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], Yue Wu[2], 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 expected 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 47 histone PTMs were identified, including a novel mark H2AZR38me1, 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 identified a correlation between the presence of specific chromatin marks and the level of gene expression.

Conclusions: The overview of histone PTMs in the brown algae presented here will provide a foundation for future studies aimed at understanding the role of chromatin modifications in the regulation of brown algal genomes.

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 chromatin but, at present, very little is known about the role of histone PTMs, for example, 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 an 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 correlated with 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

101 such as the plant GH1-HMGA or GH1-Myb families [19]. All but eight of the histone genes
102 were located in five gene clusters on chromosomes 4, 7 and 26 and on an unmapped scaffold
103 (sctg_442). The organisation of the clusters suggests multiple duplication, rearrangement and
104 fragmentation of an ancestral cluster with the organisation H4, H1, H3, H2B, H2A (Additional
105 file 2: Figure S1).
106 A search for genes encoding histone modifying enzymes identified both putative histone
107 acetyltransferases and methyltransferases, and predicted deacetylase and demethylase enzymes
108 (Additional file 6: Table S2). Most of the acetyltransferases were sufficiently similar to well-
109 characterised homologues to allow prediction of their target residues but the methyltransferases
110 tended to be less conserved at the sequence level and, in many cases, had novel domain
111 structures. Direct functional information, for example based on mutant analysis, will therefore
112 be necessary to investigate the specificity of the *Ectocarpus* methyltransferases.
113
114 **Identification of histone PTMs in *Ectocarpus***
115 Histone PTMs were detected using mass spectrometry of enzyme-digested histone
116 preparations. In addition, a broad range of commercially available antibodies were tested
117 against *Ectocarpus* histone preparations on immunoblots to further confirm the presence of a
118 subset of the PTMs identified by mass spectrometry. A total of 47 PTMs of core and variant
119 histones were detected in *Ectocarpus* (Fig. 1a, Additional file 3: Figure S2). Six of these marks
120 were only detected by immunoblotting and should therefore be treated with caution (marked
121 with an asterisk in Fig. 1a, Additional file 7: Table S3). Note that two of these PTMs, H3K9me2
122 and H3K9me3, were also not detectable using mass spectrometry in *Arabidopsis* but have been
123 detected using immunoblotting [20,21]. Figure S3 (Additional file 4) shows immunoblots that
124 detected a weak signal for these two PTMs in *Ectocarpus* chromatin.
125 Most of the histone PTMs detected in *Ectocarpus* have been reported previously in
126 species from one or more of the land plant, animal or fungal lineages, either at exactly the same
127 position or at an equivalent position in the corresponding protein (Fig. 1b, Additional file 7:
128 Table S3). However, a number of marks have only been described so far in stramenopiles. Of
129 these, some PTMs such as acetylation of lysines nine and 15 of H2A.Z were detected in both
130 *Ectocarpus* and the diatom *P. tricornutum* [22], whereas others were detected only in the
131 diatoms (e.g. acetylation of lysine 59 on H4) or only in *Ectocarpus* (methylation of arginine 38
132 on H2A.Z; Fig. 1b).
133
134 **Genome-wide distribution of selected histone PTMs**

135 To investigate the functions of histone PTMs in *Ectocarpus*, we analysed the distribution of
136 eight selected marks across the genome using chromatin immunoprecipitation and sequencing
137 (ChIP-seq). The PTMs analysed were H3K4me2, H3K4me3, H3K9ac, H3K14ac, H3K27ac,
138 H3K36me3, H3K79me2 and H4K20me3 (Fig. 2). Five of these PTMs, H3K4me2, H3K4me3,
139 H3K9ac, H3K14ac and H3K27ac, were preferentially associated with the transcription start
140 sites (TSSs) of genes but tended to be depleted from gene bodies (Fig. 2a,b,c,d). Genome-wide
141 patterns of these five PTMs exhibited a range of levels of positive correlation (Pearson's r 0.13
142 to 0.86; Fig. 2e), indicating that they tended to be co-localised in the genome. Similar PTM
143 patterns were detected at protein-coding and lncRNA genes (Fig. 2c), suggesting that histone
144 PTMs may be used in a similar manner to regulate these two types of gene. For each of the five
145 TSS-associated PTMs a peak was detected within 500 bp of the TSS for between 77% and 83%
146 of the genes in the genome. For H3K4me2, a pair of adjacent peaks was detected, one on each
147 side of the TSS (Fig. 2b,c). Genome-wide, peaks of H3K4me2, H3K4me3, H3K9ac, H3K14ac
148 and H3K27ac marks were associated with between 13, 416 and 14,423 of the 17,406 genes. We
149 noted a strong positive correlation between gene expression level, measured in transcripts per
150 million (TPM) and the strength of mapping of each mark to TSSs (Fig. 3). Taken together, these
151 observations indicate a strong association of H3K4me2, H3K4me3, H3K9ac, H3K14ac and
152 H3K27ac with transcriptionally active genes in *Ectocarpus*.
153 The three remaining PTMs, H3K36me3, H3K79me2 and H4K20me3, were depleted
154 from TSSs (Fig. 2). H3K36me3, which was detected at 12,863 genes (74% of the genome), was
155 most strongly associated with gene bodies and was depleted from both TSSs and transcription
156 end sites (TESs). As observed for the TSS-localised PTMs discussed above, presence of
157 H3K36me3 on gene bodies was positively correlated with expression, indicating that this mark
158 is also associated with gene activation (Fig. 3).
159 H4K20me3 peaks were detected principally in the non-coding part of the genome
160 (intergenic regions and introns; Fig. 2a) and occurred preferentially in intergenic regions or
161 introns that also contained a transposon (Pearson's χ^2 test with Yates' continuity correction p-
162 value < 2.2e-16 for both introns and intergenic regions). Indeed, almost all (94.6%) of the
163 H4K20me3 peaks co-localised with either a transposon or a region of short repeated sequence.
164 The distribution of H4K20me3 was therefore consistent with it playing a role in the regulation
165 of inserted transposons, possibly in the silencing of these elements. H4K20me3 peaks were
166 associated with all types of transposon, including both class I (RNA) and class II (DNA)
167 transposons (Additional file 5: Table S4). The occurrence of H4K20me3 in introns was not
168 unexpected because introns make up approximately 40% of the *Ectocarpus* genome [23], and

169 therefore represent a significant proportion of the non-coding fraction of the genome.
170 Moreover, 44.0% of the transposons (repeated elements ≥400 bp) in the genome are located in
171 introns.

172 Genes marked with H4K20me3 tended to have fewer exons, to be longer and to have
173 lower levels of expression than genes without a H4K20me3 peak (Fig. 4a,b,c). The inverse
174 relationship between the presence of H4K20me3 and gene expression level (Fig. 3) could be
175 interpreted as indicating a role in gene regulation but, given the co-localisation of this PTM
176 with transposon sequences, a more likely explanation is that the observed effect on gene
177 expression was an indirect effect of silencing of intronic transposon sequences. Interestingly,
178 genes that were marked with H4K20me3 were significantly less strongly marked with the TSS-
179 associated PTMs H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac (Fig. 4d). Several
180 hypotheses could be proposed to explain this observation. If the presence of H4K20me3 within
181 a gene directly causes a reduction in the abundance of TSS-associated PTMs, this may
182 constitute part of the mechanism that mediates the repressive effect of H4K20me3 on gene
183 expression. Alternatively, it is also possible that the reduction in the abundance of TSS-
184 associated PTMs is an indirect consequence of H4K20me3-induced repression of gene
185 expression.

186

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

188 H3K79me2 exhibited an unusual distribution pattern. This PTM was detected throughout the
189 genome, often in large, discrete, continuously marked regions of several kilobases separated by
190 H3K79me2-depleted regions (Fig. 5a, Additional file 4: Figure S4). Together the H3K79me2
191 regions covered 74.5 Mbp, 37.2% of the genome. About a third (36.6%) of the H3K79me2
192 regions were longer than 5 kbp and many of these regions spanned more than one gene (on
193 average, H3K79me2 regions of >5 kbp contained 1.85 genes). The borders of the long
194 H3K79me2 regions tended to be localised near a TSS or a TES. When H3K79me2 regions of
195 >5 kbp were considered, 49.3% of the borders were located within a kilobase of a TSS or a TES
196 and a Bonferonni-adjusted Wilcoxon test indicated that the borders were significantly closer to
197 the nearest TSS or TES than they were to random anchor points in the genome (Additional file
198 4: Figure S5).

199 Comparison of ChIP-seq data for the sporophyte and gametophyte generations of the
200 life cycle indicated that the H3K79me2 regions were stably maintained throughout the life
201 cycle. Only 0.3% or 1.2% of H3K79me2 regions longer than 5 kbp were detected uniquely in
202 the sporophyte or the gametophyte generation, respectively.

203 The H3K79me2 regions were strongly associated with peaks of H4K20me3 (Fig. 5a,b;
204 Additional file 4: Figure S4). Consistent with the co-localisation of the two PTMs, genes within
205 H3K79me2 regions were significantly shorter, possessed fewer introns and their transcripts
206 were less abundant on average than genes that were outside H3K79me2 regions (Fig. 4c-e).
207 The TSSs of genes within the H3K79me2 regions were also significantly less strongly marked
208 by H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac (Fig. 4f) and there was a negative
209 correlation between gene expression level (measured in TPM) and the quantitative presence of
210 H3K79me2 over genes (Fig. 3). To determine whether the association of H3K79me2 with
211 reduced levels of gene expression and reduced deposition of TSS-localised PTMs was an
212 indirect consequence of the co-localisation of this mark with H4K20me3, we analysed loci (Fig.
213 5b) where the two PTMs occurred independently of each other (Additional file 4: Figure S6).
214 This analysis indicated that both H3K79me2 and H4K20me3 were associated with reduced
215 levels of gene expression and reduced deposition of TSS-localised PTMs, even when one PTM
216 was present independently of the other. Moreover, the effects of H3K79me2 and H4K20me3
217 on gene expression and on the deposition of TSS-localised PTMs were additive when both
218 PTMs were present at the same locus, i.e. levels of gene expression and TSS-localised PTMs
219 were even lower when both PTMs were present. Therefore, like H4K20me3, H3K79me2
220 appears to be associated with the repression of gene expression. However, based on a similar
221 argument to that proposed for H4K20me3 above, the organisation of deposition of H3K79me2
222 across blocks of contiguous genes (Fig. 5) does not appear to be consistent with a direct role in
223 gene regulation and we therefore favour the hypothesis that, as proposed for H4K20me3, the
224 effect of H3K79me2 on gene regulation is an indirect consequence of the involvement of this
225 PTM in another, currently unknown, role in chromatin homeostasis.
226 Chromosome six contains an integrated copy of a large DNA virus (spanning 0.3 Mbp
227 from approximately 4,244,200 to 4,547,200) that is closely related to the *Ectocarpus* phaeovirus
228 EsV-1. The inserted virus has been shown to be transcriptionally silent [18,23]. A large
229 H3K79me2 region of about 0.42 Mbp was detected that spanned the entire inserted viral
230 genome (Additional file 4: Figure S7). Chromatin within this H3K79me2 region was also
231 marked with H4K20me3 but the other histone PTMs assayed were depleted from the inserted
232 viral genome (Additional file 4: Figure S7).
233
234 **Overlapping TSS regions of divergently transcribed gene pairs**
235 One unusual feature of the *Ectocarpus* genome, compared to genomes of similar size, is that
236 there is a relatively strong tendency for adjacent genes to be transcribed from opposite strands

237 of the DNA helix [23]. Consequently, the genome contains many divergently transcribed gene
238 pairs. To investigate the effect of this pattern of gene organisation on the chromatin
239 characteristics of TSSs, we searched for pairs of adjacent genes located on the same sequence
240 scaffold. This search identified 10,399 TSS-adjacent intergenic regions and 61.7% of the genes
241 flanking these regions were part of a divergently transcribed gene pair. The intergenic regions
242 of divergently transcribed gene pairs were significantly shorter than those of tandem gene pairs
243 (median 409 and 2293, respectively, Wilcoxon test, p-value<$2e^{-16}$; Fig. 6a). When the
244 intergenic regions of divergent gene pairs were shorter than about 600 bp, the two TSS
245 chromatin regions overlapped and shared the same nucleosome depleted region (NDR) based
246 on micrococcal nuclease digestion data (Fig. 6a). This overlap correlated with the presence of
247 double peaks for the PTMs H3K4me3, H3K9ac, H3K14ac and H3K27ac, one on each side of
248 the TSS (Fig. 6a,b). For H3K4me2, which was detected as a double peak at most individual
249 TSSs (Fig. 2b), the two peaks were further apart at the overlapping TSSs of divergent gene
250 pairs than they were at the single intergenic TSSs of tandem gene pairs (Fig. 6a,b). Therefore,
251 the pattern of PTMs also indicated that the TSSs of proximate divergent genes are located
252 within a shared chromatin domain. Signals for all five of the above PTMs were significantly
253 stronger for divergent than for tandem gene pairs (Fig. 6c, measured at the TSS of the second
254 gene of the gene pair, see schema in Fig. 6a). This difference appeared to be due to the overlap
255 of the two divergent TSS regions as the signal markedly increased for intergenics of divergent
256 gene pairs that were shorter than 1 kbp (Fig. 6c). Consistent with this observation, no significant
257 difference was detected between the median expression levels of divergent and tandem genes
258 (Fig. 6d). Note that transcript abundances were not correlated for the two genes of a divergent
259 gene pair (Pearson coefficients were 0.042 for all the divergent pairs and 0.054 for divergent
260 pairs closer than 600 bp; Additional file 4: Figure S8), despite the presence of a shared
261 chromatin domain at the TSSs and correlated histone PTM signals immediately downstream of
262 each TSS when divergent genes were closer than 600 bp (Additional file 4: Figure S8).
263

264 **Histone PTM patterns during the *Ectocarpus* life cycle**

265 To relate patterns of histone modification to changes in gene expression during the *Ectocarpus*
266 life cycle, ChIP-seq analysis was used to compare the distributions of histone PTMs during the
267 sporophyte and gametophyte generations. Overall, the distribution of PTMs was stable between
268 the two life cycle generations. For example, the six PTMs associated with actively expressed
269 genes, H3K4me2, H3K4me3, H3K9ac, H3K14ac, H3K27ac and H3K36me3, were either stably
270 present or stably absent for between 82.7% and 97.1% of the 17,406 genes genome-wide,

271 depending on the PTM. This analysis indicated that there were not any major, genome-wide
272 changes in patterns of histone PTMs associated with alternation between life cycle generations.
273 To focus more specifically on changes in patterns of histone PTMs associated with life-cycle-
274 related changes in gene expression, we analysed the presence and absence of marks at genes
275 that were differentially regulated between the two life cycle generations (Fig. 7a).

