Plant Mol Biol
DOI 10.1007/s11103-013-0165-6

A single CMT methyltransferase homolog is involved in CHG DNA methylation and development of *Physcomitrella patens*

Chen Noy-Malka · Rafael Yaari · Rachel Itzhaki · Assaf Mosquna · Nitzan Auerbach Gershovitz · Aviva Katz · Nir Ohad

Received: 13 November 2013 / Accepted: 12 December 2013

Abstract C-5 DNA methylation is an essential mechanism controlling gene expression and developmental programs in a variety of organisms. Though the role of DNA methylation has been intensively studied in mammals and *Arabidopsis*, little is known about the evolution of this mechanism. The chromomethylase (CMT) methyltransferase family is unique to plants and was found to be involved in DNA methylation in *Arabidopsis*, maize and tobacco. The moss *Physcomitrella patens,* a model for early terrestrial plants, harbors a single homolog of the CMT protein family designated as PpCMT. Our phylogenetic analysis suggested that the CMT family is unique to embryophytes and its earliest known member PpCMT belongs to the CMT3 subfamily. Thus, *P. patens* may serve as a model to study the ancient functions of the CMT3 family. We have generated a *ΔPpcmt* deletion mutant which demonstrated that PpCMT is essential for *P. patens* protonema and gametophore development and is involved in CHG methylation as demonstrated at four distinct genomic loci. PpCMT protein accumulation pattern correlated with proliferating cells and was sub-localized to the nucleus as predicted from its function. Taken together, our results suggested that CHG DNA methylation mediated by CMT has been employed early in land plant evolution to control developmental programs during both the vegetative and reproductive haploid phases along the plant life cycle.

Keywords Chromomethylase · DNA methylation · Epigenetic regulation · Development · *Physcomitrella patens* · Transgene copy number

Abbreviations

BAH	Bromo-adjacent homology domain
CMT	Chromomethylase
DNMT	DNA methyltransferase
DRM	Domains rearranged methyltransferase
GUS	β-Glucuronidase
MTD	C-5 DNA methyltransferase domain
MET1	DNA methyltransferase 1
RdDM	RNA directed DNA methylation

Introduction

DNA methylation, the addition of a methyl group to position 5 on a cytosine (C-5), is an epigenetic modification which contributes to the regulation of gene expression (Goll and Bestor 2005). This modification is necessary for regulating different processes including differentiation and development, silencing transposons and repetitive DNA elements, X-inactivation and parental imprinting (Goll and Bestor 2005; Jullien and Berger 2010; Law and Jacobsen

Chen Noy-Malka and Rafael Yaari have contributed equally to the work.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-013-0165-6) contains supplementary material, which is available to authorized users.

C. Noy-Malka · R. Yaari · R. Itzhaki · N. Auerbach Gershovitz · A. Katz (✉) · N. Ohad (✉)
Department of Molecular Biology and Ecology of Plants, Tel-Aviv University, 69978 Tel-Aviv, Israel
e-mail: tuliplavan@gmail.com

N. Ohad
e-mail: niro@tauex.tau.ac.il

A. Mosquna
Department of Plant Sciences, The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, Hebrew University of Jerusalem, 76100 Rehovot, Israel

Published online: 27 December 2013

2010; Zemach and Zilberman 2010; Bauer and Fischer 2011; Cedar and Bergman 2012). Cytosine methylation is present in most genomes of animals, plants, fungi, algae, protista and bacteria (Goll and Bestor 2005; Feng et al. 2010; Zemach et al. 2010). This chemical modification is enzymatically catalyzed by the C-5 DNA methyltrasferase domain (MTD) found in DNA methyltrasferases (DNMTs). DNA methylation of plant genomes occurs at the sequence contexts CG, CHG and CHH, where H is any nucleotide except G, while in animals methylation occurs almost exclusively at CG sequence context (Feng et al. 2010; Zemach et al. 2010). Accordingly, methylation in plants involves specific DNMTs, some of which are unique to plants. In mammals, three types of DNMT families are known: DNMT1, which was shown to control maintenance methylation at symmetric CG sites (Jurkowska et al. 2011), DNMT3, which acts in *de-novo* DNA methylation at CG sites (Jurkowska et al. 2011) and DNMT2, which so far was not found to take part in DNA methylation in both animals and plants (Schaefer and Lyko 2010). In *Arabidopsis*, DNA METHYLTRANSFERASE 1 (MET1), a homolog of DNMT1, maintains methylation of CG sites, however methylation of CHG and CHH sites is also affected in the *Atmet1* mutant (Cokus et al. 2008; Stroud et al. 2013) indicating that it may take part in these context as well. *Arabidopsis* DOMAINS REARRANGED METHYLTRANSFERASE 1 (AtDRM1), AtDRM2 and AtDRM3, homologs of DNMT3, are responsible for *de-novo* methylation of cytosine residues in all three sequence contexts through the RNA-directed DNA methylation pathway (RdDM) (Cao et al. 2003; Cokus et al. 2008; Henderson et al. 2010b; Stroud et al. 2013). Additionally, AtDRM1, AtDRM2 and AtDRM3 maintain CHH and CHG methylation at particular loci along the genome (Cao and Jacobsen 2002; Cokus et al. 2008; Henderson et al. 2010b; Stroud et al. 2013). Chromomethylase (CMT) represents a unique DNMT family found only in plants (Henikoff and Comai 1998). Two subfamilies, CMT2 and CMT3, were identified in flowering plants (Zemach et al. 2013). In *Arabidopsis,* AtCMT3 catalyzes methylation mainly at CHG sites and also maintains CHH methylation at some genomic loci (Cao and Jacobsen 2002; Cokus et al. 2008; Stroud et al. 2013). The CMT3 homologs, Zea methyltransferase2 (Zmet2) in *Zea mays* and NbCMT3 in *Nicotiana benthamiana,* were shown to take part in CHG methylation (Papa et al. 2001; Hou et al. 2013). Interestingly, the *Atmet1* mutant which lost methylation at almost all CG sites, showed hypermethylation of CHG sites in euchromatic regions and in gene bodies as compared to WT (Cokus et al. 2008; Stroud et al. 2013) indicating that it may affect CHG methylation as well. AtCMT2, a member of the CMT2 subfamily, catalyzes methylation mainly at CHH sites at genomic loci different from those regulated by DRMs (Zemach et al. 2013). Thus in flowering plants, each DNMT specialize in methylating particular sequence contexts, but under some conditions may act redundantly in methylating other sequence contexts as well.

Next generation bisulfite sequencing revealed that DNA methylation in flowering plants is found mostly in repetitive sequences and to some extent in euchromatic regions (Cokus et al. 2008; Lister et al. 2008; Feng et al. 2010; Zemach et al. 2010). Repetitive sequences are densely methylated in all sequence contexts although loss of CG methylation was enough to induce a massive reactivation of silenced transposons as observed in the *Arabidopsis Atmet1* mutant which may indicate a role for CG methylation in silencing transposable elements (Zhang et al. 2006). Euchromatic intergenic regions are also methylated in all three sequence contexts while genes are methylated almost exclusively at CG sites (Cokus et al. 2008; Zemach et al. 2010). Analyses of CG methylation and gene expression showed that CG hypo-methylation of transcription start and stop sites correlated with the expression of their related genes (Zemach et al. 2010). Nevertheless, in *Arabidopsis* more genes were upregulated in a *Atdrm1 Atdrm2 Atcmt3* triple mutant than in the *Atmet1* mutant (Zhang et al. 2006), indicating a role for CHG and CHH methylation in regulating gene expression as well.

CHG DNA methylation is site-specific and is regulated by additional epigenetic factors, including histone methylation, siRNA and chromatin remodeling. Genome-wide studies have shown that CHG methylation correlates with dimethylated histone tails at lysine 9 (H3K9m2) (Bernatavichute et al. 2008; Deleris et al. 2012). Moreover, a triple mutant of H3K9m2 histone methyltransferases (*kyp suvh5 suvh6*) caused loss of CHG methylation, similar to the *Atcmt3* mutant (Ebbs and Bender 2006; Stroud et al. 2013). Interestingly, some of CHG sites also lost methylation in the *Atmet1* mutant, indicating that CHG methylation also depends, in part, on CG methylation (Stroud et al. 2013). Additional studies have revealed crosstalk between CHG methylation and H3K9m2 as AtCMT3 recognizes H3K9m2 and the KRYPTONITE (KYP) an H3K9 methyltransferase recognizes CHG methylation (Feng and Jacobsen 2011). On the other hand, lack of CHG methylation at gene bodies correlates with lack of the H3K9me2 mark, which is maintained by a histone demethylase, INCREASE IN BONSAI METHYLATION 1 (IBM1) (Saze and Kakutani 2011). An additional pathway which targets DNA methylation to specific genomic regions is mediated by RdDM (Law and Jacobsen 2010). In *Arabidopsis,* genome-wide analysis of DNA methylation has shown that small RNAs direct approximately 37 % of DNA methylated regions (Ebbs and Bender 2006). DDM1 is also involved in methylation of genes as well as transposable elements probably by remodeling of histone1-bound nucleosomes and thereby

providing DNA methyltransferases access to the DNA (Lippman et al. 2004; Zemach et al. 2013).

The role of DNA methylation has been intensively studied in the flowering plant model *Arabidopsis*, but little is known about the evolution of this mechanism and the differences in epigenetic states during differentiation. We aim to understand the biological role of DNA methylation in regulating developmental processes in early terrestrial plants. To this end we chose the model bryophyte *P. patens* to reveal novel functional properties of the CMT family and to provide insight into the evolution of this regulatory mechanism. Methylation of DNA was found in all three contexts in *P. patens* and was restricted mostly to repetitive regions (Zemach et al. 2010). Additionally, *P. patens* protonema tissue treated with zebularine, a DNMT inhibitor, showed partial loss of DNA methylation and profound effects on growth and differentiation of protonema cells (Malik et al. 2012). Furthermore, *P. patens* encodes for a single CMT gene designated *PpCMT*, which was shown to be expressed in protonema and gametophore tissues and localized to the nucleus, as demonstrated by transient expression of GFP-PpCMT (Malik et al. 2012).

