5-16-2010

Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas

Scott Shaver
University of Nebraska - Lincoln

J. Armando Casas-Mollano
University of Nebraska - Lincoln

Ronald L. Cerny
University of Nebraska - Lincoln

Heriberto D. Cerutti
University of Nebraska-Lincoln, hcerutti1@unl.edu

RESEARCH PAPER

Epigenetics 5:4, 301-312; May 16, 2010; © 2010 Landes Bioscience

Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas

Scott Shaver,[1,†] J. Armando Casas-Mollano,[1,†] Ronald L. Cerny[2] and Heriberto Cerutti[1,*]

[1]School of Biological Sciences and Center for Plant Science Innovation; [2]Department of Chemistry; University of Nebraska; Lincoln, NE USA

[†]These authors contributed equally to this work.

Key words: polycomb, enhancer of zeste, H3K27 methylation, chromatin, gene silencing

Polycomb group proteins play an essential role in the maintenance of cell identity and the regulation of development in both animals and plants. The Polycomb Repressive Complex 2 (PRC2) is involved in the establishment of transcriptionally silent chromatin states, in part through its ability to methylate lysine 27 of histone H3 by the Enhancer of zeste [E(z)] subunit. The absence of PRC2 in unicellular model fungi and its function in the repression of genes vital for the development of higher eukaryotes led to the proposal that this complex may have evolved together with the emergence of multicellularity. However, we report here on the widespread presence of PRC2 core subunits in unicellular eukaryotes from the Opisthokonta, Chromalveolata and Archaeplastida supergroups. To gain insight on the role of PRC2 in single celled organisms, we characterized an E(z) homolog, EZH, in the green alga *Chlamydomonas reinhardtii*. RNAi-mediated suppression of *EZH* led to defects in the silencing of transgenes and retrotransposons as well as to a global increase in histone post-translational modifications associated with transcriptional activity, such as trimethylation of histone H3 lysine 4 and acetylation of histone H4. On the basis of the parsimony principle, our findings suggest that PRC2 appeared early in eukaryotic evolution, even perhaps in the last unicellular common ancestor of eukaryotes. One of the ancestral roles of PCR2 may have been in defense responses against intragenomic parasites such as transposable elements, prior to being co-opted for lineage specific functions like developmental regulation in multicellular eukaryotes.

Introduction

Transcriptional activity in eukaryotic organisms is often modulated by chromatin states.[1] The nucleosome, the structural unit of chromatin, consists of 146 base pairs of DNA wrapped around an octamer of four core histones: H3, H4, H2A and H2B.[1-3] The histones have protruding, N-terminal tails that can be modified post-translationally by, among others, acetylation, methylation, phosphorylation, sumoylation or ubiquitination; and these modifications play distinct roles in determining chromatin structure.[1-3] For instance, transcriptionally permissive euchromatin is often characterized by acetylation of histone H4 and trimethylation of lysine 4 of histone H3 (H3K4me3).[1-3] By contrast, histones in silent (hetero)chromatin are usually deacetylated and enriched in the methylation of lysines 9 and/or 27 of histone H3.[1,2] Histone modifications either directly influence chromatin folding, by altering histone-DNA or inter-nucleosomal interactions, or result in the recruitment of factors that can modulate chromatin structure and/or affect transcription.[1-3]

Polycomb group (PcG) proteins were first identified in *Drosophila melanogaster* as regulators of homeotic genes and thus, determinants of body patterning during development.[4] It was subsequently demonstrated that PcG proteins are required to maintain cell identity through the repression of many genes involved in alternative genetic programs.[5-7] Other studies showed that methylation of histone H3 on lysine 27 (H3K27) is carried out by a large complex of PcG proteins named Polycomb Repressive Complex 2.[5,8-10] PRC2 in Drosophila (and similar complexes in mammals) consists of four core proteins: the histone methyltransferase Enhancer of zeste, which serves as the catalytic subunit methylating H3K27 via its SET domain, the WD40 domain-containing polypeptide Extra Sex Comb (ESC), the C2H2-type zinc finger protein Suppressor of zeste 12 [Su(z)12] and the Nucleosome remodeling factor 55-kDa subunit (Nurf55).[8-14] In metazoans, PRC2 appears to function primarily in vivo as a H3K27 mono- and di-methyltransferase whereas a related complex (Pcl-PRC2), which contains additional subunits such as Polycomb-like (Pcl), functions as a H3K27 trimethyltransferase.[14-16]

Core PRC2 subunits are well conserved among metazoans and plants. Moreover, the plant proteins are also essential for the maintenance of cell fate and normal development.[7,15,17]

*Correspondence to: Heriberto Cerutti; Email: hcerutti1@unl.edu
Submitted: 12/23/09; Accepted: 02/24/10
Previously published online: www.landesbioscience.com/journals/epigenetics/article/11608

In *Arabidopsis thaliana*, there are three homologs of E(z) [CURLY LEAF (CLF), MEDEA (MEA) and SWINGER (SWN)][18-20] and three homologs of Su(z)12 [EMBRYONIC FLOWER 2 (EMF2), VERNALIZATION 2 (VRN2) and FERTILIZATION INDEPENDENT SEED 2 (FIS2)].[11,15,20,21] Proteins closely related to ESC and Nurf55, FERTILIZATION-INDEPENDENT ENDOSPERM (FIE) and MULTICOPY SUPPRESSOR OF IRA 1 (MSI1), respectively, have also been identified.[22,23] Furthermore, PRC2-like complexes in Arabidopsis have been shown to be required for trimethylation of H3K27 (H3K27me3) at several loci.[7,17,24,25] However, monomethylation of H3K27 is carried out, at least partly, by two plant-specific methyltransferases, ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 (ATXR5) and ATXR6, which contain divergent SET domains unrelated to that of E(z).[26]

H3K27 methylation is commonly associated with transcriptional repression but the molecular mechanism(s) responsible for this outcome is not clearly understood. In metazoans, another PcG complex, the Polycomb Repressive Complex 1 (PRC1), binds to H3K27me3 via its Polycomb (Pc) subunit.[14,27,28] PRC1 and a related complex (called RING-associated factor in Drosophila) contain RING finger domain proteins that mediate the monoubiquitination of histone H2A (H2Aub).[14,28,29] H3K27me3 and H2Aub play a key role in PcG-mediated gene repression but PRC2 and PRC1 also seem to have silencing functions, such as chromatin compaction, aside from their enzymatic activities.[14,16,28-30] Moreover, emerging evidence suggests that alternative versions of PRC2 and PRC1, with variant subunit compositions, exhibit distinct functions.[14,28] In plants, most PRC1 subunits are not conserved and a protein unrelated to Pc, LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), binds to H3K27me3 and is required for the repression of floral homeotic genes.[15,31,32] However, Arabidopsis homologs of animal RING1 have recently been identified, suggesting the existence of a plant-specific complex analogous to PRC1.[33] Two additional plant proteins, VERNALIZATION 1 and EMBRYONIC FLOWER 1, are also known to play a part in gene repression in cooperation with PRC2-like complexes.[15] Thus, core PRC2 appears to be widely conserved in higher eukaryotes whereas complexes acting, at least partly, downstream from PRC2 and contributing to the stable maintenance of gene silencing seem to be much more divergent.

The lack of E(z) homologs and of H3K27 methylation in model unicellular fungi, such as budding yeast,[34,35] and the function of PcG proteins in regulating the expression of genes involved in organismal development, in both metazoans and plants, initially suggested that this molecular mechanism may have evolved together with the emergence of multicellularity.[7,36] However, given the degree of conservation of PRC2 components among plants and animals, it seems possible or even likely that genes for some PcG proteins were present in their last unicellular common ancestor and were subsequently lost in certain single-celled lineages. The later hypothesis would be strengthened by the identification of PRC2 subunits and H3K27 methylation in extant unicellular eukaryotes. Interestingly, an earlier study reported H3K27 monomethylation in the unicellular green alga *Chlamydomonas reinhardtii*.[37] Although, to our knowledge, the methyltransferase(s) responsible for this modification and its biological function(s) have not been examined.

We report here on the presence of core subunits of the PRC2 complex not only in *C. reinhardtii* but also in several unicellular organisms belonging to three of the major eukaryotic supergroups. To gain insight on the role of PRC2 in single-celled eukaryotes we also characterized a Chlamydomonas Enhancer of zeste homolog, EZH. Suppression of *EZH* by RNA interference (RNAi) resulted in the release of transcriptional silencing of tandemly repeated transgenes as well as retrotransposons. RNAi-mediated depletion of EZH also led to a global increase in H3K4 trimethylation and histone H4 acetylation, modifications generally associated with transcriptional activity. Our results suggest that some PcG proteins likely appeared very early in eukaryotic evolution and that they may have played a role in the silencing of intragenomic parasites such as transposable elements.