276 A comparison of gene expression patterns in the sporophyte and gametophyte, based on
277 RNA-seq data generated using the same biological samples as were used for the ChIP-seq
278 analysis, identified 774 genes that exhibited a generation-biased pattern of expression
279 (padj<0.05, fold change>2, TPM>1; Fig. 7a, Additional file 5: Table S5). We will refer to these
280 774 genes as generation-biased genes (GBGs). Analysis of the predicted functions of the GBGs
281 using a system of manually-assigned functional categories [7], indicated significant enrichment
282 in several functional categories, in particular in cell wall and extracellular proteins and proteins
283 of unknown function (χ^2 test padj $3.73e^{-40}$ and $1.56e^{-5}$, respectively; Table S6). Prediction of
284 subcellular location using Hectar [24], indicated that the GBG protein set was also significantly
285 enriched in secreted proteins (χ^2 test, $2.20e^{-27}$; Table S7). These observations are consistent with
286 a recent analysis using an independently identified set of GBGs [7] and suggest possible
287 important roles for the cell wall and intercellular interactions in the specific functions of the
288 two generations of the life cycle. Interestingly, this set of differentially expressed genes was
289 significantly enriched in long non-coding RNAs (lncRNAs; Additional file 5: Table S8; χ^2 test,
290 $1.52e^{-15}$) and 19 of the 72 differentially expressed lncRNAs were adjacent to a differentially
291 expressed protein-coding GBG (including four lncRNA/protein-coding gene pairs with
292 overlapping transcripts; Additional file 5: Table S8). In most cases (84%), the lncRNA and its
293 adjacent protein-coding gene were co-ordinately upregulated during the same generation of the
294 life cycle (Table S9). These observations suggest a possible role for lncRNAs in the regulation
295 of adjacent genes in *Ectocarpus*.

296 To analyse changes in histone PTMs at GBGs between life cycle generations, 16
297 chromatin states were defined based on different combinations of four TSS-localised PTMs that
298 exhibited clear patterns of presence and absence: H3K4me2, H3K4me3, H3K9ac and H3K27ac
299 (Fig. 7b). In most cases when GBGs changed chromatin state during the transition between the
300 two generations, increased transcript abundance was correlated with acquisition of additional
301 activation-associated PTMs (76.1% of the GBGs that changed chromatin state). However, the
302 majority of GBGs (61.1%) did not change chromatin state between generations, although we
303 noted that most of these genes (81.4%) were marked with all four activation-associated PTMs
304 during both generations. Taken together, these observations are consistent with a correlation

305 between the presence of the four TSS-localised marks and gene activation but indicate that
306 additional mechanisms may be necessary to induce differential gene expression. Therefore, if
307 the histone PTMs have role in regulating transcription, this role would most likely be to
308 facilitate or potentiate transcription rather than directly activating gene transcription. Note,
309 however, that it is also possible that some or all of the observed histone PTM modifications
310 were a consequence, rather than a cause, of transcription.

311 We detected more sporophyte-biased (503) than gametophyte-biased (271) genes (χ^2
312 test, $7.49e^{-17}$) and the former were significantly more strongly upregulated than the latter (4.2
313 and 2.8 mean log2TPM fold changes, respectively; Kruskal-Wallis rank sum test *p*-value =
314 $4.0e^{-11}$). Overall, sporophyte-biased and gametophyte-biased genes showed similar patterns of
315 transitions to new chromatin states (Spearman's rank correlation rho 0.61, *p*-value = 0.011).
316 However, the sporophyte-biased genes exhibited a stronger tendency to switch to chromatin
317 state 1 (presence of all four activation-associated PTMs) than the gametophyte-biased genes
318 (74.8% compared with 26.4% of chromatin state transitions, respectively; z score 7.7857,
319 $p<0.00001$). This observation is consistent with the larger fold changes in transcript abundance
320 (TPM) observed for sporophyte-biased genes as transitions to chromatin state 1 were associated
321 with significantly larger fold changes than transitions to other states (Wilcoxon test, p-value
322 <0.0078 for all GBGs). We also noted that the GBGs that exhibited the highest fold changes
323 also exhibited a stronger tendency to change chromatin state between generations (Fig. 7c).
324 This was true for both sporophyte-biased and gametophyte-biased genes.
325

326 ***Ectocarpus* lacks polycomb repressive complexes**

327 Polycomb repressive complexes have been shown to play important roles in the regulation of a
328 broad range of developmental processes in both animals and land plants [25,26] and a key role
329 has been proposed for Polycomb repressive complex 2 (PRC2) in the regulation of life cycle
330 transitions in both *Arabidopsis* [10,12–15] and the moss *Physcomitrella patens* [16,17].
331 However, in contrast to the conservation and functional importance of these complexes in the
332 land plant and animal lineages, a homology search failed to identify the core proteins of PRC2
333 in the *Ectocarpus* genome, and a similar result was obtained for PRC1 ([2]; Fig. 8, Additional
334 file 5: Table S10). This analysis did identify an orthologue of the WD domain protein RbAp48
335 but this protein is also known to be a key component of other regulatory complexes in the cell,
336 such as CAF-1, NuRD and NURF [27]. Moreover, analysis of the *Cladosiphon okamuranus*
337 [28], *Nemacystus decipiens* [29] and *Saccharina japonica* [30] genomes indicated that they too
338 lacked all the PRC2 and PRC1 core proteins except RbAp48, which is consistent with absence

339 of PRC complexes being a general feature of the brown algae (Table. S10). We cannot rule out
340 the possibility that the brown algae possess highly diverged versions of the PRC2 and PRC1
341 complexes but this would seem unlikely as the evolutionary time separating brown algae from
342 plants and animals is similar to that separating the latter two lineages. Moreover, orthologues
343 of all the PRC2 and PRC1 core components were found in the genome of the diatom
344 *Phaeodactylum tricornutum*, which also belongs to the stramenopile lineage ([31]; Additional
345 file 5: Table. S10). Analysis of the limited number of complete genomes available for the
346 stramenopiles indicated that PRC1 was lost after divergence from the diatoms and PRC2 at a
347 later stage after divergence from the Pelagophyceae (Fig. 8b). Absence of PRC2 in *Ectocarpus*
348 was supported by the fact that the mass spectrometry analysis did not detect any evidence of
349 tri-methylation of H3K27. Histone immunoblots with an antibody raised against H3K27me3
350 (Additional file 5: Table S11) weakly detected a protein of the expected size but the signal was
351 only detected after a lengthy exposure (see legend to Additional file 4: Figure S3) and when
352 this antibody was employed in a ChIP-seq experiment we did not detect any specific chromatin
353 immunoprecipitation (data not shown). Similarly, an antibody raised against H2AK119ub did
354 not detect this histone PTM in *Ectocarpus* chromatin (Additional file 3: Figure S9). Therefore,
355 taken together, these analyses indicated that *Ectocarpus* lacks the PRC2 and PRC1 polycomb
356 complexes and, consequently, the histone modifications mediated by these complexes, tri-
357 methylation of H3K27 and mono-ubiquitination of H2AK119.
358
359 **Discussion**
360
361 **Histone post-translational modifications in a multicellular brown alga**
362 Brown algae are the third most developmentally complex phylogenetic group on the planet and
363 include members whose bodyplans rival those of land plants in their cellular and developmental
364 complexity. However, compared to animals and land plants, many aspects of brown algal
365 developmental biology are still very poorly understood, including epigenetic regulatory
366 mechanisms. The objective of this study was to provide a comprehensive overview of histone
367 PTMs in the model brown alga *Ectocarpus* and to investigate the relationship between patterns
368 of key PTMs across the genome and the developmental programs that mediate alternation
369 between the sporophyte and gametophyte generations during this seaweed's life cycle.
370 Mass spectrometry analysis of *Ectocarpus* histone preparations demonstrated that
371 brown algal histones are subject to a broad range of PTMs. Most of the modifications detected
372 had been reported previously for histones of organisms from other eukaryotic supergroups, such

373 as the land plants within the Archaeplastida or animals and fungi from the Opisthokonta.
374 However, the analysis also confirmed some stramenopile-specific histone PTMs such as
375 H2A.ZK9ac and H2A.ZK15ac and detected a previously unreported (and therefore possibly
376 brown-algal-specific) modification, H2A.ZR38me1. Overall, however, this study confirmed
377 the conclusion, based on an analysis of diatom histone PTMs [22], that a large number of
378 histone PTMs detected in other eukaryotic supergroups are conserved in the stramenopiles. This
379 conclusion is consistent with the observation that histone molecules have been highly conserved
380 during eukaryotic evolution [32,33] and with evidence that many of the proteins involved in
381 post-translational modification of histones can be traced back to the last eukaryotic common
382 ancestor [34]. Indeed, genes encoding histone modifying enzymes with highly conserved
383 domain structures and domain sequences, such as the histone acetyltransferases, were detected
384 in the *Ectocarpus* genome. For other families, it was difficult to assign predicted proteins to
385 specific functions due to sequence divergence. Experimental analysis of protein function will
386 therefore be necessary to identify the specific writers and readers of many histone PTMs in
387 brown algae.

388 The *Ectocarpus* genome does not appear to encode the core components of the
389 polycomb complexes PRC1 and PRC2. Absence of polycomb complexes in a complex
390 multicellular organism may seem surprising from the perspective of animal and land plant
391 model systems but the components of PRC2 are known to exhibit a patchy distribution across
392 the diverse eukaryotic supergroups [35], indicating that they have been repeatedly lost over
393 evolutionary time. In this respect, the absence of polycomb complexes in brown algae
394 represents a striking example of how comparative analysis of groups that have evolved complex
395 multicellularity independently of plants and animals can be used to test the degree to which
396 specific genetic systems (in this case polycomb regulation) have been essential for the evolution
397 of complex multicellularity.

398 Analysis of ChIP-seq data allowed inferences to be made about the possible functional
399 roles of eight histone PTMs in brown algae. H3K4me2, H3K4me3, H3K9ac, H3K14ac and
400 H3K27ac were all detected predominantly at TSSs and the degree to which genes were marked
401 with these PTMs was proportional to their level of expression (transcript abundance),
402 suggesting a role in promoting or facilitating gene transcription. Note, however, that we cannot
403 exclude the opposite hypothesis, that deposition of these PTMs is promoted by gene
404 transcription.

405 Further evidence for a correlation between gene expression and the presence of the five
406 TSS PTMs was provided by a comparative analysis of chromatin states at GBGs between life

407 cycle generations, which indicated a broad correlation between accumulation of H3K4me2,
408 H3K4me3, H3K9ac and H3K27ac at the TSS and increased expression of GBGs during the life
409 cycle. However, patterns of histone PTMs at TSSs were not strictly correlated with gene
410 expression indicating that, if these marks do facilitate transcription, they are not sufficient alone
411 to induce gene expression. This conclusion was supported by a genome-wide analysis of closely
412 spaced (<600 bp) pairs of divergently transcribed genes, which showed that the TSS regions of
413 gene pairs overlapped but the two genes were nonetheless independently regulated. At present,
414 it is not clear what processes might be necessary, in addition to the deposition of histone PTMs
415 at the TSS, to induce gene expression in *Ectocarpus*, but these could include, for example, the
416 presence of specific transcription factors and the transcription machinery or the presence of
417 additional histone PTMs, including alternative acyl marks such as propionylation, butyrylation
418 and crotonylation which have been shown to stimulate transcription in cell-free assays [36].
419 Peaks of H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac are detected near the
420 TSSs of active genes in both animals and land plants and all of these marks have been associated
421 with active gene expression in both lineages ([37–48]; with the possible exception of
422 H3K4me2, which may represent a repressive mark on plants; [48,49]). Overall, therefore, these
423 five histone PTMs appear to have similar roles associated with gene activation at promoters in
424 animals, land plants and brown algae, indicating that at least part of the "histone code" may be
425 universal and therefore have a very deep evolutionary origin.
426 H3K36me3 marked gene bodies and was absent from TSSs and TESs but its presence
427 was also positively correlated with the level of gene expression. H3K36me3 also marks the
428 bodies of active genes in both animals and land plants, although its distribution along the gene
429 body differs in the two lineages. In *Arabidopsis* and rice a strong peak was observed at the 5'
430 end of the gene [44,47,50], whereas strong signals have been reported at the 3' ends [51] and
431 over exons [52] of animal genes. H3K36me3 is catalysed by SETD2, which is associated with
432 Pol II and is therefore directly linked to transcript elongation [53]. H3K36me3 has also been
433 shown to be involved in alternative splicing in animals and plants [54,55]. The position of
434 H3K36me3 within *Ectocarpus* genes and the positive correlation with transcript abundance are
435 consistent with an evolutionarily conserved role associated with transcript elongation.
436 H4K20me3, was associated with transposons and repeat sequences, both between and
437 within genes. The presence of this mark at intronic transposons was inversely correlated with
438 levels of gene expression but its effect on gene expression may be indirect, for example as a
439 consequence of silencing intronic transposons. The distribution of H4K20me3 suggests a
440 similar role to that observed in animals, where H4K20me3 is enriched at transposons and in

441 heterochromatin [56–58] and where it has been shown to repress transposons [58]. Note that
442 H4K20me3 appears to have a different role in land plants as it localises to euchromatin and is
443 associated with transcriptional activation [59].

444 Heterochromatin is marked by methylation of lysine 9 of histone H3 in animals and land
445 plants [60]. H3K9me3 is the most abundant heterochromatin marker in animals [61,62]. In
446 plants, heterochromatin is silenced by a feedback loop involving H3K9me2 and DNA
447 methylation [63], whereas H3K9me3 is associated with euchromatin and appears to have a
448 different function [50]. H3K9me2 and H3K9me3 were detected in *Ectocarpus* but appear to be
449 present at low abundance and ChIP-seq profiling did not provide any exploitable information
450 about the distribution of these PTMs (unpublished results, data available at NCBI Gene
451 Expression Omnibus accession GSE146369). More sensitive analysis methods will be
452 necessary to investigate the roles of H3K9me2 and H3K9me3 in brown algae but it is possible
453 that other PTMs, such as H4K20me3 and H3K79me3, carry out at least some of the functions
454 that have been attributed to these PTMs in other lineages.

455 H3K79me2 exhibited an unusual distribution pattern, marking discrete regions of the
456 genome that often spanned multiple genes. The function of the H3K79me2 is not clear, although
457 H3K79me2-marked regions were strongly correlated with the presence of H4K20me3 and
458 commonly included transposons (88.6% of H3K79me2 regions of >5 kbp included a ≥400 bp
459 repeated element). As observed for H4K20me3, the presence of H3K79me2 was correlated with
460 repression of gene expression (lower levels of transcript abundance). Interestingly, genes
461 marked with both H3K79me2 and H4K20me3 were expressed at significantly lower levels than
462 genes marked with only one of the two PTMs, indicating that the mechanisms linking
463 H3K79me2 and H4K20me3 with gene repression act independently.