In this study, we deleted the gene encoding for PpCMT in *P. patens* and by employing bisulfite DNA sequencing we were able to determine in vivo that it is involved in methylation of CHG sites. This function correlates with the phylogenetic classification of PpCMT as a member of the CMT3 subfamily. Transgenic plants expressing a PpCMT–GUS fusion protein showed GUS accumulation in the nucleus, mainly in dividing cells at all developmental stages. The *ΔPpcmt* deletion mutant displayed abnormal growth of protonemata, which developed malformed gametophores lacking reproductive organs thus leading to sterility. Taken together, our results suggested a regulatory role for CHG DNA methylation during moss development.

Materials and methods

Protein alignments and construction of phylogenetic trees

AtCMT1/2/3 protein sequences were obtained from the TAIR website (www.tair.org). The PpCMT sequence was obtained from www.cosmoss.org, version 1.6 (Zimmer et al. 2013). Alignment of *Arabidopsis* and *P. patens* CMT protein family homologs was constructed using Muscle (Edgar 2004) showing identity and similarity shading based on the BLOSUM62 matrix (Henikoff and Henikoff 1992) at a 75 % cut off (Vector NTI software suite, Invitrogen, USA). Predicted protein sequences from other plants were obtained from Phytozome v8.0 (http://www.phytozome.net) (Goodstein et al. 2012) via BLAST using AtCMT3 protein sequence as a reference (Table S1). Sequences were aligned using either Muscle (Edgar 2004), Prank (Loytynoja and Goldman 2010) or Mafft (Katoh et al. 2009). The alignments were cured using Gblocks (Castresana 2000) and the trees were constructed by PhyML (Guindon and Gascuel 2003) and visualized by TreeDyn (Chevenet et al. 2006) using the www.phylogeny.fr website (Dereeper et al. 2008). Different alignment algorithms resulted in the same tree topology.

Plant material, culture conditions and treatments

The 'Gransden 2004' strain of *P. patens* (Ashton and Cove 1977; Rensing et al. 2008) was propagated on BCD and BCDAT media (Nishiyama et al. 2000) at 25 °C under a 16-h light and 8-h dark cycle (Frank et al. 2005). For induction of gametangia the plants were transferred to 16 °C under cycles of 8 h light and 16 h of dark. Germination of spores was carried as described by (Sakakibara et al. 2001). For unidirectional growth, plants were grown in a box with one side either opened or covered with a 3-mm-thick red acrylic transparent plastic sheet to allow unidirectional white or red light, respectively.

Statistical analysis

Wilcoxon signed-rank test at a 95 % confidence level was used to compare WT and mutant data. Two-sided *P* values of <0.05 were considered to be statistically significant. Data was analyzed using R software (v. 3.0.2, http://www.R-project.org/) and boxplot graphs were built using ggplot2 R package.

Generation of PpCMT deletion lines

ΔPpcmt deletion mutant lines were constructed by replacing *PpCMT* genomic coding region between position 640–3,700 bp (relative to the translation start site), which encode the BAH and choromo domains, with a hygromycin resistance cassette (*hptII*). Genomic fragments containing the 5′ and 3′ flanking regions (Fig. S1b) were amplified using the following primers: CMT 5′ Fw and CMT 5′ RV Rev, CMT 3′ RV Fw and CMT 3′ Rv (Table S2), then cloned into pTZ57 (Fermentas, Lithuania) and sequenced to ensure their integrity. Next, the hygromycin resistance cassette (*hptII*) from pMHubi (Bezanilla et al. 2003) and the 3′ CMT KO fragment were cloned successively into pTZ57 + 5′ CMT KO resulting in the final *ΔPpcmt* deletion vector. Following transformation to protoplasts, hygromycin resistant lines were screened by PCR to verify correct integration, by amplifying the junction regions between the insert and *PpCMT* locus at both the 5′ and 3′ ends (Fig. S1). Transgene copy number was determined by real-time PCR (Fig. S2) and loss of *PpCMT* coding regions

was confirmed by RT-PCR analysis in six independent mutant deletion lines (Fig. S1).

Generation of PpCMT–GUS fusion reporter plants

For the construction of the C-terminal translational fusion of *PpCMT* with the *uidA* (*GUS*) gene (Jefferson et al. 1987) at the endogenous genomic context, the *uidA* gene was fused in-frame at the carboxyl terminus of *PpCMT* coding sequence replacing the stop codon (Fig. S3). To this end, the 5′ flanking region excluding the stop codon was amplified using the following primers: CMT GUS HindIII 5′ Fw and CMT 5′ no stop Rv and the 3′ flanking region was amplified using the following primers: CMT GUS 3′ sph FW and CMT GUS 3′ sph Rv (Table S2). The PpCMT–GUS fusion vector contained the 5′ *PpCMT* genomic fragment ligated in-frame to the *uidA* reporter gene, followed by a nopaline synthase polyadenylation signal (NOS-ter), *hptII* cassette and by the 3′ *PpCMT* genomic fragment in pMBL5 vector (GenBank: DQ228130.1). Following transformation to protoplasts, hygromycin resistant lines were screened by PCR to verify correct integration, by amplifying the junction regions between the insert and the PpCMT locus at both the 5′ and 3′ ends and were further analyzed for transgene copy number by real-time PCR (Fig. S4).

PEG-mediated transformation of *P. patens*

PEG transformation was performed as described (Nishiyama et al. 2000) using 15 μg of linearized plasmid. Six days after regeneration, transformants were selected on BCDAT medium containing 25 mg/l of Hygromycin (Sigma-Aldrich, USA).

DNA extraction

The genomic DNA samples used for the quantitative qPCR analyses and bisulfite sequencing were extracted from 100 mg of 7 days old protonema, grown on BCD medium and purified using a DNeasy plant mini kit (Qiagen, Germany) following the manufacturer's instructions. DNA integrity was analyzed by 1 % agarose gel electrophoresis in 0.5 × TBE with ethidium bromide staining from which 0.5 μl were used per qPCR reaction and 10 μl for bisulfite reaction.

RNA extraction and RT-PCR

Total RNA was extracted using SV Total RNA Isolation System (Promega, USA) from 7 days old protonema grown on BCD medium. The cDNA was synthesized from 2.0 μg of total RNA with the SuperScript® III First-Strand Synthesis System (Life Technologies, USA) using the oligo (dT) 20 primer followed by PCR as described (Katz et al. 2004), using the following gene-specific primers: CMT EXON 5 Fw and CMT EXON 7 Rv (Table S2). Cycling conditions: 94 °C for 2 min, and 40 cycles of 94 °C for 15 s, 57 °C for 15 s and 72 °C for 30 s.

Real-time RT-PCR analysis

Primers were designed with the Probe Finder version 2.45 (Roche Applied Science, Germany). All primers were synthesized by Hylabs (Israel). Real-time quantitative PCR analysis was performed by the ΔΔCt method of relative quantification with a StepOne™ Thermal Cycler (Life Technologies), using Fast SYBR Green Master Mix (Life Technologies) to monitor dsDNA synthesis. Cycling conditions: 95 °C for 20 s, and 40 cycles of 95 °C for 1 s and 60 °C for 20 s. The following primers were used to detect expression of R1 and R2: R1- Sca 256 RT Fw, Sca 256 RT Rv, Sca 256 RTM Fw and Sca 256 RTM Rv; R2- Sca 332 RT Fw and Sca 332 RT Rv (Table S2). The following primers were used to amplify normalizing genes: PpHistone- PpHistone3 157 Fw, PpHistone3 211 Rv, PpHistone3 100 Fw and PpHistone3 143 Rv; PpTATA- TATA binding p Left and TATA binding p Right; PpFie- PpFIE -245 Left and PpFIE -213 Right (Table S2). Real-time amplification data was statistically analyzed by StepOne™ software v2.1 using the manufacture default parameters (Life Technologies).

Methylation sensitive DNA restriction followed by qPCR

Genomic DNA was subjected to overnight digestion using methylation-sensitive DNA restriction enzymes *HpaII* or *MspI* (Fermentas). The restriction mix contained 1.5 μl of genomic DNA, 1.5 μl of Tango buffer and 0.8 μl of restriction enzyme and 11.2 μl ddH_20. 1 μl of each restricted DNA from WT or *ΔPpcmt* mutant was used as a template for quantitative real-time PCR (Metivier et al. 2008) using the following primers Sca 256 RTM Fw and Sca 256 RTM Rv which enclose the CCGG site. Uncut DNA served as a reference for each background. PpHistone3 gene was used to normalize the template DNA quantity (using PpHistone3 157 Fw and PpHistone3 211 Rv primers). Amplification of a nearby flanking region which is not affected by restriction cleavage, served as an internal control to quantify the template (using Sca 256 RT Fw and Sca 256 RT Rv primers).

Real time qPCR assay for quantification of transgene copy number

Transgene copy number was determined using quantitative PCR amplifying two regions, 5′ genomic region and 3′ genomic region that were cloned into the vector to serve

as sequence templates for homologous recombination. Following homologous recombination these regions replace the corresponding sequence in the genomic loci (Fig. S3a). If the recombination took place only at the homologous site and not in a tandem fashion (Kamisugi et al. 2006), it should be present in the transgenic line in a single copy as in the wild type. Therefore, the wild type genome was used as a reference for monitoring a single copy of the desired locus. In order to normalize the initial amount of template DNA, any gene can be used. In our study we used either primers for *PpCLF* (PpCLF mix 7 5915 Fw and PpCLF mix 7 5981 Rv primers) or *PpFIE* (PpFIE -245 Left and PpFIE -213 Right primers). As an internal control for quantification, an additional set of primers (PpCLF mix 8 7739 Fw and PpCLF mix 8 7804 Rv primers) amplifying a different region of the *PpCLF* gene was used. This control would be expected to give the same relative ratio for a single copy sequence.

A second approach for relative quantitation of transgene copy number was applied by amplifying the hygromycin resistant gene *hptII* (using Hyg 687 Left and Hyg 730 Right primers). In this case, we used as a reference the transgenic line *PpCMT–GUS2* which was shown to have a single copy of the hygromycin resistance cassette (Fig. S4).

In order to amplify the 5′ genomic region in *PpCMT–GUS* lines the CMT 5′ GUS Left and CMT 5′ GUS Right primers were used; for 3′ region the CMT 3′ GUS Left and CMT 3′ GUS Right were used. To amplify the 5′ region in deletion mutants, the CMT 3′ KO 43 Fw and CMT 3′ KO 105 Rv pair was used.

