



**HAL**  
open science

## **H3K27me3 natural variation selectively marks genes predicted to be important for differentiation in unicellular algae**

Xue Zhao, Achal Rastogi, Anne Flore Deton Cabanillas, Ouardia Ait Mohamed, Catherine Cantrel, Berangère Lombard, Omer Murik, Auguste Genovesio, Chris Bowler, Daniel Bouyer, et al.

► **To cite this version:**

Xue Zhao, Achal Rastogi, Anne Flore Deton Cabanillas, Ouardia Ait Mohamed, Catherine Cantrel, et al.. H3K27me3 natural variation selectively marks genes predicted to be important for differentiation in unicellular algae. 2020. hal-02998293

**HAL Id: hal-02998293**

**<https://hal.science/hal-02998293>**

Preprint submitted on 10 Nov 2020

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

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

1 H3K27me3 natural variation selectively marks genes predicted to be important for  
2 differentiation in unicellular algae

3

4 Xue Zhao<sup>1,3</sup>, Achal Rastogi<sup>1†</sup>, Anne Flore Deton Cabanillas<sup>1</sup>, Ouardia Ait Mohamed<sup>1</sup>, Catherine  
5 Cantrel<sup>1</sup>, Berangère Lombard<sup>2</sup>, Omer Murik<sup>1‡</sup>, Auguste Genovesio<sup>1</sup>, Chris Bowler<sup>1</sup>, Daniel  
6 Bouyer<sup>1</sup>, Damarys Loew<sup>2</sup>, Xin Lin<sup>1§</sup>, Alaguraj Veluchamy<sup>4</sup>, Fabio Rocha Jimenez Vieira<sup>1</sup> and  
7 Leila Tirichine<sup>1,3\*</sup>

8

9 <sup>1</sup>Institut de Biologie de l'ENS (IBENS), Département de biologie, École normale supérieure,  
10 CNRS, INSERM, Université PSL, 75005 Paris, France

11

12 <sup>2</sup>Institut Curie, PSL Research University, Centre de Recherche, Laboratoire de Spectrométrie  
13 de Masse Protéomique, 26 rue d'Ulm 75248 Cedex 05 Paris, France

14 <sup>3</sup>CNRS UMR6286, UFIP UFR Sciences et Techniques, Université de Nantes, 2 rue de la  
15 Houssinière 44322, Nantes Cedex 03

16 <sup>4</sup>Laboratory of Chromatin Biochemistry, BESE Division Building 2, Level 3, Office B2-3327,  
17 4700 King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900,  
18 Kingdom of Saudi Arabia

19 <sup>†</sup>Present address: Corteva Agriscience, Madhapur, Hyderabad 500 081, Telangana, India

20

21 <sup>‡</sup>Present address: Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel

22 <sup>§</sup>Present address: State Key Laboratory of Marine Environmental Science, Xiamen University

23

24

25

26

27

28 **Abstract**

29 In multicellular organisms H3K27me3 has been shown to be deposited by Polycomb Repressive  
30 Complex 2 (PRC2) to establish and maintain gene silencing, critical for cell fate and  
31 developmentally regulated processes. PRC2 complex is absent in both yeasts *Saccharomyces*  
32 *cerevisiae* and *Schizosaccharomyces pombe*, which initially suggested that PRC2 arose with  
33 the emergence of multicellularity. However, its discovery in several unicellular species  
34 including microalgae questions its role in unicellular eukaryotes. Here, we show in the model  
35 diatom *Phaeodactylum tricornutum* (Pt), using mutants in the homologue of the catalytic  
36 subunit of PRC2, *enhancer of zeste E(z)*, that Pt *E(z)* is responsible for di and tri-methylation  
37 of lysine 27 of histone H3. H3K27me3 depletion abolishes cell morphology in Pt providing  
38 evidence for a role of H3K27me3 in cell differentiation in unicellular species. Genome wide  
39 profiling of H3K27me3 in fusiform and triradiate cells further revealed genes that may specify  
40 cell identity. These results suggest a role for PRC2 and its associated histone mark in cell  
41 differentiation in unicellular species and highlights their ancestral function in a broader  
42 evolutionary context than is currently appreciated.

## 43 **Introduction**

44 Tri-methylation of lysine 27 of histone H3 (H3K27me3) is a mark deposited by Polycomb  
45 Repressive Complex 2 (PRC2), which mediates silencing of gene expression during  
46 differentiation and development in both animals and plants<sup>1-3</sup>. PRC2 is comprised of four core  
47 proteins, highly conserved among multicellular organisms: the histone methyltransferase  
48 (HMTase) enhancer of zeste *E(z)*, the WD40 domain containing polypeptide Extra Sex Comb  
49 *Esc*, the C2H2 type zinc finger protein Suppressor of zeste 12 *Su(z)12* and the Nucleosome  
50 remodeling factor 55 kDa subunit *Nurf-55*<sup>4,5</sup>. The absence of PRC2 in the unicellular yeast  
51 models *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* initially led to suggestions  
52 that it arose to regulate cell differentiation in multicellular organisms<sup>6</sup>. This hypothesis has  
53 recently been questioned because components of PRC2 and the associated mark H3K27me3  
54 are found in several unicellular species that belong to different lineages including, but not only,  
55 Stramenopiles, Alveolates and Rhizaria (SAR)<sup>7-9</sup> (Supplementary Fig. 1a-d), thus questioning  
56 the function of such a widespread complex in single celled organisms.

57 Attempts to understand the role of H3K27me3 in the unicellular green alga *Chlamydomonas*  
58 *reinhardtii* were not conclusive because tri-methylation of lysine 27 could not be assessed  
59 reliably due to its nominal mass which was found similar to acetylation of lysine 27 of the same  
60 histone<sup>7</sup>. However, we have previously identified H3K27me3 by mass spectrometry and  
61 mapped its localization genome-wide in the pennate diatom *Phaeodactylum tricornutum*<sup>9</sup> (Pt),  
62 which belongs to the stramenopile group of eukaryotes, only distantly related to the animal  
63 (Opisthokonta) and plant (Archaeplastida) eukaryotic crown groups. Pt has different  
64 morphotypes, fusiform (FM hereafter), which is the most prevailing morphology among the  
65 sampled accessions known so far, triradiate (TM hereafter), oval (OM hereafter) and cruciform  
66 (CM hereafter)<sup>10-13</sup> (Fig. 1a). Each morphotype can switch reversibly into a different  
67 morphology in response to several growth and environmental cues<sup>10</sup>. FM is the most stable  
68 morphotype while switching is more prominent in TM, CM and OM, which tend to convert to  
69 FM in the growth conditions used in this study<sup>10,11</sup>.

70 Interestingly, western blot analysis using a monoclonal antibody against the mark in FM, TM,  
71 CM and OM cells revealed a strong correlation between the complexity of the morphology  
72 (branching of the cell) and the absolute quantity of H3K27me3, which is higher in both CM  
73 and TM cells compared to FM and OM cells (Supplementary Fig. 1e), suggesting that PRC2  
74 activity controls cell differentiation in *P. tricornutum*. Cell differentiation is often orchestrated  
75 by H3K27me3-mediated silencing that underlies the establishment and maintenance of cellular

76 identity in multicellular model species<sup>14</sup>. These results prompted us to investigate the putative  
77 role of H3K27me3 in cell differentiation in a unicellular model. In silico annotation of  
78 polycomb complex members and identification of H3K27me3 by mass spectrometry<sup>15</sup> in a well-  
79 developed experimental model such as *P. tricornutum* (Supplementary Fig. 1a-c) might give  
80 new insights for the unsolved questions in the study of PRC2 complex in multicellular species,  
81 especially the availability of different morphotypes which is unique to *P. tricornutum* among  
82 stramenopiles, present an opportunity to decipher its role in single celled organisms with respect  
83 to its potential contribution to establish morphotype switches as well as its function in an  
84 evolutionary context.

85 To gain insights into the function of  $E(z)$  and its associated H3K27me3 mark in Pt, we generated  
86 two CRISPR/cas9 knockouts of the gene in each of the three morphotypes (FM, TM and CM)  
87 leading to putative loss of function mutations including premature stop codon and frameshifts  
88 (Fig. 1b). Light microscopy analysis of  $E(z)$  knockouts shows a change in cell morphology  
89 which becomes shorter in the FM background. Both triradiate and cruciform morphologies were  
90 abolished in TM and CM, respectively while transgenic lines carrying cas9 control vectors in  
91 each morphotype remain unchanged (Fig. 1c). Whereas *P. tricornutum* transgenic lines with  
92 the Cas9 control vector show similar H3K27me3 enrichment than the wild type (Supplementary  
93 Fig. 2a). knockout of  $E(z)$  led to an overall depletion of H3K27me3 shown by western blot  
94 using a monoclonal antibody against the mark (Fig. 1d), suggesting that the loss or diminution  
95 of  $E(z)$  activity, and hence H3K27me3, causes the observed changes in cell morphology (Fig.  
96 1c; Supplementary Fig. 2a; Table S1). Overall, this suggests that  $E(z)$  and its associated mark  
97 are required for morphotype switch/to establish specific cell identity.

98 Mass spectrometry analysis of histones extracted from both wild type and  $E(z)$  knockout  
99 confirmed the loss of H3K27me3 and revealed a depletion of di-methylation of H3K27,  
100 corroborating the role of  $E(z)$  in di and tri-methylation of lysine 27 of histone H3  
101 (Supplementary Fig. 2b,c,d). This is similar to fungi<sup>16</sup> and mammals<sup>17</sup> but different from *A.*  
102 *thaliana* where PRC2 loss of function leads to specific depletion of H3K27me3<sup>18</sup>, although in  
103 vitro assays with reconstituted *A. thaliana* PRC2 components showed mono, di and tri-  
104 methylation of lysine 27 of histone H3<sup>19,20</sup>. Western blot analysis of  $E(z)$  knockout mutants  
105 using a monoclonal antibody against H3K27me2 (Supplementary Fig. 2e) confirmed the  
106 depletion of the mark from the mutants, supporting further the mass spectrometry analysis and  
107 the role of  $E(z)$  in di-methylation of lysine 27 of histone H3 in *P. tricornutum*.

108 Similar to animals, plants and *Neurospora crassa*, H3K27me3 appears not to be essential for  
109 cell survival in *P. tricornutum*, as indicated by the overall growth of *E(z)* knockout lines, which  
110 are only slightly retarded compared to wild type lines (Supplementary Fig. 2f).

111 To further investigate the role of H3K27me3 and its targets in different morphotypes, we carried  
112 out a genomic approach and performed Chromatin Immuno-Precipitation (ChIP) on two  
113 biological replicates of TM morphotypes using an antibody against H3K27me3 followed by  
114 DNA sequencing (ChIP-Seq) to generate a map of H3K27me3 distribution, which we compared  
115 to the one previously generated in FM<sup>21</sup>. ChIP-Seq data analysis revealed a similar H3K27me3  
116 enrichment profile between TM and FM that localizes principally on transposable elements  
117 (TEs), with 58% and 60% of the reads overlapping with TE annotations for FM<sup>21</sup> and TM,  
118 respectively (Fig. 2a). The mark was found to occupy, on average, ~11.6% of the genome in  
119 FM cells, targeting approximately 15% of genes (consistent with<sup>15</sup>) and ~13.2% of the genome  
120 within TM, targeting 19% of genes in agreement with the absolute amount of H3K27me3 to be  
121 elevated in TM compared to FM as detected by western blot (Supplementary Fig. 1e). Indeed,  
122 more genes are marked by H3K27me3 in TM than in FM (Fig. 2a, Supplementary Fig. 3a),  
123 although most of the PRC2 targets are shared between both morphotypes (Fig. 2a) and exhibit  
124 globally broad coverage over the annotation (Supplementary Fig. 3b). Among the PRC2 target  
125 genes, 635 and 297 genes are found to be specifically marked by H3K27me3 in TM and FM,  
126 respectively (Fig. 2a). We used ChIP followed by quantitative PCR (ChIP-qPCR) to validate  
127 H3K27me3 enrichment over specifically marked loci in both backgrounds, which corroborated  
128 the genome wide data for most of the tested genes (Fig. 2b, c). Common marked loci and  
129 unmarked loci were tested as internal controls (Supplementary Fig. 3c). To test for the loss of  
130 H3K27me3 in *E(z)* knockouts, we performed ChIP-qPCR in FM and TM as well as the  
131 respective mutants, which confirmed the depletion of H3K27me3 in these mutants (Fig. 1e,f).

132 To gain insights into the functional categories enriched in H3K27me3 target genes that are  
133 shared between the two morphotype or specific, we applied to the updated Phatr3 annotation a  
134 gene ontology analysis using DAMA<sup>22</sup> and CLADE<sup>23</sup> which is a machine learning methodology  
135 that uses pHMMs, positive score matrix and support vector machines to infer the corresponding  
136 most probable GO category to genes. DAMA and Clade allow a more sensitive remote  
137 homology that permits to assign to genes with no or poor domain conservation the  
138 corresponding GO categories that would have been missed by other methods. Are considered  
139 statistically significant, only GO classes that are represented by at least 3 standard deviations  
140 above the average of observed entries.