464 Taken together, the above observations would be consistent with a role for H3K79me2
465 in the silencing of repeated elements (based on a similar argument to that proposed above for
466 H4K20me3). If H3K79me2 does have a role in the silencing of transposons, it is not clear why
467 the H3K79me2 regions often cover much larger regions than the transposon(s) itself. One
468 possible explanation may be that the regions extend to neighbouring TSSs in order to regulate
469 transcriptional activity in a genomic region containing a transposon. We also noted that
470 H3K79me2 tended to mark convergent gene pairs, even in regions that contained little repeated
471 sequence. A possible additional role of H3K79me2 at these loci may be to regulate transcription
472 in some manner to limit or respond to the formation of double-stranded RNA due to transcript
473 overlap. Such double-stranded RNA could be problematic within the cell if it activates small

474 RNA pathways but note that an earlier analysis did not find any evidence that convergent gene
475 pairs were a preferential source of sRNAs in *Ectocarpus* [64].
476 The distribution pattern observed for H3K79me2 in *Ectocarpus* is very different from
477 the pattern observed in animals, where H3K79me2 marks the 5' part of the gene body as part
478 of a mutually exclusive pattern with H3K36me3, which marks the 3' part [52,65,66]. The
479 H3K79me2/H3K36me3 boundary corresponds to the first internal exon [52]. In animals, both
480 H3K79me2 and H3K36me3 are considered to be associated with active transcription [65–67],
481 and more specifically transcript elongation, although conflicting evidence has been reported
482 [51]. In addition, a role has been proposed for H3K79me2 in the regulation of alternative
483 splicing [68], although again there is evidence to the contrary [52]. In fungi H3K79me2 has
484 also been found to be associated with euchromatin and to be excluded from heterochromatin
485 [69]. H3K79 methylation has not been detected in land plants [21]. Based on the different
486 distribution patterns, H3K79me2 would appear to have different roles in animals and in brown
487 algae.
488 We also observed that a region of chromosome six, containing a large, inserted DNA
489 virus genome, was extensively marked with both H3K79me2 and H4K20me3. The viral genes
490 at this locus have been shown to be transcriptionally silent in a range of tissues and life cycle
491 stages [18], suggesting possible roles for H3K79me2 and H4K20me3 in defending the alga
492 against lysogenic viral infections by silencing inserted viral genomes.
493
494 **Histone PTM patterns during the *Ectocarpus* life cycle**
495 To our knowledge this is the first study to have compared patterns of histone PTMs across the
496 two generations of a haploid-diploid life cycle (although previous studies have looked at PTMs
497 associated with the repression of germline genes during the sporophyte generation; [70,71]).
498 Overall, genome-wide patterns of histone PTM were found to be stable between the *Ectocarpus*
499 sporophyte and gametophyte generations, with the marks deposited at most genes remaining
500 unchanged. The *Ectocarpus* sporophyte and gametophyte are very similar morphologically
501 (near-isomorphy) and it is possible that more marked changes in histone PTM patterns occur
502 during the life cycles of brown algal species (such as kelps) that exhibit greater differences
503 between the sporophyte and gametophyte generations. However, note that there does not appear
504 to be a strong correlation between the degree of morphological heteromorphy between life cycle
505 generations and the proportion of GBGs in the genome [72]. Analysis of the epigenetic status
506 of GBGs during the life cycle indicated that the differential expression patterns of these genes
507 were correlated with modifications of multiple TSS-localised histone PTMs. Again, as stated

508 above, is not clear at present whether these histone PTM modifications are involved in
509 facilitating or potentiating gene upregulation or whether, on the contrary, changes in the
510 transcription status of these genes leads to changes in the pattern of histone PTMs at the TSS.
511
512 **Conclusion**
513 This study provides a first overview of the nature and functions of histone PTMs in the brown
514 algae, a group that has evolved complex multicellularity independently of the animal and land
515 plant lineages. The general, emerging picture is that epigenetic regulatory mechanisms are
516 broadly conserved with those of other eukaryotic supergroups with some key differences that
517 will potentially provide important insights into the epigenetic regulation of developmental
518 processes, including novel histone modifications, unusual patterns of deposition of known
519 PTMs such as H3K79me2 and the absence of polycomb complexes. The data and resources
520 generated by this study will provide a foundation for future studies aimed at understanding
521 epigenetic processes in the brown algae.
522
523
524 **Material and Methods**
525
526 **Strains and growth conditions**
527 The *Ectocarpus* sp. strain used in this study was Ec32 [73]. *Ectocarpus* was cultivated as
528 described previously [74]. Cultures were grown at 13°C with a 12h/12h day/night cycle and 20
529 µmol photons.m^{-2}.s^{-1} irradiance. Gametophytes were obtained from germinating meio-spores,
530 the sporophyte generation corresponded to partheno-sporophytes derived from parthenogenetic
531 gamete development.
532
533 **Detection of histone PTMs using mass spectrometry**
534 *Ectocarpus* histone proteins were isolated using the method described by Tirichine *et al*. [75].
535 Briefly, histones were extracted from isolated nuclei in 1 M $CaCl_2$, 20 mM Tris HCl pH 7.4 in
536 the presence of a cocktail of protease inhibitors. After precipitation of the acid-insoluble
537 fraction in 0.3 M HCl, the histones were precipitated by dropwise addition of TCA, centrifuged
538 and the pellet washed with 20% TCA and 0.2% HCl.
539 Gel purification and digestion of histones and mass spectrometry analysis were carried
540 out essentially as described by Veluchamy *et al*. [22]. Briefly, histone proteins, excised from a
541 14% SDS-polyacryamide gel, were digested overnight with endoproteinase (12.5 ng/µl), trypsin

542 (Promega), chymotrypsin (12.5 ng/µl, Promega), ArgC (12.5 ng/µl, Promega) or elastase (20
543 ng/µl Sigma-Aldrich). Peptides were analysed using a Q Exactive HF-X (Thermo Scientific) or
544 a TripleTOF™ 6600 (ABSciex) mass spectrometer coupled to a RSLCnano system (Ultimate
545 3000, Thermo Scientific, Bremen, Germany). Spectra were generated using Xcalibur (version
546 2.0.7) and analysed with Mascot™ (version 1.4, Thermo Scientific) using an in-house database
547 consisting of the complete *Ectocarpus* proteome available at Orcae [76]. The mass spectrometry
548 proteomics data have been deposited in the PRIDE database [77] via ProteomeXchange with
549 identifier PXD013535. Details of the liquid chromatography-tandem mass spectrometry
550 parameters are provided in the Supplementary Methods.
551
552 **Detection of histone PTMs using immunoblots**
553 Commercially available antibodies (Additional file 5: Table S11) for a broad range of histone
554 PTMs were tested against *Ectocarpus* chromatin preparations using immunoblots as described
555 previously [78]. For antibodies that gave no signal or a weak signal with *Ectocarpus* extracts,
556 human HeLa cell extracts were used as a positive control. Briefly, HeLa cells were harvested
557 and centrifuged at room temperature at 1500 rpm for 5 min and then lysed in ice for 20 minutes
558 using lysis buffer A (5 mM PIPES pH8.0, 85 mM KCl, 0.5% IGEPAL® CA-630). This cell
559 extract was used for the immunoblot. An anti-histone H4 antibody (Millipore 05-858) was used
560 as a loading control.
561
562 **Genome-wide detection of histone PTMs**
563 The ChIP-seq experiments were carried out in two batches, the first for H3K4me3, H3K9ac,
564 and H3K27ac, the second for H3K4me2, H3K14ac, H3K36me3, H4K20me3, H3K79me2 and
565 H3K27me3. RNA-seq data (see below) was generated for both batches to ensure that the histone
566 PTM and gene expression data were fully compatible. For ChIP-seq, *Ectocarpus* tissue was
567 fixed for five minutes in seawater containing 1% formaldehyde and the formaldehyde
568 eliminated by rapid filtering followed by incubation in PBS containing 400 mM glycine. Nuclei
569 were isolated by grinding in liquid nitrogen and in a Tenbroeck Potter in nuclei isolation buffer
570 (0.1% triton X-100, 125 mM sorbitol, 20 mM potassium citrate, 30 mM MgCl2, 5 mM EDTA,
571 5 mM β-mercaptoethanol, 55 mM HEPES at pH 7.5 with cOmplete ULTRA protease
572 inhibitors), filtering through Miracloth and then washing the precipitated nuclei in nuclei
573 isolation buffer with, and then without, triton X-100. Chromatin was fragmented by sonicating
574 the purified nuclei in nuclei lysis buffer (10 mM EDTA, 1% SDS, 50 mM Tris-HCl at pH 8
575 with cOmplete ULTRA protease inhibitors) in a Covaris M220 Focused-ultrasonicator (duty

576 25%, peak power 75, cycles/burst 200, duration 900 sec at 6°C). The chromatin was incubated
577 with an anti-histone PTM antibody (Additional file 5: Table S11) overnight at 4°C and the
578 immunoprecipitation carried out using Dynabeads protein A and Dynabeads protein G.
579 Following immunoprecipitation and washing, a reverse cross-linking step was carried out by
580 incubating for at least six hours at 65°C in 200 mM NaCl and the samples were then digested
581 with Proteinase K and RNAse A. Purified DNA was analysed on an Illumina HiSeq 4000
582 platform with a single-end sequencing primer over 50 cycles. At least 20 million reads were
583 generated for each immunoprecipitation (Additional file 5: Table S12).

584 Nuclei were prepared for MNase digestion using the same procedure as that used for
585 ChIP samples except that the isolation buffer did not contain EDTA. Each sample of nuclei was
586 digested with 400 agarose gel units of MNase for 10 min at 37°C in MNase reaction buffer (5
587 mM $CaCl_2$, 60 mM KCl, 0.5 mM DTT, 15 mM NaCl, 125 mM sorbitol, 50 mM Tris-HCl pH
588 7.5). Following lysis of the nuclei, reverse cross-linking and digestion with Proteinase K and
589 RNAse A, 120-210 bp fragments were excised from a 2% agarose gel and sequenced on a
590 HiSeq 4000 platform (Additional file 5: Table S12).

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

594 Quality control of the sequence data was carried out using FASTQC [79]. Poor quality
595 sequence was removed and the high quality sequence was trimmed with Flexcat [80]. Illumina
596 reads were mapped onto the *Ectocarpus* genome ([18]; available at Orcae; [76]) using Bowtie
597 [81]. Peaks corresponding to regions enriched in PTMs were identified using the MACS2
598 (version 2.1.1) callpeak module (minimum FDR of 0.01) and refined with the MACS2
599 bdgpeakcall and bdgbroadcall modules [82]. Additional file 5: Table S13 summarises the
600 correspondence between PTM peaks and genes across the *Ectocarpus* genome. Co-localised
601 peaks corresponding to regions enriched in several PTMs were detected using MACS2 outputs
602 in BedTools multiinter [83]. Overlaps of PTM peaks with other genomic features were analysed
603 using BedTools intercept [83]. The analysis of association of H3K79me2 domain borders with
604 TSSs and TESs was carried out in R, using the function gr.rand from the package gUtils to
605 generate the random anchors. Heatmaps, average tag graphs and coverage tracks were plotted
606 using EaSeq software [84], deepTools [85] or pyGenomeTracks
607 (https://pypi.org/project/pyGenomeTracks/). Circos graphs were generated using Circos
608 software [86]. These analyses were carried out for two biological replicates for each PTM
609 during both the sporophyte and gametophyte generations of the life cycle. Pearson correlation

610 analysis between replicates was performed with deepTools 3.2.0 [85]. Replicate samples were
611 strongly correlated (Pearson correlations of 0.87 to 0.99; Additional file 4: Figure S10).
612 Normalisation was carried out using the input chromatin data (Additional file 5: Table S12).
613 Genome-wide analyses (e.g. Fig. 2) used data from sporophytes unless stated otherwise.
614
615 **Comparisons of sporophyte and gametophyte transcriptomes using RNA-seq**
616 RNA for transcriptome analysis was extracted from the same duplicate sporophyte and
617 gametophyte cultures that were analysed for each of the two ChIP-seq experiments using the
618 Qiagen RNeasy plant mini kit with an on-column Deoxyribonuclease I treatment. RNA quality
619 and concentration was then analysed on a Qubit® 2.0 Fluorometer using Qubit RNA BR assay
620 kit (Invitrogen, Life Technologies, Carlsbad, CA, USA). Between 49 and 107 million sequence
621 reads were generated for each sample on an Illumina HiSeq 4000 platform with a single-end
622 sequencing primer over 150 cycles (Additional file 5: Table S12). Quality trimming of raw
623 reads was carried out with Flexcat [80] and reads of less than 50 nucleotides after trimming
624 were removed. Tophat2 [87] was used to map reads to the *Ectocarpus* genome [18] and the
625 mapped sequencing data was processed with featureCounts [88] to obtain counts for sequencing
626 reads mapped to exons. Expression values were represented as TPM. Differential expression
627 was detected using the DESeq2 package (Bioconductor; [89]) using an adjusted *p*-value cut-off
628 of 0.05 and a minimal fold-change of 2. The set of GBGs corresponded to genes that were
629 significantly differentially regulated between life cycle generation in both of the ChIP-seq
630 experiments.
631
632 **Searches for histone and histone modifying enzyme encoding genes in *Ectocarpus***
633 Histone and histone modifier genes were detected in the *Ectocarpus* genome [18] using Blast
634 [90] and manually reannotated when necessary, using GenomeView for structural reannotation
635 [91].

636 **Supplementary information**

637

638 **Additional file 1:** Supplementary methods

639 **Additional file 2: Figure S1.** Histone gene clusters in the *Ectocarpus* genome.

640 **Additional file 3: Figure S2.** Representative MS/MS spectra showing the identification of
641 histone modifications.

642 **Additional file 4: Figure S3.** Immunoblots of histone PTMs. **Figure S4.** Genomic distribution
643 of the histone PTMs H3K79me2 and H4K20me3.**Figure S5.** Boundaries of H3K79me2 regions
644 are preferentially located near TSSs and TESs. **Figure S6.** Comparison of genes marked with
645 H3K79me2 alone, H4K20me3 alone or both H3K79me2 and H4K20me3. **Figure S7.**
646 H3K79me2, H4K20me3, H3K36me3, H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac
647 signals for a region of chromosome 6 spanning an inserted viral genome. **Figure S8.**
648 Correlations of transcript abundances and histone PTM signals for divergently transcribed pairs
649 of genes. **Figure S9.** The histone PTM H2AK119ub1 was not detected in *Ectocarpus*. **Figure**
650 **S10.** Pearson correlation scores for comparisons of the genomic distributions of ChIP-seq signal
651 peaks for duplicate assays of the eight histone PTMs during both the sporophyte and
652 gametophyte generations.

653 **Additional file 5: Table S1.** Histone proteins in *Ectocarpus*. **Table S4.** H4K20me3 peaks
654 associated with different transposon families. **Table S5.** Genes with significantly different
655 transcript abundances during the gametophyte and sporophyte generations. **Table S6**.
656 Identification of enriched functional categories in the proteins encoded by the generation-biased
657 gene set. **Table S7**. The generation-biased gene set is enriched in secreted proteins. **Table S8.**
658 Analysis of long non-coding RNA genes that were differentially expressed during the
659 gametophyte and sporophyte generations. **Table S9**. lncRNA GBGs with an adjacent protein-
660 coding GBG. **Table S10.** Presence and absence of polycomb complex proteins in animals, land
661 plants and stramenopiles. **Table S11**. Anti-histone-PTM antibodies used in this study. **Table**
662 **S12.** Sequence data generated by this study. **Table S13**. Presence or absence of histone PTMs
663 and RNA-seq TPMs for all *Ectocarpus* genes in the sporophyte and the gametophyte
664 generations.