Results

Distribution of PRC2 core subunits among eukaryotic organisms. Eukaryotes have recently been classified into six major groups: the Opisthokonta (animals and fungi), the Amoebozoa (amoeba and slime molds), the Excavata (diplomonads and heterotrophic flagellates), the Rhizaria (Foraminifera and Cercozoa), the Archaeplastida (red algae, green algae and plants), and the Chromalveolata (Alveolata and Stramenopiles).[38] In order to evaluate the taxonomic distribution of the PRC2 core components, E(z), ESC, Su(z)12 and Nurf55, we surveyed the genomes from organisms belonging to five of these eukaryotic supergroups. Reciprocal BLAST or PSI-BLAST searches of protein and/or translated genomic DNA databases, using as queries conserved domains or full-length sequences, revealed the presence of homologs of E(z), ESC and Su(z)12 in members of the Opisthokonta, Archaeplastida and Chromalveolata (**Table 1**); whereas Nurf55 homologs were found in all organisms examined. However, since several of the analyzed genomes are in draft stage, an important caveat in our analyses is that some proteins may be missing from the databases whereas others may have errors in the predicted gene structure.

The widespread distribution of Nurf55 among eukaryotes (**Table 1**) is not surprising since this polypeptide is shared by several distinct protein machineries, including PRC2, the nucleosome remodeling complex NURF and the chromatin assembly complex CAF-1.[39,40] Thus, to avoid confusing interpretations, only the taxonomic distribution of E(z), ESC and Su(z)12 was considered when making inferences about the origin of the PRC2 complex. Recent evidence supports the placement of the root of the eukaryotic tree between unikonts (Opisthokonta and Amoebozoa) and bikonts (Archaeplastida, Chromalveolata, Rhizaria and Excavata)[41,42] or, alternatively, on the branch leading to the common ancestor of Diplomonadida/Parabasalia (within Excavata).[43] In either case, the pervasive presence of E(z), ESC and Su(z)12 homologs in organisms from the Opisthokonta, Chromalveolata and Archaeplastida (**Table 1**) suggests that PRC2 appeared early in eukaryotic evolution. Moreover, the existence

Table 1. Distribution of PRC2 complex components in eukaryotes

Species	Genome[a]	E(z)[b]	ESC[b]	Su(z)12[b]	Nurf55[b]	H3K27me[bc]
Excavata						
Giardia lamblia	Assembly	-	-	-	+	?
Trypanosoma brucei	Assembly	-	-	-	+	?
Trypanosoma cruzi	Complete	-	-	-	+	?
Leishmania major	Assembly	-	-	-	+	?
Naegleria gruberi	In Progress	-	-	-	+	?
Trichomonas vaginalis	Assembly	-	-	-	+	?
Chromalveolata						
Paramecium tetraurelia	Assembly	+	-	-	+	?
Tetrahymena thermophila	Assembly	+	-	-	+	+
Plasmodium falciparum	Assembly	-	-	-	+	?
Phytophthora sojae	Assembly	+	+	+	+	?
Thalassiosira pseudonana	Assembly	+	+	+	+	?
Emiliania huxleyi	In Progress	+	+	-	+	?
Aureococcus anophagefferens	In Progress	+	-	+	+	?
Rhizaria						
Data not available						
Archaeplastida						
Cyanidioschyzon merolae	Complete	+	+	+	+	?
Ostreococcus tauri	In Progress	+	+	+	+	?
Coccomyxa sp. C169	In Progress	+	+	-	+	?
Volvox carteri	In Progress	+	+	-	+	?
Chlamydomonas reinhardtii	Assembly	+	+	-	+	+
Arabidopsis thaliana	Complete	+	+	+	+	+
Zea mays	In Progress	+	+	+	+	+
Oryza sativa	Complete	+	+	+	+	+
Amoebozoa						
Dictyostelium discoideum	Assembly	-	-	-	+	?
Entamoeba histolytica	Complete	-	-	-	+	?
Opisthokonta						
Monosiga brevicollis	Assembly	+	+	-	+	?
Saccharomyces cerevisiae	Complete	-	-	-	+	-
Schizosaccharomyces pombe	Complete	-	-	-	+	-
Neurospora crassa	Assembly	+	+	+	+	+
Gibberella zeae	Assembly	+	-	+	+	?
Magnaporthe grisea	Assembly	+	-	+	+	?
Aspergillus nidulans	Assembly	-	-	-	+	?
Cryptococcus neoformans	Assembly	+	+	-	+	?
Nematostella vectensis	Assembly	+	+	+	+	?
Caenorhabditis elegans	Complete	+	+	-	+	+
Drosophila melanogaster	Complete	+	+	+	+	+
Strongylocentrotus purpuratus	Assembly	+	+	+	+	?
Homo sapiens	Complete	+	+	+	+	+

[a]Status of genome sequencing projects taken from http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi. [b]Data is expressed as: +, Present; -, Absent; ?, Data not available. [c]Inferred from published data on immunoblot analyses with specific antibodies or mass spectrometry of histone H3.

of PRC2 core subunits in unicellular species belonging to each of these three eukaryotic supergroups indicates that the advent of this complex is unlikely to have been associated with the emergence of multicellularity. However, E(z), ESC and Su(z)12 (or a subset of these proteins) also appear to have been lost independently several times during eukaryotic evolution. For instance, PRC2 seems to be entirely absent in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, both belonging to the Opisthokonta, in the malaria parasite, *Plasmodium falciparum*, classified as a Chromalveolata and in several members of the Amoebozoa (**Table 1**).

Phylogenetic analyses of PRC2 core components. The greatest conservation among the examined polypeptides (with the exception of Nurf55) corresponded to E(z) proteins, which were clearly identifiable in all species where H3K27 methylation has been experimentally demonstrated (**Table 1**). However, a phylogenetic tree of E(z) polypeptides, constructed based on the alignment of their SET domains, did not allow resolving the relationship among most of these proteins (**Fig. 1A**). Moreover, none of the best phylogenetic trees for E(z), ESC or Su(z)12 polypeptides (see below) strictly coincided with the consensus tree of Eukaryota[44] nor could reconstruct the monophyly of some higher-order eukaryotic groups, as previously reported for individual gene or protein trees.[43,45] Besides the usual problems of weakness of phylogenetic signal, lateral gene transfers, hidden paralogy and tree reconstruction artifacts,[45] incorrectly predicted protein models (since some sequences were extracted from draft genomes) may have contributed to greater than usual divergences making the relationship among deep branches difficult to resolve. Despite these caveats, plant and metazoan (with the exception of nematodes) E(z) sequences formed well-supported clades (**Fig. 1A**). Interestingly, plant *E(z)* genes underwent considerable expansion, at least partly prior to the divergence of monocots and dicots.[46] In contrast, most animals appear to encode a single E(z) sequence, save for some vertebrate lineages that contain two.

In spite of the poor statistical support for a single clustering of E(z) proteins (**Fig. 1A**), a common origin of these polypeptides is suggested by their comparable domain organization (**Fig. 1B**). All E(z)-like proteins have a cysteine-rich domain (the Pre-SET motif) preceding the SET domain (**Fig. 1B**). They also contain another cysteine-rich region that corresponds to the previously described Enhancer of zeste domain 2 (EZD2)[46] (**Figs. 1B and S1A**). Some E(z) polypeptides also possess one or two SANT domains (**Figs. 1B and S1B**), a motif implicated in histone binding or protein-protein interactions, but this region appears to have been lost or may have undergone considerable sequence divergence in several lineages.

A phylogenetic tree of ESC-like proteins, constructed by aligning their WD40 repeat domains (**Fig. S2**), suggests a monophyletic origin for the polypeptides found in most eukaryotes (**Fig. 2**). Comparison of the ESC protein architecture in different organisms also supports this conclusion, since nearly all sequences include seven WD40 repeat domains (**Fig. S2**). Interestingly, in nematodes such as *Caenorhabditis elegans*, ESC- and E(z)-like proteins are currently so divergent that they do not reconstruct the monophyly of animals (**Figs. 1A and 2**). In addition, Su(z)12 appears to have been entirely lost from the nematode genome or its sequence might be so dissimilar that it is no longer recognizable in BLAST searches (**Table 1**). These observations are consistent with an unusually fast evolution of PRC2 subunits within the nematode lineage, which may have contribute to the difficulty in resolving the phylogenetic relationship among deep tree branches.

Su(z)12 proteins are not as widely distributed as E(z) and ESC, having been apparently lost from several organisms that encode the other PRC2 core components (**Table 1**). The most notable example occurs in the green alga lineage in which Su(z)12 homologs are present in *Ostreococcus tauri* and *O. lucimarinus* but absent in the volvocine algae, *C. reinhardtii* and *Volvox carteri*. A phylogenetic tree of Su(z)12 proteins, constructed based on the alignment of the C2H2 zinc finger motif and of the VEFS box (found in the C-terminal region of the proteins and characterized by an acidic cluster and a tryptophan/methionine-rich sequence), did not resolve the relationship among most of these polypeptides (**Fig. S3**). Although, as with E(z) proteins, plant and metazoan Su(z)12 sequences formed well-supported clusters and the corresponding genes underwent expansion within the plant lineage (**Fig. S3**). In addition, most Su(z)12 polypeptides share the same domain architecture, a C2H2 zinc finger motif followed by a VEFS box, suggesting that they likely derive from a common ancestor (data not shown). In summary, our results point to a monophyletic origin for ESC and a likely common origin, based on their conserved domain architectures, for both E(z) and Su(z)12; consistent with the notion that the PRC2 subunits may have evolved in an ancestral unicellular eukaryote.