141 Out of 1,640 H3K27me3 marked genes, 753 could not be assigned to a more specific GO  
142 category and are therefore marked as unknown. The genes that are marked in both morphotypes  
143 show top enrichment in RNA related biological processes such as RNA-dependent DNA  
144 biosynthetic process, RNA phosphodiester bond hydrolysis and RNA-DNA hybrid  
145 ribonuclease activity. Genes marked by H3K27me3 specifically in TM displayed top  
146 enrichment exclusively in (1) glycoprotein biosynthetic processes involved in the transfer of  
147 sugar moieties that might determine different sugar composition of the cell wall which is known  
148 to be sugar rich in *P. tricornutum*<sup>24</sup>, (2) Peptidyl tyrosine dephosphorylation processes with  
149 Ankyrin repeats proteins known to act as scaffold for connecting molecular interactions, likely  
150 important for development of the numerous signaling pathways associated generally to more  
151 complex multicellular organisms<sup>25</sup> (Fig. 2d; Supplementary Table S2). Genes that are  
152 specifically marked in FM cells exhibit enrichment in categories such as peptidyl-tyrosine  
153 phosphorylation containing genes with central roles as modulators of cell differentiation and  
154 cell fate decisions<sup>26</sup> (Fig. 2d; Supplementary Table S2). Interestingly, additional genes  
155 specifically marked in each FM or TM share categories with predicted functions in positive  
156 regulation of (1) GTPase activity with a role in cell morphology changes, and neurite outgrowth  
157 and guidance<sup>27</sup> as well as the differentiation of many cell types, including neurons, T  
158 lymphocytes and myocytes<sup>28</sup>; (2) protein ubiquitination shown to play a role in the complex  
159 regulation of the levels and function of many proteins and signaling pathways involved in  
160 determining cell fate<sup>29</sup>. Overall, the genes that are specifically marked in the TM or FM  
161 morphotypes reflect processes related to cell growth, proliferation and differentiation.

162 We have reported previously that H3K27me3 marked genes in FM are characterized by low  
163 expression<sup>15</sup>, consistent with the role of H3K27me3 as a repressive mark. Interestingly, when  
164 genes are marked by H3K27me3 in both FM and TM, their expression is lower compared to  
165 the genes that are uniquely marked in FM and to a lesser extent in TM (Fig. 3a,b). This suggests  
166 that specifically marked genes are kept under less stringent and tight repression which might be  
167 due to their putative role in morphotype switch, which is known to be a dynamic process<sup>10</sup>.

168 Considering the conserved role of H3K27me3 in repression, we tested the effect of  $E(z)$   
169 knockout on gene expression. Therefore, RNA sequencing (RNA-seq) of two biological  
170 replicates of the  $E(z)$  mutant (Del6) was carried out and compared to previously generated  
171 RNA-seq in the wild type (FM). Around 1/4 of all genes are (23%, 2795 out of 12152  
172 annotations) differentially expressed in the  $E(z)$  mutant (P-value. < 0.05), (Supplementary Fig.  
173 3d; Supplementary Table S2), indicating an essential role in gene regulation by PRC2 in Pt.

174 We further monitored by RT-qPCR the expression of 27 specifically marked genes in the  $E(z)$   
175 knockout of the TM and found that 18 genes out of 27 showed a gain of expression in the mutant  
176 compared to the TM background, demonstrating further that depletion of H3K27me3 likely  
177 releases the repression of target genes and correlates with the loss of the triradiate morphology  
178 (Fig. 3c). Although the remaining genes showed no change or even a gain in expression, these  
179 genes can be targets of other repressive or active marks as shown previously with co-occurrence  
180 of several repressive marks over genic regions<sup>9</sup>. The analysis of the R value<sup>30</sup> which reflects  
181 the entropy, and therefore the variability in expression of genes between FM, TM and loci  
182 marked in both morphotypes, showed a higher value in specifically marked genes compared to  
183 commonly marked ones (Fig. 3d). This supports further the finding that specific enrichment in  
184 each of the morphotypes are less silenced and potentially more dynamic compared to genes  
185 marked by H3K27me3 in both TM and FM cells whereas commonly H3K27me3 marked genes  
186 are globally silenced.

187 To substantiate the assumption that phenotypic plasticity and morphotype switch are regulated  
188 by PRC2 in *P. tricornutum*, we took advantage of the lack of stability of the TM phenotype and  
189 its tendency to switch to FM. Specifically, we used clonal cell samples with FM and TM  
190 morphologies from the same genetic background (TM), which switches habitually to fusiform  
191 and therefore contains a mixture of FM and TM cells. We reasoned that the activity of  $E(z)$   
192 should correlate with H3K27me3 levels in the following way: (1) a pure triradiate population  
193 isolated from TM-N (named here TM-T1): highest level of H3K27me3), (2) a population of  
194 cells from TM after N generations (N is  $60 \pm 5$ ) of cell division containing a mixture of triradiate  
195 and fusiform morphotypes (TM-N): medium level of H3K27me3) and (3) fusiform cells  
196 isolated from the triradiate background TM (TM-Fusi): lowest level of H3K27me3 (Fig. 4a).  
197  $E(z)$  transcript levels show a clear decrease in TM-N and TM-Fusi compared to TM-T1 (Fig.  
198 4b) which correlates with the switch from TM to FM, reflecting a lower activity of  $E(z)$  and  
199 H3K27me3 levels. We then asked whether specifically H3K27me3-marked loci in TM lose the  
200 mark upon cell switching to FM after multiple generations of sub-culturing leading to TM-N  
201 and in transformed fusiform cells (TM-Fusi). As expected, ChIP-qPCR showed clearly a loss  
202 of the mark in a population containing a mixture of fusiform and triradiate cells (TM-N) as well  
203 as in TM-Fusi compared to TM-T1 (Fig. 4c,d), which contains only triradiate cells, thus  
204 correlating the morphology with the level of enrichment in H3K27me3 over specific genes.

205 In summary, we have demonstrated in this study the role in *P. tricornutum* of  $E(z)$  as a histone  
206 methyltransferase responsible for di and tri-methylation of lysine 27 of histone H3. Knockout

207 of  $E(z)$  causes H3K27me3 depletion and loss of triradiate cell shape maintenance, providing  
208 evidence for the involvement of  $E(z)$  and its associated mark in establishing and/or maintaining  
209 cell morphology in unicellular species. We showed the dynamic nature of the mark, depending  
210 on the specific morphology between and within *P. tricornutum* accessions that correlate with  
211 the level of H3K27me3 enrichment. We showed differential marking in two different  
212 accessions of *P. tricornutum*, FM versus TM, which identified genes related to cell fate  
213 decisions compared to commonly marked genes. This is the first evidence of the involvement  
214 of H3K27me3 in cell differentiation in unicellular eukaryotes only distantly related to animals  
215 and plants. Our study points to the emerging function of PRC2 and its H3K27me3 associated  
216 mark as a determinant of the establishment and maintenance of cell morphology in single celled  
217 species such as *P. tricornutum* that shows signs of differentiation of the cell into diverse  
218 morphologies. This same function likely diversified with the emergence of multicellularity with  
219 PRC2 orchestrating development in plants and animals.

220

## 221 **Methods**

### 222 **Strains and growth conditions**

223 *Phaeodactylum tricornutum* Bohlin Clone Pt1 8.6 (CCMP2561) (referred as FM) and Clone  
224 Pt8Tc (referred as TM) cells were grown as described previously <sup>31</sup>.

### 225 **Isolation and immunoprecipitation of chromatin**

226 Chromatin isolation and immunoprecipitation were performed as described previously <sup>32</sup>. The  
227 following antibodies were used for immunoprecipitation: H3K27me3 (07-449) from Millipore  
228 and H3K27me3 from cell signaling technology. qPCR on recovered DNA was performed as  
229 described previously <sup>32</sup>

### 230 **CRISPR/Cas9 plasmid construction**

231 hCAS9n (Cas9 from *Streptococcus pyogenes*, adapted to human codon usage, fused to SV40  
232 nuclear localization sequence, and containing a D10A mutation) was amplified from  
233 pcDNA3.3-TOPO-hCAS9n (kindly received from Dr. Yonatan B. Tzur, Hebrew University of  
234 Jerusalem), using the primers 5'-CAC CAT GGA CAA GAA GTA CTC-3' and 5'- TCA CAC  
235 CTT CCT CTT CTT CTT-3'. The PCR product was first cloned into pENTR using pENTR/D-  
236 TOPO cloning kit (ThermoFisher Scientific), and then sub-cloned into a PT pDest, containing

237 an N-terminal HA-tag<sup>31</sup>, following the manufacturer's protocol, which was named pDest-HA-  
238 hCAS9n.

239 For the sgRNA vector we first cloned the snRNA U6 promoter<sup>33</sup> from *P. tricornutum* genomic  
240 DNA using the primers 5'- AAA CGA CGG CCA GTG AAT TCT CGT TTC TGC TGT CAT  
241 CAC C-3' and 5'- TCT TTA ATT TCA GAA AAT TCC GAC TTT GAA GGT GTT TTT TG-  
242 3'. PU6::unc-119\_sgRNA (kindly received from Dr. Yonatan B. Tzur) backbone was amplified  
243 using the primers 5'-CAA AAA ACA CCT TCA AAG TCG GAA TTT TCT GAA ATT AAA  
244 GA-3' and 5'-GGT GAT GAC AGC AGA AAC GAG AAT TCA CTG GCC GTC GTT T-3'.  
245 The two PCR products were used as template for a second round fusion PCR reaction as  
246 described in<sup>34</sup>. We further transformed the resulting product into *E. coli*, and extracted the  
247 ligated plasmid. The terminator sequence of the *P. tricornutum* U6 was amplified using the  
248 primers 5'-CATTCTAGAAGAACCGCTCACCCATGC-3' and 5'-  
249 GTTAAGCTTGAAAAGTTCGTCGAGACCATG-3', digested by XbaI/HindIII and ligated  
250 into XbaI/HindIII digested pU6::unc-119. The resulting vector, ptU6::unc-119-sgRNA, was  
251 used as template to replace the target sequence to *E(Z)* target by PCR using primers  
252 32817TS12fwd GTG TCG GAG CCC GCC ATA CCG TTT TAG AGC TAG AAA TAG C  
253 and 32817TS12rev GGT ATG GCG GGC TCC GAC ACC GAC TTT GAA GGT GTT TTT  
254 TG. Target sequences were picked using PhytoCRISP-Ex<sup>35</sup>.

### 255 **Transformation of *P. tricornutum* cells and screening for mutants**

256 Wild type cells of the reference strain FM and the TM were transformed with three plasmids  
257 (pPhat1, Cas9 and guide RNA with the target sequence) as described previously<sup>36</sup>. Positive  
258 transformants were validated by triple PCR screen for pPhaT1 shble primers (ACT GCG  
259 TGCACTTCGTGGC/TCGGTCAGTCCTGCTCCTC), sgRNA  
260 (GAGCTGGAAATTGGTTGTC/GACTCGGTGCCACTTTTTCAAGTT) and CAS9n  
261 (GGGAGCAGGCAGAAAACATT/TCACACCTTCCTCTTCTTCTT). For each colony, a  
262 rapid DNA preparation was performed as described previously and fragment of 400 bp was  
263 amplified with primers flanking the target sequence in the *E(z)* gene. The forward primer used  
264 is 5'-TAAGATGGAGTATGCCGAAATTC-3' and reverse primer is 5'-  
265 AGGCATTTATTCGTGTCTGTTCG-3' PCR product was run in 1% agarose gel and a single  
266 band was extracted using Machery Nagle kit and according to the manual manufacturer. PCR  
267 product was sequenced using the primer 5'-AGCCACCCTGCGTAACTGAAAAT-3'.

268 To make sure that the fusiform cells originating from the switch of TM are not contaminants  
269 from the FM, each of the TM-T1, TM-Fusi and TM-N were checked for their genetic  
270 background whether it is FM or TM using a molecular marker designed around a 400 bp  
271 insertion in the FM background (Supplementary Fig. 4g), identified from genome sequencing  
272 of FM and TM strains of *P. tricornutum*<sup>37</sup>. The PCR check confirmed that all the cell samples  
273 described above are in the TM genetic background.

