *Ectocarpus* sp. protein names should be prefixed with Phatr3 or Ec-, respectively, e.g. Phatr3_J17948, Ec-21_004560. **b** Schematic phylogenetic tree indicating PRC2 and PRC1 genes present in *Homo sapiens, Arabidopsis thaliana* and four selected stramenopiles. Red bars indicate predicted approximate timepoints for loss of the PRC1 and PRC2 complexes during stramenopile evolution. Approximate divergence times (±100 My) are based on Brown and Sorhannus [81].

Supplementary figures

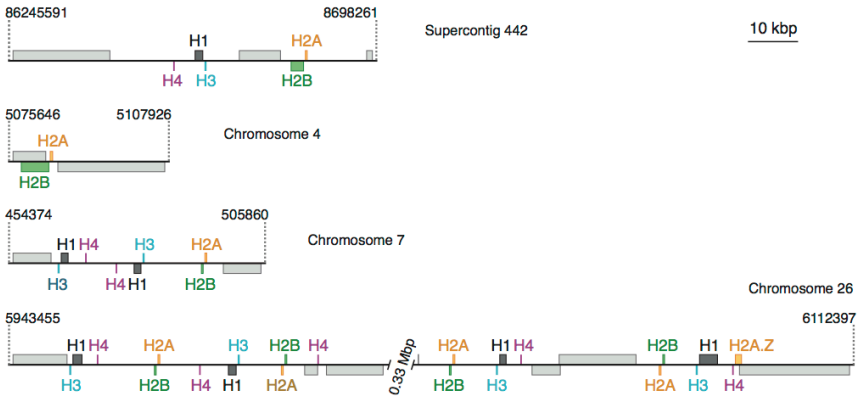


Fig. S1 Histone gene clusters in the *Ectocarpus* genome. Only regions with two or more histone genes are shown. Coding regions from the start to the stop codon are shown as boxes. Histone genes are colour coded, flanking, non-histone genes are in grey. Genes above the line are transcribed to the right, genes below the line to the left. Dotted lines indicate chromosomal or scaffold coordinates.

(see additional file 3 for figure)

Fig. S2 Representative MS/MS spectra showing the identification of histone modifications. Trypsin-peptide-derived LC-ESI MS/MS fragmentation spectra for modified H2, H3 and H4 peptides. The inset shows the peptide sequence and the observed ions. Tandem mass spectra are labelled to show singly and/or doubly charged b and y ions, as well as ions corresponding to neutral loss of water (°) or NH_3 (*). M is the parent ion mass and Im the immonium ion mass. Me, methylation; DiMe, dimethylation: TriMe, trimethylation, Prop, propionylation; Ac, Acetylation, Ub, ubiquitination; N-ter, N-terminal.