Detection of H3K27 methylation in *Chlamydomonas reinhardtii*. In order to assess the functional relevance of the single E(z) homolog encoded by this unicellular alga (**Table 1**), we first examined the post-translational modifications associated with lysine 27 of histone H3. Chlamydomonas histones were purified by acid extraction of isolated nuclei and separated on a 15% SDS-polyacrylamide gel. A protein band corresponding to histone H3 was excised from the gel, digested with trypsin and subjected to mass spectrometry analyses. Modifications were identified by the observation of mass increments of tryptic peptides in MS spectra; and the sites of modification were determined based on mass changes of individual amino acids in MS/MS fragmentation spectra, which cause specific shifts of *b*- and *y*-ions. For instance, mono- and dimethylation of peptides involve characteristic mass gains of 14 and 28 Da, respectively. These analyses revealed the presence of several peptides containing modifications on Chlamydomonas H3K27 (**Fig. 3**).

From the MS/MS spectrum of the precursor ion at *m/z* 436.67, the fragmentation ions indicated the peptide sequence $^{27}K_{me}$TPATGGVK in which monomethylation at lysine 27 was assigned (**Fig. 3A**). Likewise, fragmentation of the precursor ion at *m/z* 443.77 was consistent with peptide $^{27}K_{me2}$TPATGGVK where H3K27 is dimethylated (**Fig. 3B**). Mascot searches also identified a precursor ion at *m/z* 450.76 whose fragmentation ions (mostly *y* ions) corresponded to the peptide sequence $^{27}K_{ac/me3}$TPATGGVK, either acetylated or trimethylated on K27. Acetylation and trimethylation both increase a peptide nominal mass by ~42 Da

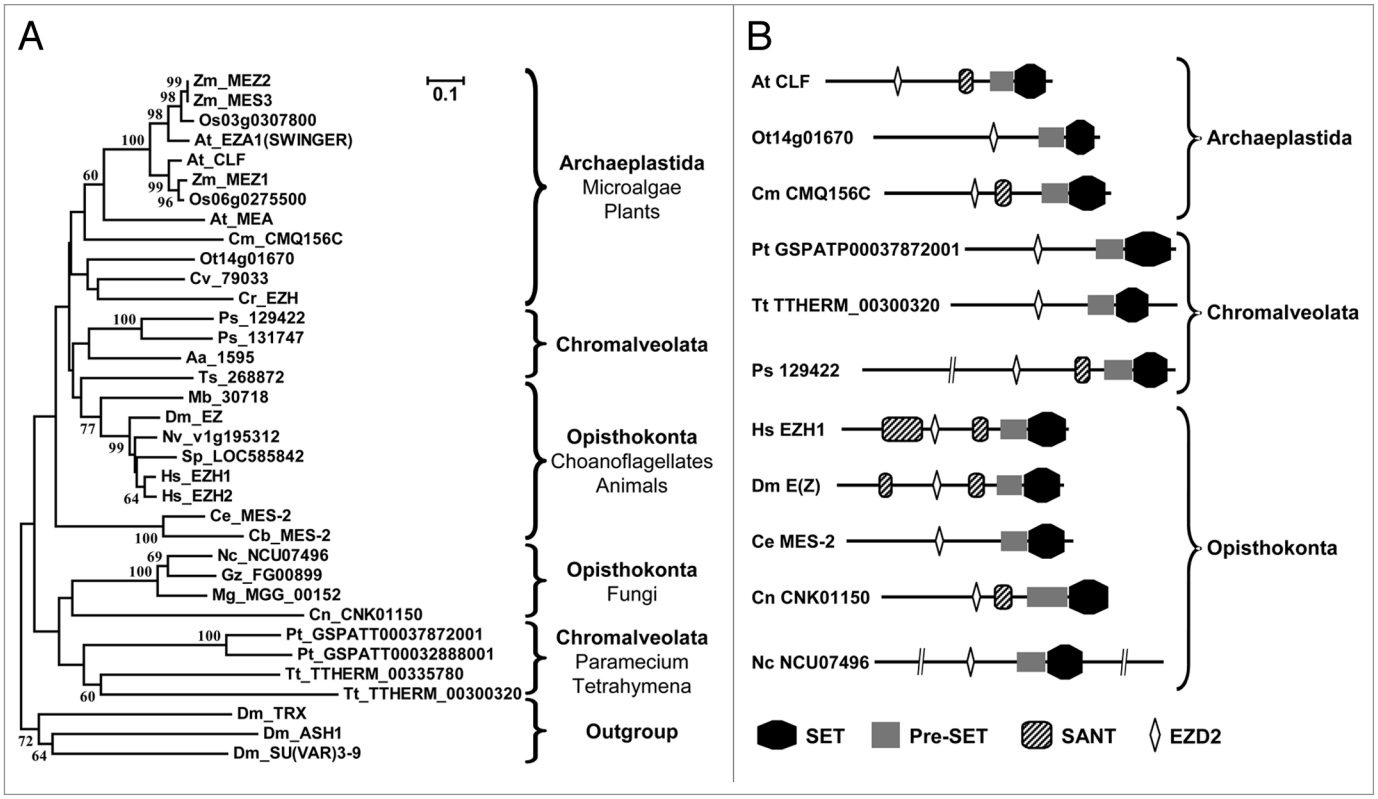

Figure 1. Phylogenetic relationship and protein architecture of E(z) homologs. (A) Neighbor-Joining tree showing the phylogenetic relationship among E(z) proteins. Sequences corresponding to the SET domain were aligned using the ClustalX program and the tree was drawn using the MEGA 4.0 program. Numbers on the branches indicate bootstrap values, as percentage, based on 1,000 pseudoreplicates (only values >60% are shown). Eukaryotic supergroups and the outgroup sequences are indicated on the right of the tree. Species are designated by a two-letter abbreviation preceding the name of each protein: Aa, *Aureococcus anophagefferens*; At, *Arabidopsis thaliana*; Cb, *Caenorhabditis briggsae*; Ce, *Caenorhabditis elegans*; Cm, *Cyanidioschyzon merolae*; Cn, *Cryptococcus neoformans*; Cr, *Chlamydomonas reinhardtii*; Cv, Coccomyxa sp. C169; Dm, *Drosophila melanogaster*; Gz, *Gibberella zeae*; Hs, *Homo sapiens*; Mb, *Monosiga brevicollis*; Mg, *Magnaporthe grisea*; Nc, *Neurospora crassa*; Nv, *Nematostella vectensis*; Os, *Oryza sativa*; Ot, *Ostreococcus tauri*; Ps, *Phytophthora sojae*; Pt, *Paramecium tetraurelia*; Sp, *Strongylocentrotus purpuratus*; Ts, *Thalassiosira pseudonana*; Tt, *Tetrahymena thermophila* and Zm, *Zea mays*. Accession numbers of proteins used in the analysis are as follows: Aa_15951, 15951 at http://genome.jgi-psf.org/Auran1/Auran1.home.html; At_CLF, NP_179919; At_EZA1(SWINGER), NP_567221; At_MEA, NP_563658; Cb_MES-2, CAP30312; Ce_MES-2, AAC27124; Cm_CMQ156C, CMQ156C at http://www.genome.jp/dbget-bin/www_bget?cme:CMQ156C; Cn_CNK01150, XP_567801; Cr_EZH, 154761 at http://genome.jgi-psf.org/Chlre4/Chlre4.home.html; Cv_79033, 79033 at http://genome.jgi-psf.org/Coc_C169_1/Coc_C169_1.home.html; Dm_EZ, P42124; Dm_ASH1, Q9VW15; Dm_Su(var)3-9, CAA56376; Dm_TRX, P20659; Gz_FG00899, XP_381075.1; Hs_EZH1, NP_001982; Hs_EZH2, NP_004447; Mb_30718, XP_001742056; Mg_MGG_00152, XP_369092; Nc_NCU07496, XP_965043; Nv_v1g195312, XP_001622422; Os03g0307800, NP_001049899; Os06g0275500, NP_001057379; Ot_Ot14g01670, CAL56913; Ps_129422, 129422 at http://genome.jgi-psf.org/Physo1_1/Physo1_1.home.html; Ps_131747, 131747 at http://genome.jgi-psf.org/Physo1_1/Physo1_1.home.html; Pt_GSPATT00032888001, XP_001430434 or CAK63036; Pt_GSPATT00037872001, XP_001436830; Sp_LOC585842, XP_790741; Ts_268872, XP_002290191; Tt_TTHERM_00300320, XP_001024559; Tt_TTHERM_00335780, XP_001017546; Zm_MEZ1, NP_001105078; Zm_MEZ2, NP_001105650; Zm_MEZ3, NP_001105079. (B) Domain organization of representative E(z) homologs. The SET, Pre-SET and SANT domains were delimited on each protein by comparison to the SMART and PFAM databases. The EZD2 domain was delimited by using a seed alignment of conserved plant and animal E(z) homologs.