GUS assay

The histochemical assay for GUS activity was performed as described by (Nishiyama et al. 2000). The incubation time was adjusted from 2 to 24 h, depending on the tissues examined.

Hoechst staining

Protonema tissue was incubated for 10 min in 5 μg/ml solution of Hoechst33342 in BCD medium, supplemented with 0.05 % TritionX-100.

Bisulfite assay

Bisulfite conversion was performed using EZ DNA Methylation-Gold™ kit (Zymo Research, USA) according to the manufacturer's instructions. 2 μl of bisulfite converted DNA was used as template to amplify the analyzed loci (R1–R4) by using sets of specific primers flanking the loci analyzed for methylation status. The primers were designed to bind to regions that do not contain cytosines in the sense strand without preference for methylated or unmethylated DNA as described in (Henderson et al. 2010a): R1-Sca. 256 509,582 Fw and Sca. 256 510,013 Rv, R2-Sca. 332 15,158 Fw and Sca. 332 15,602 Rv, R3-Sca. 15 Fw and Sca. 15 Rv and R4-Sca. 60 Fw and Sca. 60 Rv (Table S2). PCR products were amplified using TaKaRa Ex-Taq DNA polymerase (Takara bio, Japan) and cloned into pTZ vector (Fermentas) following the manufacturer's instructions, and 5–15 independent clones were sequenced and analyzed for each genetic background as compared to WT (Table S3).

Microscopy

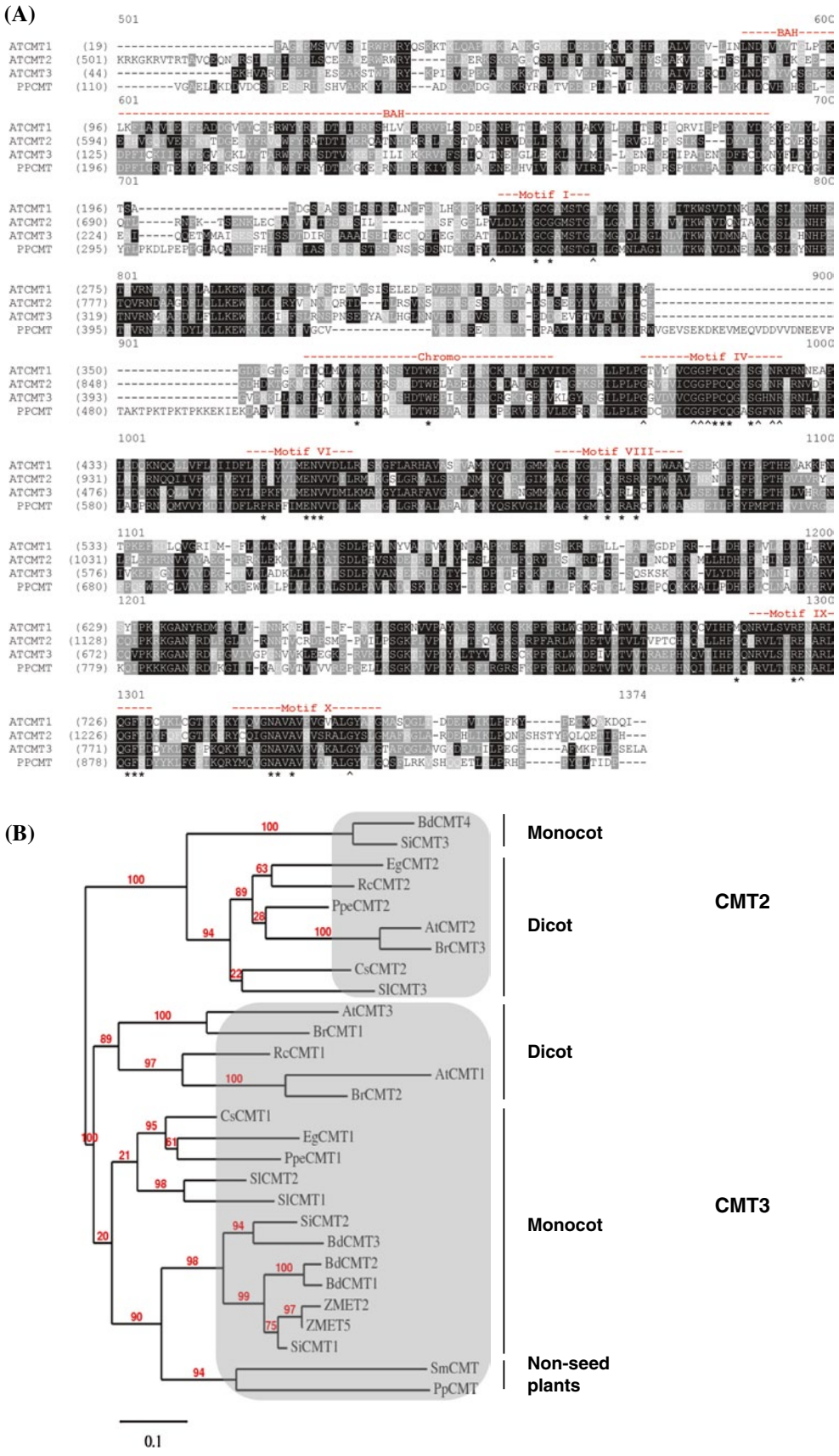
Mutant and WT morphology along their development and histochemical staining of GUS were detected using stereomicroscope Stemi SV11 Apo (Carl Zeiss AG, Germany), SZRX-ILLB2-200 (Olympus, Japan) and Axioplan2 (Zeiss) microscope equipped with Olympus DP71 and Coolpix P5100 cameras (Nikon, Japan). Hoechst staining was visualized using Axioplan2 epifluorescence microscope.

Results

Phylogenetic analysis of CMT protein family

The CMT DNA methyltrasferase protein family is found only in plants (Goll and Bestor 2005). This family is characterized by the presence of a chromo domain located within the C-5 DNA methyltrasferase domain (MTD) and by a C-terminal regulatory bromo-adjacent homology domain (BAH) (Henikoff and Comai 1998). Alignment of *P.patens* and *Arabidopsis* CMT proteins revealed that all of the domains which characterize the CMT family show a high degree of sequence conservation, particularly in residues known to take part in the enzymatic activity (indicated by asterisks in Fig. 1a) as described by (Goll and Bestor 2005; Pavlopoulou and Kossida 2007; Malik et al. 2012). Using AtCMT3 as a query sequence in the NCBI protein database we could identify CMT homologs only in land plants. Previously, three predicted algal DNMT proteins were defined as CMT proteins (Feng et al. 2010; Zemach et al. 2010). However, a multiple sequence alignment suggested that all three algal proteins differ from PpCMT and AtCMT3 and lack the chromodomain and some conserved residues within the MTD (Fig. S5). In agreement, analysis of the three algal proteins using the conserved domain database algorithm (Marchler-Bauer et al. 2011) did not detect the distinguishing chromo domain. Hence, we suggest that the reported algal proteins are not related to CMT family and that the CMT protein family is unique to land plants.

To understand the evolutionary relationships between PpCMT and the CMT protein family, candidate members


(A)
ATCMT1
ATCMT2
ATCMT3
PPCMT
BAH
Motif I
Chromo
Motif IV
Motif VI
Motif VIII
Motif IX
Motif X
(B)
BdCMT4
SiCMT3
EgCMT2
RcCMT2
PpeCMT2
AtCMT2
BrCMT3
CsCMT2
SlCMT3
AtCMT3
BrCMT1
RcCMT1
AtCMT1
BrCMT2
CsCMT1
EgCMT1
PpeCMT1
SlCMT2
SlCMT1
SiCMT2
BdCMT3
BdCMT2
BdCMT1
ZMET2
ZMET5
SiCMT1
SmCMT
PpCMT
Monocot
Dicot
Dicot
Monocot
Non-seed plants
CMT2
CMT3
0.1

◂**Fig. 1** Phylogenetic analysis of PpCMT. **a** Sequence alignment of *Arabidopsis* and *P. patens* CMT protein family homologues using Muscle (Edgar 2004) showing identity and similarity indicated by *shading* based on the BLOSUM62 matrix (Henikoff and Henikoff 1992) at a 75 % cut off. Identical residues are *shaded in black* and similar residues are *shaded in grey*. *Dashed line* denotes conserved domains. *Asterisks* indicates conserved amino acids among eukaryotic DNMTs. *Hat* indicates conserved amino acids among plant DNMTs. **b** CMT protein sequences from a subset of plants were used to construct a phylogenetic tree by PhyML (Guindon and Gascuel 2003), branch support values are indicated. Protein accessions are detailed in Table S1. Organism abbreviations: Pp, *Physcomitrella patens;* Sm, *Selaginella moellendorffii;* Bd, *Brachypodium distachyon;* Si, *Setaria italica;* Z, *Zea mays;* Sl, *Solanum lycopersicum;* At, *Arabidopsis thaliana;* Br, *Brassica rapa;* Eg, *Eucalyptus grandis;* Rc, *Ricinus communis;* Cs, *Cucumis sativus;* Ppe, *Prunus persica*

were selected from fully sequenced genomes representing the main plant phyla (Table S1; Fig. 1b). Using the selected proteins we constructed a phylogenetic tree which clustered into two clades corresponding to the CMT2 and CMT3 subfamilies (Fig. 1b) previously described in angiosperms (Zemach et al. 2013). The early land plants representatives, *P. patens* and *Selaginella moellendorffii*, each encode for a single CMT which belongs to the CMT3 subfamily (Fig. 1b). Tree topology indicated that the CMT family has evolved from a single CMT3-type gene in early land plants and diverged into two subfamilies in flowering plants. Thus, *P. patens* may serve as a model to study the ancient functions of the CMT3 subfamily.