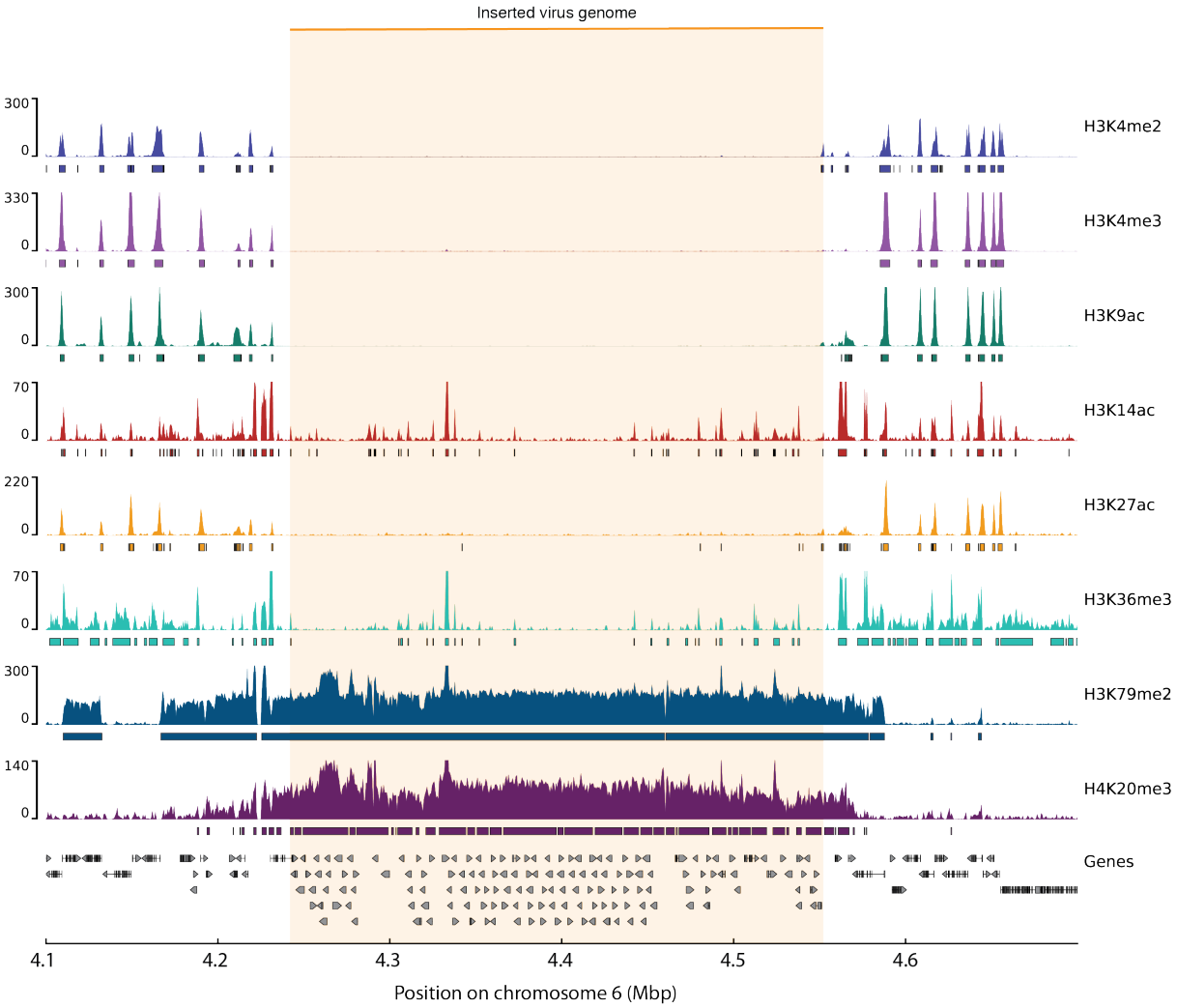


Fig. S3 H3K79me2, H4K20me3, H3K36me3, H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac signals for a region of chromosome 6 spanning an inserted viral genome (orange box). Coloured boxes indicate histone PTM peaks, grey boxes and arrows represent exons superimposed on genes represented by a black line.

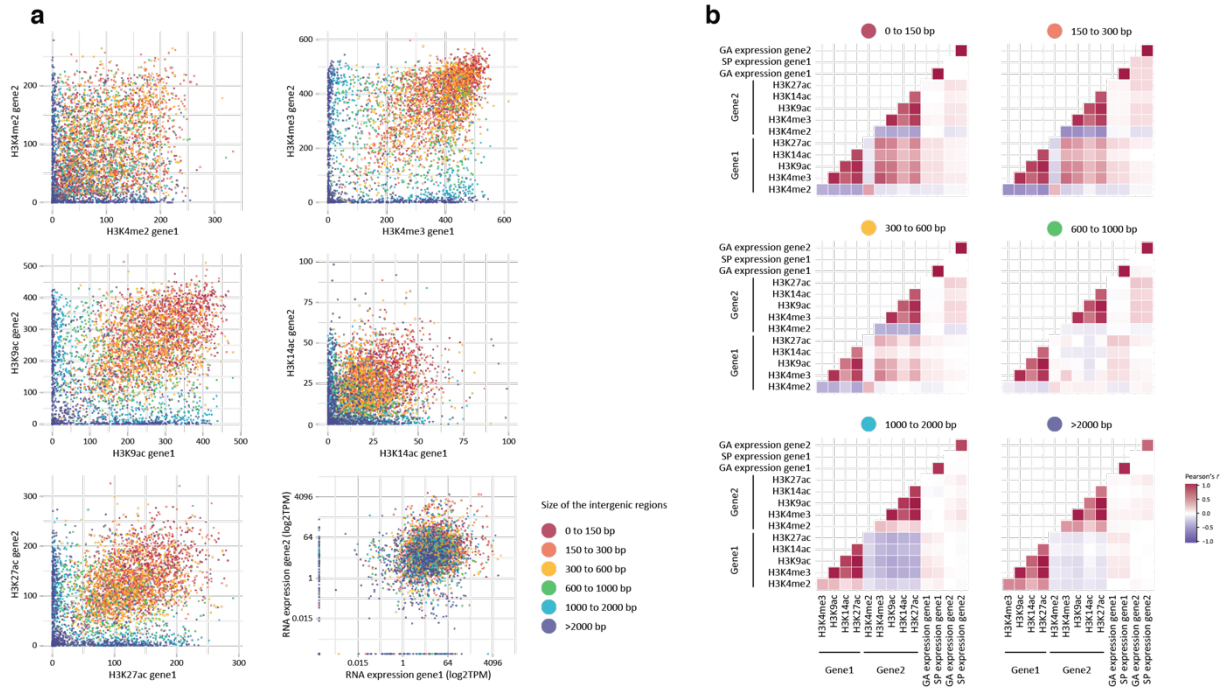


Fig. S4 Correlations of transcript abundances and histone PTM signals for divergently transcribed pairs of genes. **a** Pairwise comparisons of transcript abundances (log2TPM) and signal for the five TSS-localised histone PTMs for pairs of divergently transcribed genes (gene 1 and gene 2). **b** Pearson correlation scores for comparisons of transcript abundances (log2TPM) and signal for the five TSS-localised histone PTMs for pairs of divergently transcribed genes. Histone PTM signals were measured for the regions 500 bp downstream of each TSS. The schema in Fig. 5a indicates the relative organisation of gene 1 and gene 2.