665 **Additional file 6: Table S2.** Putative PTM writers and erasers in *Ectocarpus* sp.

666 **Additional file 7: Table S3.** Presence or absence of histone post-translational modifications in
667 seven species from diverse eukaryotic supergroups.

668

669 **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 and Houda Benhelli-Mokrani for providing us with HeLa cells.

Funding

This work was supported by the Centre National de la Recherche Scientifique, Sorbonne University (including a PhD grant for SB), the Agence Nationale de la Recherche projects Epicycle and Idealg (ANR-19-CE20-0028-01 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 S12) 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, YW, SB and LT carried out the immunoblots, 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.

704 **Consent for publication**

705

706 Not applicable.

707

708 **Competing interests**

709

710 The authors declare that they have no competing interests.

711

712 **References**

713

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

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

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

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

722 5. Macaisne N, Liu F, Scornet D, Peters AF, Lipinska A, Perrineau M-M, et al. The
723 *Ectocarpus IMMEDIATE UPRIGHT* gene encodes a member of a novel family of cysteine-
724 rich proteins with an unusual distribution across the eukaryotes. Development. 2017;144:409–
725 18.

726 6. Godfroy O, Uji T, Nagasato C, Lipinska AP, Scornet D, Peters AF, et al.
727 DISTAG/TBCCd1 Is Required for Basal Cell Fate Determination in *Ectocarpus*. Plant Cell.
728 2017;29:3102–22.

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

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

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

737 10. Ohad N, Margossian L, Hsu YC, Williams C, Repetti P, Fischer RL. A mutation that
738 allows endosperm development without fertilization. Proc Natl Acad Sci U A. 1996;93:5319–
739 24.

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

743 12. Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ. Fertilization-
744 independent seed development in *Arabidopsis thaliana*. Proc Natl Acad Sci U A.
745 1997;94:4223–8.

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

749 14. Guitton AE, Page DR, Chambrier P, Lionnet C, Faure JE, Grossniklaus U, et al.
750 Identification of new members of Fertilisation Independent Seed Polycomb Group pathway
751 involved in the control of seed development in *Arabidopsis thaliana*. Development.
752 2004;131:2971–81.

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

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

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

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

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

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

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

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

775 23. Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, et al. The *Ectocarpus*
776 genome and the independent evolution of multicellularity in brown algae. Nature.
777 2010;465:617–21.

778 24. Gschloessl B, Guermeur Y, Cock J.M. HECTAR: a method to predict subcellular
779 targeting in heterokonts. BMC Bioinf. 2008;9:393.

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

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

784 27. Philpott A, Krude T, Laskey RA. Nuclear chaperones. Semin Cell Dev Biol. 2000;11:7–
785 14.

786 28. Nishitsuji K, Arimoto A, Iwai K, Sudo Y, Hisata K, Fujie M, et al. A draft genome of the
787 brown alga, *Cladosiphon okamuranus*, S-strain: a platform for future studies of "mozuku"
788 biology. DNA Res. 2016;23:561–70.

789 29. Nishitsuji K, Arimoto A, Higa Y, Mekaru M, Kawamitsu M, Satoh N, et al. Draft genome
790 of the brown alga, *Nemacystus decipiens*, Onna-1 strain: Fusion of genes involved in the
791 sulfated fucan biosynthesis pathway. Sci Rep. 2019;9:4607.

792 30. Ye N, Zhang X, Miao M, Fan X, Zheng Y, Xu D, et al. *Saccharina* genomes provide
793 novel insight into kelp biology. Nat Commun. 2015;6:6986.

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

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

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

801 34. Iyer LM, Anantharaman V, Wolf MY, Aravind L. Comparative genomics of transcription
802 factors and chromatin proteins in parasitic protists and other eukaryotes. Int J Parasitol.
803 2008;38:1–31.

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

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

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

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

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

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

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

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

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

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

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

830 46. Jurkowska RZ, Qin S, Kungulovski G, Tempel W, Liu Y, Bashtrykov P, et al. H3K14ac is
831 linked to methylation of H3K9 by the triple Tudor domain of SETDB1. Nat Commun.
832 2017;8:2057.

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

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

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

841 50. Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, et al. Integrative
842 epigenomic mapping defines four main chromatin states in *Arabidopsis*. EMBO J.
843 2011;30:1928–38.

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

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

848 53. Krogan NJ, Kim M, Tong A, Golshani A, Cagney G, Canadien V, et al. Methylation of
849 Histone H3 by Set2 in *Saccharomyces cerevisiae* Is Linked to Transcriptional Elongation by
850 RNA Polymerase II. Mol Cell Biol. 2003;23:4207.

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

853 55. Pajoro A, Severing E, Angenent GC, Immink RGH. Histone H3 lysine 36 methylation
854 affects temperature-induced alternative splicing and flowering in plants. Genome Biol.
855 2017;18:102.

856 56. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-
857 wide maps of chromatin state in pluripotent and lineage-committed cells. Nature.
858 2007;448:553–60.

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

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

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

865 60. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and function. Nat
866 Rev Mol Cell Biol. 2018;19:229–44.

867 61. Peters AHFM, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AAHA, Perez-Burgos L, et
868 al. Partitioning and plasticity of repressive histone methylation states in mammalian
869 chromatin. Mol Cell. 2003;12:1577–89.

870 62. Rice JC, Briggs SD, Ueberheide B, Barber CM, Shabanowitz J, Hunt DF, et al. Histone
871 methyltransferases direct different degrees of methylation to define distinct chromatin
872 domains. Mol Cell. 2003;12:1591–8.

873 63. Jiamu Du, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, et al. Dual binding of
874 chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in
875 plants. Cell. 2012;151:167–80.

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

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

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

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

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

889 69. Ng HH, Ciccone DN, Morshead KB, Oettinger MA, Struhl K. Lysine-79 of histone H3 is
890 hypomethylated at silenced loci in yeast and mammalian cells: A potential mechanism for
891 position-effect variegation. Proc Natl Acad Sci. 2003;100:1820.

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

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

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

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

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

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

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

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

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

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

916 80. Hansen P, Hecht J, Ibn-Salem J, Menkuec BS, Roskosch S, Truss M, et al. Q-nexus: a
917 comprehensive and efficient analysis pipeline designed for ChIP-nexus. BMC Genomics.
918 2016;17:873.

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

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

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

925 84. Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. An interactive environment for
926 agile analysis and visualization of ChIP-sequencing data. Nat Struct Mol Biol. 2016;23:349–
927 57.

928 85. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a
929 next generation web server for deep-sequencing data analysis. Nucleic Acids Res.
930 2016;44:W160-165.

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

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

936 88. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for
937 assigning sequence reads to genomic features. Bioinforma Oxf Engl. 2014;30:923–30.

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

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

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

945 92. Brown JW, Sorhannus U. A molecular genetic timescale for the diversification of
946 autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages.
947 PLoS One. 2010;5:e12759.

948

949

950

Figures

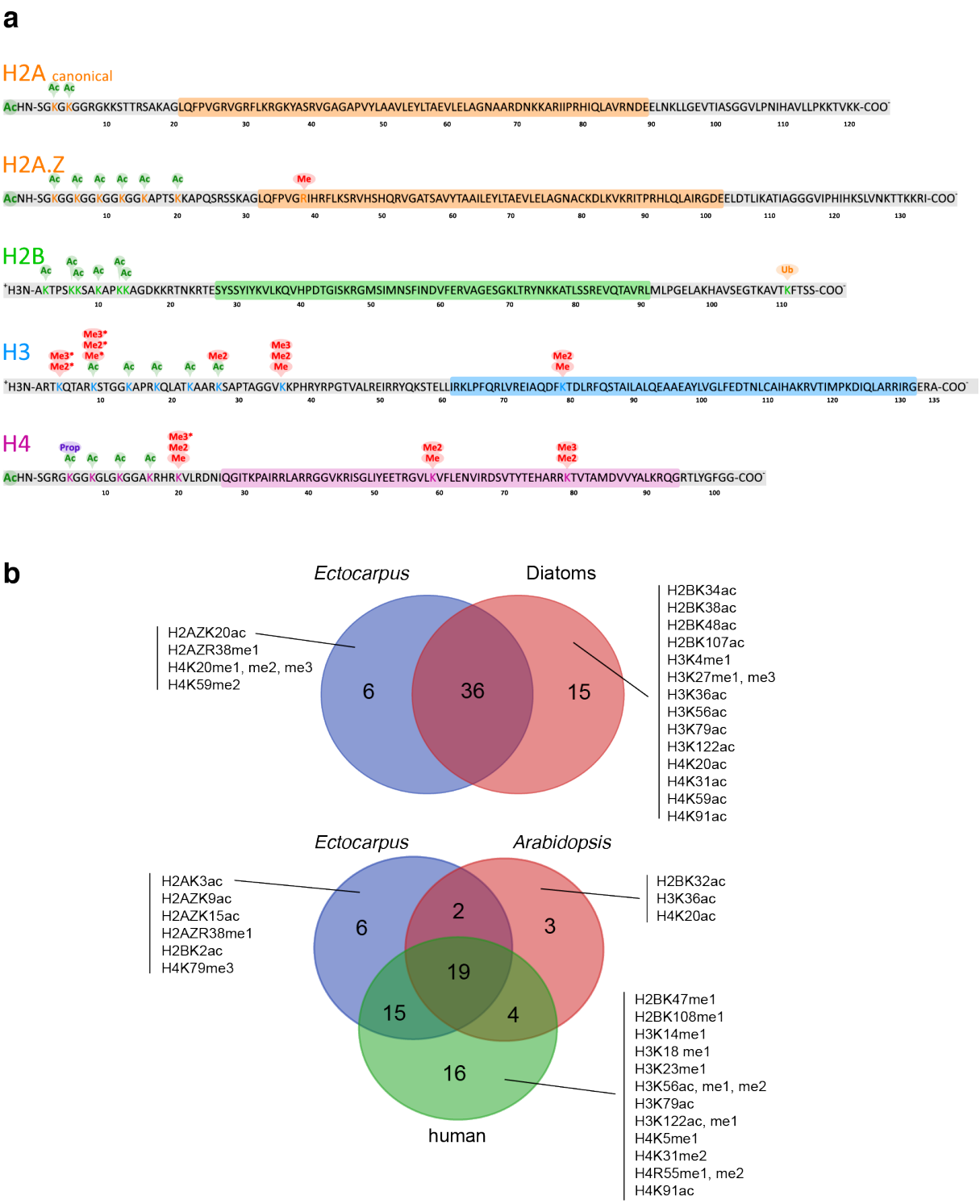


Fig. 1 a Post-translational modifications of *Ectocarpus* histones. Acetylation, methylation, ubiquitylation and propionylation modifications of core and variant histones identified in this study. Coloured boxes indicate globular core domains and grey boxes N- and C-terminal tails.

958 Amino acid positions are indicated below the protein sequence. Asterisks indicate histone
959 modifications that were only detected using antibodies. All other modifications were detected
960 by mass spectrometry. **b** Comparison of the histone PTMs detected in *Ectocarpus* with PTMs
961 reported for diatoms (combined data for *Phaeodactylum tricornutum* and *Thalassiosira*
962 *pseudonana*, upper panel) and for humans and the land plant *Arabidopsis* (lower panel). See
963 Additional file 5: Table S3 for details.
964

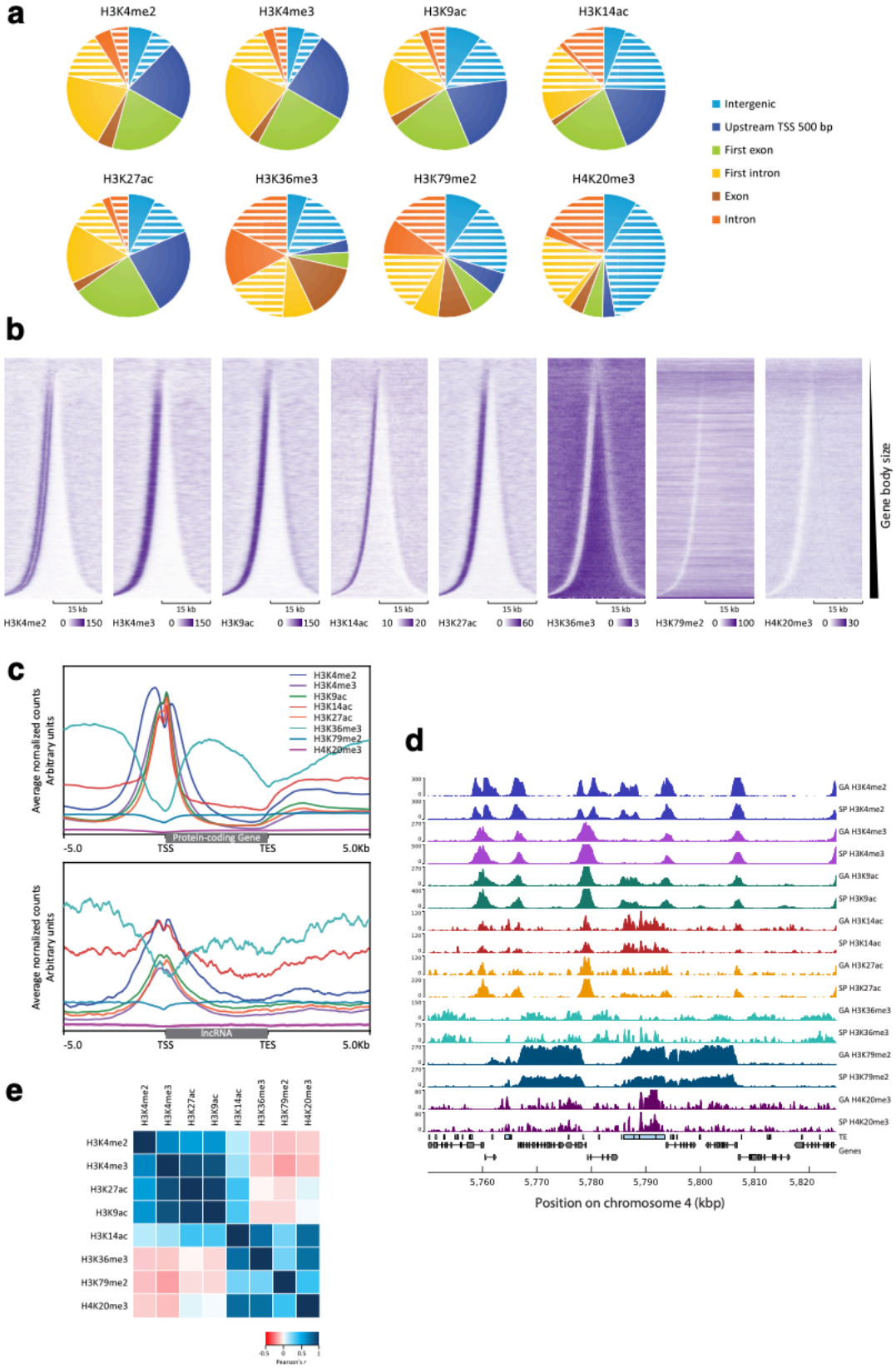


Fig. 2 Distribution of histone PTMs in the *Ectocarpus* genome. **a** Distribution of eight histone PTMs across six genome feature classes. Intergenic, all intergenic regions except the TSS