PpCMT is involved in CHG DNA methylation

In order to study PpCMT function, we generated deletion mutant plants in which the DNA sequence encoding for BAH and chromo domains was deleted (Fig. S1). To determine the contribution of PpCMT to DNA methylation, we scanned the published *P. patens* methylome (Zemach et al. 2010) in 300 bp windows in order to identify non-repetitive sequences with a high level of DNA methylation which allows to detect loss of DNA methylation. Four genomic regions designated R1 to R4 (Table S3) were chosen. R1 and R2 loci are enriched in methylated CG and CHG sites and located within predicted ORFs which were not reported to be expressed in any tissue (Zimmer et al. 2013). R3 and R4 loci are enriched for methylated CHH sites and located in intragenic regions. Bisulfite analysis of methylation levels from protonema tissue of WT and *ΔPpcmt* mutant (lines *ΔPpcmt103*, *ΔPpcmt157* and *ΔPpcmt281*) lacking the BAH and chromo domains revealed no significant changes in CG methylation at R1 and R2 loci but an almost complete loss of CHG methylation in the *ΔPpcmt* mutant (Fig. 2). In contrast, R3 and R4 loci which are enriched for CHH methylation in WT, showed similar methylation levels in *ΔPpcmt* (Fig. 3). This indicates that PpCMT is involved in CHG methylation at the loci tested. To note, both CHG sites at R3 and the single CHG site in R4 loci showed complete loss of DNA methylation in *ΔPpcmt*, while both CG sites present at R4 retained methylation as in WT.

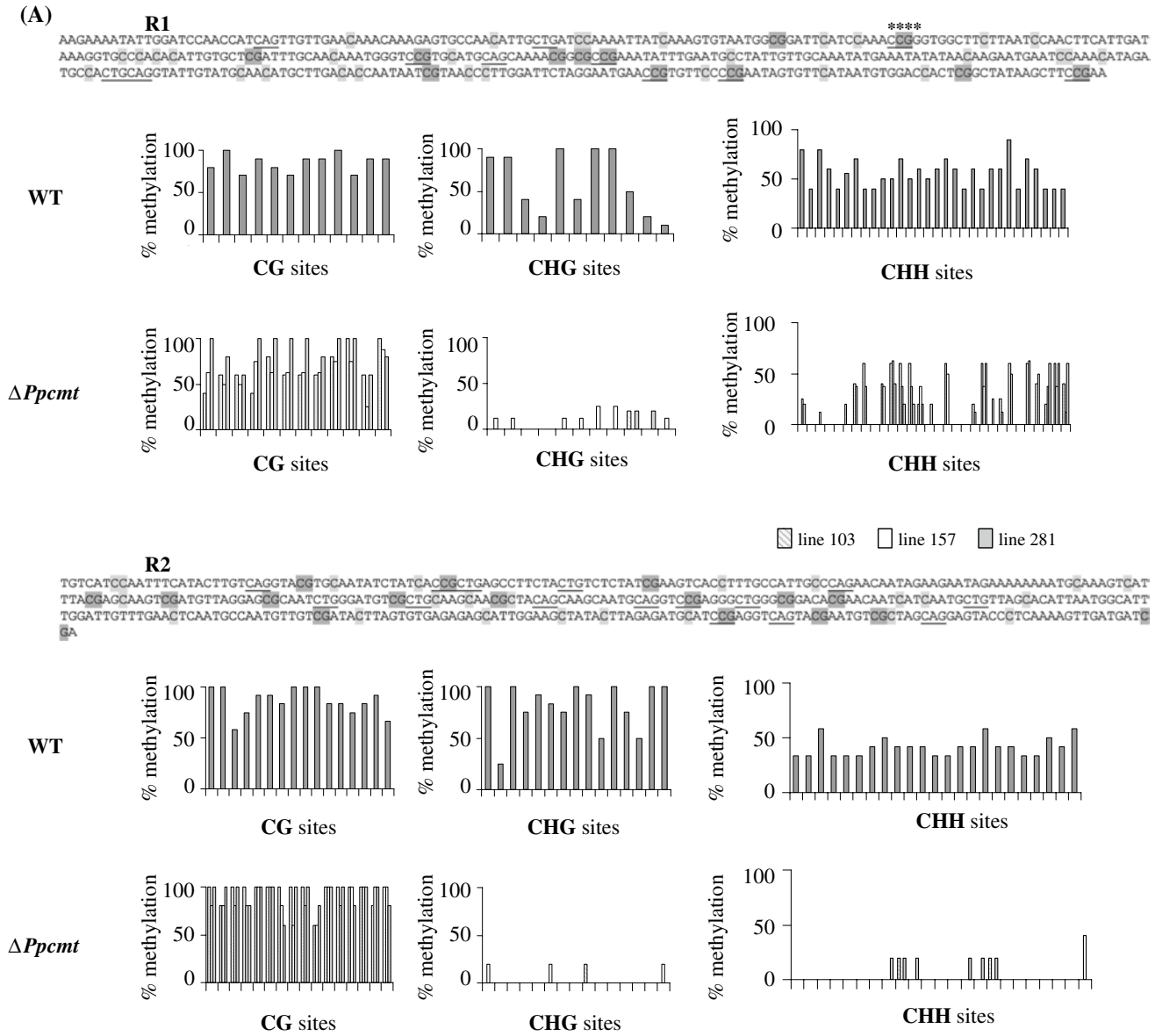
To corroborate these results, the methylation status was evaluated using methylation sensitive DNA restriction assay which measured DNA methylation at one specific genomic site within the R1 loci, in many molecules (Table S3 depicts the location of the site within the genome, asterisks indicate the position of the site in Fig. 2a). To this end, genomic DNA from *ΔPpcmt* and WT were digested with either *HpaII* or *MspI* followed by quantitative real-time PCR using primers flanking a CCGG site in the R1 loci (Fig. 2b). Both enzymes recognize the same CCGG sequence, however *MspI* is unable to digest when the first cytosine is methylated, making it sensitive to CHG methylation. *HpaII* is unable to restrict when the second cytosine is methylated, making it sensitive to CG methylation. While in WT approximately half of the molecules were protected from *MspI* restriction, almost all molecules were digested in *ΔPpcmt* mutant background (Fig. 2b), indicating tissue-wide loss of CHG methylation at this site. Restriction with *HpaII* showed no significant change between WT and *ΔPpcmt* plants, indicating that CG methylation levels at this site were not impaired by the absence of PpCMT (Fig. 2b).

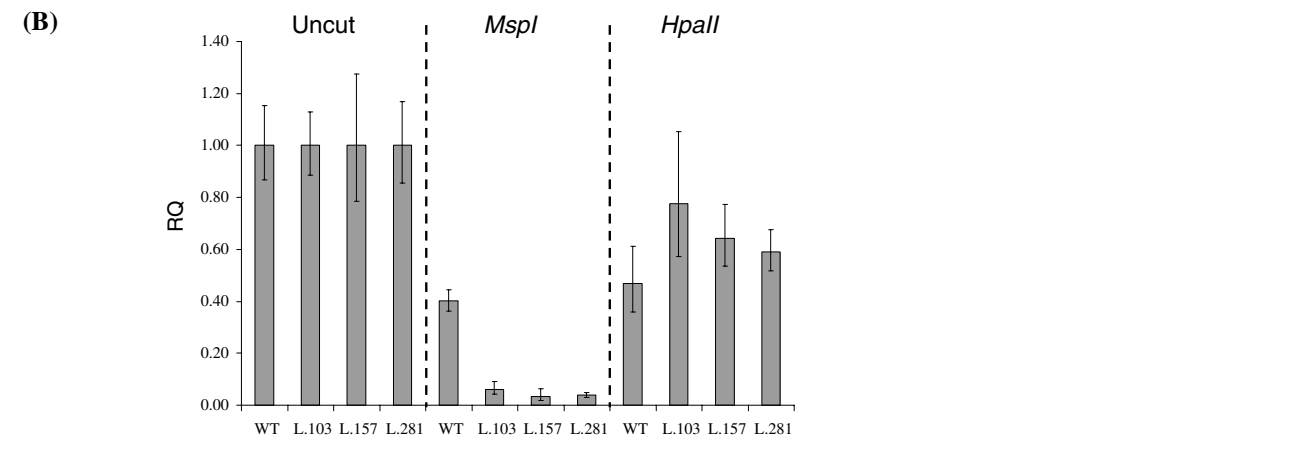
As R1 and R2 loci lie within putative ORFs, we tested the effect of CHG methylation loss on their expression. Using quantitative RT-PCR as well as semi-quantitative PCR, the expression levels of both ORFs in WT and mutant were found to be below detection threshold.

To test whether loss of CHG methylation is associated with loss of CHH methylation, the levels of CHH methylation at the R1 and R2 loci was determined. While almost all CHG sites were methylated at R1 and R2 loci (Fig. 2), approximately half of the CHH sites were methylated above 30 percent in WT (30 out of 61 sites at R1 loci and 23 out of 51 sites at R2 loci) (Fig. 2). Comparison of CHH methylation levels at the methylated sites revealed a decrease in CHH methylation in *ΔPpcmt*, namely 22 % at R1 and 3 % at R2 in *ΔPpcmt* as compared to 56 and 41 % in WT, respectively (Fig. 2). Thus, loss of CHG methylation is associated with partial loss of CHH methylation in the tested loci enriched for CHG methylation but not in CHH methylated enriched loci.