since the exact mass difference between these modifications is only 0.03639 Da. The distinction between acetylation and trimethylation requires either high accuracy mass measurements or detection of modification-specific marker ions or neutral losses. Indeed, collision induced dissociation of peptides containing acetylated lysine residues results in a modification-specific ion of *m/z* 126.[47] Manual inspection of the MS/MS spectrum derived from the precursor ion at *m/z* 450.76 revealed the presence of the characteristic fragmentation ion at m/z 126, thus indicating that the H3K27 residue is acetylated in the examined peptide (**Fig. 3C**). These findings demonstrated that lysine 27 of histone H3 can be mono- or dimethylated in Chlamydomonas, in agreement with an earlier report of H3K27me1 based on Edman degradation protein sequencing and analysis of phenylthiohydantoin-amino acid derivatives by HPLC.[37] Moreover, the results are consistent with the existence of a functional H3K27 methyltransferase(s) in *C. reinhardtii*, very likely EZH, since this alga does not encode homologs of the plant-specific H3K27 mono-methyltransferases ATXR5 and ATXR6.[26]

Expression of the chlamydomonas enhancer of zeste homolog. To gain insight into the biological role of E(z) in unicellular organisms, we decided to characterize the Chlamydomonas *EZH* gene. Reverse transcriptase (RT)-PCR, using primers designed to anneal to exons of computationally predicted *EZH*

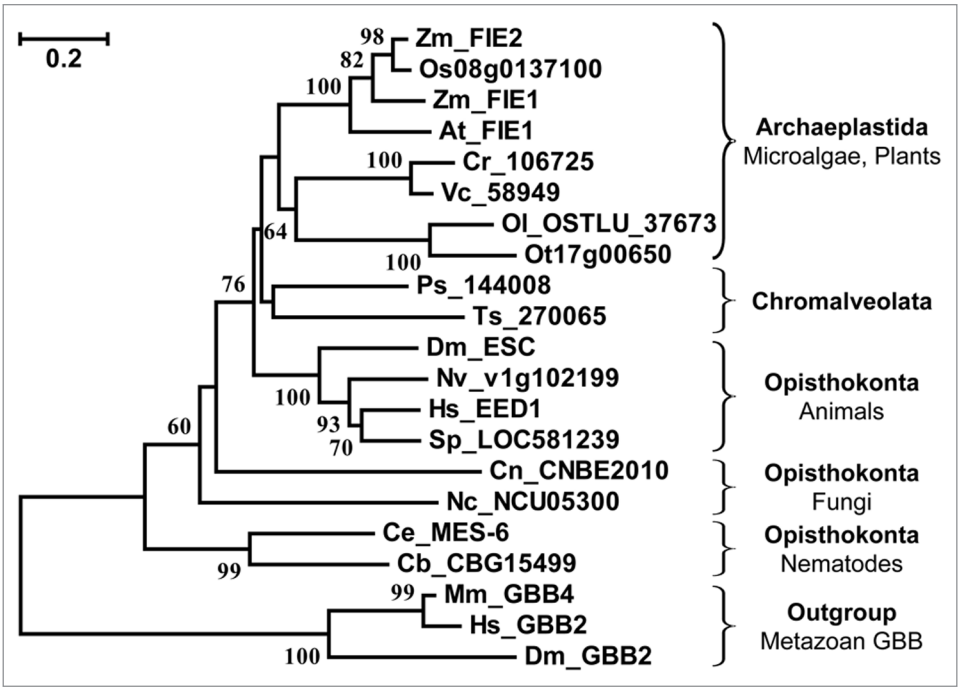

Figure 2. Neighbor-Joining tree showing the phylogenetic relationship among ESC homologs. Sequences corresponding to the WD40 repeat domains were aligned using the ClustalX program and the tree was drawn using the MEGA 4.0 program. Numbers on the branches indicate bootstrap values, as percentage, based on 1,000 pseudoreplicates (only values >60% are shown). The eukaryotic supergroups and the outgroup clade are indicated to the right of the tree. We employed as outgroup sequences the Guanine nucleotide Binding protein subunit Beta2 (GBB2), characterized, as ESC, for containing seven WD40 repeat domains. Species are designated by a two-letter abbreviation preceding the name of each protein, as indicated in the legend to Figure 1, but also including Mm, *Mus musculus*; Ol, *Ostreococcus lucimarinus* and Vc, *Volvox carteri*. Accession numbers of proteins used in the alignment are as follows: At_FIE1, NP_188710; Cb_CBG15499, XP_001677417; Ce_MES-6, NP_001021320; Cn_CNBE2010, XP_775487; Cr_106725, 106725 at http://genome.jgi-psf.org/Chlre3/Chlre3.home.html; Dm_ESC, Q24338; Dm_GBB2, P29829; Hs_EED1, AAD08714; Hs_GBB2, P11016; Mm_GBB4, P29387; Nc_NCU05300, XP_962071; Nv_v1g102199, XP_001634083; Ol_OSTLU_37673, XP_001421975; Os08g0137100, NP_001060953; Ot17g00650, CAL57792; Ps_144008, 144008 at http://genome.jgi-psf.org/sojae1/sojae1.home.html; Sp_LOC581239, XP_78634; Ts_270065, XP_002294990; Vc_58949, 58949 at http://genomeportal.jgi-psf.org/Volca1/Volca1.home.html; Zm_FIE1, AAL35973; Zm_FIE2, AAL35974.

gene models (518348 and 154761, at http://genome.jgi-psf.org/Chlre4/Chlre4.home.html), allowed us to amplify an ~500-bp fragment corresponding to the 3' end of the coding sequence (**Fig. 4A**). This PCR fragment was sequenced and the translated amino acid sequence showed a high degree of identity to the Pre-SET and SET domains of Arabidopsis CLF and SWN. These observations substantiated that Chlamydomonas encodes an E(z) homolog that is expressed in vegetative cells.

Silencing of multiple-copy *RbcS2:aadA:RbcS2* transgenes and of the *TOC1* retrotransposon by EZH. To evaluate a potential role of Chlamydomonas *EZH* in gene repression, we tested whether its RNAi-mediated suppression affected the expression of transcriptionally silenced repetitive sequences. Chlamydomonas strain 1-P(300),[48] which contains more than 100 repressed copies of the *Rbcs2:aadA:Rbcs2* gene (conferring resistance to spectinomycin), was transformed with an inverted repeat (IR) transgene designed to produce double-stranded RNA homologous to the *EZH* transcript.[49] Several of the recovered transformants showed downregulation of expression of *EZH* and two strains, EZH-IR7 and EZH-IR15, were selected for subsequent analyses. RT-PCR assays demonstrated that both strains exhibited significantly reduced steady-state levels of the *EZH* mRNA (**Fig. 4A**). Moreover, EZH-IR7 and EZH-IR15 contain small RNAs hybridizing to the sequence used to build the *EZH* inverted repeats (data not shown), strongly supporting that *EZH* expression is indeed suppressed by RNAi.

Interestingly, EZH-IR7 and EZH-IR15 were able to grow quite well on spectinomycin-containing medium, whereas the parental 1-P[300] strain is very sensitive to this antibiotic (**Fig. 4B**). Northern blot analyses revealed that the EZH-IR strains displayed enhanced levels of *RbcS2:aadA:RbcS2* RNAs (**Fig. 4C**). The full-length ~1.2 kb *RbcS2:aadA:RbcS2* transcript was clearly increased in the RNAi strains, relative to 1-P[300], but we also detected greater accumulation of smaller, seemingly truncated RNAs (**Fig. 4C**). All these observations are consistent with a defect in the transcriptional silencing of the repetitive *RbcS2:aadA:RbcS2* sequences, which are integrated in the Chlamydomonas genome at two chromosomal loci as tandem arrays of head-to-head, tail-to-tail and head-to-tail concatameres interspersed with truncated transgenes.[48] We also analyzed the effect of RNAi-mediated *EZH* suppression on the expression of another repetitive sequence, the *TOC1* retrotransposon. The EZH-IR strains also showed increased accumulation of *TOC1* RNAs (**Fig. 4D**), characterized, as previously described, by a smear of transcripts of various sizes originating from different genomic regions.[50,51] These findings strongly support a role

for Chlamydomonas *EZH* in the silencing of repetitive sequences, including transposable elements.

Effect of *EZH* downregulation on DNA methylation and on global post-translational histone modifications. To gain further insight on the mechanism(s) of *EZH*-mediated gene silencing, we next examined whether DNA (cytosine) methylation of the *RbcS2:aadA:RbcS2* transgenes was affected in the EZH-IR strains. The methylation sensitive isoschizomers HpaII and MspI recognize the same DNA sequence (5'-CCGG-3'), but HpaII is inhibited by methylation of either cytosine whereas MspI is only sensitive to methylation of the outer cytosine residue. Thus, digestion of an unmethylated *RbcS2:aadA:RbcS2* transgene with these enzymes and HindIII generates a fragment of 530 bp and three fragments smaller than 160 bp that can be detected by hybridization with a probe encompassing the *aadA* coding sequence.[52] In contrast, if some of the HpaII/MspI sites become methylated, the inability of the enzymes to cleave will result in the appearance of DNA fragments of higher molecular weight. By using this approach, we observed that the multiple copies of the *RbcS2:aadA:RbcS2* transgene were nearly fully digested with MspI but only partly cleaved with HpaII in the 1-P[300] parental strain (**Fig. S4**), suggesting the presence of CpG DNA methylation in the *aadA* coding sequence and, possibly, in the upstream *RbcS2* promoter. However, we did not find any significant change in the patterns of DNA digestion among 1-P[300] and the EZH-IR strains, indicating that cytosine DNA methylation was unaffected by the RNAi-mediated suppression of *EZH* (**Fig. S4**).