970 regions; Upstream TSS 500 bp, 500 bp region 5' to the transcription start site; Exon, all exon
971 sequence except first exons; Intron, all intron sequence except first introns. The sectors for
972 intergenic, intron and first intron regions have been separated in two to indicate values for
973 regions that contain TEs (white stripes) and regions that do not (full colour). **b** Distribution of
974 the eight histone PTMs across the complete set of *Ectocarpus* genes (17,447) sorted by gene
975 body size and centred on the gene body. The black brackets correspond to 15 kilobases (kb). **c**
976 Enrichment profiles of the eight histone PTMs for all protein-coding (top panel) and all lncRNA
977 (bottom panel) genes. Gene bodies are plotted as proportional lengths, upstream and
978 downstream intergenic regions in kb. **d** Representative region of the chromosome 4 showing
979 profiles of mapped ChIP-seq reads for the eight histone PTMs in the sporophyte (SP) and
980 gametophyte (GA) generations. Light blue boxes represent transposons (TEs). Grey boxes and
981 arrows represent exons superimposed on genes represented by a black line. **e** Pearson
982 correlation scores for comparisons of the genomic distributions of ChIP-seq signal peaks for
983 the eight histone PTMs.
984

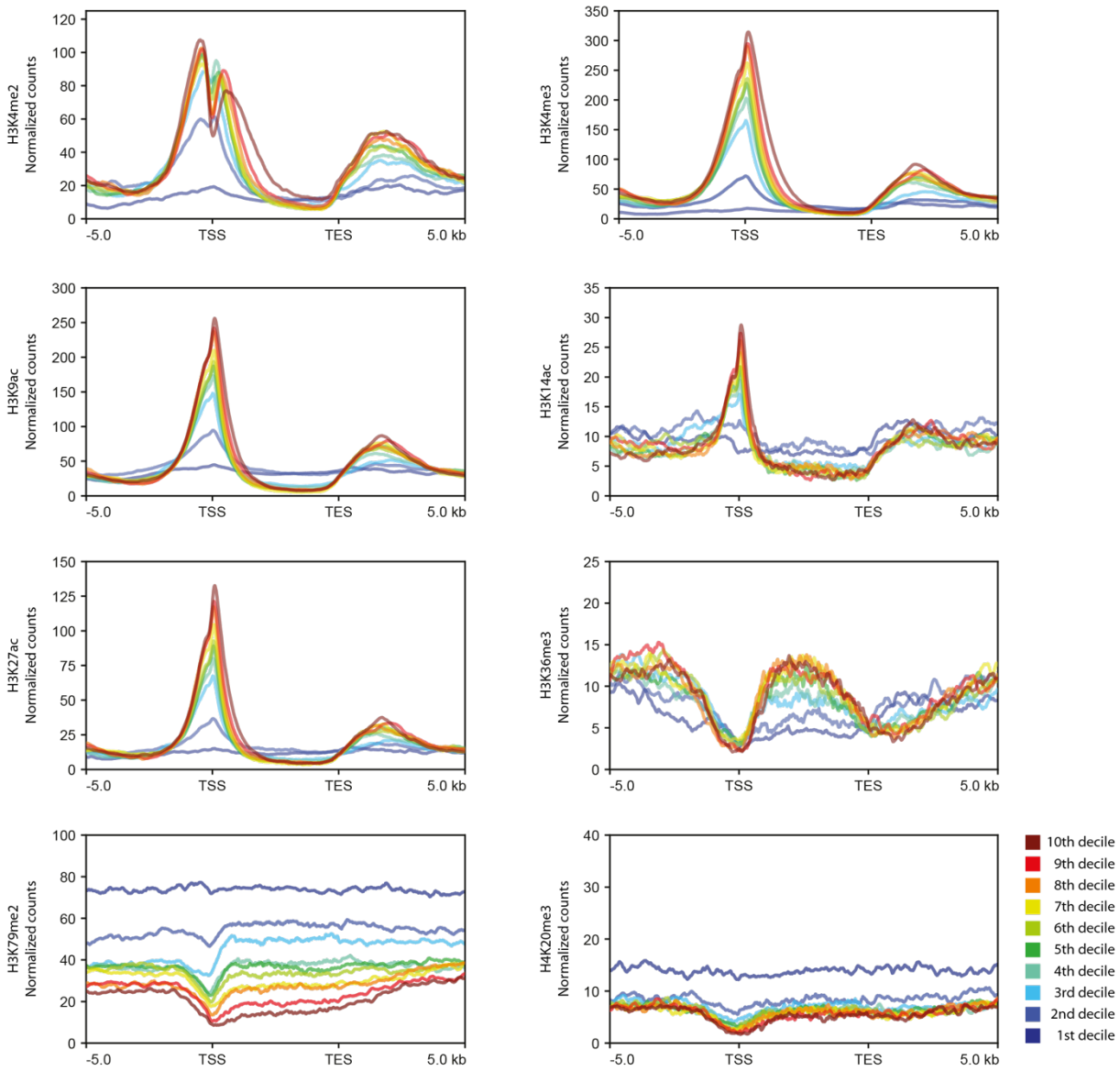


Fig. 3 Enrichment profiles for the histone PTMs H3K4me2, H3K4me3, H3K9ac, H3K14ac, H3K27ac, H3K36me3, H3K79me2 and H4K20me3 across genes expressed at different levels (TPM deciles). Gene bodies are plotted as proportional lengths, upstream and downstream intergenic regions in kilobases. TSS, transcription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

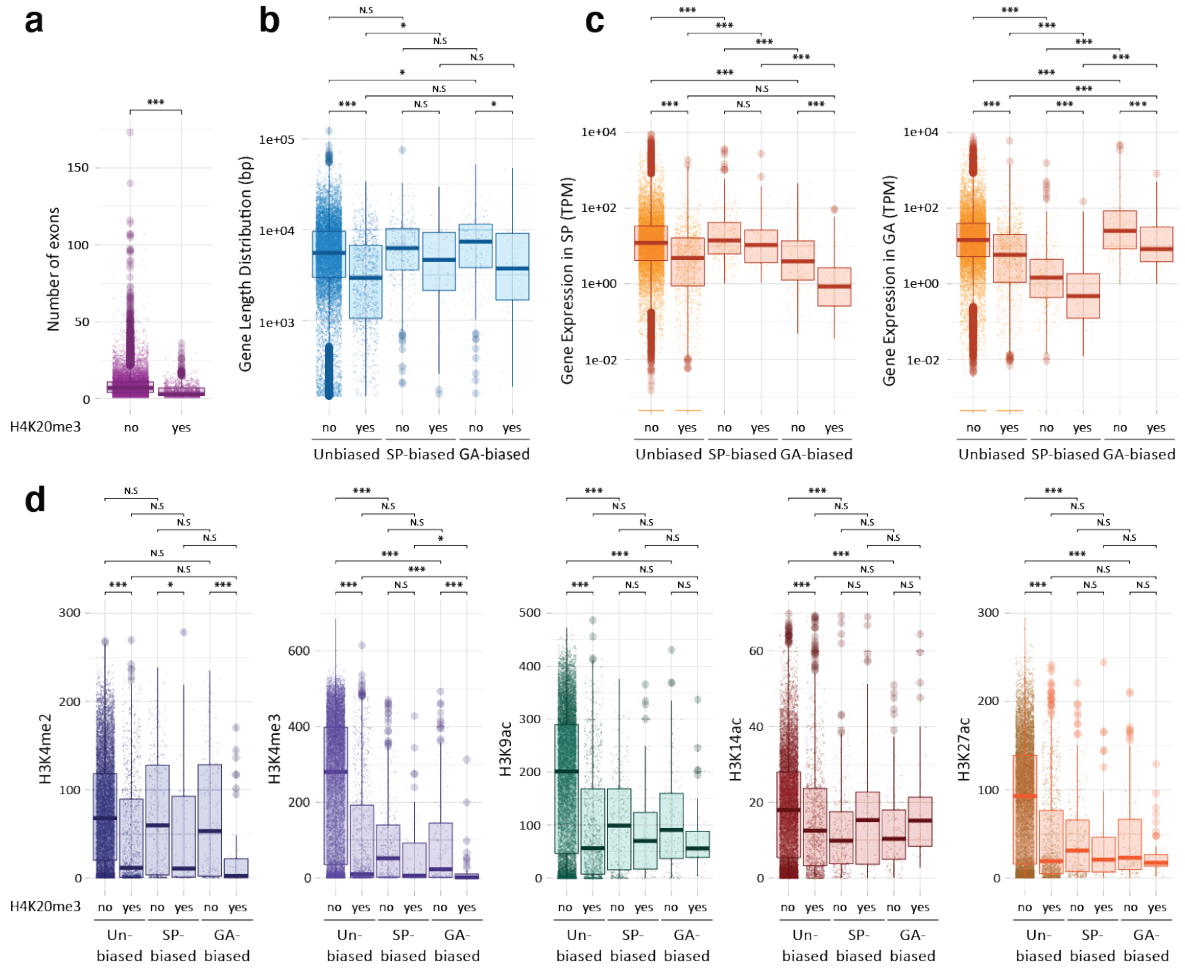


Fig. 4 Comparison of genes that contained or did not contain a H4K20me3 peak. **a** Genes with at least one H4K20me3 peak (yes) have fewer exons than genes without a H4K20me3 peak (no). **b** Lengths of genes with or without a H4K20me3 peak. **c** Abundances of transcripts for genes with or without a H4K20me3 peak. **d** Histone PTMs on genes with or without a H4K20me3 peak. SP, sporophyte; GA, gametophyte. Asterisks indicate significant differences (Bonferonni-adjusted Wilcoxon tests; ***, $p < 0.0001$; *, $p < 0.05$).

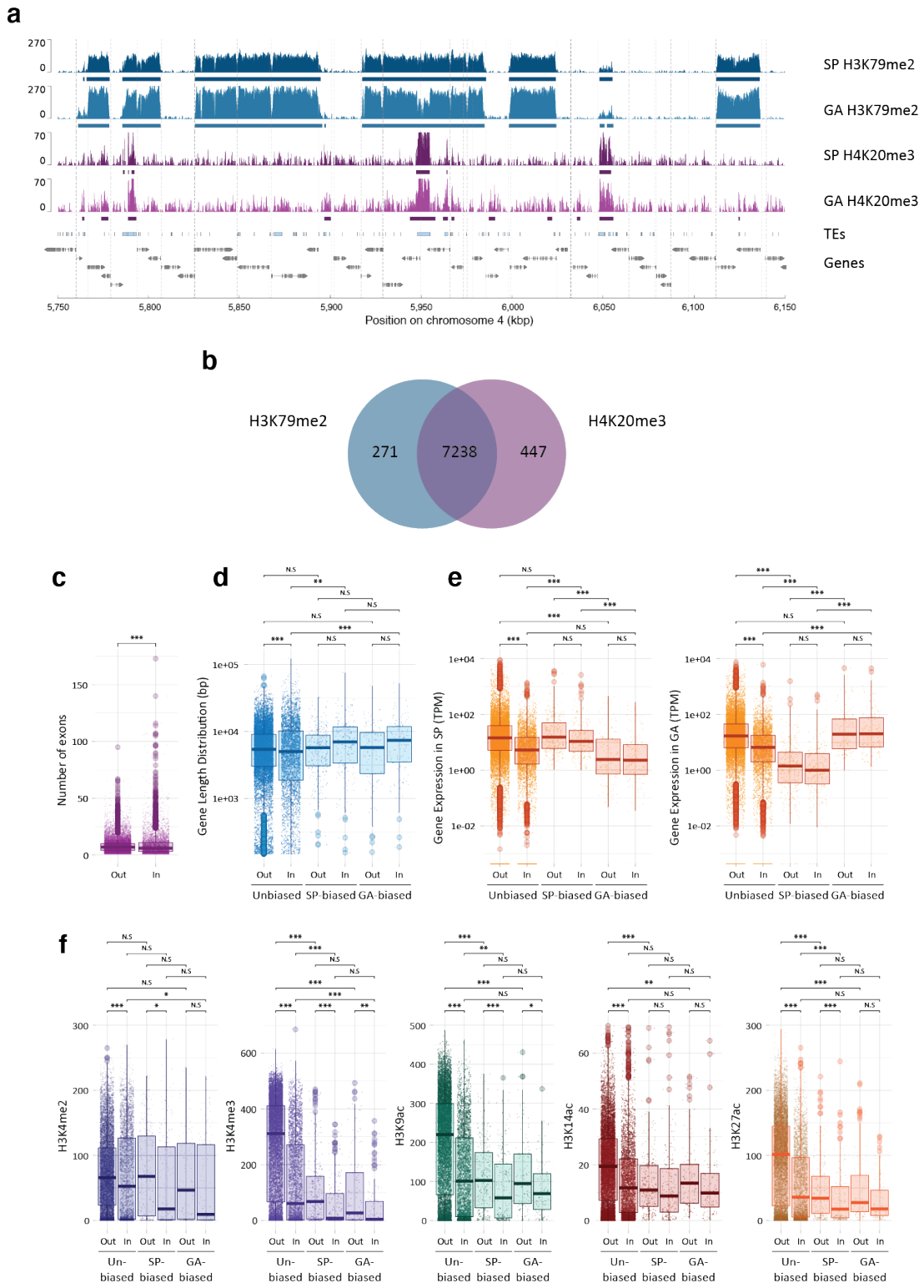
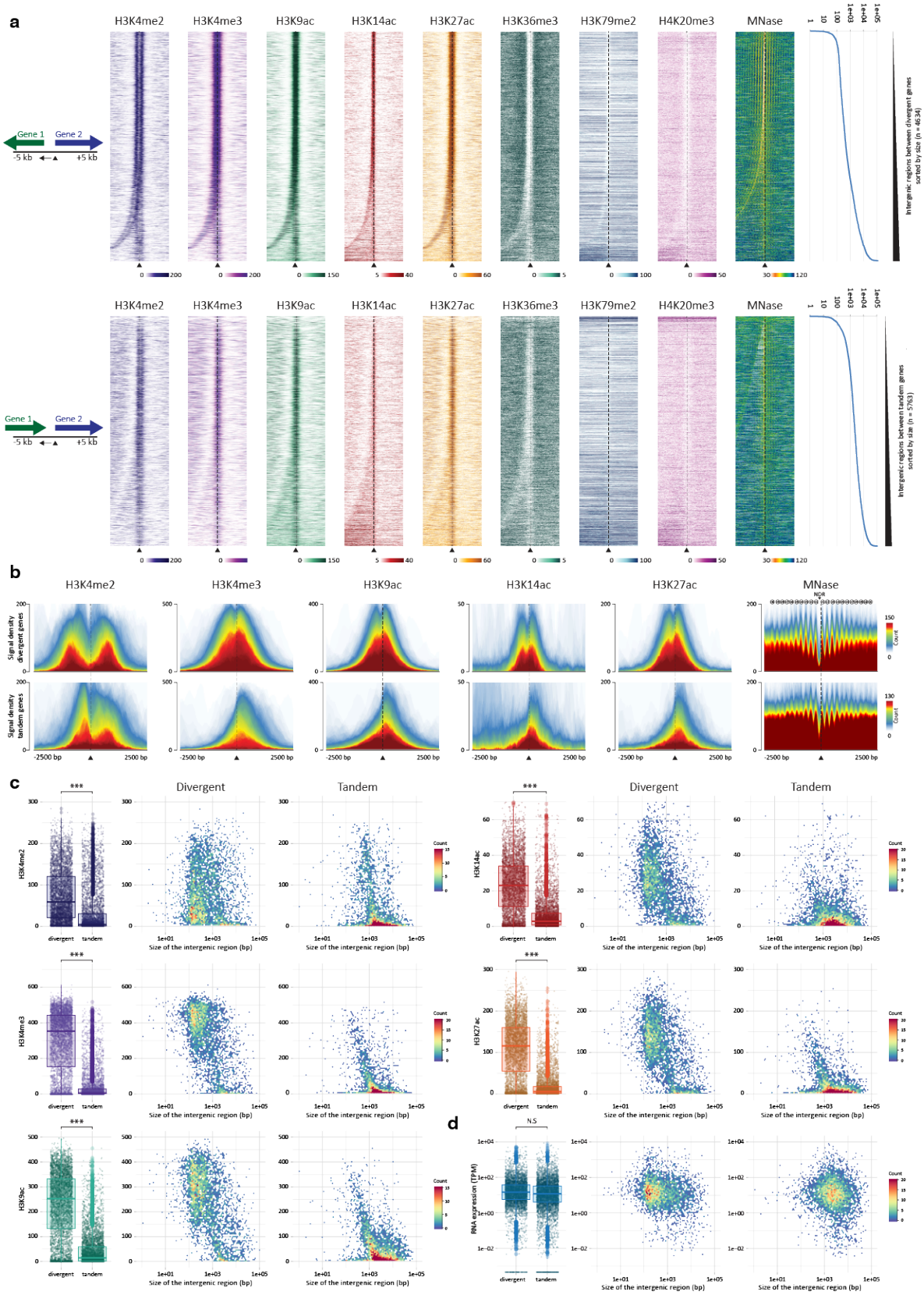
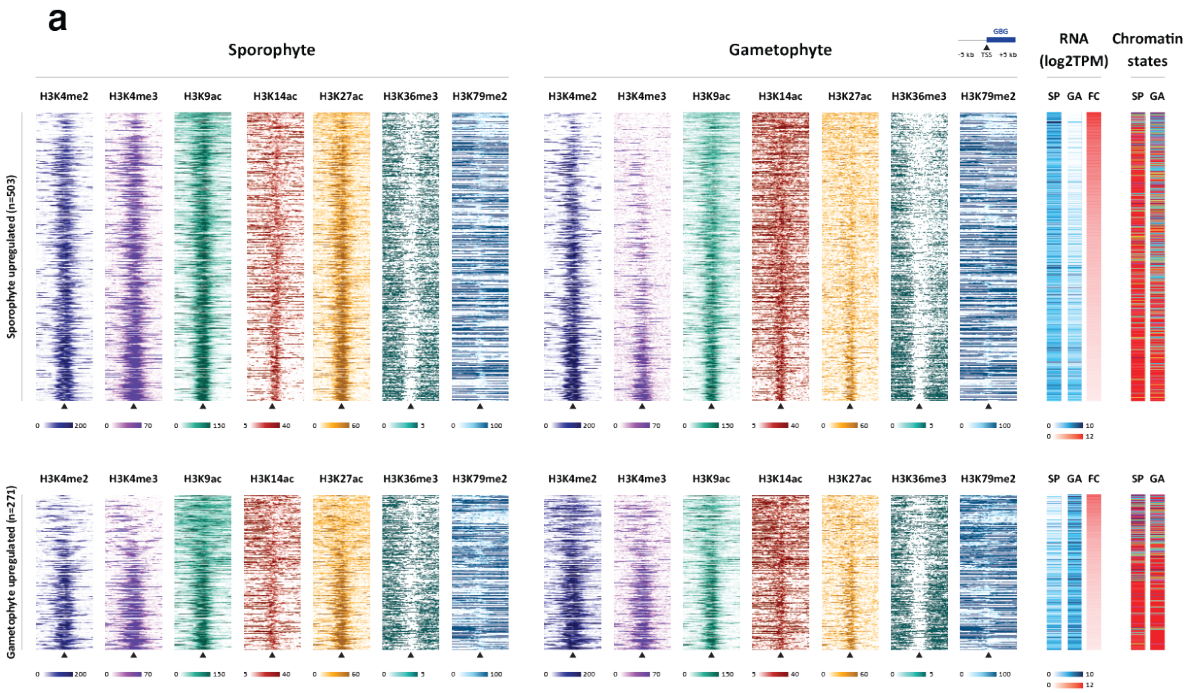