PpCMT accumulates in dividing cells

Translational fusion of PpCMT with the *uidA* (*GUS*) (Jefferson et al. 1987) reporter gene was used to determine the protein accumulation pattern throughout the *P.patens* life cycle (Cove 2005; Frank et al. 2005). In these plants,


(A)
R1
****
WT
ΔPpcmt
% methylation
CG sites
CHG sites
CHH sites
line 103
line 157
line 281
R2


(B)
Uncut
MspI
HpaII
RQ
WT
L.103
L.157
L.281

◂**Fig. 2** Comparison of DNA methylation levels in WT versus *ΔPpcmt* deletion lines at the R1 and R2 loci. **a** Methylation levels at R1 and R2 loci were analyzed in WT and three independent *ΔPpcmt* deletion *lines* (lines 103, 157 and 281). CG sites are marked by *dark grey* and CHG sites are *underlined*. Cytosines of CHH sites which were methylated over 30 % in WT are marked by *light grey* and are presented in the *bar chart*. *Bar charts* are separated into CG, CHG or CHH-context groups for each locus. *Bar charts* indicate the methylation levels for individual C residues (*X-axis*) in the order they appear in the sequence, as compared to the methylation potential considered as 100 % (if all reads were methylated). Methylation levels were lost at CHG sites in *ΔPpcmt* and partially in CHH sites but unaffected at CG sites. **b** Quantification of DNA methylation levels at a specific site within R1 loci (indicated in the R1 sequence by *asterisks*) was obtained by methylation sensitive restriction assay using *MspI* (sensitive to CHG methylation) or *HpaII* (sensitive to CG methylation), followed by quantitative PCR. Relative quantification (RQ) of WT and three *ΔPpcmt* mutant lines are presented where uncut DNA served as a reference in each genetic background. The results were normalized to the *PpHistone3* gene. The errors bars indicate standard deviation of three technical replicates. Restriction with *MspI* indicated that about half of the DNA molecules were CHG methylated in WT which was lost in all three *ΔPpcmt* mutant lines tested. While half of the CHG sites were protected by methylation from *MspI* in WT, almost all CHG sites were sensitive to *MspI* in all three *ΔPpcmt* mutant lines

PpCMT–GUS is expressed under the control of its native promoter within the endogenous genomic context. Three independent transgenic lines were generated and their development was no different than WT. Protein accumulation pattern was indistinguishable in all three lines. Mature spores did not show PpCMT accumulation (Fig. 4a) however, in germinating spores GUS staining was localized to the three dividing nuclei, in agreement with its nuclear function (Fig. 4b). GUS accumulation was detected during the filamentous stage in the nuclei of apical and sub-apical cells (Fig. 4c) and in the nuclei of filament cells, which divide to form either side branches or buds (Fig. 4d). During bud formation accumulation was observed in nuclei of all bud cells (Fig. 4e, f) and as the bud developed (Fig. 4g–i) to form a gametophore the staining became limited to the dividing apex (Fig. 4i). GUS accumulation was detected at the base of young leaves (Fig. 4j), but not in mature leaves (Fig. 4k). Lateral gametophores showed similar accumulation pattern restricted to the apex (Fig. 4i, insertion). GUS accumulation was also observed in apical cells of rhizoids but gradually decayed as the cells matured (Fig. 4l, m).

The accumulation pattern of PpCMT in dividing cells was also observed upon transition to the reproductive phase. GUS accumulation was detected during antheridia development in the dividing spermatozoids (Fig. 5a), whereas no accumulation was detected upon antheridia maturation (Fig. 5b). During archegonia development, GUS accumulation was detected at early stages and during gametogenesis in the young egg cell and canal cells (Fig. 5c–e). GUS accumulation was absent from the mature archegonium and the egg cell before fertilization (Fig. 5f, g), as well as after fertilization (Fig. 5h) as staged by (Landberg et al. 2013).

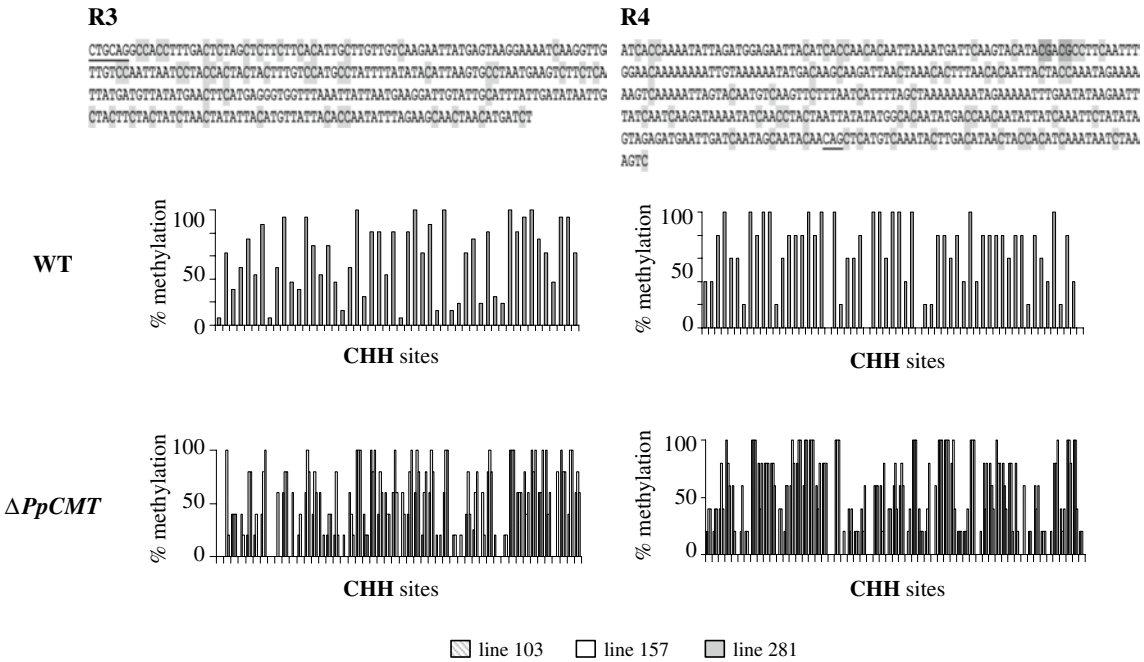


Fig. 3 Comparison of CHH DNA methylation levels in WT versus *ΔPpcmt* deletion lines at the R3 and R4 loci. Cytosines within the CHH context are marked by *light grey*, CG sites are marked by *dark grey* and CHG sites are *underlined*. *Bar charts* indicate the methylation levels for individual C residues at each site as compared to the methylation potential considered as 100 % (if all reads were methylated). CHH methylation levels were similar to WT in all three independent *ΔPpcmt* lines (103, 157 and 281)

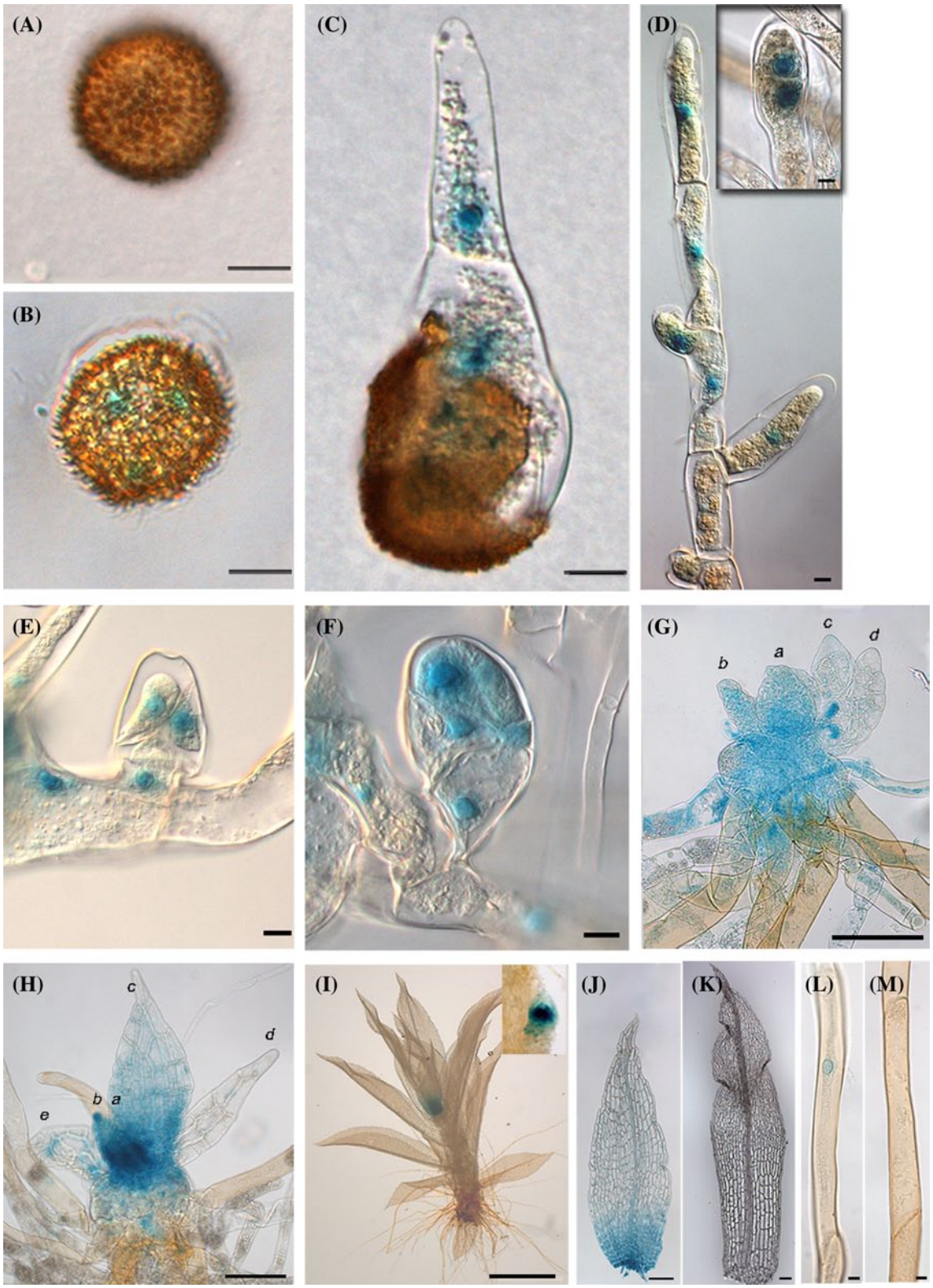


Fig. 4 Accumulation pattern of PpCMT–GUS fusion protein during the vegetative phase. **a** Mature spore. **b** Germinating spore. **c**, **d** Protonema filaments. Insertion in (**d**) shows a dividing apical cell. **e**, **f** buds. **g**, **h** Young gametophores. In (**g**) the gametophore was spread in order to visualize the young leaves. Small letters designate leaves on the same apex in the order of leaf maturation (where "**a**" designates the youngest leaf and "**d**" designates the oldest leaf). **i** Mature gametophore. Insertion shows lateral gametophore. **j** and **k** Young and mature leaves. **l** and **m** Developing and differentiated rhizoids cells. *Scale bars* **a**–**f**, **l**, **m** 10 μm; **g**, **h**, **j**, **k** 100 μm; **i** 2,000 μm. In small dividing cells the strong GUS staining seems to leak to the cytoplasm

Sporophyte development begins when the diploid zygote and the haploid archegonium divide to produce an embryo enclosed in the epigonium (Sakakibara et al. 2008). GUS staining appeared in the dividing epigonium, starting at the base, (Fig. 5h) and continued during embryo development (Fig. 5i, j). GUS staining within the epigonium gradually decayed when the divisions ceased and the elongation of the seta caused the epigonium to rupture into two parts, an upper calyptra and lower vaginula, which supports the developing embryo (Fig. 5k–m). GUS accumulation could be detected throughout all stages of embryo development until the sporangium was formed (Fig. 5p–v), while staining disappeared after sporogenesis (Fig. 5w). During early stages of embryo development, GUS accumulation was detected throughout the young embryo (Fig. 5p–r). Later, the staining disappeared from the no-longer dividing apical region (Sakakibara et al. 2008) and became gradually restricted to two regions: the center of the developing sporangium and in the developing foot (Fig. 5s–u). In later stages of development, accumulation was detected in the foot, endothecium and sporogenous tissues (Fig. 5v), as well as in the tetrads formed following meiosis (Fig. 5v insertion). After spore maturation, GUS accumulation was observed in the foot but not in the sporangium (Fig. 5w). GUS staining was not detected in the mature sporophyte or spores (Figs. 4a, 5x, y).