By using antibodies specific for certain histone modifications, we also analyzed if the downregulation of *EZH* caused alterations in post-translational histone marks and, presumably, chromatin states. We examined the steady-state levels of several modifications often linked with silent chromatin such as H3K9me1 and H3K9me3 as well as others usually associated with transcriptionally active chromatin such as H3K4me3 and acetylated H4.[1-3] The immunoblot analyses revealed that the amounts of H3K9me1, H3K9me3, H3K4me1 and H3K4me2 were not affected by *EZH* suppression (**Fig. 5A**). In contrast, H3K4me3 and H4 acetylation levels were increased to varying degrees in the EZH-IR strains compared to the control strains (**Fig. 5A and B**). While E(z) homologs are responsible for H3K27 methylation in both plants and animals,[7,14,15,28] we could not assess directly possible changes in H3K27me associated with downregulation of Chlamydomonas *EZH*. Unfortunately, the amino acid context around K27 in *C. reinhardtii* histone H3 (AAR[27]KTPAT)[37] differs from that in the animal and plant proteins (AAR[27]KSAPAT), rendering commercial antibodies ineffective for the identification of H3K27me1 or H3K27me3 in this green alga (data not shown). In addition, based on precursor ion counts in MS spectra, the level of H3K27 methylation is relatively low in

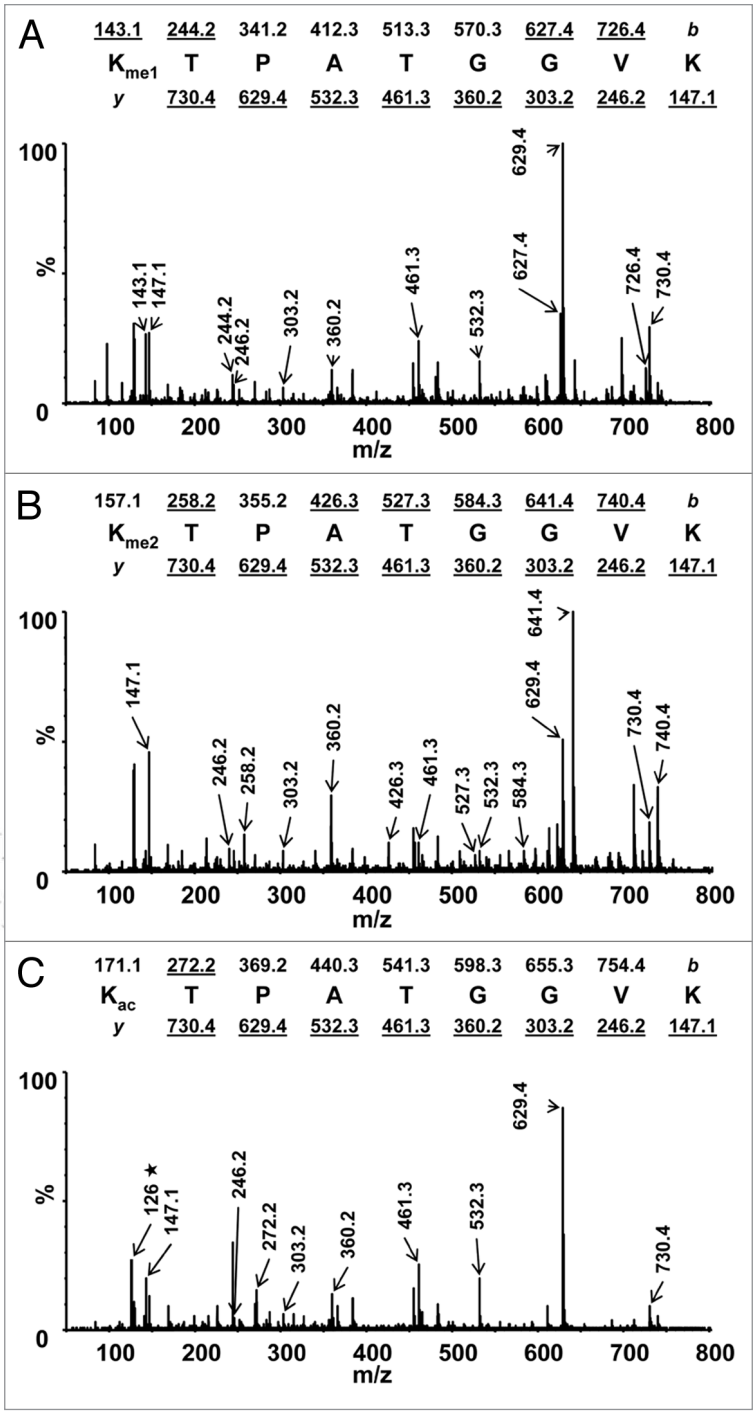

Figure 3. Identification of histone H3 lysine 27 post-translational modifications by mass spectrometry analysis of the $[M + 2H]^{2+}$ peptide spanning residues 27-35(KTPATGGVK) from Chlamydomonas. Predicted nominal masses of *b* and *y* fragment ions are indicated above and below the peptide sequence, respectively. Observed masses on the MS/MS spectra fragment ions are underlined. (A) MS/MS spectrum of the doubly-charged precursor ion at *m/z* 436.67 showing mono-methylation at K27. (B) MS/MS spectrum of the doubly-charged precursor ion at *m/z* 443.77 showing di-methylation at K27. (C) MS/MS spectrum of the doubly-charged precursor ion at *m/z* 450.76 showing acetylation at K27. The ion at *m/z* 126, specific of acetylated peptides, is indicated by an asterisk.

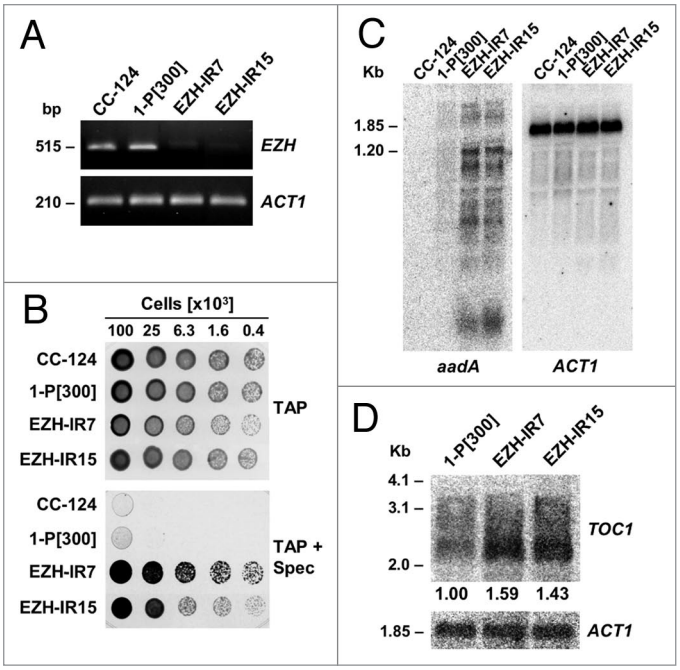

Figure 4. Reactivation of expression of the *RbcS2:aadA:RbcS2* transgene and the *TOC1* retrotransposon in EZH-IR strains. (A) RT-PCR analysis of *EZH* expression in the indicated strains. Amplification of Actin (*ACT1*) transcripts was used as a control for equivalent RNA input. CC-124, wild type strain; 1-P[300], strain containing over 100 copies, transcriptionally silenced, of the *RbcS2:aadA:RbcS2* transgene (conferring resistance to spectinomycin); EZH-IR7 and EZH-IR15, strains derived from 1-P[300] transformed with an IR transgene designed to induce RNAi of *EZH*. (B) Growth and survival on TAP medium or on TAP medium containing spectinomycin (TAP + Spec) of the strains described above. (C) Northern blot analysis of the indicated strains. Total cell RNA was isolated from each strain, separated under denaturing conditions, and hybridized to the *aadA* coding sequence (left). The same blot was reprobed with the coding sequence of Actin (*ACT1*) as a loading control (right). (D) Northern blot of total cell RNA probed sequentially for *TOC1* (upper), to examine transcript levels, and for *ACT1* (lower), to test for equivalent loading of the lanes. The numbers below the upper blot indicate relative *TOC1* transcript abundance.

Chlamydomonas making quantification by mass spectrometry analyses quite unreliable (data not shown). Nevertheless, our observations, taken together, strongly suggest that *EZH* suppression results in defective chromatin-mediated silencing since post-translational histone modifications usually associated with transcriptional activation were globally enhanced in the EZH-IR strains.