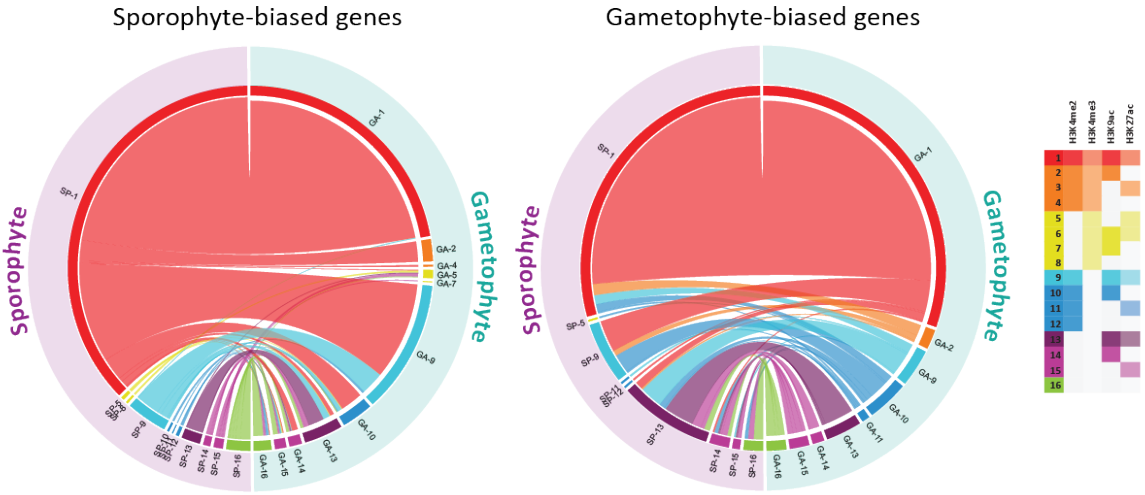
Fig. 5 Genomic distribution of the histone PTM H3K79me2. **a** Representative region of 0.4 Mbp from chromosome 4 (spanning the genes with LocusIDs Ec-04_005800 to Ec-04_006210) showing alternating regions marked or not with H3K79me2 (blue peaks) during both the sporophyte and gametophyte generations of the life cycle. Blue bars indicate H3K79me2 regions. The distribution of H4K20me3 (purple peaks) is also shown for comparison (peaks called by MACS2 are indicated by purple bars). Light blue boxes represent transposons (TEs).

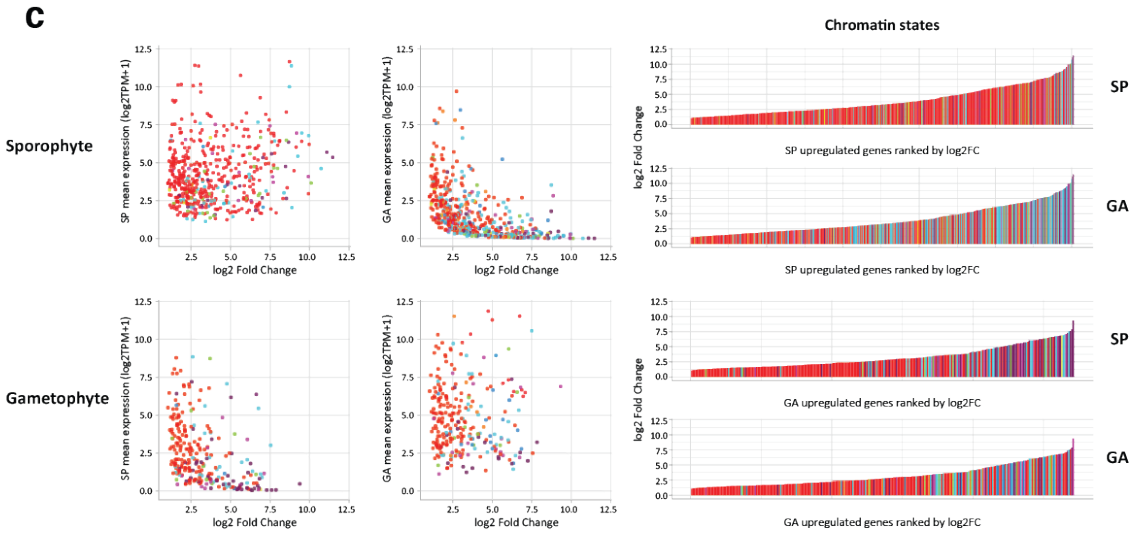
1016 Grey boxes and arrows represent exons superimposed on genes represented by a black line.
1017 Vertical dotted lines indicate the positions of TSSs. **b** Genome-wide overlap between peaks of
1018 H3K79me2 and H4K20me3. **c** Genes in H3K79me2 regions have fewer exons. **d** Lengths of
1019 genes within (In) or outside (Out) H3K79me2 regions. **e** Abundances of transcripts for genes
1020 within (In) or outside (Out) H3K79me2 regions. **f** Histone PTMs on genes within (In) or outside
1021 (Out) H3K79me2 regions. A gene was considered to be "In" if more than 50% of the gene body
1022 (TSS-TES) was inside a H3K79me2 region. SP, sporophyte; GA, gametophyte; kb, kilobase
1023 pairs. Asterisks indicate significant differences (Bonferonni-adjusted Wilcoxon tests; ***, $p <$
1024 0.0001; **, $p < 0.001$; *, $p < 0.05$).
1025


a
H3K4me2
H3K4me3
H3K9ac
H3K14ac
H3K27ac
H3K36me3
H3K79me2
H4K20me3
MNase
Gene 1
Gene 2
-5 kb
+5 kb
Intergenic regions between divergent genes sorted by size (n = 4634)
Intergenic regions between tandem genes sorted by size (n = 5763)
b
Signal density divergent genes
Signal density tandem genes
-2500 bp
2500 bp
NDR
Count
c
Divergent
Tandem
***
divergent
tandem
Size of the intergenic region (bp)
d
N.S
RNA expression (TPM)

1027
1028 **Fig. 6** Histone PTM distributions in intergenic regions between divergent and tandemly
1029 organised pairs of genes. **a** Distribution of eight histone PTMs and MNase-sensitive sites in
1030 regions of 10 kb surrounding the TSS of the right-hand gene for pairs of either divergently
1031 (upper panel) or tandemly (lower panel) organised genes. The lengths of the intergenic regions
1032 are plotted on the right in bp. **b** Signal distributions for five TSS-localised histone PTMs and
1033 MNase-sensitive sites in regions of 5 kb surrounding the TSS of the right-hand gene for pairs
1034 of either divergently (upper plots) or tandemly (lower plots) organised genes on chromosome
1035 4. **c** Abundance of five histone PTMs at the TSS of Gene 2 (see schema in A) in the intergenic
1036 regions of divergently or tandemly organised gene pairs together with plots indicating
1037 abundance in relation to the size of the intergenic region. **d** Transcript abundances for Gene 2.
1038 For **c** and **d**, heat score (count) indicates the number of genes with both the same PTM or
1039 transcript abundance and the same size intergenic region. MNase, micrococcal nuclease; kb,
1040 kilobases; bp, base pairs.
1041

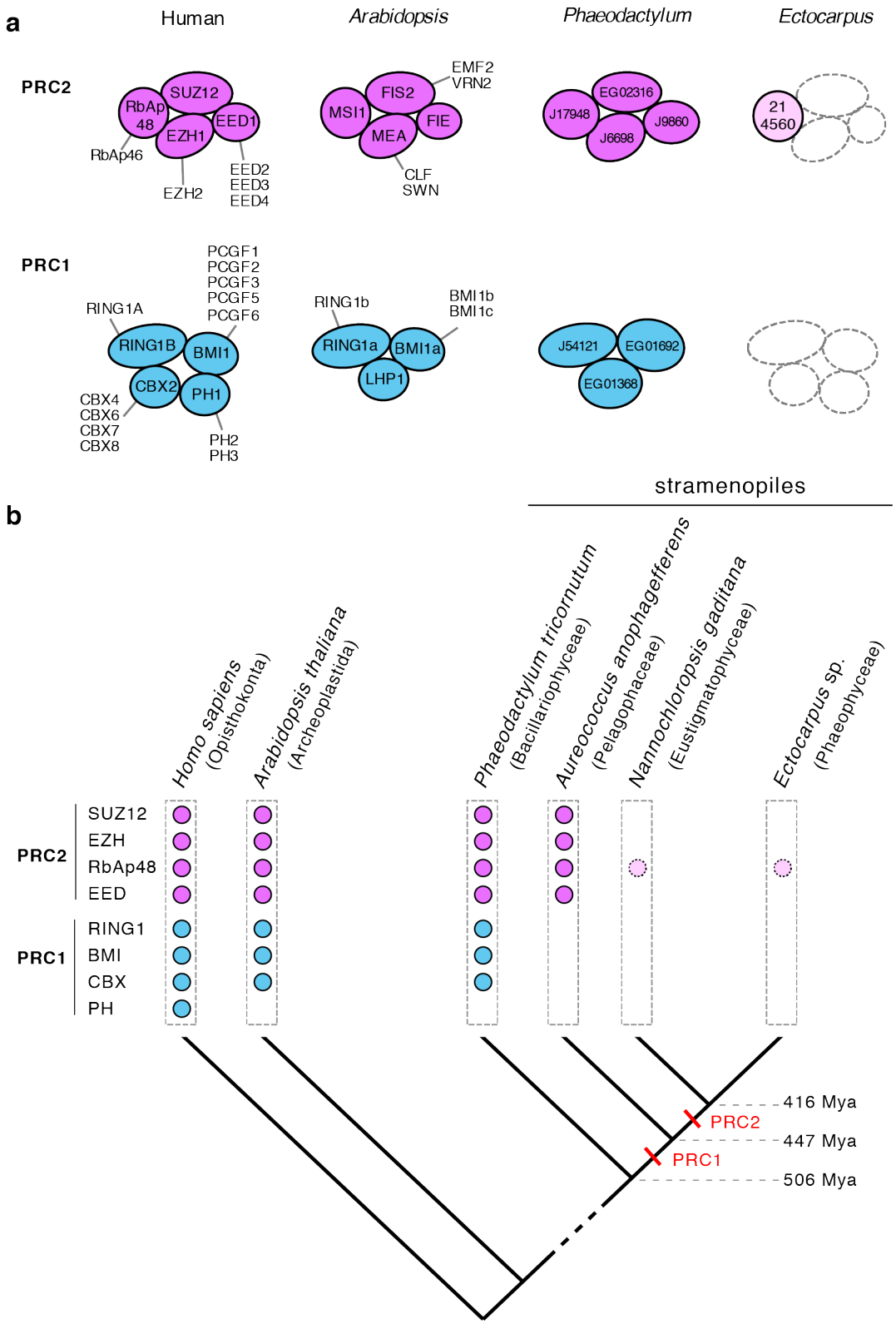

a
Sporophyte
Gametophyte
RNA (log2TPM)
Chromatin states
H3K4me2
H3K4me3
H3K9ac
H3K14ac
H3K27ac
H3K36me3
H3K79me2
SP GA FC
SP GA
Sporophyte upregulated (n=503)
Gametophyte upregulated (n=271)