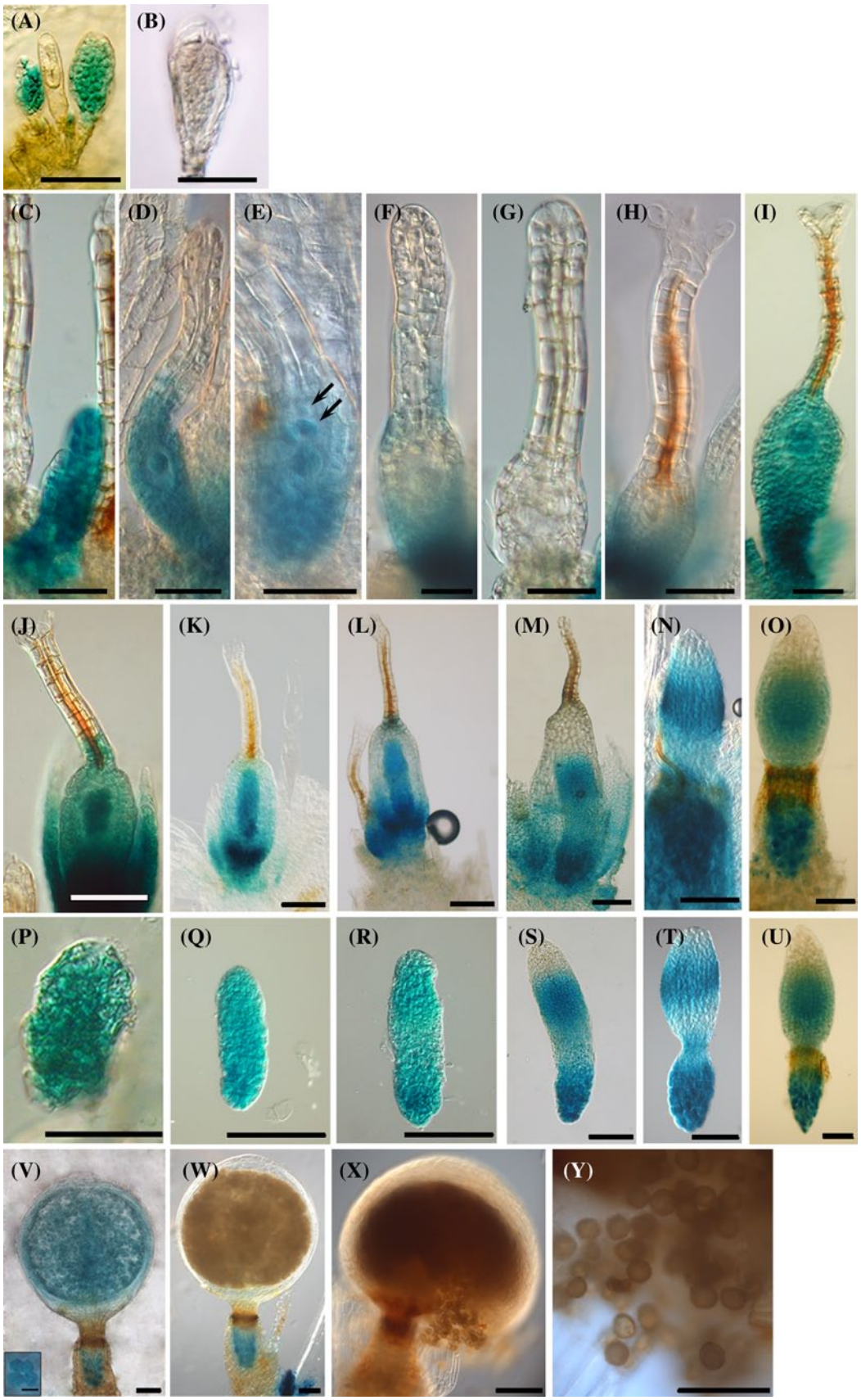
The above results indicate that PpCMT–GUS fusion protein is present mainly in dividing cells and is lacking from fully differentiated cells, which is consistent with its proposed function to maintain DNA methylation patterns. To further examine the correlation between cell divisions and PpCMT presence, we took advantage of the unique feature of moss to regenerate, which allows differentiated leaf cells to reactivate the cell cycle, divide and give rise to chloronema filaments, following tissue damage (Ishikawa et al. 2011). The regeneration was induced by detaching leaves from mature gametophores of *PpCMT–GUS* transgenic plants and placing them on BCD media for increasing periods of time up to 88 h and then staining for GUS (Fig. 6). Up to 24 h after detachment, no GUS staining could be detected (Fig. 6a, b). At 48 h after induction, PpCMT–GUS accumulation was observed in the nucleus of single cells of the detached leaves (Fig. 6c). GUS staining was observed in dividing cells containing two nuclei before cytokinesis (Fig. 6d-single arrow) and after cytokinesis (Fig. 6d-two arrows). Some of the cells divided to form protonema filaments (Fig. 6e, f). At Fig. 6e and its schematic presentation Fig. 6f GUS staining is detected in the nuclei of a leaf cell that divided once (Fig. 6f, arrow a) and gave rise to the first filament cell, which expressed CMT–GUS as well (Fig. 6f, arrow b). Following several protonema apical cell divisions, the leaf cell which gave rise to the protonema filament did not show PpCMT–GUS (Fig. 6f, arrow c), indicating that during the regeneration process PpCMT expression is restricted to dividing cells. *ΔPpcmt* mutant leaves were able to regenerate protonema (data not shown), indicating that PpCMT is not vital for the regeneration process.

PpCMT is essential for *P. patens* development

To study the role of PpCMT during *P. patens* development, morphological analysis of six independent *ΔPpcmt* deletion mutant lines (Fig. S1, S2) was conducted throughout their lifecycle as compared to WT. All six mutant lines were similar in growth and morphology. *ΔPpcmt* developed plants reduced in size without foraging filaments (Fig. 7a vs. b). *ΔPpcmt* mutant grown towards unidirectional white light developed protonema filaments which barely produced secondary filaments (Fig. 7c vs. d and Fig. S6a). *ΔPpcmt* grown towards unidirectional red light, exhibited cell division rates of about five times slower than WT (Fig. S6b) producing cells that were about 1.5 times longer (Fig. S6c) with abnormal morphology (Fig. S7). When tetrahedral shaped buds were formed in *ΔPpcmt*, the cells had a distinct globular morphology and did not elongate as in WT (Fig. 7g, h vs. m, n). Almost all buds failed to develop leaf initials and continued to divide resulting in globular architectures (Fig. 7i, j vs. o, p). However, few buds did escape this fate and developed leaf initials (Fig. 7q–s). Some of these escapees (which could reach up to 10 % of buds) continued to develop into stunted gametophores (Fig. 7t vs. u), which had very small, malformed leaves and at times contained apoptotic patches (Fig. 7v vs. w). Upon transition to the reproductive phase these developing stunted gametophores did not produce gametangia (Fig. 7x vs. y). In one case, a single abnormal archegonium formed at the apex (Fig. 7z). In accordance, no sporophytes were observed even after 1 year of growth under gametangia inductive conditions.

Discussion

Physcomitrella patens is a model for studying epigenetic processes during the haploid gametophytic generation as well as the transition to the sporophytic phase, and serves as a representative for early terrestrial plants to study the evolution of developmental processes (Mosquna et al. 2009; Okano et al. 2009). Our phylogenetic analysis indicated *P. patens* as the earliest diverged plant in which a *CMT* gene can be identified. Previously, it was argued that CMT is present earlier in evolution, in green algae. It was claimed, that the green alga *Volvox carteri* posses a CMT homolog however no CHG methylation was detected in its genome (Zemach et al. 2010). Low levels of


(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)
(J)
(K)
(L)
(M)
(N)
(O)
(P)
(Q)
(R)
(S)
(T)
(U)
(V)
(W)
(X)
(Y)

◂**Fig. 5** Accumulation pattern of PpCMT–GUS fusion protein during the reproductive phase. **a** and **b** Young and mature antheridia, respectively. **c–f** Developmental stages of young archegonia. Arrows indicate canal cells. **g** Mature archegonia before fertilization. **h** Archegonia after fertilization. **i** Early stages of epigonium development. **j–o** and **p–u** Epigonium development and corresponding isolated embryos. **v** and **w** Late stages of sporophyte maturation. Insertion in (**v**) is magnification of a tetrad after meiosis. **x** and **y** Mature sporophyte and spores. *Scale bars* **v** *insertion* 10 μm; **a–d**, **f–i** 50 μm; **j–y** 100 μm; **e** 200 μm

CHG methylation were found in two species, *Chlorella* sp. *NC64A* and *Chlamydomonas reinhardtii* (2.2 and 2.6 %, respectively) (Feng et al. 2010; Zemach et al. 2010). However, the two studied algal putative CMT proteins lack the characteristic CMT family chromo domain (Fig. S5). The presence of CHG methylation in a particular genome does not necessarily imply that a CMT protein catalyzed this modification, as evident from the presence of low CHG methylation (1.22 %) in *Danio rerio* (Zebrafish), which does not encode for a *CMT* (Feng et al. 2010). Furthermore, in *C. reinhardtii* CHG methylation was shown to be catalyzed by a MET1 homolog (Nishiyama et al. 2002). In view of the above, we postulate that low CHG methylation found in algae could be mediated by MET1 homologs and not by CMT.