274

## 275 **Validation of enrichment and expression of target genes**

276 **qPCR:** Total RNA was extracted from TM and FM cells as described previously<sup>31</sup> and cDNA  
277 was synthesized with cDNA high yield synthesis kit according to the manufacturer user manual.  
278 Quantitative PCR was performed as described previously<sup>31</sup> using the primer list in  
279 Supplementary Table S3

## 280 **Proteomics and PRM Measurements**

281 Three independent histone purifications recovered from FM wild type cells as well as *E(z)*  
282 knockout mutant Del6 in FM genetic background were simultaneously separated by SDS-  
283 PAGE and stained with colloidal blue (LabSafe Gel Blue GBiosciences). Three gel slices were  
284 excised for each purification and in-gel digested by using trypsin/LysC (Promega). Peptide  
285 extracted from each set were pooled and analyzed by nanoLC-MS/MS using an Ultimate 3000  
286 system (Thermo Scientific) coupled to a TripleTOFTM 6600 mass spectrometer (AB Sciex).  
287 Peptides were first trapped onto a C18 column (75  $\mu$ m inner diameter  $\times$  2 cm; nanoViper  
288 Acclaim PepMapTM 100, Thermo Scientific) with buffer A (2/98 MeCN/H<sub>2</sub>O in 0.1% formic  
289 acid) at a flow rate of 2.5  $\mu$ L/min over 4 min. Separation was performed on a 50 cm x 75  $\mu$ m  
290 C18 column (nanoViper C18, 3  $\mu$ m, 100 $\text{\AA}$ , Acclaim PepMapTM RSLC, Thermo Scientific)  
291 regulated to 50°C and with a linear gradient from 1% to 30% buffer B (100 MeCN in 0.085%  
292 formic acid) at a flow rate of 400 nL/min over 90 min. The mass spectrometer was operated in  
293 PRM top30 high sensitivity mode with 100 ms acquisition time for MS1 and MS2 scans  
294 respectively with included precursor mass list for 600 sec (see Supplementary Table S4)

## 295 **PRM Data Analysis**

296 The PRM data were analyzed using Skyline version 3.7.0.11317 MacCoss Lab Software,  
297 Seattle, WA; <https://skyline.ms/project/home/software/Skyline/begin.view>, fragment ions for  
298 each targeted mass were extracted and peak areas were integrated. The peptide areas were log<sub>2</sub>

299 transformed and the mean log<sub>2</sub>- area was normalized by the mean area of peptide STDLLIR  
300 using software R version 3.1.0. On each peptide a linear model was used to estimate the mean  
301 fold change between the conditions, its 97.5% confidence interval and the p-value of the two  
302 sided associated t-test. The p-values were adjusted with the Benjamini-Hochberg procedure<sup>38</sup>.  
303 The mass spectrometry proteomics data have been deposited to the ProteomeXchange  
304 Consortium via the PRIDE [1] partner repository with the dataset identifier PXD012347.

305

306 **Western blot analysis:** Chromatin was extracted from wild type as well as mutants of both TM  
307 and FM cells and western blot performed as described previously<sup>32</sup>.

308

### 309 **Sequencing and computational data analysis**

310 **ChIP-Seq:** Chromatin Immunoprecipitation (ChIP) was done with monoclonal cell cultures  
311 grown using single triradiate cell from Pt8 population (referred as TM). CHIP-Seq was  
312 performed as described previously<sup>15,39</sup>. Two replicates were performed and showed a good  
313 Pearson correlation (Supplementary Fig. 5). Raw reads were filtered and low quality read-pairs  
314 were discarded using FASTQC with a read quality (Phred score) cutoff of 30. Using the genome  
315 assembly published in 2008 as reference (Pt1 8.6), we performed reference-assisted mapping  
316 of filtered reads using BOWTIE. We then performed the processing and filtering of the  
317 alignments using SAMTOOLS and BEDTOOLS. SICER<sup>40</sup> was then used to identify  
318 significant enriched H3K27me3 peaks by comparing it with the INPUT. Differential  
319 H3K27me3 peak enrichment analysis between FM and TM backgrounds was also done using  
320 SICER-df plugin. Peaks with Padj < 0.05 differential enrichment or depletion were considered  
321 significant. Functional inferences were obtained by overlapping the differentially enriched  
322 peaks over structural annotations from Phatr3 genome annotation<sup>41</sup>.

323 **RNA sequencing (RNA-Seq):** Total RNA was extracted from FM, TM, and FM Ez-KO (Del6)  
324 cell lines. RNA expression and differential gene expression analysis was performed using  
325 Eoulsan version 1.2.2 with default parameters<sup>42</sup>. Genes having at least 2 folds expression  
326 change with P-value < 0.05 were considered as significant different expressed genes (DEGs).

327 **Whole genome sequencing (WGS):** Whole genome sequencing was performed using DNA  
328 extracted from monoclonal cell cultures grown using single triradiate cell taken from Pt8 and  
329 Pt1 accession, referred to as Pt8tc and Pt1 8.6, respectively. At least 6 µg of genomic DNA  
330 from each accession was used to construct a sequencing library following the manufacturer's

331 instructions (Illumina Inc.). Paired-end sequencing libraries with a read size of 100 bp and an  
332 insert size of approximately 400 bp were sequenced on an Illumina HiSeq 2000 sequencer at  
333 Berry Genomics Company (China) and Fasteris for Pt1 8.6 and Pt8tc, respectively. Low quality  
334 read-pairs were discarded using FASTQC with a read quality (Phred score) cutoff of 30. Using  
335 the genome assembly published previously<sup>43</sup>, we performed reference-assisted assembly of all  
336 the accessions. We used BOWTIE (-n 2 -X 400) for mapping the high quality NGS reads to  
337 the reference genome followed by the processing and filtering of the alignments using  
338 SAMTOOLS and BEDTOOLS. For estimating the genetic diversity between Pt1 8.6 and Pt8Tc  
339 genome, GATK<sup>44</sup> configured for diploid genomes, was used for variant calling, which included  
340 single nucleotide polymorphisms (SNPs), small insertions and deletions ranging between 1 and  
341 300 base pairs (bp). The genotyping mode was kept default (genotyping mode =  
342 DISCOVERY), Emission confidence threshold (-stand\_emit\_conf) was kept 10 and calling  
343 confidence threshold (-stand\_call\_conf) was kept at 30. The minimum number of reads per base  
344 to be called as a high quality SNP was kept at 4 (i.e., read-depth  $\geq 4x$ ). SNPEFF was used to  
345 annotate the functional nature of the polymorphisms.

346 **GO enrichment analysis:** GO categories were grouped by 3 different levels of expression  
347 according to a simple density clustering algorithm (also confirmed by iterative k-means  
348 clustering).

#### 349 **Data Availability:**

350 All data are available through NCBI Sequence Read Archive with accession number  
351 PRJNA565539.

352

#### 353 **References**

354

- 355 1 Aldiri, I. & Vetter, M. L. PRC2 during vertebrate organogenesis: a complex in transition. *Dev*  
356 *Biol* **367**, 91-99, doi:10.1016/j.ydbio.2012.04.030 (2012).
- 357 2 Fragola, G. *et al.* Cell reprogramming requires silencing of a core subset of polycomb targets.  
358 *PLoS Genet* **9**, e1003292, doi:10.1371/journal.pgen.1003292 (2013).
- 359 3 Surface, L. E., Thornton, S. R. & Boyer, L. A. Polycomb group proteins set the stage for early  
360 lineage commitment. *Cell Stem Cell* **7**, 288-298, doi:10.1016/j.stem.2010.08.004 (2010).
- 361 4 Schwartz, Y. B. & Pirrotta, V. A new world of Polycombs: unexpected partnerships and  
362 emerging functions. *Nature reviews. Genetics* **14**, 853-864, doi:10.1038/nrg3603 (2013).
- 363 5 Martinez-Balbas, M. A., Tsukiyama, T., Gdula, D. & Wu, C. Drosophila NURF-55, a WD repeat  
364 protein involved in histone metabolism. *Proceedings of the National Academy of Sciences of*  
365 *the United States of America* **95**, 132-137 (1998).

- 366 6 Kohler, C. & Villar, C. B. Programming of gene expression by Polycomb group proteins. *Trends*  
367 *Cell Biol* **18**, 236-243, doi:10.1016/j.tcb.2008.02.005 (2008).
- 368 7 Shaver, S., Casas-Mollano, J. A., Cerny, R. L. & Cerutti, H. Origin of the polycomb repressive  
369 complex 2 and gene silencing by an E(z) homolog in the unicellular alga *Chlamydomonas*.  
370 *Epigenetics* **5**, 301-312 (2010).
- 371 8 Mikulski, P., Komarynets, O., Fachinelli, F., Weber, A. P. M. & Schubert, D. Characterization of  
372 the Polycomb-Group Mark H3K27me3 in Unicellular Algae. *Front Plant Sci* **8**, 607,  
373 doi:10.3389/fpls.2017.00607 (2017).
- 374 9 Veluchamy, A. *et al.* An integrative analysis of post-translational histone modifications in the  
375 marine diatom *Phaeodactylum tricornutum*. *Genome biology* **16**, 102, doi:10.1186/s13059-  
376 015-0671-8 (2015).
- 377 10 De Martino, A. *et al.* Physiological and Molecular Evidence that Environmental Changes Elicit  
378 Morphological Interconversion in the Model Diatom *Phaeodactylum tricornutum*. *Protist*  
379 **162**, 462-481, doi:S1434-4610(11)00006-X [pii]  
10.1016/j.protis.2011.02.002 (2011).
- 380 11 De Martino, A. M., A. Juan Shi, K.P. Bowler, C. Genetic and phenotypic characterization of  
381 *Phaeodactylum tricornutum* (Bacillariophyceae) accessions. *J. Phycol.* **43**, 992–1009 (2007).
- 382 12 He, L., Han, X. & Yu, Z. A rare *Phaeodactylum tricornutum* cruciform morphotype: culture  
383 conditions, transformation and unique fatty acid characteristics. *PLoS one* **9**, e93922,  
384 doi:10.1371/journal.pone.0093922 (2014).
- 385 13 Borowitzka, M. A., Volcani, B.E. The polymorphic diatom *Phaeodactylum tricornutum*:  
386 Ultrastructure of its morphotypes. *J. Phycol.* **14**, 10-21 (1978).
- 387 14 Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**,  
388 343-349, doi:10.1038/nature09784 (2011).
- 389 15 Veluchamy, A. *et al.* An integrative analysis of post-translational histone modifications in the  
390 marine diatom *Phaeodactylum tricornutum*. *Genome biology* **16**, 102, doi:10.1186/s13059-  
391 015-0671-8 (2015).
- 392 16 Jamieson, K., Rountree, M. R., Lewis, Z. A., Stajich, J. E. & Selker, E. U. Regional control of  
393 histone H3 lysine 27 methylation in *Neurospora*. *Proceedings of the National Academy of*  
394 *Sciences of the United States of America* **110**, 6027-6032, doi:10.1073/pnas.1303750110  
395 (2013).
- 396 17 Ferrari, K. J. *et al.* Polycomb-dependent H3K27me1 and H3K27me2 regulate active  
397 transcription and enhancer fidelity. *Molecular cell* **53**, 49-62,  
398 doi:10.1016/j.molcel.2013.10.030 (2014).
- 399 18 Lafos, M. *et al.* Dynamic regulation of H3K27 trimethylation during *Arabidopsis*  
400 differentiation. *PLoS Genet* **7**, e1002040, doi:10.1371/journal.pgen.1002040 (2011).
- 401 19 Jacob, Y. *et al.* ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin  
402 structure and gene silencing. *Nat Struct Mol Biol* **16**, 763-768, doi:10.1038/nsmb.1611  
403 (2009).
- 404 20 Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks.  
405 *Molecular cell* **42**, 330-341, doi:10.1016/j.molcel.2011.03.025 (2011).
- 406 21 Veluchamy, A. *et al.* An integrative analysis of post-translational histone modifications in the  
407 marine diatom *Phaeodactylum tricornutum*. *Genome Biol* **16**, 102 (2015).
- 408 22 Bernardes, J. S., Vieira, F. R., Zaverucha, G. & Carbone, A. A multi-objective optimization  
409 approach accurately resolves protein domain architectures. *Bioinformatics* **32**, 345-353,  
410 doi:10.1093/bioinformatics/btv582 (2016).
- 411 23 Bernardes, J., Zaverucha, G., Vaquero, C. & Carbone, A. Improvement in Protein Domain  
412 Identification Is Reached by Breaking Consensus, with the Agreement of Many Profiles and  
413 Domain Co-occurrence. *PLoS Comput Biol* **12**, e1005038, doi:10.1371/journal.pcbi.1005038  
414 (2016).
- 415