Discussion

The most likely point of origin of the E(z), ESC and Su(z)12 protein families was inferred from the patterns of taxonomic distribution and phylogenetic tree topology and on the basis of the parsimony principle.[53] For instance, if a particular protein family is widely represented in all eukaryotic supergroups, the most parsimonious scenario points to its presence in the last common ancestor of eukaryotes. PRC2 core subunits showed widespread distribution among members of three eukaryotic supergroups, the Opisthokonta, the Archaeplastida and the Chromalveolata. Thus, a parsimonious interpretation of the data, reinforced by some of the phylogenetic analyses, is consistent with the appearance of E(z), ESC and Su(z)12 early in eukaryotic evolution, at least in the unicellular common ancestor of these three supergroups. Moreover, an emerging hypothesis is that the earliest evolutionary divergence within eukaryotes (and the root of the eukaryotic tree) falls between unikonts and bikonts.[41,42] If true, then the presence of PRC2 core subunits in animals and plants (which would include organisms derived from each branch of the earliest divergence) would be sufficient to propose that E(z), ESC and Su(z)12 likely evolved in the last unicellular common ancestor of all eukaryotes and were subsequently lost from certain lineages.

In both plants and animals, PRC2 plays a pivotal role in determining cell identity and repressing homeotic genes that are crucial for multicellular organismal development.[5-7,15] These observations and the lack of PRC2 core subunits in several unicellular model fungi provided the basis for the proposal that PRC2 evolved in close association with the emergence of multicellularity.[7,36] However, we found clear evidence for the widespread existence of PRC2 components in unicellular eukaryotes from the Opisthokonta, Archaeplastida and Chromalveolata supergroups. Moreover, H3K27 methylation, the outcome of the E(z) enzymatic activity, has now been demonstrated to occur in single celled species such as Chlamydomonas and *Tetrahymena thermophila*, in the later organism associated with heterochromatin formation and DNA elimination.[54] Thus, PRC2 seems to have appeared and provided a function in an ancestral unicellular eukaryote, and may have been co-opted later for lineage-specific roles like the regulation of genes involved in multicellular development.

To learn more about the possible function(s) of PRC2 in unicellular organisms we characterized the Chlamydomonas E(z) homolog. RNAi-mediated suppression of this gene revealed that it is involved in the transcriptional repression of repetitive transgenes and retrotransposons. In Chlamydomonas, concatameric repeats of the *RbcS2:aadA:RbcS2* transgene are found in a (hetero)chromatic silenced state.[48,52] Similarly to findings in other species, transgenic repeat arrays appear to become silenced by condensation into heterochromatin, an intrinsic property of tandem concatamers not attributable to position effects of nearby sequences.[55-57] The *TOC1* retrotransposon is also present in multiple copies but dispersed throughout the Chlamydomonas genome, mostly as single insertions in intergenic regions (http://genome.jgi-psf.org/Chlre4/Chlre4.home.html).[51] The increased expression of both (hetero)chromatic transgenes and dispersed, presumably euchromatically silenced, retrotransposons in strains suppressed for *EZH* implicates this gene broadly in a form of repression that may reflect a genomic defense response against invading foreign sequences.

Methylation of H3K27 carried out by the SET domain of E(z) is a hallmark of silencing mediated by PRC2.[14,15,28] In Chlamydomonas, mass spectrometry analyses indicated that histone H3 lysine 27 can be mono- or dimethylated. However, we did not identify H3K27me3, a modification that may be either

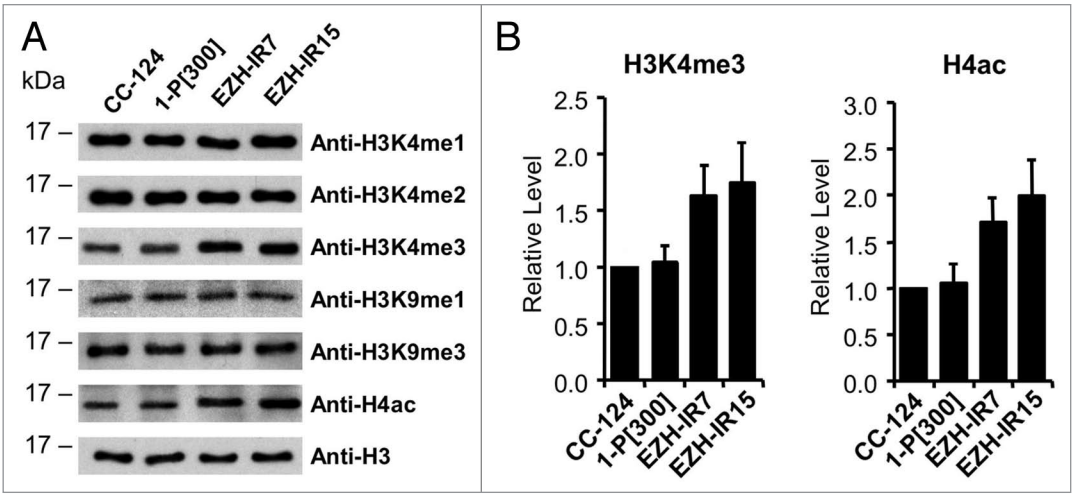

Figure 5. Effect of RNAi-mediated *EZH* suppression on histone post-translational modifications. (A) Immunoblot analyses of global histone modifications in the indicated strains. Whole cell protein extracts were separated by SDS-PAGE, transferred to nitrocellulose and probed with antibodies raised against mono-, di- or trimethyl H3K4; mono- or trimethyl H3K9; or acetylated H4 (H4ac). Sample loading was adjusted based on immunoblot detection with a modification insensitive anti-H3 antibody. (B) Global levels of trimethyl H3K4 (H3K4me3) or acetylated H4 (H4ac) in the examined strains. The histograms show relative amounts of H3K4me3 or H4ac normalized to histone H3 for each strain. For illustration purposes, the H3K4me3 and H4Ac levels in the wild-type strain (CC-124) were set to 1.0 and the remaining samples adjusted accordingly in the bar graphs. The results represent the mean ± the standard deviation of three independent experiments.

entirely absent or present at very low levels in this green alga, and whose detection may be obscured by the more abundant peptides of similar nominal mass that contain acetylated K27. Interestingly, in Arabidopsis, an *atxr5 atxr6* double mutant defective in H3K27 monomethylation shows reactivation of heterochromatic transposable elements,[26] suggesting that lower degrees of H3K27 methylation may be enough to mediate gene silencing. Unfortunately, because of technical issues, we were unable to test directly the effect of *EZH* downregulation on H3K27 methylation in Chlamydomonas. Yet, EZH appears to be the only homolog of a known H3K27 methyltransferase encoded in the *C. reinhardtii* genome and, thus, likely responsible, at least in part, for the observed H3K27me1 and H3K27me2. In addition, a role for EZH in modulating repressive chromatin states is supported by the reactivation of transgenes and transposons and by the changes in global histone post-translational modifications that occur in the EZH-IR strains.

Although the interplay among PRC2, H3K27me3 and DNA methylation is not clearly understood in metazoans,[58-60] EZH-mediated silencing of the *RbcS2:aadA:RbcS2* transgenes in Chlamydomonas appears to occur without affecting DNA methylation. Likewise, no correlation between H3K27 methylation and DNA methylation has been detected in Arabidopsis.[26,32,61] Indeed, Chlamydomonas EZH seems to be involved in maintaining a repressive chromatin conformation, independently of DNA methylation, and its depletion led to an increase in H3K4me3 and H4 acetylation, modifications generally associated with transcriptional activation.[1] Silencing by PcG proteins in both metazoans and plants is counteracted by Trithorax group regulators, which act through several chromatin modifications including H3K4 methylation, histone acetylation and ATP-dependent nucleosome remodeling.[5-7,28,62,63] In animals, direct loss of H3K27me3 results in increases of H3K4me3 and histone acetylation at PcG target loci.[64-67] Moreover, in mouse ES cells, PRC2 has been demonstrated to recruit to target genes a H3K4 demethylase, Retinoblastoma Binding Protein 2.[68] Robust silencing by PRC2, at least in mice, may involve H3K27 methylation along with demethylation of H3K4me3.[28,68] Our findings are consistent with a similar PRC2-mediated mechanism(s) in Chlamydomonas.

In summary, our results strongly suggest that PRC2 appeared early during eukaryotic evolution, in an ancestral unicellular organism, and may have played a general role in gene silencing, perhaps as part of a defense response against intragenomic parasites such as transposable elements. We propose that this core complex alone may have been able to provide silencing through H3K27 methylation, recruitment of H3K4 demethylases and even chromatin compaction, like PRC2-EZH1 in mammals.[28,30] Several other PcG components (including most subunits of PRC1 or analogous complexes) show a more limited taxonomic distribution in living eukaryotes and may have evolved more recently in specific lineages. Moreover, PRC2 or E(z) also appear to have acquired specialized functions in certain lineages such as regulation of genes involved in maintaining cell identity and controlling development in metazoans and plants or participation in the pathway mediating DNA elimination in Tetrahymena. In multicellular organisms part of this specialization may have been achieved through gene expansion and functional diversification.