In land plants, the emergence of the CMT protein family is in correlation with high levels of CHG DNA methylation (Feng et al. 2010; Zemach et al. 2010). In angiosperms, the CMT proteins cluster into two subfamilies: the CMT3 subfamily, which catalyze CHG DNA methylation

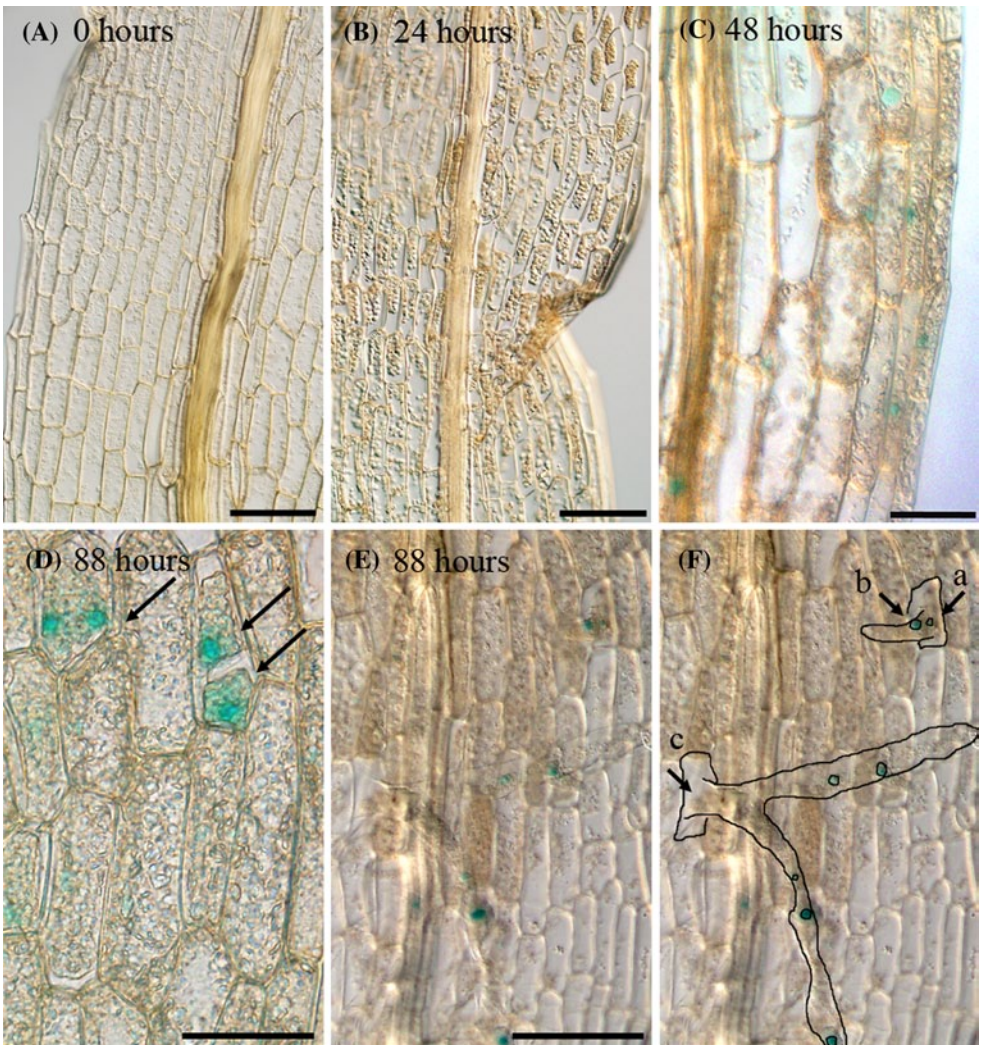


Fig. 6 Accumulation pattern of PpCMT–GUS fusion protein during dedifferentiation of leaf cells to protonema. Detached leaves were incubated on BCD agar medium for 24, 48 and 88 h. **a** and **b** 0 and 24 h after detachment, respectively. **c** 48 h after induction. **d–f** 88 h after detachment. Arrows indicate dividing nuclei. **f** Schematic drawing emphasizing the differentiated protonema filaments in **e**. *Scale bars* 100 μm

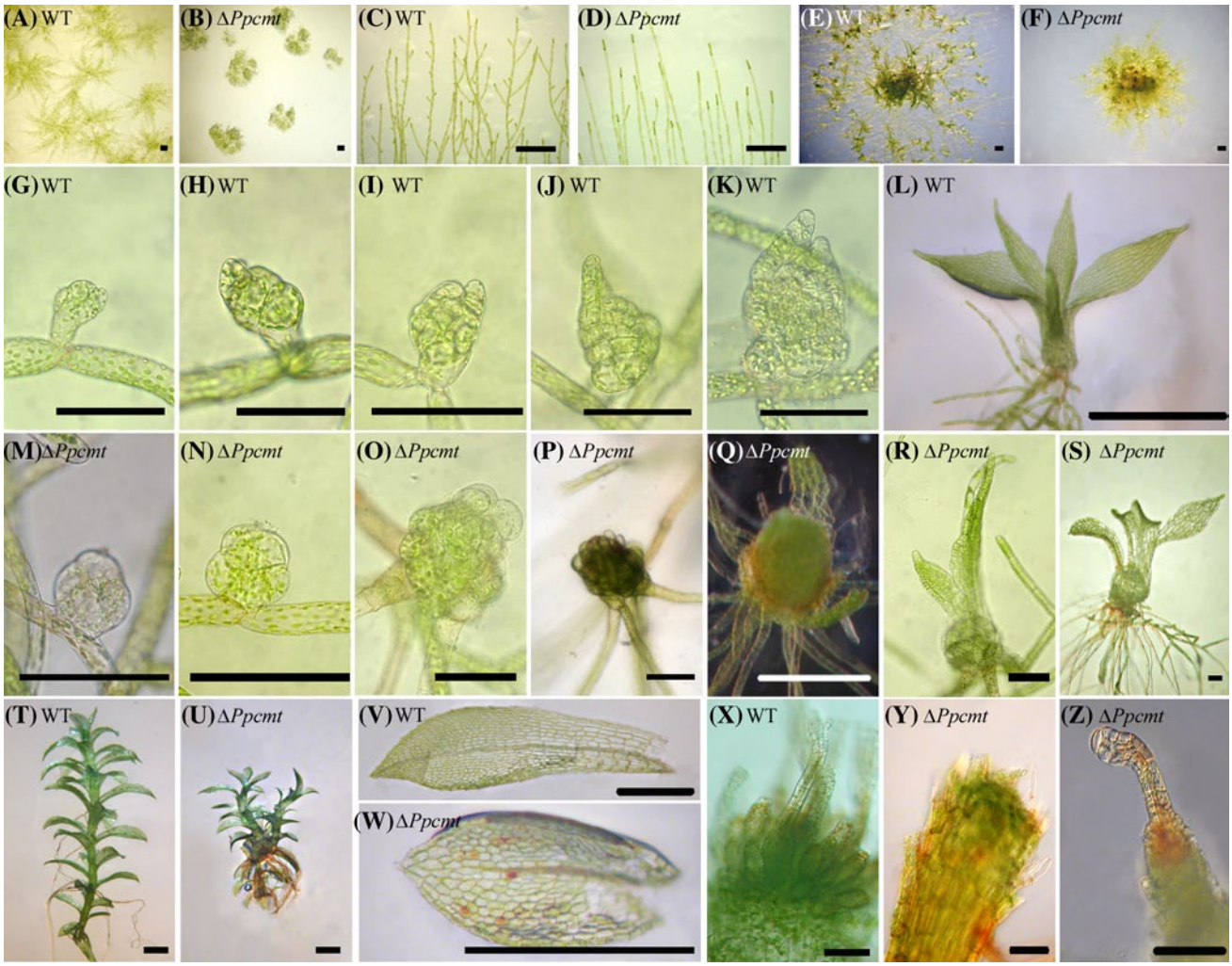


Fig. 7 Morphological analysis of *ΔPpcmt* deletion mutant and WT. **a** and **b** WT and *ΔPpcmt* 2 week old protonema grown from protoplasts. *ΔPpcmt* formed aberrant plants which failed to develop foraging filaments. **c** and **d** WT and *ΔPpcmt* protonema grown for 2 weeks towards unidirectional white light. *ΔPpcmt* barely produced secondary filaments. **e** and **f** WT and *ΔPpcmt* 3 weeks old plants grown after grinding. *ΔPpcmt* developed mostly brown miss-formed buds and seldom stunted green gametophores. **g–l** Developmental stages of WT growth from young bud up to young gametophore. **m** and **n** *ΔPpcmt* abnormal bud cells displaying a globular morphology. **o–q** Almost all *ΔPpcmt* buds resulted in globular architectures. **q–s** Few *ΔPpcmt* buds developed leaf initials.**t** and **u** Mature WT and stunted *ΔPpcmt* gamethophores. **v** and **w** WT and *ΔPpcmt* leaves. *ΔPpcmt* leaves are smaller than WT and at times develop only a partial midrib with necrotic patches. **x** and **y** WT and *ΔPpcmt* gametophore apexes 8 weeks after gametangia induction. *ΔPpcmt* did not develop reproductive organs. **z** A rare single aberrant archegonium on *ΔPpcmt* gametophore apex. *Scale bars* **g–k**, **m–p**, **x–z** 100 μm; **c**, **d** 500 μm; the rest are 1 mm

and also maintains CHH methylation at some genomic loci (Cao and Jacobsen 2002; Cokus et al. 2008; Stroud et al. 2013) and the CMT2 subfamily, which is specialized in CHH methylation (Zemach et al. 2013). Our phylogenetic analysis has classified PpCMT as a member of the CMT3 subfamily (Fig. 1b). DNA methylation analysis of four genomic loci in *ΔPpcmt* demonstrated that PpCMT is involved in DNA methylation specifically in the context of CHG sites (Figs. 2, 3). Furthermore, loss of PpCMT correlates with partial loss of CHH methylation in loci associated with CHG methylation as shown in our analysis of two distinct loci (Fig. 2). Taken together, the above may suggests that CHG methylation is the ancient functions of the CMT protein family. We propose that during land plant evolution, CMT genes duplicated and subsequently diversified into two subfamilies, allowing specialization in CHH methylation function by CMT2 subfamily proteins.