- 416 24 Le Costaouëc, T., Unamunzaga, C., Mantecon, L. and Helbert, W. New structural insights into  
417 the cell-wall polysaccharide of the diatom *Phaeodactylum tricornutum*. *Algal research* **26**  
418 172-179 (2017).
- 419 25 Marcotte, E. M., Pellegrini, M., Yeates, T. O. & Eisenberg, D. A census of protein repeats. *J*  
420 *Mol Biol* **293**, 151-160, doi:10.1006/jmbi.1999.3136 (1999).
- 421 26 Yu, J. S. & Cui, W. Proliferation, survival and metabolism: the role of PI3K/AKT/mTOR  
422 signalling in pluripotency and cell fate determination. *Development* **143**, 3050-3060,  
423 doi:10.1242/dev.137075 (2016).
- 424 27 Etienne-Manneville, S. & Hall, A. Rho GTPases in cell biology. *Nature* **420**, 629-635,  
425 doi:10.1038/nature01148 (2002).
- 426 28 Bryan, B. A., Li, D., Wu, X. & Liu, M. The Rho family of small GTPases: crucial regulators of  
427 skeletal myogenesis. *Cell Mol Life Sci* **62**, 1547-1555, doi:10.1007/s00018-005-5029-z (2005).
- 428 29 Thompson, S. J., Loftus, L. T., Ashley, M. D. & Meller, R. Ubiquitin-proteasome system as a  
429 modulator of cell fate. *Curr Opin Pharmacol* **8**, 90-95, doi:10.1016/j.coph.2007.09.010 (2008).
- 430 30 Maheswari, U. *et al.* Digital expression profiling of novel diatom transcripts provides insight  
431 into their biological functions. *Genome biology* **11**, R85, doi:gb-2010-11-8-r85 [pii]  
432 10.1186/gb-2010-11-8-r85 (2010).
- 433 31 Siaut, M. *et al.* Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*.  
434 *Gene* **406**, 23-35, doi:S0378-1119(07)00275-2 [pii]  
435 10.1016/j.gene.2007.05.022 (2007).
- 436 32 Lin, X., Tirichine, L. & Bowler, C. Protocol: Chromatin immunoprecipitation (ChIP)  
437 methodology to investigate histone modifications in two model diatom species. *Plant*  
438 *methods* **8**, 48, doi:10.1186/1746-4811-8-48 (2012).
- 439 33 Rogato, A. *et al.* The diversity of small non-coding RNAs in the diatom *Phaeodactylum*  
440 *tricornutum*. *BMC genomics* **15**, 698, doi:10.1186/1471-2164-15-698 (2014).
- 441 34 Hobert, O. PCR fusion-based approach to create reporter gene constructs for expression  
442 analysis in transgenic *C. elegans*. *Biotechniques* **32**, 728-730 (2002).
- 443 35 Rastogi, A., Murik, O., Bowler, C. & Tirichine, L. PhytoCRISP-Ex: a web-based and stand-alone  
444 application to find specific target sequences for CRISPR/CAS editing. *BMC bioinformatics* **17**,  
445 261, doi:10.1186/s12859-016-1143-1 (2016).
- 446 36 Falcatore, A., Casotti, R., Leblanc, C., Abrescia, C. & Bowler, C. Transformation of  
447 nonselectable reporter genes in marine diatoms. *Mar Biotechnol (NY)* **1**, 239-251, doi:MBT30  
448 [pii] (1999).
- 449 37 Rastogi, A. V. F., Deton-Cabanillas AF, Veluchamy A., Cantrel, C., Wang, G., Vanormelingen,  
450 P., Bowler, C., Piganeau, G., Hu, H. and Leila Tirichine. A genomics approach reveals the  
451 global genetic polymorphism, structure and functional diversity of ten accessions of the  
452 marine model diatom *Phaeodactylum tricornutum*  
453  
454 *ISME J* (2019).
- 455 38 Benjamini, Y., and Yekutieli, D. . The control of the false discovery rate in multiple testing  
456 under dependency. *Annals of Statistics* **29**, 1165–1188 (2001).
- 457 39 Veluchamy, A. *et al.* Insights into the role of DNA methylation in diatoms by genome-wide  
458 profiling in *Phaeodactylum tricornutum*. *Nat Commun* **4**, doi:10.1038/ncomms3091 (2013).
- 459 40 Zang, C. *et al.* A clustering approach for identification of enriched domains from histone  
460 modification ChIP-Seq data. *Bioinformatics* **25**, 1952-1958,  
461 doi:10.1093/bioinformatics/btp340 (2009).
- 462 41 Rastogi, A. *et al.* Integrative analysis of large scale transcriptome data draws a  
463 comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of  
464 diatoms. *Sci Rep* **8**, 4834, doi:10.1038/s41598-018-23106-x (2018).

- 465 42 Jourdren, L., Bernard, M., Dillies, M. A. & Le Crom, S. Eoulsan: a cloud computing-based  
466 framework facilitating high throughput sequencing analyses. *Bioinformatics* **28**, 1542-1543,  
467 doi:10.1093/bioinformatics/bts165 (2012).  
468 43 Bowler, C. *et al.* The Phaeodactylum genome reveals the evolutionary history of diatom  
469 genomes. *Nature* **456**, 239-244, doi:nature07410 [pii]  
470 10.1038/nature07410 (2008).  
471 44 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-  
472 generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110  
473 (2010).

474

#### 475 **Acknowledgements**

476 Hanhua Hu from the Chinese Academy of Science is acknowledged for the gift of CM  
477 morphotype. ). LT acknowledges funds from the CNRS and the region of Pays de la Loire  
478 (ConnecTalent EPIALG project). CB acknowledges funding from the ERC Advanced Award  
479 ‘Diatomite. XZ was supported by a PhD fellowship from the Chinese Scholarship Council  
480 (CSC-201604910722). AR was supported by an International PhD fellowship from the MEMO  
481 LIFE Program.

482

483

#### 484 **Competing interests**

485 The authors declare no competing interests.

486

487 **Figure legends**

488 **Figure 1. *Phaeodactylum tricornutum* morphotypes and enhancer of zeste knockout**  
489 **mutants. (a)** *P. tricornutum* morphotypes (top left fusiform, scale bar = 3 $\mu$ m, top right oval,  
490 bottom left triradiate, scale bar = 2 $\mu$ m, bottom right cruciform). Scale bars correspond to 1  $\mu$ m  
491 in OM and CM. **(b)** Sequence chromatograms of PCR product from WT cells and CRISPR cas9  
492 mutants of enhancer of zeste showing the different indels in FM, TM and CM. **(c)** Light  
493 microscopy images of WT, E(z) KO and Cas9 control cells. Empty vector controls containing  
494 Cas9 and Shble antibiotic resistance gene show no loss of H3K27me3 and retain the wild type  
495 morphology suggesting that morphology distortion is not due to the transformation but to the  
496 absence of H3K27me3. **(d)** Western blots of WT and two E(z)KOs from each morphotype  
497 using a monoclonal antibody against H3K27me3. Histone H4 was used as a loading control.  
498 ChIP-QPCR enrichment levels of H3K27me3 on genes in WT and E(z) KO in FM **(e)** and TM  
499 **(f)** backgrounds.

500

501 **Figure 2. Genomic features of H3K27me3 targets in FM and TM cells. (a)** Venn diagrams  
502 showing the number of common and specific genomic features [Genes, Transposable elements  
503 (TEs), and Intergenic Regions (IGRs)] targeted by H3K27me3 in TM (orange circles) and FM  
504 (blue circles). ChIP-QPCR validation of H3K27me3 specifically marked genes in **(b)** TM and  
505 **(c)** FM morphotypes. **(d)** Distribution of the most frequent GO terms on genes marked with  
506 H3K27me3. The distribution was sub-divided into different categories, where TM and FM  
507 represent the GOs observed exclusively on triradiate and fusiform genes, respectively. A third  
508 category (Both) also presents a GO distribution for genes observed on both morphotypes.

509

510

511 **Figure 3. Gene expression profiles of H3K27me3 specific and common targets in FM and**  
512 **TM. (a)** The box plot represents mean enrichment of H3K27me3 (Y-axis), with log<sub>10</sub> scaling,  
513 over genes marked specifically in FM (blue), TM (orange), and also on genes marked in both  
514 (commonly marked) morphotypes. The enrichment profile is generated using number of genes  
515 marked by H3K27me3 specifically in each morphotype and also in both. The significant  
516 H3K27me3 enrichment difference between specifically marked and commonly marked is  
517 estimated using two-tailed t-test with P value < 0.0001, denoted by “\*\*\*\*\*”. **(b)** Expression of

518 genes marked specifically in FM (blue), TM (orange), and in both phenotypic backgrounds with  
519 same principle aesthetics and categorical genes, used in 2f. The significance/non-significance  
520 of the variability of expression between specifically and commonly marked genes is estimated  
521 using two-tailed t-test with P value = 0.0433, as denoted by “\*”. “ns” denote non-significant.  
522 (c) Relative expression level of H3K27me3 targets genes in the TM and enhancer of zeste knock  
523 out M2-11. In the plot, fold change  $\log_2(\text{PtM2-11}/\text{Pt8Tc})$  values are shown (d) Boxplot  
524 showing the entropy value distribution of H3K27me3 marked genes. Entropy values measures  
525 the differential expression of genes under different experimental conditions. Entropy values are  
526 derived from expression data (fragments per kilobase of exon per million fragments mapped)  
527 under five different experimental conditions. Genes marked specifically by H3K27me3 in TM  
528 (Triradiate) shows higher variation in expression followed by FM (Fusiform) specific  
529 H3K27me3 marked genes. Horizontal lines represent the median entropy values. The  
530 significant of FM/TM-specific H3K27me3 marked genes and genes marked on both conditions  
531 are estimated using two-tailed t-test with P value < 0.0001.

532

533

534 **Figure 4. H3K27me3 enrichment levels and morphotype switch.** (a) Schematic diagram  
535 showing how TM-N, TM-T1 and TM-Fusi were generated. After generations of culture in lab  
536 growth condition (ASW media, 19°C, 12h/12h light dark period), some triradiate cells switch  
537 to fusiform forming a mixture of FM and TM cells, named TM-N. A single fusiform cell from  
538 TM-N was picked and propagated clonally giving rise to TM-Fusi. Similarly, single triradiate  
539 cell was isolated from TM-N and its clonal propagation gave a population of pure triradiate  
540 cells named TM-T1. (b) Relative expression level of Enhancer of Zest in FM, TM-N, TM-T1  
541 and TM-Fusi respectively, two pairs of primer were designed at N-terminal of Enhancer of  
542 Zeste gene, and at the CXC domain respectively. (c) ChIP-QPCR enrichment levels of  
543 H3K27me3 in TM-T1 and TM-Fusi. (d) ChIP-QPCR showing enrichment levels of H3K27me3  
544 in TM-T1 and TM-N.

545

546 **Supplementary Table 1.** Cell counts of different morphologies and cell size measurements in  
547 each wild type morphotype and knock out of E(z)

548 **Supplementary Table 2.** List of genes marked by H3K27me3 with their annotation and GOs.

549 **Supplementary Table 3.** List of the primers used in this study

550 **Supplementary Table 4.** Mass spectrometry quantification of Di- and Tri-methylation of  
551 H3K27me3

552 **Supplementary Figure 1.** Phylogeny (a-c). (d) Western blots using a monoclonal antibody  
553 against H3K27me3 on protein or chromatin extracts of different species representative of the  
554 super SAR lineage (1: *Bigelowiella\_natans*, 2: *Gymnophora dimorpha*, 3: *Skeletonema*  
555 *marinoi*, 4: *Thalassiosira pseudonana*, 5: *Raphoneis sp*, 6: *Synedra sp*, 7:  
556 *Asterionellopsisglacialis*, 8: *Thalassiosira rotula*, 9: *Phaeodactylum tricornutum*, 10:  
557 *Isocrhysis lutea*, 11: *Amhedinium klebselii*, 12: *Amhedinium carteri*). (e) Western blot on  
558 chromatin extracts from OM, FM, TM and CM with a monoclonal antibody against H3K27me3  
559 showing a significant difference in enrichment levels of H3K27me3 in TM and CM which are  
560 higher compared to FM and OM.

561

562 **Supplementary Figure 2.** (a) Western blot of chromatin extracts of Cas9 control lines from  
563 each of FM, TM and CM. H4 histone antibody was used as a loading control. (b, c) Mass  
564 spectrometry quantification of di and tri-methylation of lysine 27 of histone H3 in both wild  
565 type and enhancer of zeste knockout mutant. MS/MS spectrum of the  $[M + 2H]^{2+}$  precursor ion  
566 of histone H3 (27–36 residue peptide) tri-methylated or di-methylated on K27. Broken bonds  
567 above and below sequence denote b and y ions, respectively, that were annotated from the  
568 spectrum. (d) Abundance of H3 K27 di- and tri-methylated KSAPATGGVK peptide. Y axis  
569 shows normalized  $\log_2$  (WT/Mutant) of the di-methylated and tri-methylated peptides. All  
570 measurements have been performed in triplicate, and error bars indicated the 97.5% confidence  
571 interval (see supplementary Table 1). (e) Western blot of chromatin extracts from wild type FM  
572 and TM as well as E(z) knockouts in both backgrounds with a monoclonal antibody against  
573 H3K27me2. H4 histone antibody was used as loading control. (f) Growth curves of wild type,  
574 enhancer of zeste mutants and cas9 control line in each of FM, TM and CM. (g) Gel picture of  
575 a molecular marker distinguishing FM and TM and amplifying an insertion in one allele present  
576 in FM but absent in TM.