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


c
Chromatin states
Sporophyte
Gametophyte
SP mean expression (log2TPM+1)
GA mean expression (log2TPM+1)
log2 Fold Change
SP
GA
SP upregulated genes ranked by log2FC
GA upregulated genes ranked by log2FC

1043
1044 **Fig. 7** Alterations of histone PTM patterns at generation-biased genes during the two
1045 generations of the life cycle. **a** Patterns of histone PTMs over 10 kb regions surrounding the
1046 TSSs of sporophyte-biased (upper panel) and gametophyte-biased (lower panel) genes during
1047 the sporophyte (left) and gametophyte (middle) generations. The heatmaps on the right show
1048 transcript abundance during the two generations and log2 fold-change between generations. **b**
1049 Circos plots comparing chromatin states (combinations of histone PTMs) at the transcription
1050 start sites (TSS) of sporophyte-biased (left) and gametophyte-biased (left) genes in chromatin
1051 from the sporophyte (mauve) and gametophyte (green) generations of the life cycle. Colours
1052 correspond to chromatin states one to 16 as indicated in the key. **c** Plots of fold change against
1053 transcript abundance (TPM) for sporophyte- and gametophyte-biased genes during the two
1054 generations (left panel) along with ranked plots of fold change (right panel). Colour coding
1055 corresponds to chromatin states.
1056

1057



1058
1059
1060 **Fig. 8** Loss of polycomb complexes from brown algae. **a** Core components of the polycomb
1061 complexes PRC2 (mauve) and PRC1 (blue) in *Homo sapiens, Arabidopsis thaliana*,
1062 *Phaeodactylum tricornutum* and *Ectocarpus* sp.. Mauve or blue circles indicate the presence of
1063 a PRC2 or PRC1 gene, respectively. RbAp48 is indicated in pink when it probably represents
1064 a component of other cellular complexes (see text for details). Alternative proteins are shown

1065 for each component. Empty, dotted circles indicate components that are absent from
1066 *Ectocarpus*. All *P. tricornutum* and *Ectocarpus* sp. protein names should be prefixed with
1067 Phatr3 or Ec-, respectively, e.g. Phatr3_J17948, Ec-21_004560. **b** Schematic phylogenetic tree
1068 indicating PRC2 and PRC1 genes present in *Homo sapiens, Arabidopsis thaliana* and four
1069 selected stramenopiles. Red bars indicate predicted approximate timepoints for loss of the
1070 PRC1 and PRC2 complexes during stramenopile evolution. Approximate divergence times
1071 (±100 My) are based on Brown and Sorhannus [92].

1072
1073

Supplementary figures

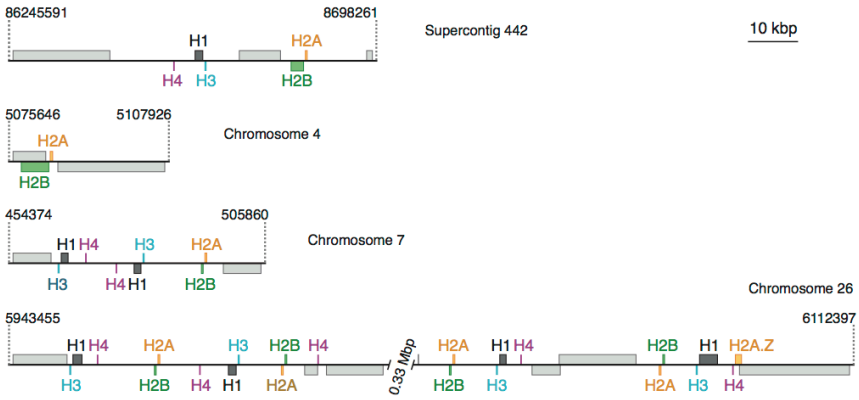
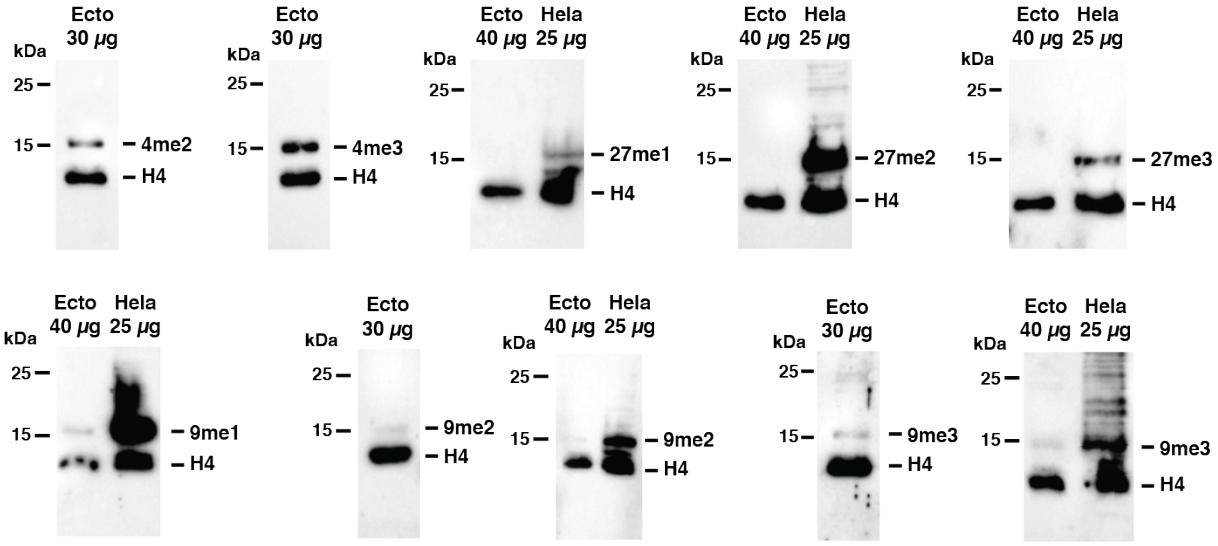


Fig. S1 Histone gene clusters in the *Ectocarpus* genome. Only regions with two or more histone genes are shown. Coding regions from the start to the stop codon are shown as boxes. Histone genes are colour coded, flanking, non-histone genes are in grey. Genes above the line are transcribed to the right,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

1086
1087
1088 (see additional file 3 for figure)
1089
1090
1091 **Fig. S2** Representative MS/MS spectra showing the identification of histone modifications.
1092 Trypsin-peptide-derived LC-ESI MS/MS fragmentation spectra for modified H2, H3 and H4
1093 peptides. The inset shows the peptide sequence and the observed ions. Tandem mass spectra
1094 are labelled to show singly and/or doubly charged b and y ions, as well as ions corresponding
1095 to neutral loss of water (°) or NH_3 (*). M is the parent ion mass and Im the immonium ion mass.
1096 Me, methylation; DiMe, dimethylation: TriMe, trimethylation, Prop, propionylation; Ac,
1097 Acetylation, Ub, ubiquitination; N-ter, N-terminal.
1098

1099



1100
1101
1102 **Fig. S3** Immunoblots of histone PTMs. Detection of histone PTMs H3K4me2 (4me2),
1103 H3K4me3 (4me3), H3K27me1 (27me1), H3K27me2 (27me2), H3K27me3 (27me3),
1104 H3K9me1 (9me1), H3K9me2 (9me2) and H3K9me3 (9me3) in either 30 µg or 40 µg of
1105 *Ectocarpus* chromatin extract (Ecto) or in 25 µg of human HeLa cell chromatin extract (HeLa)
1106 as a positive control. Histone H4 (H4) was detected as a loading control. Note that a weak, but
1107 probably non-specific, signal was obtained for H3K27me3 in *Ectocarpus* extracts following
1108 long exposure times (10 min rather than the standard 50 sec).
1109
1110
1111

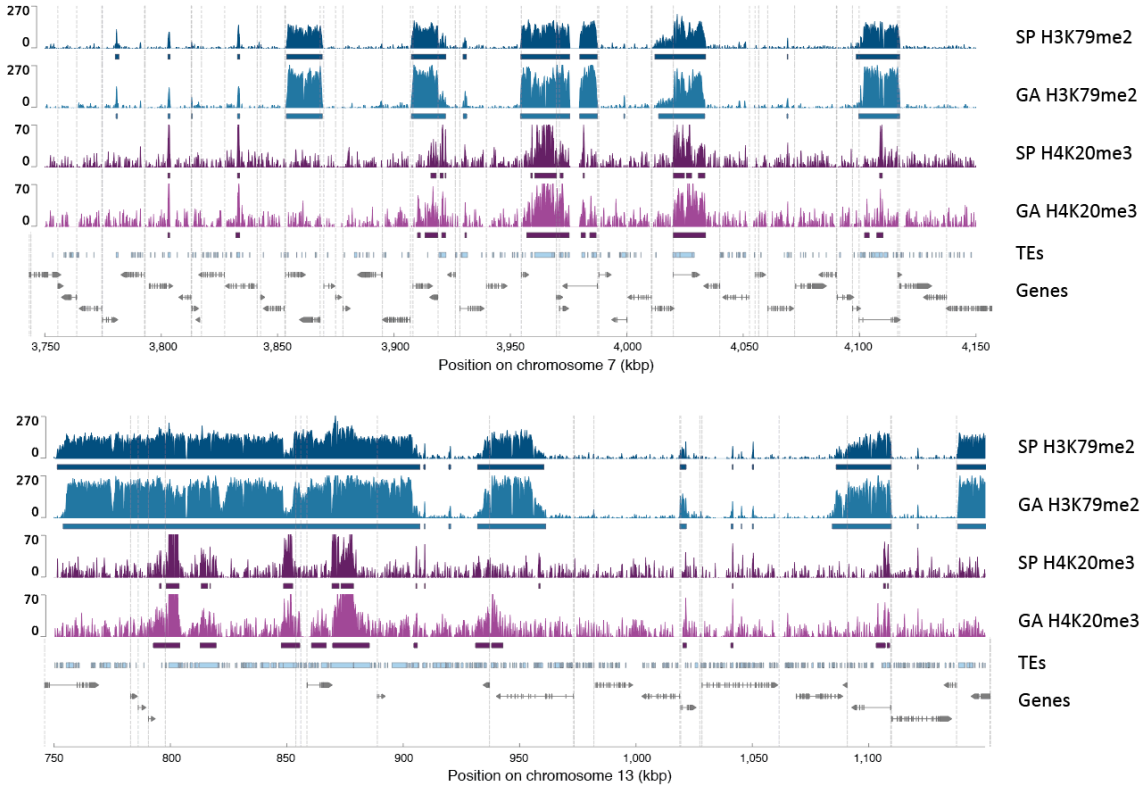
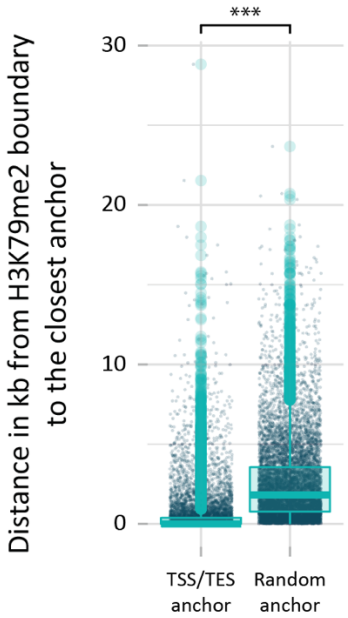


Fig. S4 Genomic distribution of the histone PTMs H3K79me2 and H4K20me3. Representative regions of chromosomes 7 and 13 (spanning the genes with LocusIDs Ec -07_003700 to Ec-07_004180 and Ec-13_000140 to Ec-13_000350, respectively) showing alternating regions marked or not with H3K79me2 (blue peaks) during both the sporophyte and gametophyte generations of the life cycle. Dark blue bars indicate H3K79me2 regions. The distribution of H4K20me3 (purple peaks) is also shown for comparison (peaks called by MACS2 are indicated by purple bars). Light blue boxes represent transposons (TEs). Grey boxes and arrows represent exons superimposed on genes represented by a black line. Vertical dotted lines indicate the positions of TSSs.

1124



1125
1126
1127 **Fig. S5** Boundaries of H3K79me2 regions are preferentially located near TSSs and TESs. The
1128 asterisks indicate a significant difference (Bonferonni-adjusted Wilcoxon test; ***, $p < 0.0001$).
1129

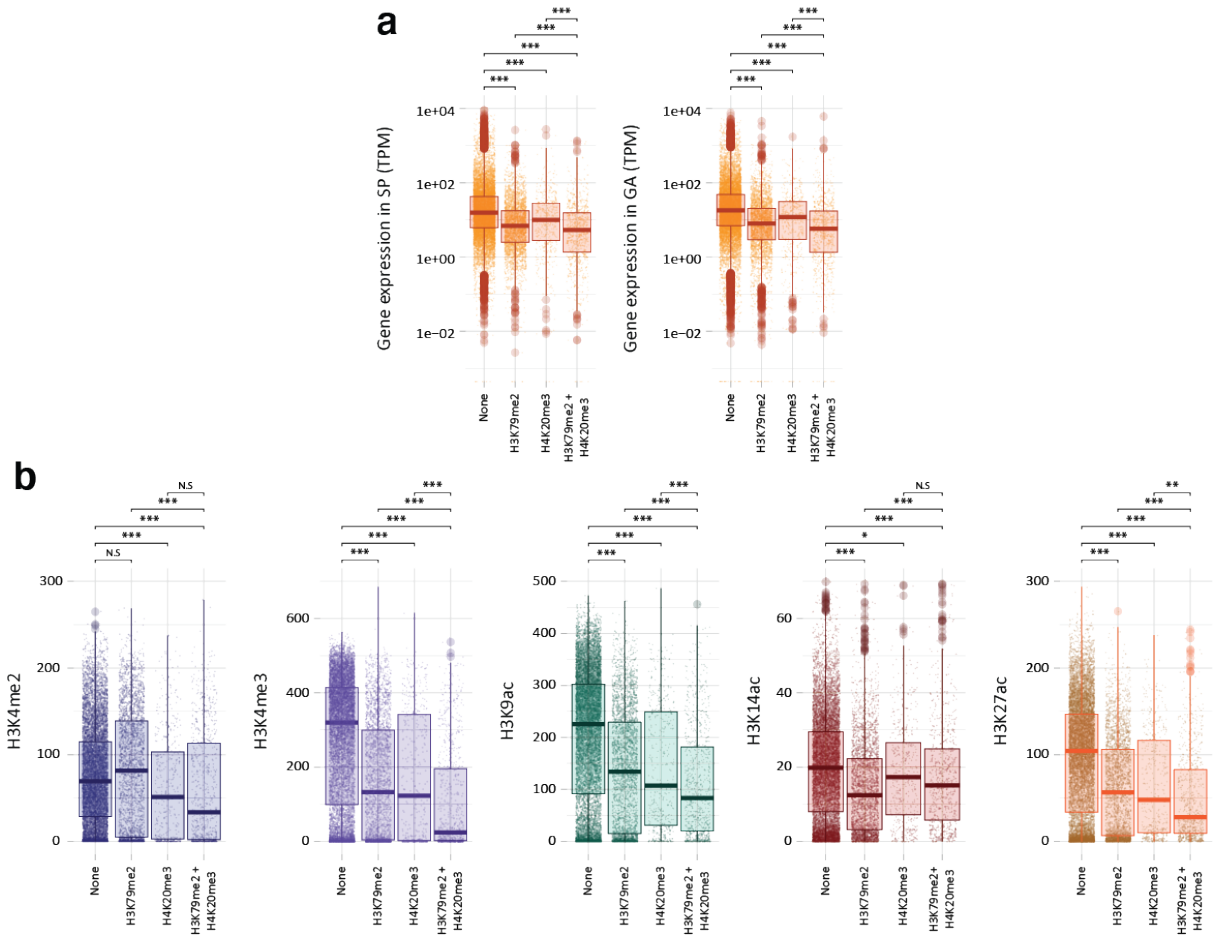


Fig. S6 Comparison of genes marked with H3K79me2 alone, H4K20me3 alone or both H3K79me2 and H4K20me3. **a** Abundances of transcripts for genes marked with H3K79me2 alone, H4K20me3 alone or both H3K79me2 and H4K20me3. **b** Histone PTMs on genes marked with H3K79me2 alone, H4K20me3 alone or both H3K79me2 and H4K20me3. SP, sporophyte; GA, gametophyte. Asterisks indicate significant differences (Bonferonni-adjusted Wilcoxon tests; ***, $p < 0.0001$; **, $p < 0.001$; *, $p < 0.05$).