Materials and Methods

Homology searches. The Arabidopsis CLF (NP_179919) or Drosophila E(z) (P42124) SET domain sequences were used to perform BLAST or PSI-BLAST searches of either the NCBI

database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) or organism specific databases (*Aureococcus anophagefferens* at http://genome.jgi-psf.org/Auran1/Auran1.home.html; Coccomyxa sp. C169 at http://genome.jgi-psf.org/Coc_C169_1/Coc_C169_1.home.html; *Cyanidioschyzon merolae* at http://merolae.biol.s.u-tokyo.ac.jp/; *Emiliania huxleyi* at http://genome.jgi-psf.org/Emihu1/Emihu1.home.html; *Naegleria gruberi* at http://genome.jgi-psf.org/Naegr1/Naegr1.home.html; *Phytophthora sojae* at http://genome.jgi-psf.org/Physo1_1/Physo1_1.home.html; *Volvox carteri* at http://genome.jgi-psf.org/Volca1/Volca1.home.html). The highest scoring protein from each target organism was then used for a BLAST search against the *Arabidopsis thaliana* or the *Drosophila melanogaster* genomes, to confirm a reciprocal best match to the original protein. For the other members of PRC2, BLAST or PSI-BLAST searches were performed using FIE (AAD23584) or ESC (NP_477431), MSI1 (NP_200631) or Nurf55 (NP_524354), and EMF2 (NP_001119413) or Su(z)12 (NP_730465) full-length amino acid sequences. Sequences corresponding to the VEFS box were also used to confirm some hits obtained with EMF2 and Su(z)12.

Phylogenetic analyses. The SMART database was employed to identify conserved domains present in E(z), ESC, Su(z)12 and Nurf55 from different organisms.[69] Amino acid sequence alignments were produced using CLUSTAL X version 1.81,[70] and manual corrections were made using the GENEDOC program (http://www.psc.edu/biomed/genedoc). The phylogenetic relationship among related proteins was deduced by using the neighbor-joining (NJ) method.[71] NJ trees were created using the MEGA program version 4.0,[72] using Poisson-corrected amino acid distances, and including the bootstrap support values for 1,000 pseudoreplicates.

***Chlamydomonas reinhardtii* strains, transgene construction and culture conditions.** The 1-P(300) strain, containing over 100 silenced copies of the *RbcS2:aadA:RbcS2* transgene, and the CC-124 wild-type strain have been previously described.[48,73] *C. reinhardtii* cells were routinely grown in Tris–Acetate–Phosphate (TAP) medium[73] under moderate light conditions at 21°C.[50] The EZH-IR strains are derivatives of 1-P(300), transformed by the glass-beads procedure[74] with an inverted repeat construct designed to produce dsRNA homologous to the *EZH* mRNA. To generate the IR transgene a ~200-bp fragment corresponding to part of the *EZH* coding sequence was amplified by RT-PCR with primers E(Z)F1 (5'-GCT GGG CAC ATC CGA CAT ACC-3') and E(Z)R1 (5'-GGC GTC CAC CAC CCA CTC TGT-3'). This PCR product was cloned in sense and antisense orientation, flanking a 200 bp DNA spacer,[49] by using an inverted repeat-PCR approach.[75] The *EZH* IR cassette was then inserted, by blunt end ligation, into the *Eco*RI sites of the *MAA7/X* IR vector.[49] The resulting construct contains tandem inverted repeats with homology to the *MAA7* and the *EZH* genes so that cells undergoing efficient RNAi can be selected on medium containing 5-fluoroindole.[49] Reactivation of the *RbcS2:aadA:RbcS2* transgenes in the obtained transformants was assayed by spotting serial dilutions of cells on TAP-agar plates with or without 40 mg/L of spectinomycin.

DNA and RNA analyses. Standard protocols were used for DNA isolation, fractionation by gel electrophoresis, and hybridization with ^{32}P-labeled probes.[48,76] Total cell RNA was isolated using TRI reagent, according to the manufacturer's instructions (Molecular Research Center). RNA was resolved in formaldehyde-agarose denaturing gels, blotted to Hybond-NX membranes (GE Healthcare) and then sequentially hybridized with appropriate probes.[76]

Reverse transcriptase-polymerase chain reaction (RT-PCR) analyses. Total RNA was treated with DNase-I (Ambion) to remove contaminant DNA. First-strand cDNA synthesis and PCR reactions were performed as previously described.[49] PCR products were resolved in 1.5% agarose gels and visualized by ethidium bromide staining. The numbers of cycles showing a linear relationship between input RNA and the final product were determined in preliminary experiments. Controls included the use as template of reactions without the reverse transcriptase enzyme (data not shown). The primer sequences were as follows: for *EZH*, E(Z)RTF6 (5'-TGA GAA GTT CTG CGC CTG CAG C-3') and E(Z)RTR2 (5'-CAA AGA TGG CGA TGC GGC TC-3'); and for *Actin*, Act-codF (5'-GAC ATC CGC AAG GAC CTC TAC-3') and Act-codR (5'-GAT CCA CAT TTG CTG GAA GGT-3').

DNA methylation analysis. Genomic DNA from Chlamydomonas was isolated, digested, resolved in agarose gels, blotted and hybridized according to standard procedures.[48,76] The probes used for Southern hybridization corresponded to the coding sequence of the bacterial *aadA* gene and to the 3' end of the chloroplast *psbA* gene.[48,52]

Immunoblot analyses. Histone acetylation and methylation status in vivo was examined by western blotting with modification-specific antibodies.[51,52] H3K4 methylation states were detected with antibodies against H3K4me1 (ab8895; Abcam), H3K4me2 (07-030; Millipore) or H3K4me3 (ab8580; Abcam). H3K9 methylation states were detected with antibodies against H3K9me1 (07-395, Millipore; or ab9045, Abcam) or H3K9me3 (ab1186 or ab8898, Abcam). Histone H4 acetylation was detected with an antibody that recognizes H4 acetylated on lysines 5, 8, 12 and 16 (06-598; Millipore). Sample loading was adjusted based on the signal obtained with a modification-insensitive anti-H3 antibody (ab1791; Abcam).

Histone isolation and identification of histone H3 modifications by mass spectrometry. Histones were purifed from nuclei isolated from a cell wall-less strain[77] following a previously described protocol.[78] Purified histones were separated by SDS PAGE and visualized by Coomassie blue staining. Bands corresponding to histone H3 were excised from the polyacrylamide gel and digested, in gel, with trypsin.[47,79] Mass spectrometry analysis of the tryptic peptides was carried out as previously described.[47,51,79] The MS/MS data obtained was searched, using the MASCOT search engine (Matrix Science), against a local database consisting of Chlamydomonas core histone sequences.

Acknowledgements

We thank members of the Cerutti lab for critical reading of the manuscript. This work was supported in part by grants from the National Institutes of Health and the National Science Foundation (to H.C.). We also acknowledge the support of the Nebraska EPSCoR program.

Note

Supplementary materials can be found at:
www.landesbioscience.com/supplement/ShaverEPI5-4-Sup.pdf