577

578 **Supplementary Figure 3.**

579 (a) Total genome coverage of H3K27me3 within TM (orange) and FM (blue) showing a higher  
580 mapping of H3K27me3 in TM compared to FM. (b) mean distribution of H3K27me3 over 500  
581 bp upstream, gene body, and 500 bp downstream region of all the gene targets in TM (orange  
582 line) and FM (blue line). (c) Unmarked genes (J01910, J31617), commonly marked genes  
583 (J34600, J44413) were chosen as internal controls. FM (EG00164) and TM specifically marked  
584 (J49062) genes were used as controls for the reproducibility of independent ChIP-QPCR  
585 results. (d) violin plot represents the mean fold change of gene expression in Ez(KO) lines in  
586 TM compared to the wild-type (WT). The significant expression difference between Ez(KO)  
587 and WT is estimated using two-tailed t-test with P value < 0.0001

588

589 **Supplementary Figure 4.** Scatter plots with Pearson correlation coefficient displaying the  
590 relationship between TM and E(z) knock out RNA Seq replicates.

591 **Supplementary Figure 5.** Scatter plots with Pearson correlation coefficient displaying the  
592 relationship between TM ChIP-Seq replicates

593

594

595 **Supplementary Table 1.** Cell counts of different morphologies and cell size measurements in  
596 each wild type morphotype and knock out of E(z)

597 **Supplementary Table 2.** List of genes marked by H3K27me3 with their annotation and GOs.

598 **Supplementary Table 3.** List of the primers used in this study

599 **Supplementary Table 4.** Mass spectrometry quantification of Di- and Tri-methylation of  
600 H3K27me3

601 **Supplementary Figure 1.** Phylogeny (a-c). (d) Western blots using a monoclonal antibody  
602 against H3K27me3 on protein or chromatin extracts of different species representative of the  
603 super SAR lineage (1: *Bigelowiella natans*, 2: *Gymnophora dimorpha*, 3: *Skeletonema*  
604 *marinoi*, 4: *Thalassiosira pseudonana*, 5: *Raphoneis sp*, 6: *Synedra sp*, 7:  
605 *Asterionellopsis glacialis*, 8: *Thalassiosira rotula*, 9: *Phaeodactylum tricornutum*, 10:  
606 *Isocrhysis lutea*, 11: *Amhedinium klebselii*, 12: *Amhedinium carteri*). (e) Western blot on  
607 chromatin extracts from OM, FM, TM and CM with a monoclonal antibody against H3K27me3

608 showing a significant difference in enrichment levels of H3K27me3 in TM and CM which are  
609 higher compared to FM and OM.

610

611 **Supplementary Figure 2.** (a) Western blot of chromatin extracts of Cas9 control lines from  
612 each of FM, TM and CM. H4 histone antibody was used as a loading control. (b, c) Mass  
613 spectrometry quantification of di and tri-methylation of lysine 27 of histone H3 in both wild  
614 type and enhancer of zeste knockout mutant. MS/MS spectrum of the  $[M + 2H]^{2+}$  precursor ion  
615 of histone H3 (27–36 residue peptide) tri-methylated or di-methylated on K27. Broken bonds  
616 above and below sequence denote b and y ions, respectively, that were annotated from the  
617 spectrum. (d) Abundance of H3 K27 di- and tri-methylated KSAPATGGVK peptide. Y axis  
618 shows normalized  $\log_2$  (WT/Mutant) of the di-methylated and tri-methylated peptides. All  
619 measurements have been performed in triplicate, and error bars indicated the 97.5% confidence  
620 interval (see supplementary Table 1). (e) Western blot of chromatin extracts from wild type FM  
621 and TM as well as E(z) knockouts in both backgrounds with a monoclonal antibody against  
622 H3K27me2. H4 histone antibody was used as loading control. (f) Growth curves of wild type,  
623 enhancer of zeste mutants and cas9 control line in each of FM, TM and CM. (g) Gel picture of  
624 a molecular marker distinguishing FM and TM and amplifying an insertion in one allele present  
625 in FM but absent in TM.

626

627 **Supplementary Figure 3.**

628 (a) Total genome coverage of H3K27me3 within TM (orange) and FM (blue) showing a higher  
629 mapping of H3K27me3 in TM compared to FM. (b) mean distribution of H3K27me3 over 500  
630 bp upstream, gene body, and 500 bp downstream region of all the gene targets in TM (orange  
631 line) and FM (blue line). (c) Unmarked genes (J01910, J31617), commonly marked genes  
632 (J34600, J44413) were chosen as internal controls. FM (EG00164) and TM specifically marked  
633 (J49062) genes were used as controls for the reproducibility of independent ChIP-QPCR  
634 results. (d) violin plot represents the mean fold change of gene expression in Ez(KO) lines in  
635 TM compared to the wild-type (WT). The significant expression difference between Ez(KO)  
636 and WT is estimated using two-tailed t-test with P value < 0.0001

637

638 **Supplementary Figure 4.** Scatter plots with Pearson correlation coefficient displaying the  
639 relationship between TM and E(z) knock out RNA Seq replicates.

640 **Supplementary Figure 5.** Scatter plots with Pearson correlation coefficient displaying the  
641 relationship between TM ChIP-Seq replicates