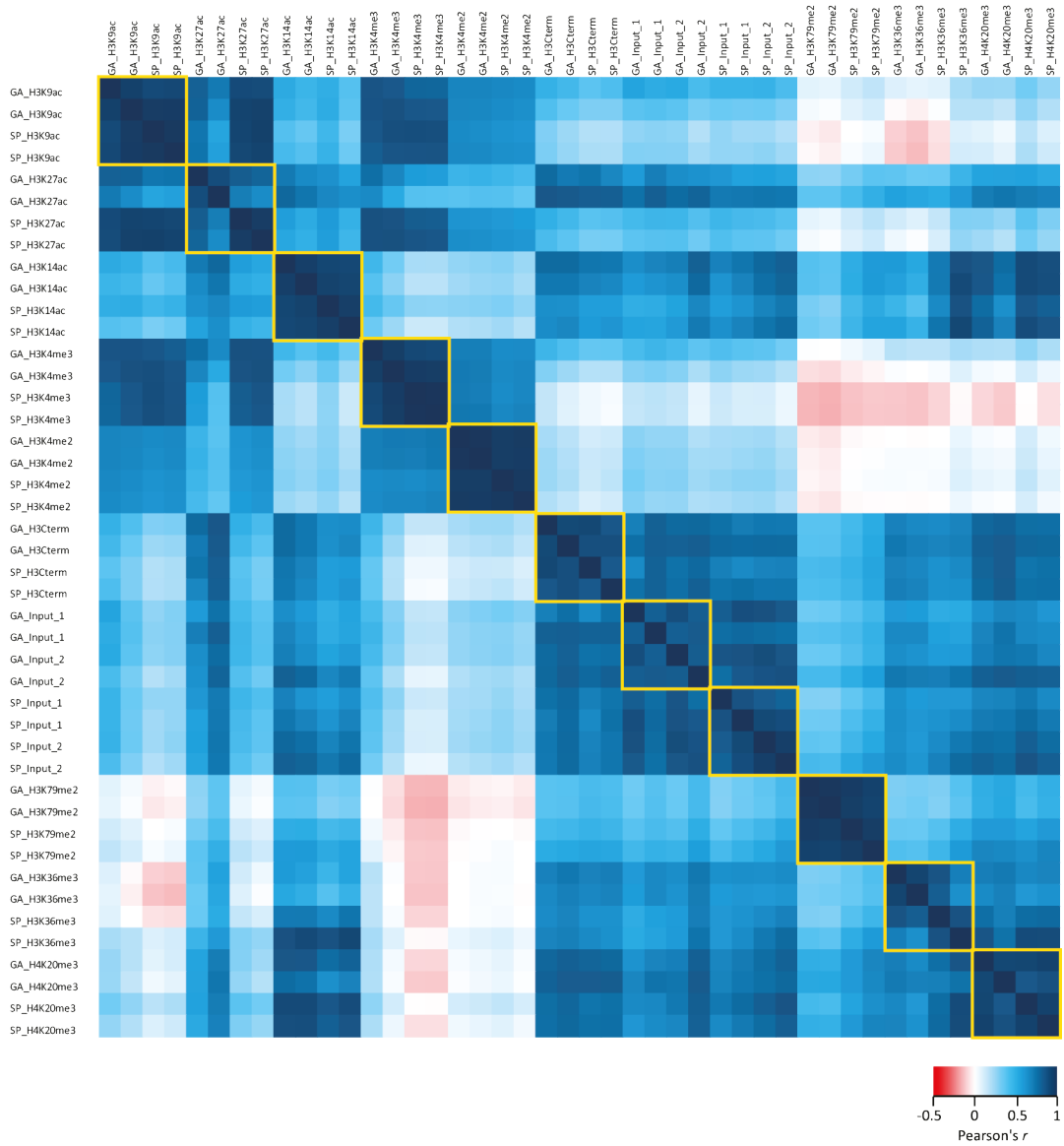


Fig. S5 Pearson correlation scores for comparisons of the genomic distributions of ChIP-seq signal peaks for duplicate assays of the eight histone PTMs during both the sporophyte and gametophyte generations.