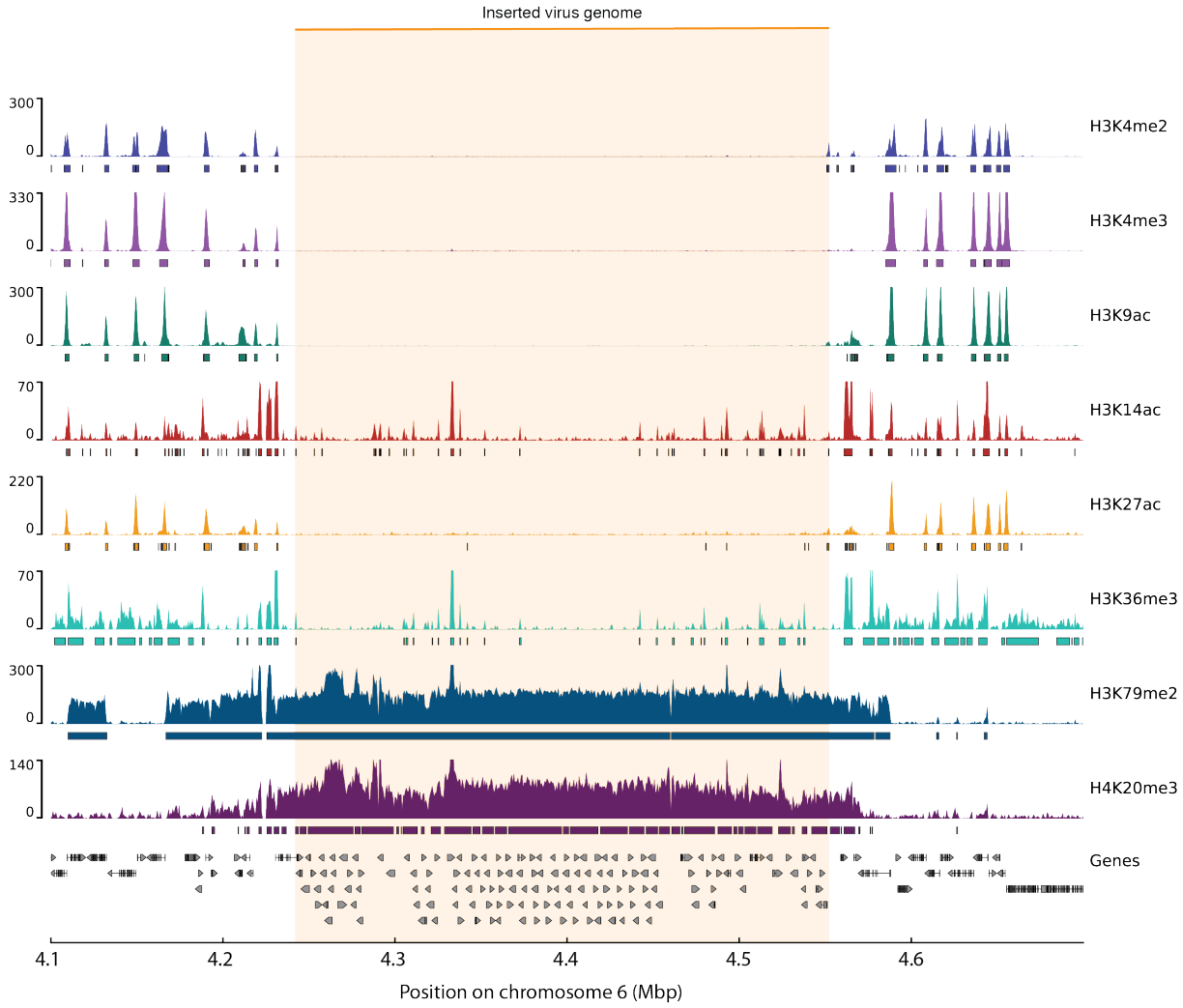


Fig. S7 H3K79me2, H4K20me3, H3K36me3, H3K4me2, H3K4me3, H3K9ac, H3K14ac and H3K27ac signals for a region of chromosome 6 spanning an inserted viral genome (orange box). Coloured boxes indicate histone PTM peaks, grey boxes and arrows represent exons superimposed on genes represented by a black line.

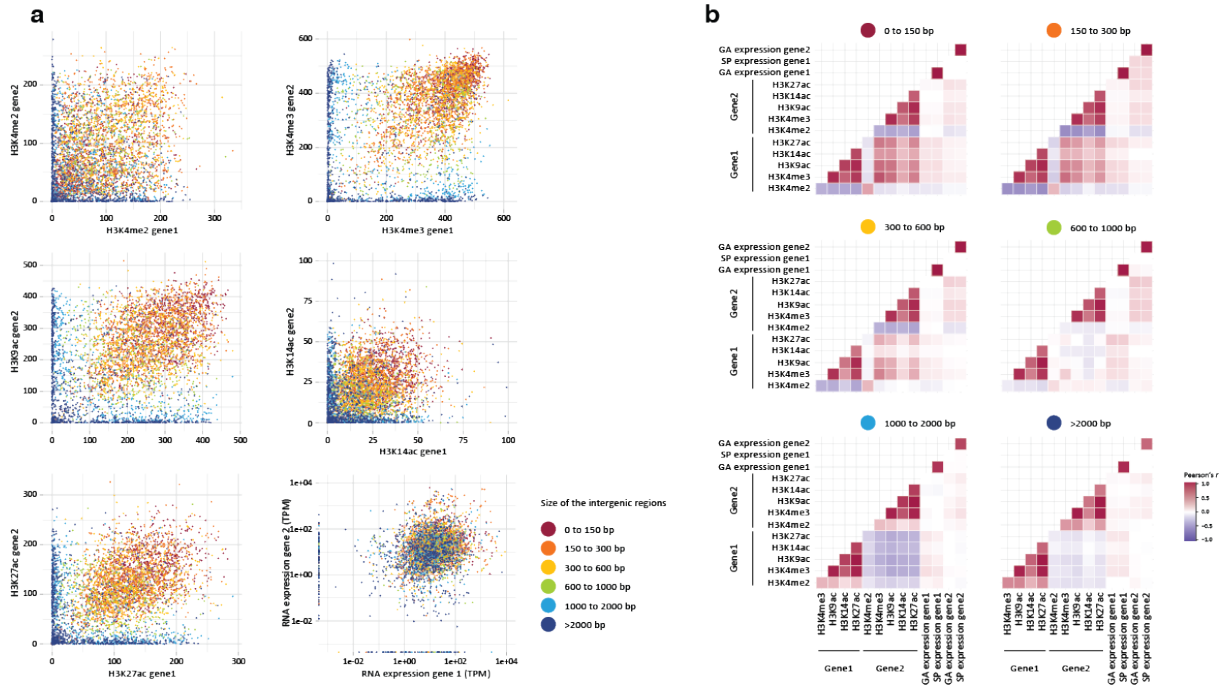
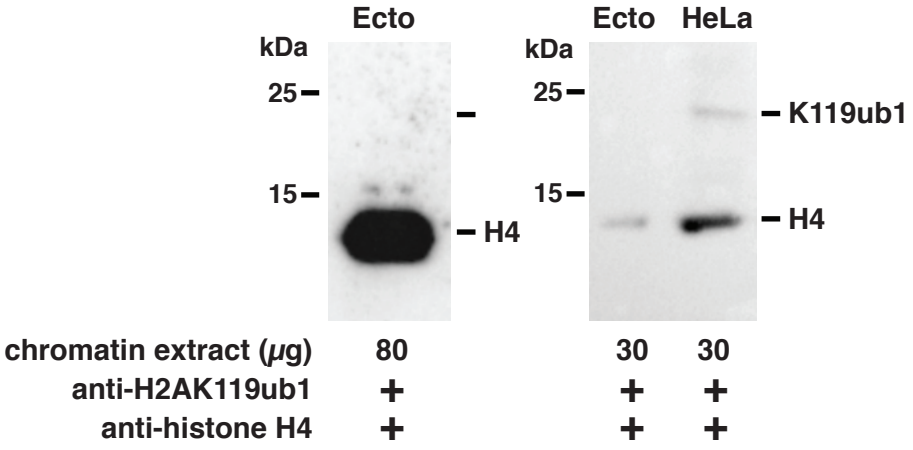


Fig. S8 Correlations of transcript abundances and histone PTM signals for divergently transcribed pairs of genes. **a** Pairwise comparisons of transcript abundances (TPM) and signal for the five TSS-localised histone PTMs for pairs of divergently transcribed genes (gene 1 and gene 2). **b** Pearson correlation scores for comparisons of transcript abundances (TPM) 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.

1165



1166
1167
1168 **Fig. S9** The histone PTM H2AK119ub1 was not detected in *Ectocarpus*. H2AK119ub1
1169 (K119ub1) was detected in a human HeLa cell chromatin extract (HeLa) as a positive control.
1170 Histone H4 (H4) was detected as a loading control. The *Ectocarpus* chromatin extracts (Ecto)
1171 were the same as those used for the immunoblots shown in Figure S3.
1172

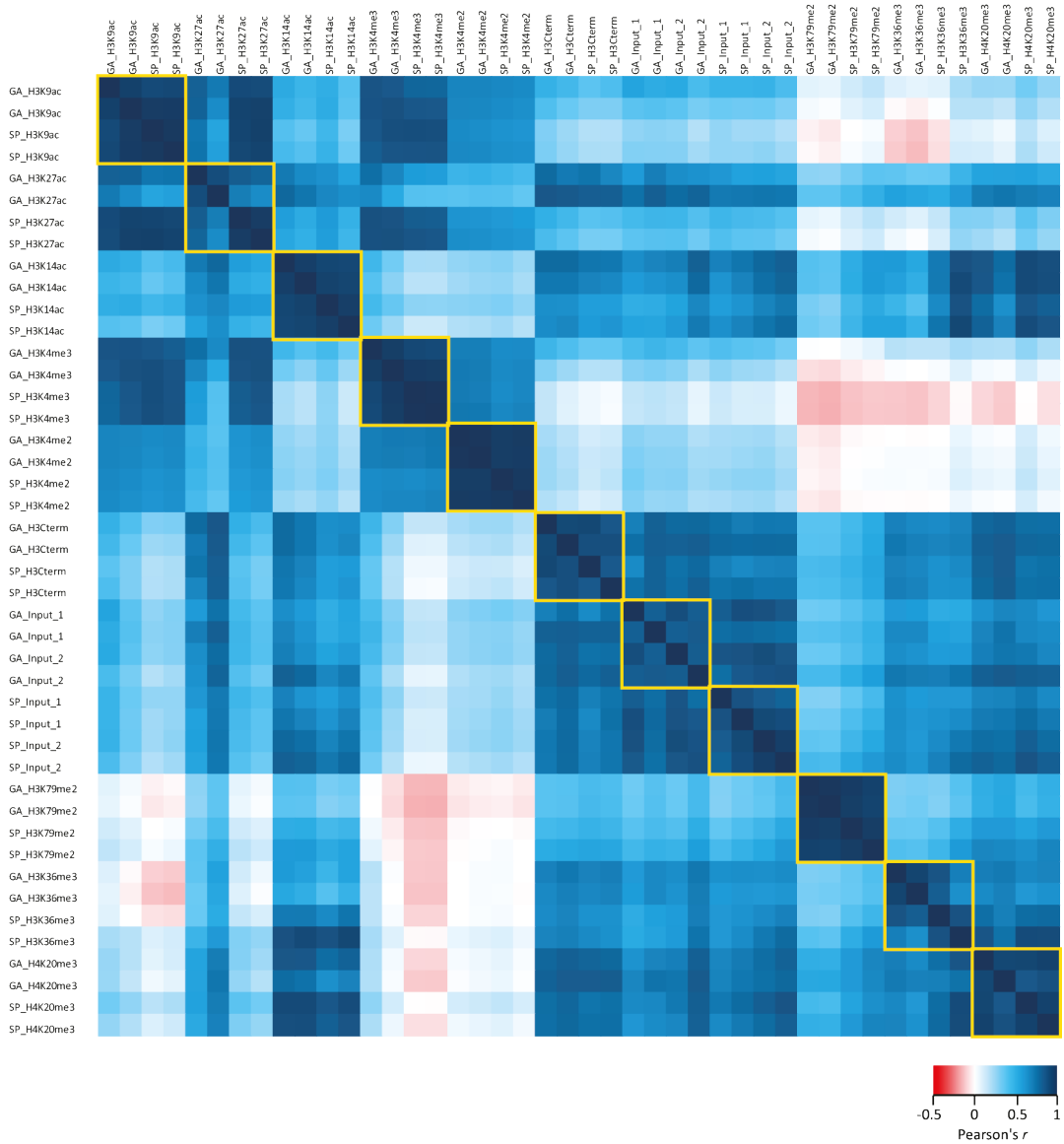


Fig. S10 Pearson correlation scores for comparisons of the genomic distributions of ChIP-seq signal peaks for duplicate assays of the eight histone PTMs during both the sporophyte and gametophyte generations.