642

643

644

645

646

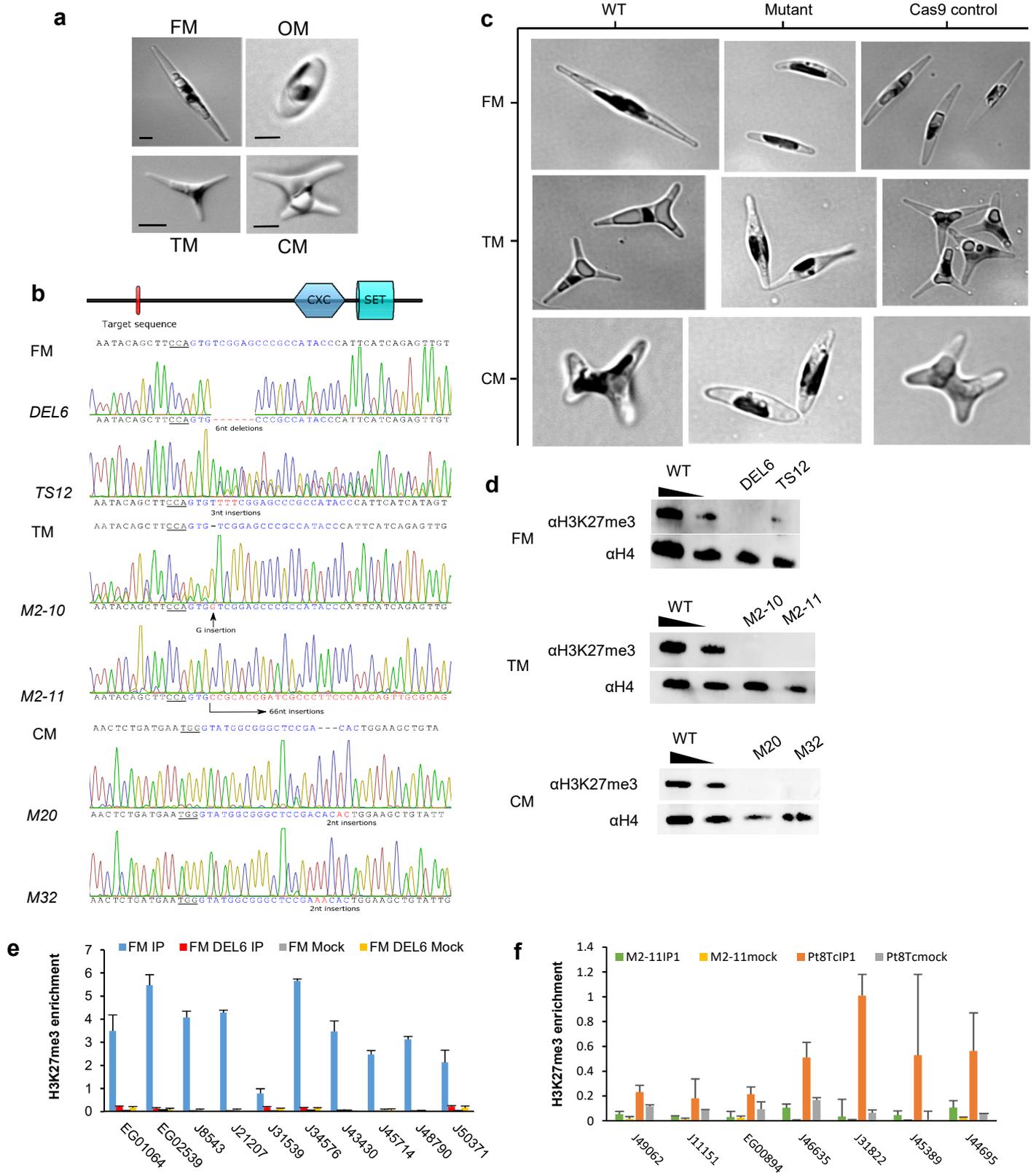


Figure 1

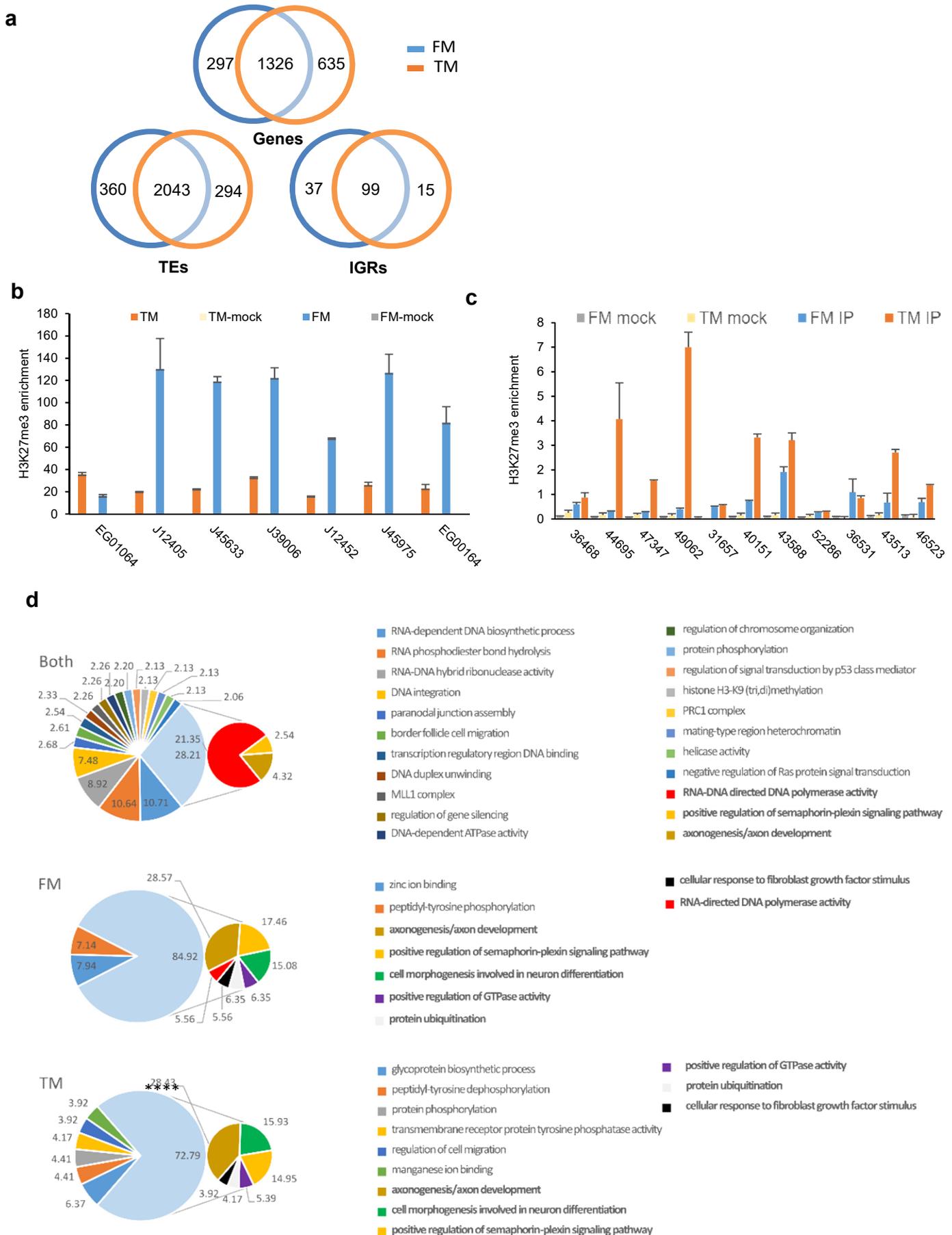


Figure 2

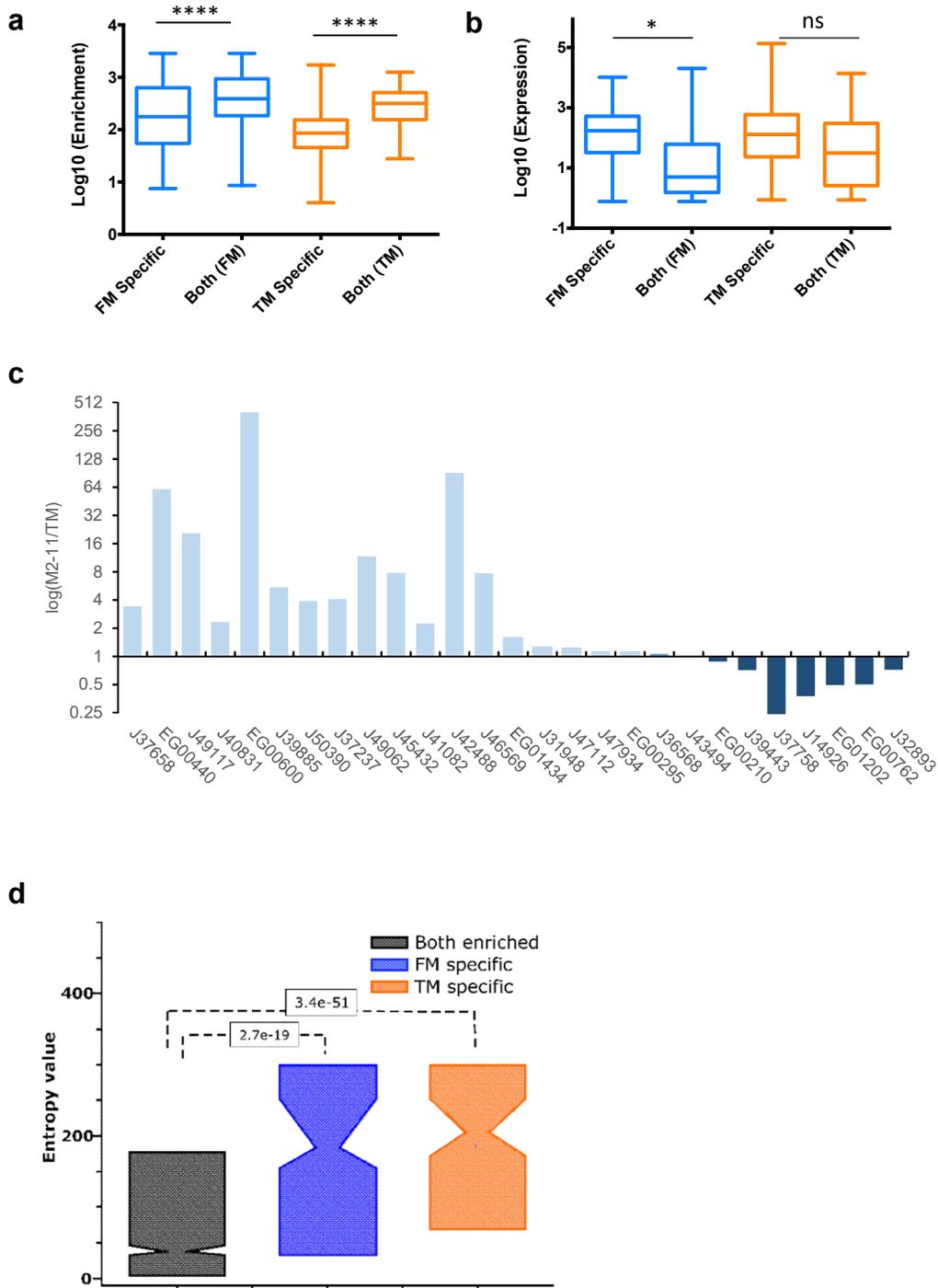


Figure 3

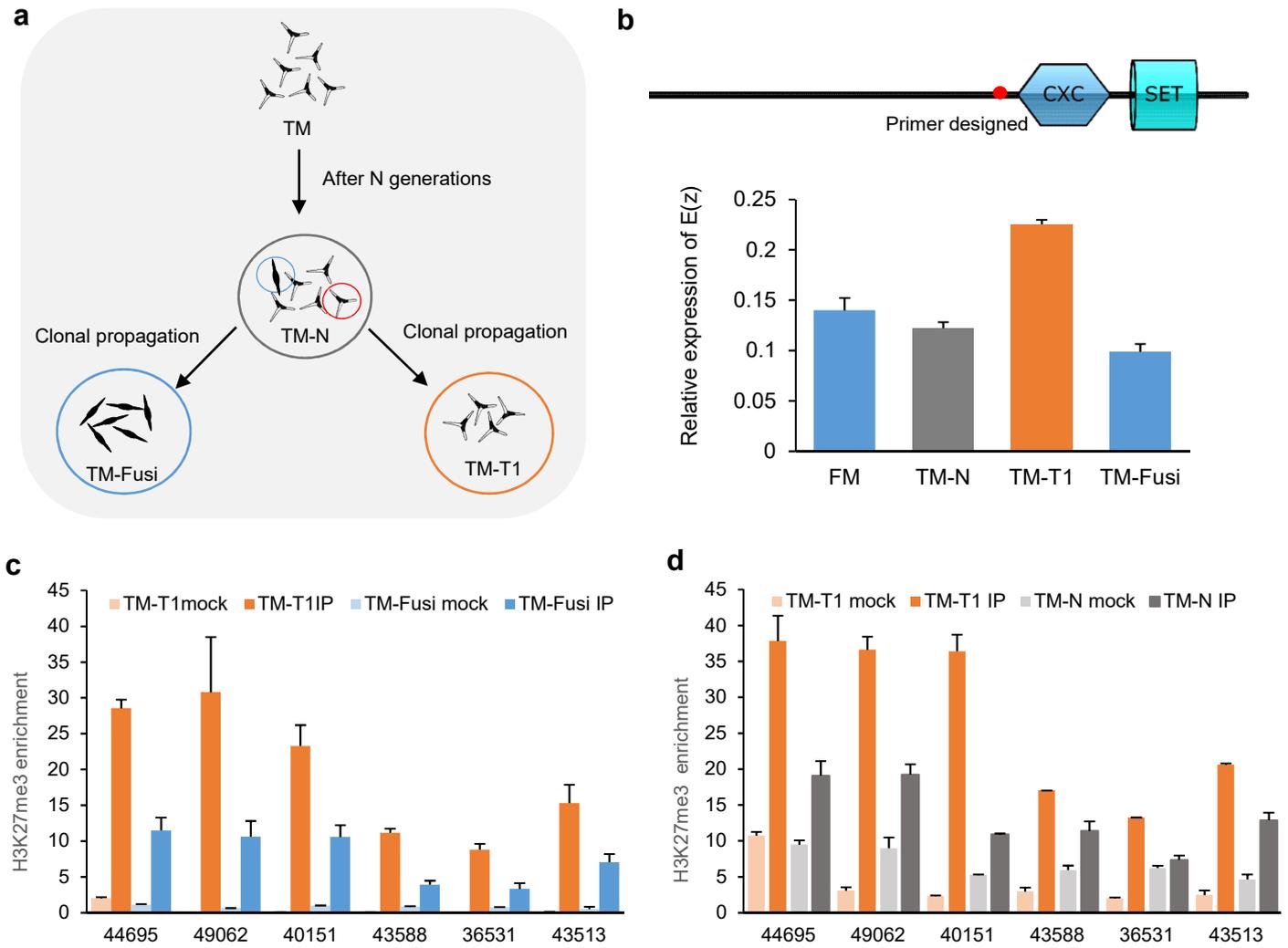