References

1. Kouzarides T. Chromatin modifications and their function. Cell 2007; 128:693-705.
2. Grewal SI, Jia S. Heterochromatin revisited. Nat Rev Genet 2007; 8:35-46.
3. Martin C, Zhang Y. The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol 2005; 6:838-49.
4. Struhl G. A gene product required for correct initiation of segmental determination in Drosophila. Nature 1981; 293:36-41.
5. Papp B, Müller J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev 2006; 20:2041-54.
6. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. Genome regulation by polycomb and trithorax proteins. Cell 2007; 128:735-45.
7. Köhler C, Villar CBR. Programming of gene expression by Polycomb group proteins. Trends Cell Biol 2008; 18:236-43.
8. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in polycomb-group silencing. Science 2002; 298:1039-43.
9. Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. Drosophila Enhancer of zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal polycomb sites. Cell 2002; 111:185-96.
10. Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of zeste protein. Genes Dev 2002; 16:2893-905.
11. Birve A, Sengupta AK, Beuchle D, Larsson J, Kennison JA, Rasmuson-Lestander A, et al. Su(z)12, a novel Drosophila polycomb group gene that is conserved in vertebrates and plants. Development 2001; 128:3371-9.
12. Nekrasov M, Wild B, Müller J. Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO Rep 2005; 6:348-53.
13. Tie F, Stratton CA, Kurzhals RL, Harte PJ. The N terminus of Drosophila ESC binds directly to histone H3 and is required for E(Z)-dependent trimethylation of H3 lysine 27. Mol Cell Biol 2007; 27:2014-26.
14. Müller J, Verrijzer P. Biochemical mechanisms of gene regulation by polycomb group protein complexes. Curr Opin Genet Dev 2009; 19:150-8.
15. Hennig L, Derkacheva M. Diversity of Polycomb group complexes in plants: same rules, different players? Trends Genet 2009; 25:414-23.
16. Nekrasov M, Klymenko T, Fraterman S, Papp B, Oktaba K, Köcher T, et al. Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO J 2007; 26:4078-88.
17. Schubert D, Primavesi L, Bishopp A, Roberts G, Doonan J, Jenuwein T, et al. Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. EMBO J 2006; 25:4638-49.
18. Goodrich J, Puangsomlee P, Martin M, Long D, Meyerowitz EM, Coupland G. A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. Nature 1997; 386:44-51.
19. Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB. Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. Science 1998; 280:446-50.
20. Chanvivattana Y, Bishopp A, Schubert D, Stock C, Moon YH, Sung ZR, et al. Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. Development 2004; 131:5263-76.
21. Yoshida N, Yanai Y, Chen L, Kato Y, Hiratsuka J, Miwa T, et al. EMBRYONIC FLOWER2, a novel polycomb group protein homolog, mediates shoot development and flowering in Arabidopsis. Plant Cell 2001; 13:2471-81.
22. Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, et al. Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. Plant Cell 1999; 11:407-16.
23. Köhler C, Hennig L, Bouveret R, Gheyselinck J, Grossniklaus U, Gruissem W. Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. EMBO J 2003; 22:4804-14.
24. De Lucia F, Crevillen P, Jones AM, Greb T, Dean C. A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. Proc Natl Acad Sci USA 2008; 105:16831-6.
25. Makarevich G, Leroy O, Akinci U, Schubert D, Clarenz O, Goodrich J, et al. Different Polycomb group complexes regulate common target genes in Arabidopsis. EMBO Rep 2006; 7:947-52.
26. Jacob Y, Feng S, LeBlanc CA, Bernatavichute YV, Stroud H, Cokus S, et al. ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat Struct Mol Biol 2009; 16:763-8.
27. Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev 2003; 17:1870-81.
28. Simon JA, Kingston RE. Mechanisms of Polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol 2009; 10:697-708.
29. Leeb M, Wutz A. Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. J Cell Biol 2007; 178:219-29.
30. Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, et al. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol Cell 2008; 32:503-18.
31. Exner V, Aichinger E, Shu H, Wildhaber T, Alfarano P, Caflisch A, et al. The chromodomain of LIKE HETEROCHROMATIN PROTEIN 1 is essential for H3K27me3 binding and function during Arabidopsis development. PLoS ONE 2009; 4:5335.
32. Zhang X, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, Goodrich J, et al. Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol 2007; 5:129.
33. Xu L, Shen WH. Polycomb silencing of KNOX genes confines shoot stem cell niches in Arabidopsis. Curr Biol 2008; 18:1966-71.
34. Lachner M, Sengupta R, Schotta G, Jenuwein T. Trilogies of histone lysine methylation as epigenetic landmarks of the eukaryotic genome. Cold Spring Harb Symp Quant Biol 2004; 69:209-18.
35. Veerappan CS, Avramova Z, Moriyama EN. Evolution of SET-domain protein families in the unicellular and multicellular Ascomycota fungi. BMC Evol Biol 2008; 8:190.
36. Trojer P, Reinberg D. Histone lysine demethylases and their impact on epigenetics. Cell 2006; 125:213-7.
37. Waterborg JH, Robertson AJ, Tatar DL, Borza CM, Davie JR. Histones of *Chlamydomonas reinhardtii*. Synthesis, acetylation and methylation. Plant Physiol 1995; 109:393-407.
38. Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 2005; 52:399-451.
39. Tyler JK, Bulger M, Kamakaka RT, Kobayashi R, Kadonaga JT. The p55 subunit of Drosophila chromatin assembly factor 1 is homologous to a histone deacetylase-associated protein. Mol Cell Biol 1996; 16:6149-59.
40. Martínez-Balbás MA, Tsukiyama T, Gdula D, Wu C. Drosophila NURF-55, a WD repeat protein involved in histone metabolism. Proc Natl Acad Sci USA 1998; 95:132-7.
41. Stechmann A, Cavalier-Smith T. The root of the eukaryote tree pinpointed. Curr Biol 2003; 13:665-6.
42. Simpson AG, Roger AJ. The real 'kingdoms' of eukaryotes. Curr Biol 2004; 14:693-6.
43. Arisue N, Hasegawa M, Hashimoto T. Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. Mol Biol Evol 2004; 22:409-20.
44. Medina M. Genomes, phylogeny and evolutionary systems biology. Proc Natl Acad Sci USA 2005; 102:6630-5.
45. Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, et al. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Philos Trans R Soc Lond B Biol Sci 2000; 267:1213-21.
46. Springer NM, Danilevskaya ON, Hermon P, Helentjaris TG, Phillips RL, Kaeppler HF, et al. Sequence relationships, conserved domains and expression patterns for maize homologs of the polycomb group genes E(z), esc and E(Pc). Plant Physiol 2002; 128:1332-45.
47. Bergmüller E, Gehrig PM, Gruissem W. Characterization of post-translational modifications of histone H2B-variants isolated from *Arabidopsis thaliana*. J Proteome Res 2007; 6:3655-68.
48. Cerutti H, Johnson AM, Gillham NW, Boynton JE. A eubacterial gene conferring spectinomycin resistance on *Chlamydomonas reinhardtii*: integration into the nuclear genome and gene expression. Genetics 1997; 145:97-110.
49. Rohr J, Sarkar N, Balenger S, Jeong BR, Cerutti H. Tandem inverted repeat system for selection of effective transgenic RNAi strains in Chlamydomonas. Plant J 2004; 40:611-21.
50. Jeong Br BR, Wu-Scharf D, Zhang C, Cerutti H. Suppressors of transcriptional transgenic silencing in Chlamydomonas are sensitive to DNA-damaging agents and reactivate transposable elements. Proc Natl Acad Sci USA 2002; 99:1076-81.
51. van Dijk K, Marley KE, Jeong BR, Xu J, Hesson J, Cerny RL, et al. Monomethyl histone H3 lysine 4 as an epigenetic mark for silenced euchromatin in Chlamydomonas. Plant Cell 2005; 17:2439-53.

52. Casas-Mollano JA, van Dijk K, Eisenhart J, Cerutti H. SET3p monomethylates histone H3 on lysine 9 and is required for the silencing of tandemly repeated transgenes in Chlamydomonas. Nucleic Acids Res 2007; 35:939-50.
53. Anantharaman V, Koonin EV, Aravind L. Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res 2002; 30:1427-64.
54. Liu Y, Taverna SD, Muratore TL, Shabanowitz J, Hunt DF, Allis CD. RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. Genes Dev 2007; 21:1530-45.
55. Henikoff S. Conspiracy of silence among repeated transgenes. Bioessays 1998; 20:532-5.
56. Probst AV, Fransz PF, Paszkowski J, Mittelsten Scheid O. Two means of transcriptional reactivation within heterochromatin. Plant J 2003; 33:743-9.
57. Martienssen RA. Maintenance of heterochromatin by RNA interference of tandem repeats. Nat Genet 2003; 35:213-4.
58. Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, et al. The polycomb group protein EZH2 directly controls DNA methylation. Nature 2006; 439:871-4.
59. Lindroth AM, Park YJ, McLean CM, Dokshin GA, Persson JM, Herman H, et al. Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. PLoS Genet 2008; 4:1000145.
60. Rush M, Appanah R, Lee S, Lam LL, Goyal P, Lorincz MC. Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not de novo DNA methylation. Epigenetics 2009; 4:404-14.
61. Mathieu O, Probst AV, Paszkowski J. Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. EMBO J 2005; 24:2783-91.
62. Aichinger E, Villar CB, Farrona S, Reyes JC, Hennig L, Köhler C. CHD3 proteins and polycomb group proteins antagonistically determine cell identify in Arabidopsis. PLoS Genet 2009; 5:1000605.
63. Avramova Z. Evolution and pleiotropy of TRITHORAX function in Arabidopsis. Int J Dev Biol 2009; 53:371-81.
64. Agger K, Cloos PA, Christensen J, Pasini D, Rose S, Rappsilber J, et al. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature 2007; 449:731-4.
65. Hong S, Cho YW, Yu LR, Yu H, Veenstra TD, Ge K. Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases. Proc Natl Acad Sci USA 2007; 104:18439-44.
66. Lee MG, Villa R, Trojer P, Norman J, Yan KP, Reinberg D, et al. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science 2007; 318:447-50.
67. Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, et al. CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development 2009; 136:3131-41.
68. Pasini D, Hansen KH, Christensen J, Agger K, Cloos PAC, Helin K. Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. Genes Dev 2008; 22:1345-55.
69. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, et al. SMART 4.0: towards genomic data integration. Nucleic Acids Res 2004; 32:142-4.
70. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997; 25:4876-82.
71. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987; 4:406-25.
72. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 2007; 24:1596-9.
73. Harris EH. The Chlamydomonas sourcebook—A comprehensive guide to biology and laboratory use. San Diego, CA: Academic Press 1989.
74. Kindle KL. High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA 1990; 87:1228-32.
75. Pawloski LC, Deal RB, McKinney EC, Burgos-Rivera B, Meagher RB. Inverted repeat PCR for the rapid assembly of constructs to induce RNA interference. Plant Cell Physiol 2005; 46:1872-8.
76. Sambrook J, Russell DW. Molecular cloning—A laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press 2001.
77. Davies DR, Plaskitt A. Genetical and structural analyses of cell-wall formation in *Chlamydomonas reinhardi*. Genet Res 1971; 17:33-43.
78. Morris RL, Keller LR, Zweidler A, Rizzo PJ. Analysis of *Chlamydomonas reinhardtii* histones and chromatin. J Protozool 1990; 37:117-23.
79. Rohila JS, Chen M, Cerny R, Fromm ME. Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. Plant J 2004; 38:172-81.