Figure 4

a

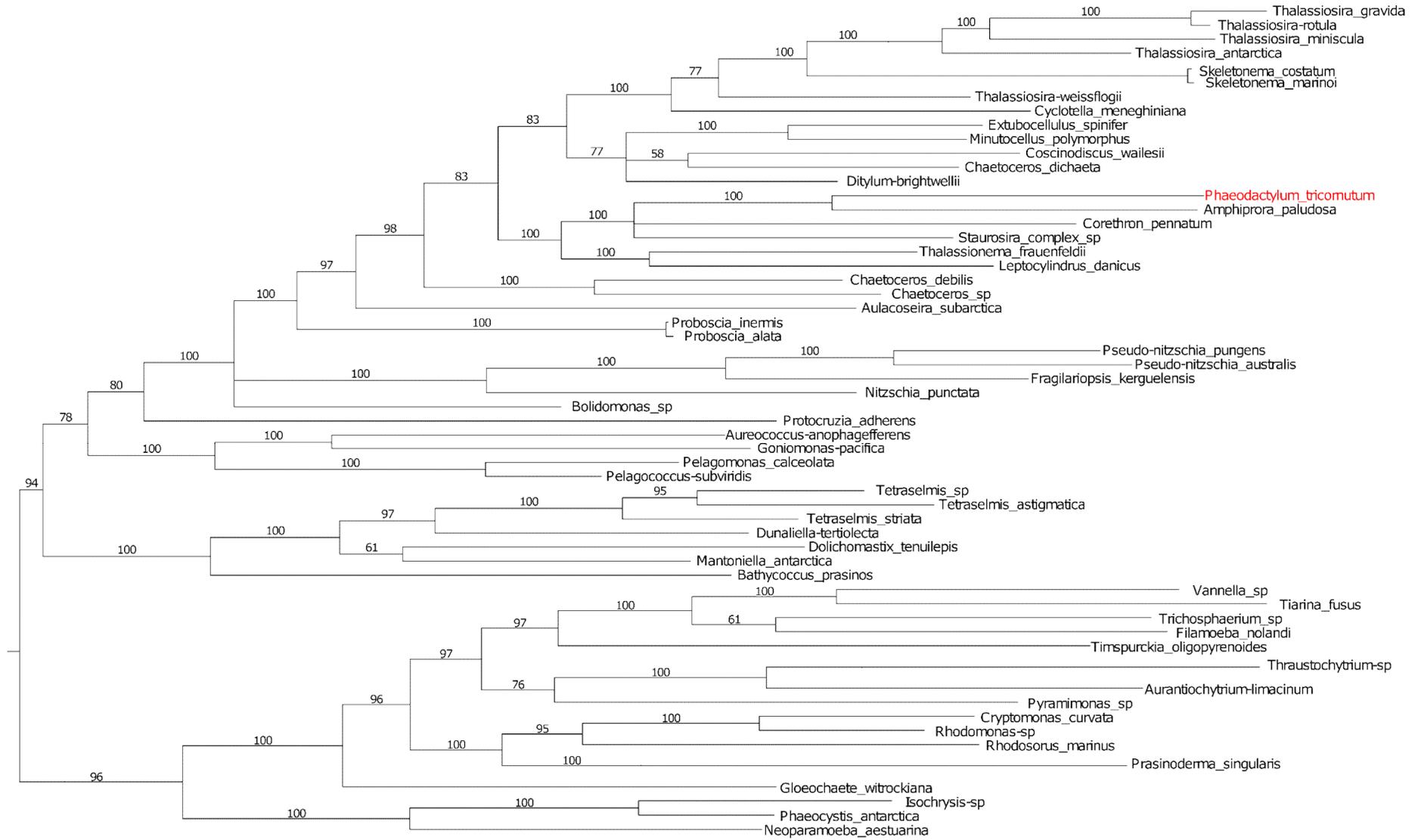


Figure S1

**b**

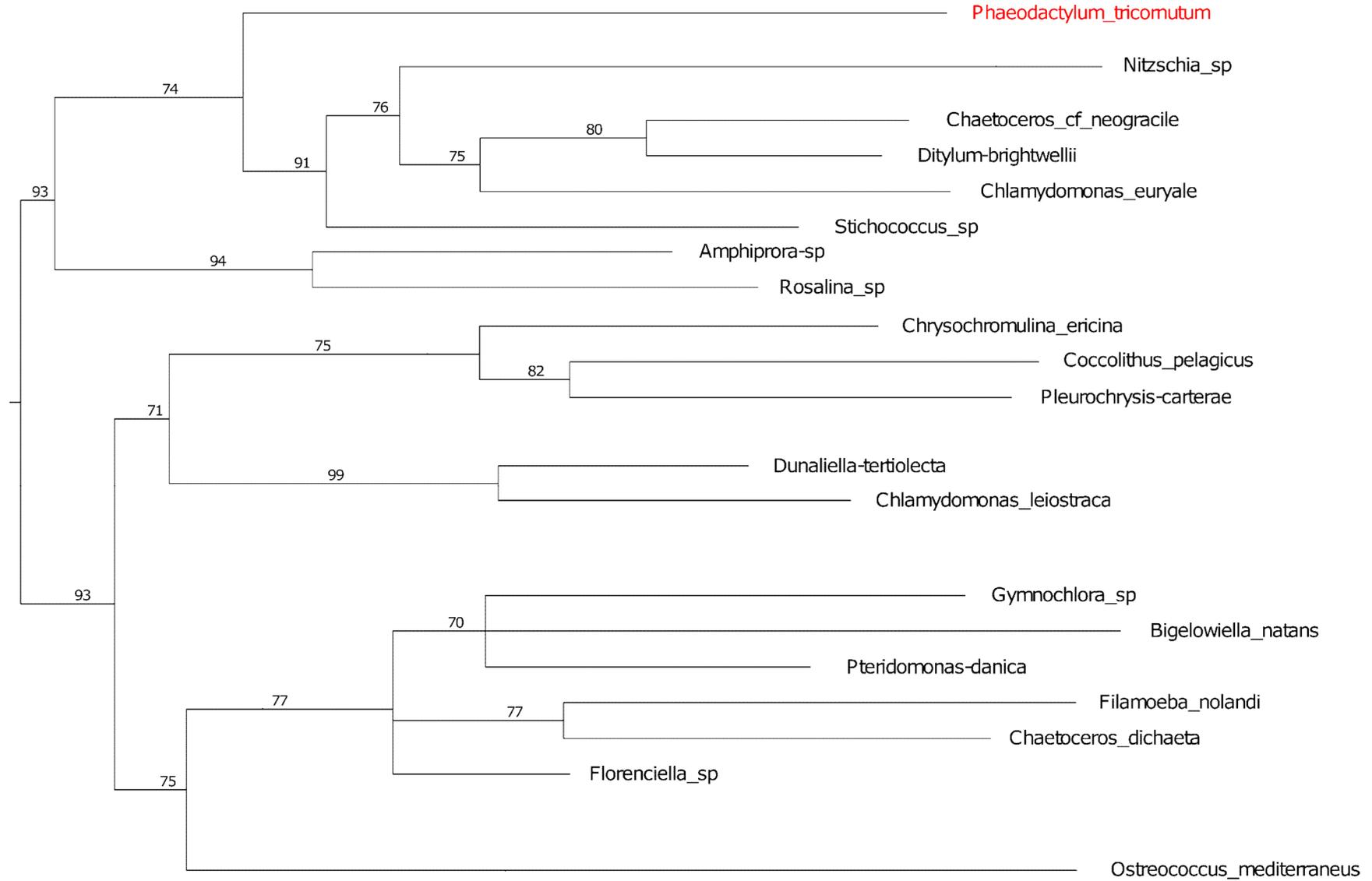


Figure S1

c

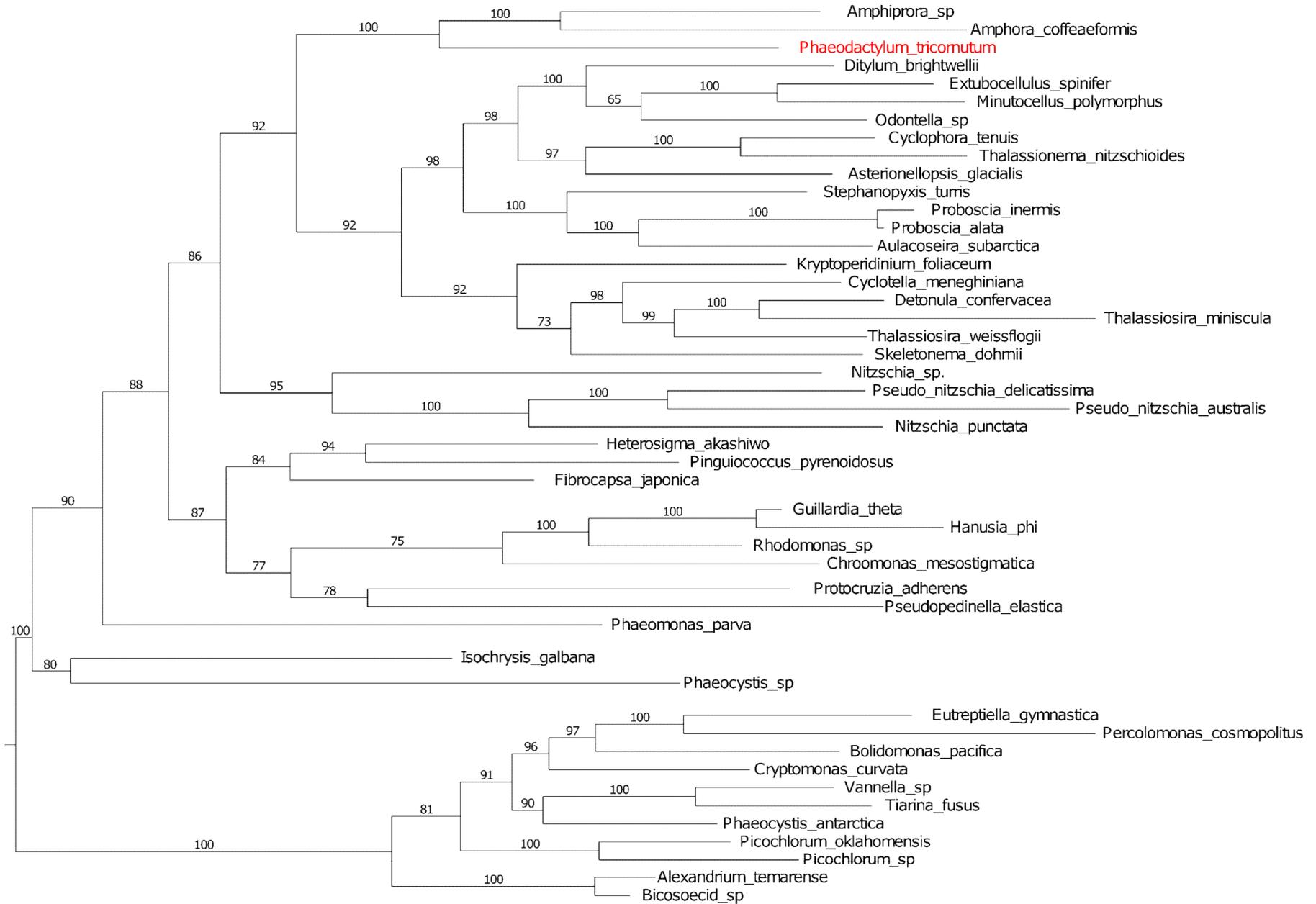
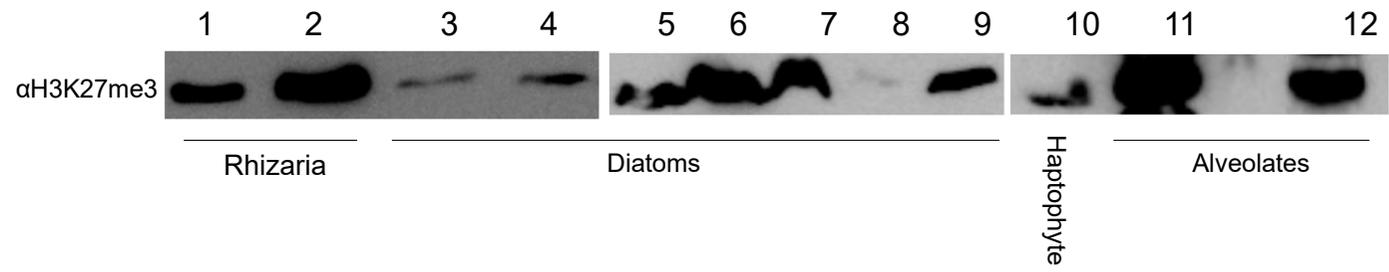


Figure S1

d



e

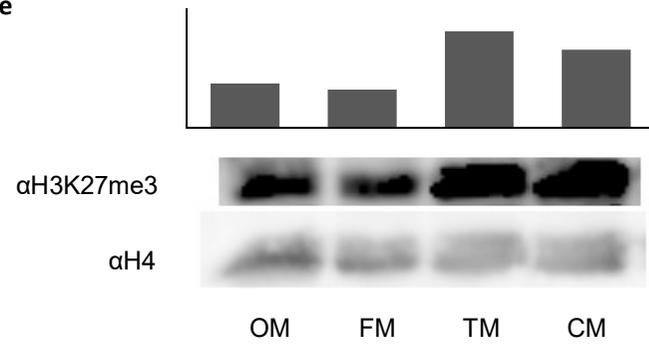


Figure S1

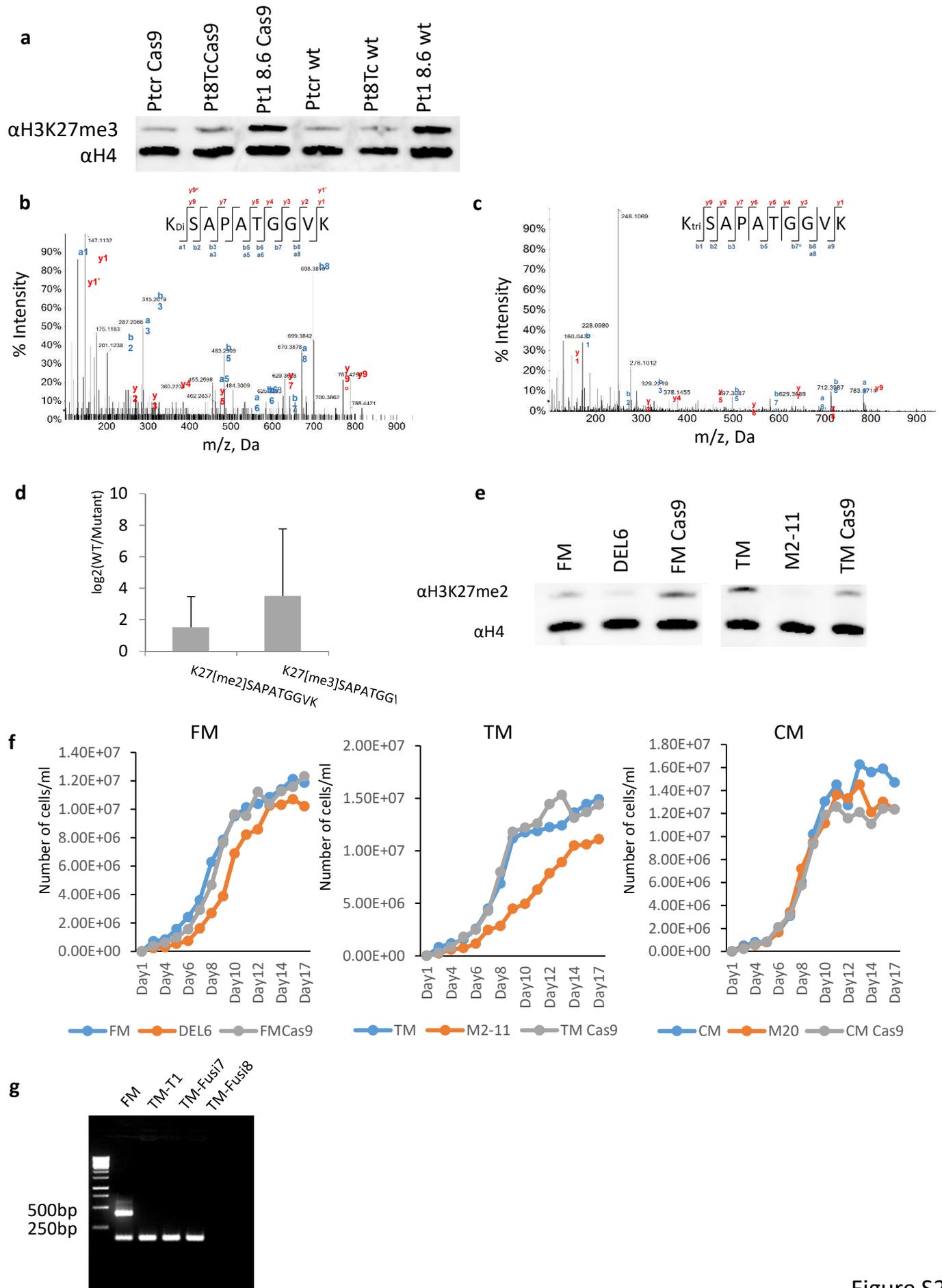


Figure S2

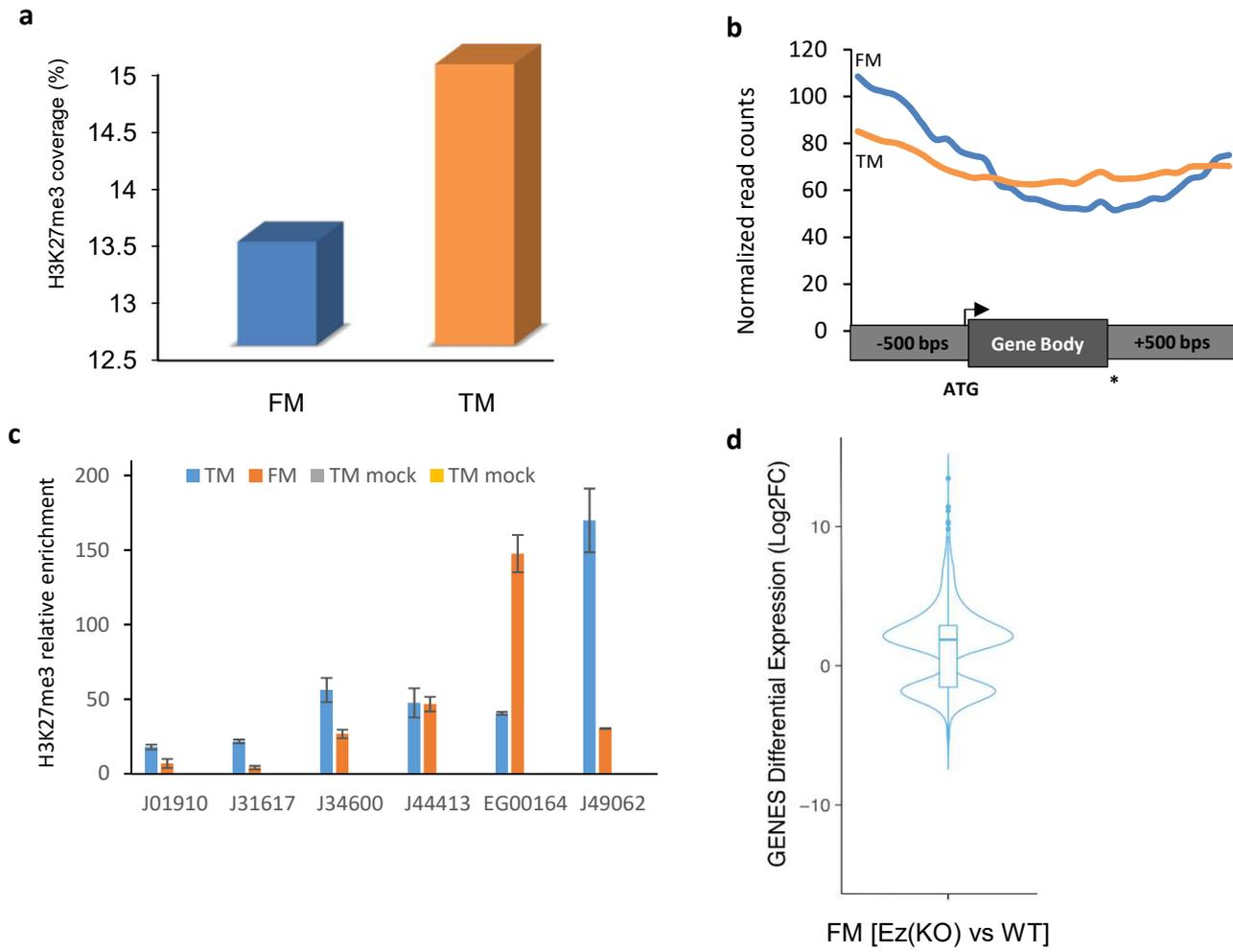


Figure S3

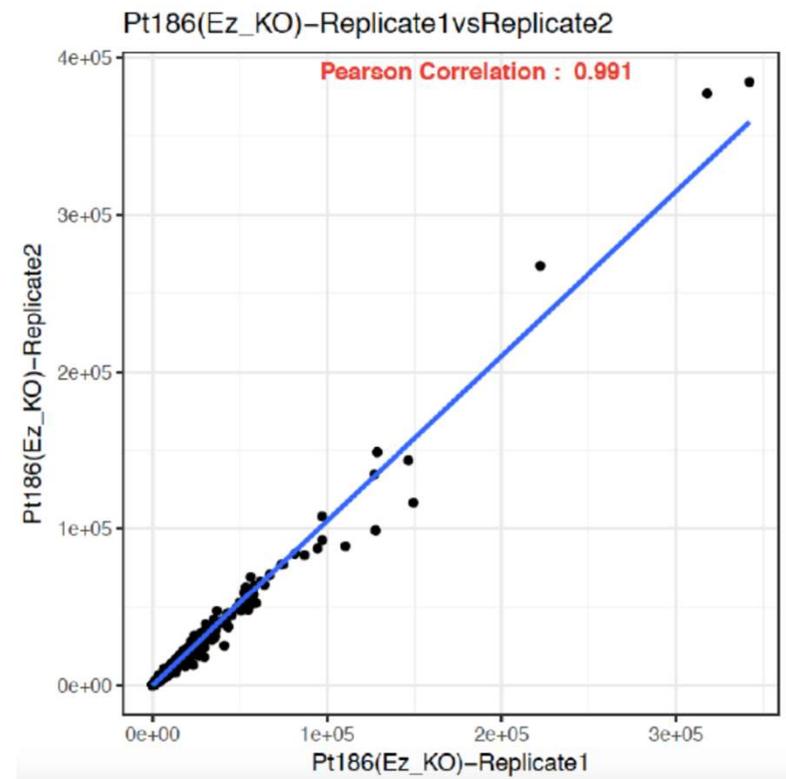
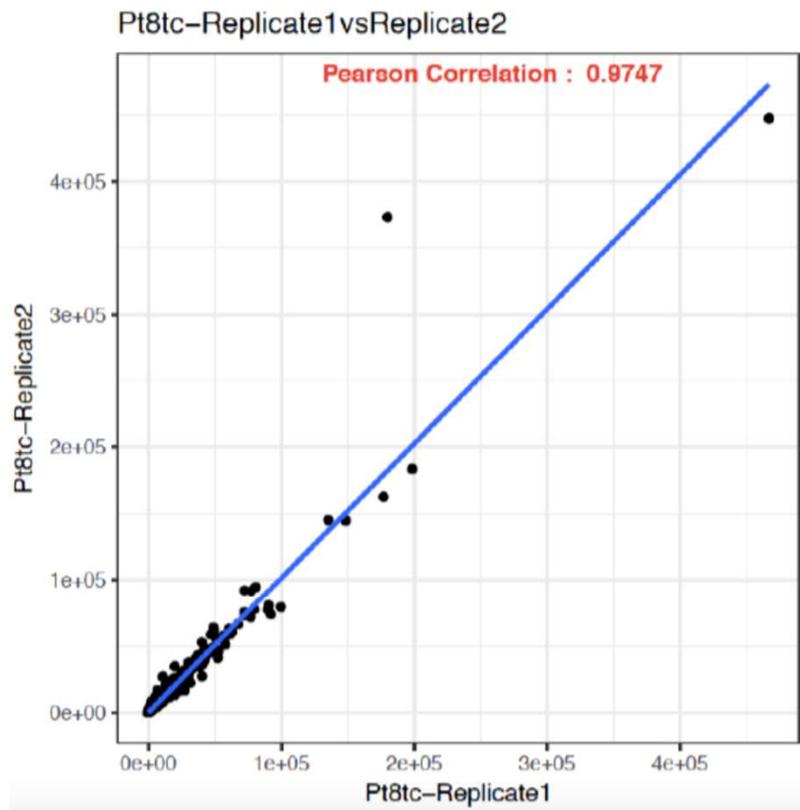


Figure S4

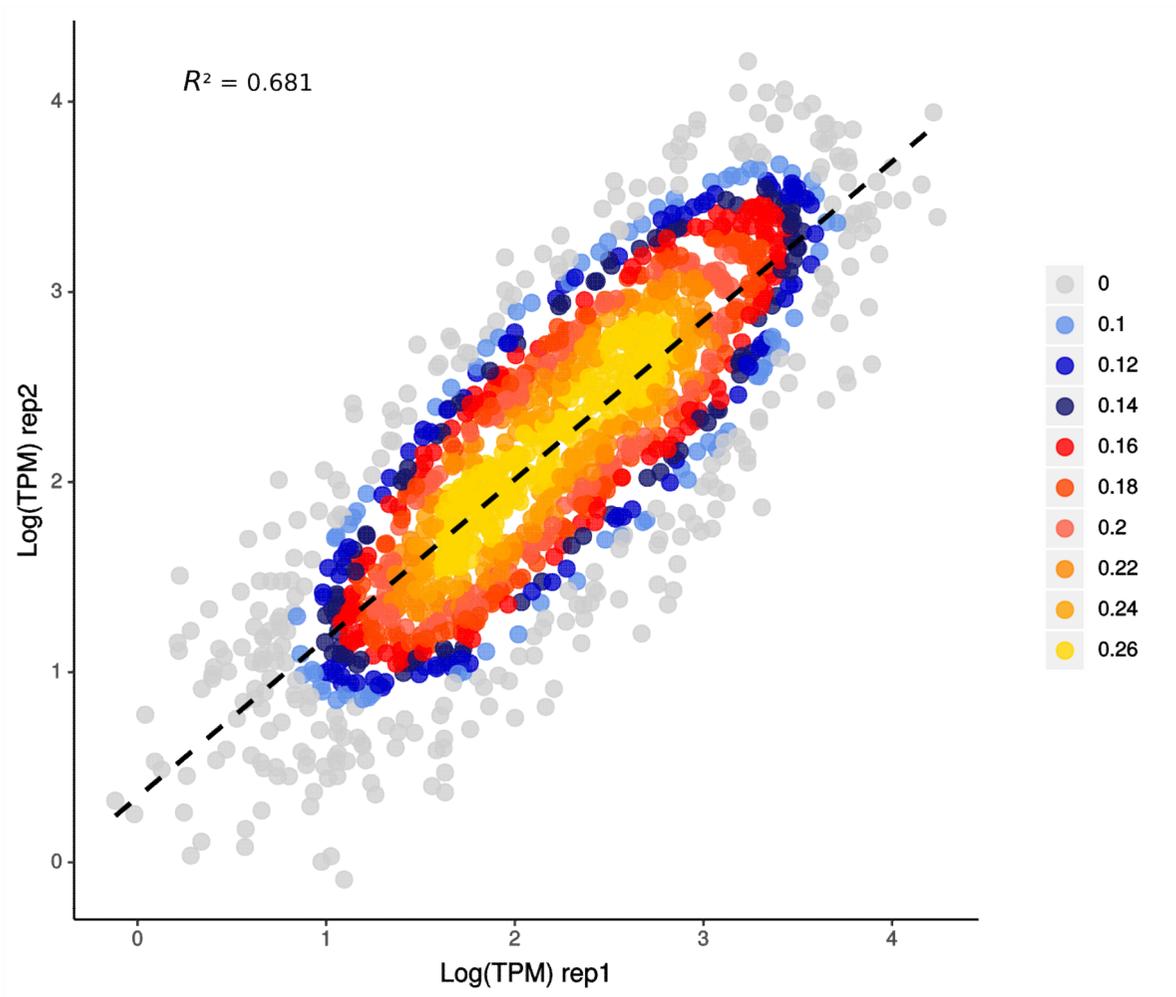


Figure S5

**Table 1. Cell counts of different morphologies and cell size measurements in each wild type morphotype and knock out of *E(z)***

	FM	TM	CM	OM	Majority	Percentage	Size
FM	107	0	0	0	FM	100%	26.58 $\mu$ m
TM	0	100	0	0	TM	100%	L:13.59 $\mu$ m; W:11.48 $\mu$ m
CM	4	31	66	0	CM	67.34%	-
OM	0	0	0	100	OM	100%	-
TM-N	63	55	0	0	FM	53.38%	-
T1	0	219	0	0	TM	100%	-
DEL6	74	0	0	0	FM	100%	19.91 $\mu$ m
TS12	41	0	0	0	FM	100%	-
M2-10	65	0	0	0	FM	100%	-
M2-11	74	0	0	0	FM	100%	21.04 $\mu$ m
M20	32	0	0	0	FM	100%	23.59 $\mu$ m

**Table 2. Mass spectrometry quantification of Di- and Tri-methylation of H3K27me3**

peptide sequence	Protein Modification (without Nterm Met)	peptide used for normalization	log <sub>2</sub> (WT/Mutant)	CI 2.5%	CI 97.5%	adjusted p value
KSAPATGGVK	Dimethyl (K:27)	STDLLIR	1.52038639	1.09023167	1.95054112	2.10E-07
KSAPATGGVK	Trimethyl (K:27)	STDLLIR	3.5091078	2.7625983	4.25561731	7.82E-10