PpCMT accumulation pattern during *P. patens* development as well as during the regeneration of protonema from excised leaves described in this study, showed a stringent localization of PpCMT in dividing cells and its absence from differentiated cells. Thus, PpCMT accumulation

pattern suggests that it could be used as a marker for dividing cells in *P. patens*. PpCMT accumulation in dividing cells is in accordance with its expected function in mediating symmetric methylation of CHG sites, thus supplementing DNA methylation in newly formed DNA strands following genomic DNA replication. The severe phenotype of the *ΔPpcmt* mutant in the vegetative phase highlights the importance of maintaining CHG DNA methylation patterns. The mutant phenotype correlates with PpCMT–GUS accumulation pattern observed in dividing cells at all developmental stages (Figs. 4, 5).

ΔPpcmt protonema tissue showed aberrant cell morphology, division and branching under normal and unidirectional white and red light growth conditions (Fig. 7c vs. d and Fig. S6, S7). While *ΔPpcmt* developed abnormal protonema, differentiation of protonema cells to buds did not depend on PpCMT. However, bud development was impaired resulting in globular architectures (Fig. 7). In *Arabidopsis,* the *Atcmt3* mutant showed no developmental abnormalities, even though most of the CHG methylation was lost. Only when *Atdrm1, Atdrm2* and *Atcmt3* were compromised in an *Arabidopsis* triple mutant, phenotypic abnormalities were observed, which indicates functional redundancy between these genes. Loss of CHG and CHH methylation in the triple *Atdrm1 Atdrm2 Atcmt3* mutant correlated with misexpression of hundreds of genes (Zhang et al. 2006). This shows that in *Arabidopsis,* gene expression is regulated by a cooperation of different DNMTs. In contrast, deletion of the single *PpCMT* resulted in loss of CHG methylation, and a high penetrance of its aberrant phenotypes, indicating that in moss CHG methylation could be sufficient to regulate gene expression. Sole dependency on PpCMT rather than on multiple DNMTs may enable expression plasticity in moss. We postulate that CHG methylation plays an important role in regulating specific genes under the control of PpCMT, thus linking the evolutionary emergence of CMT protein family in land plants with cell differentiation. This in turn, may underlie mechanisms which take part in execution of complex developmental programs facilitating the adaptation of plants to harsh environments. Identification of genomic regions subjected to CHG methylation if correlated with altered expression in *ΔPpcmt,* may allow identification of potential PpCMT targets.

Few of *ΔPpcmt* abnormal buds were able to overcome the genetic lesion and develop leaf initials and even fewer were able to give rise to stunted gametophores (Fig. 7). Similarly, few of *ΔPpcmt* stunted gametophores overcame this lesion and produced a single archegonium at their apex. Among the possible mechanisms that can allow such recovery involves other DNMTs which may partially complement DNA methylation of particular targets, compensating for the lack of PpCMT function. Partial complementation may allow some cells to continue their normal course of development. This can be achieved by CHG methylation mediated for example by DRM homologs in *P. patens* (Malik et al. 2012). In *Arabidopsis* DRMs were shown to methylate cytosine residues in all sequence contexts (Cao et al. 2003; Cokus et al. 2008; Henderson et al. 2010b; Stroud et al. 2013). Alternatively, compensatory hypermethylation of CG and CHH contexts in PpCMT targets may affect their expression, as was shown in the *Arabidopsis Atcmt3* mutant, where loss of CHG DNA methylation led to CG hyper-methylation of AtCMT3 targets (Cokus et al. 2008). *ΔPpcmt* mutant phenotype following sole loss of CHG methylation places *P. patens* as an emerging model organism to study the principles of gene regulation by CHG methylation in land plants.

Acknowledgments We thank Professors Tal Pupko, Eran Bacharach, Shaul Yalovsky and Daniel Chamovitz for critical reading and helpful comments. C.N.M and R.Y were supported in part by a matching Tel-Aviv University Deans doctoral fellowship and the Manna foundation. This research was supported by the Israeli Science Foundation Grant #767/09, and by the Israel Korea program # 3-824 financed by the Ministry of Science and Technology, both granted to N.O. Transgenic lines described in this study were deposited in the International Moss Stock Center (http://www.moss-stock-center.org/) with the accessions IMSC 40736, 40737, 40738 (*ΔPpcmt* 103, 157 and 281) and 40739, 40740 (*PpCMT–GUS* 7 and 8).

Conflict of interest The authors declare that they have no conflict of interest.

References

Ashton NW, Cove DJ (1977) The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants of the moss Physcomitrella patens. Mol Gen Genet 154:87–95

Bauer MJ, Fischer RL (2011) Genome demethylation and imprinting in the endosperm. Curr Opin Plant Biol 14(2):162–167

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS ONE 3(9):e3156

Bezanilla M, Pan A, Quatrano RS (2003) RNA interference in the moss Physcomitrella patens. Plant Physiol 133(2):470–474

Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, Matzke M, Jacobsen SE (2003) Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. Curr Biol 13(24):2212–2217

Cao X, Jacobsen SE (2002) Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci USA 99(Suppl 4):16491–16498

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17(4):540–552

Cedar H, Bergman Y (2012) Programming of DNA methylation patterns. Annu Rev Biochem 81:97–117

Chevenet F, Brun C, Banuls AL, Jacq B, Christen R (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 7:439

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun

bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452(7184):215–219

Cove D (2005) The moss Physcomitrella patens. Annu Rev Genet 39:339–358

Deleris A, Stroud H, Bernatavichute Y, Johnson E, Klein G, Schubert D, Jacobsen SE (2012) Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana. PLoS Genet 8(11):e1003062

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36(Web Server issue):W465–W469

Ebbs ML, Bender J (2006) Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. Plant Cell 18(5):1166–1176

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797

Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, Ukomadu C, Sadler KC, Pradhan S, Pellegrini M, Jacobsen SE (2010) Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci USA 107(19):8689–8694

Feng S, Jacobsen SE (2011) Epigenetic modifications in plants: an evolutionary perspective. Curr Opin Plant Biol 14(2):179–186

Frank W, Decker EL, Reski R (2005) Molecular tools to study Physcomitrella patens. Plant Biol (Stuttg) 7(3):220–227

Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. Annu Rev Biochem 74:481–514

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40(Database issue):D1178–D1186

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52(5):696–704

Henderson IR, Chan SR, Cao X, Johnson L, Jacobsen SE (2010a) Accurate sodium bisulfite sequencing in plants. Epigenetics 5(1):47–49

Henderson IR, Deleris A, Wong W, Zhong X, Chin HG, Horwitz GA, Kelly KA, Pradhan S, Jacobsen SE (2010b) The de novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in Arabidopsis thaliana. PLoS Genet 6(10):e1001182

Henikoff S, Comai L (1998) A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in Arabidopsis. Genetics 149(1):307–318

Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA 89(22):10915–10919

Hou PQ, Lee YI, Hsu KT, Lin YT, Wu WZ, Lin JY, Nam TN, Fu SF (2013) Functional characterization of *Nicotiana benthamiana* chromomethylase 3 in developmental programs by virus-induced gene silencing. Physiol Plant 150(1):119–132

Ishikawa M, Murata T, Sato Y, Nishiyama T, Hiwatashi Y, Imai A, Kimura M, Sugimoto N, Akita A, Oguri Y, Friedman WE, Hasebe M, Kubo M (2011) Physcomitrella cyclin-dependent kinase A links cell cycle reactivation to other cellular changes during reprogramming of leaf cells. Plant Cell 23(8):2924–2938

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J 6(13):3901–3907

Jullien PE, Berger F (2010) DNA methylation reprogramming during plant sexual reproduction? Trends Genet 26(9):394–399

Jurkowska RZ, Jurkowski TP, Jeltsch A (2011) Structure and function of mammalian DNA methyltransferases. ChemBioChem 12(2):206–222

Kamisugi Y, Schlink K, Rensing SA, Schween G, von Stackelberg M, Cuming AC, Reski R, Cove DJ (2006) The mechanism of gene targeting in Physcomitrella patens: homologous recombination, concatenation and multiple integration. Nucleic Acids Res 34(21):6205–6214

Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol 537:39–64

Katz A, Oliva M, Mosquna A, Hakim O, Ohad N (2004) FIE and CURLY LEAF polycomb proteins interact in the regulation of homeobox gene expression during sporophyte development. Plant J 37(5):707–719

Landberg K, Pederson ER, Viaene T, Bozorg B, Friml J, Jonsson H, Thelander M, Sundberg E (2013) The moss Physcomitrella patens reproductive organ development is highly organized, affected by the two SHI/STY genes and by the level of active auxin in the SHI/STY expression domain. Plant Physiol 162(3):1406–1419.

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11(3):204–220

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430(6998):471–476

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133(3):523–536

Loytynoja A, Goldman N (2010) webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. BMC Bioinformatics 11:579

Malik G, Dangwal M, Kapoor S, Kapoor M (2012) Role of DNA methylation in growth and differentiation in Physcomitrella patens and characterization of cytosine DNA methyltransferases. FEBS J 279(21):4081–4094

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39(Database issue):D225–D229

Metivier R, Gallais R, Tiffoche C, Le Peron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F, Salbert G (2008) Cyclical DNA methylation of a transcriptionally active promoter. Nature 452(7183):45–50

Mosquna A, Katz A, Decker EL, Rensing SA, Reski R, Ohad N (2009) Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. Development 136(14):2433–2444

Nishiyama R, Ito M, Yamaguchi Y, Koizumi N, Sano H (2002) A chloroplast-resident DNA methyltransferase is responsible for hypermethylation of chloroplast genes in Chlamydomonas maternal gametes. Proc Natl Acad Sci USA 99(9):5925–5930

Nishiyama T, Hiwatashi Y, Sakakibara I, Kato M, Hasebe M (2000) Tagged mutagenesis and gene-trap in the moss, Physcomitrella patens by shuttle mutagenesis. DNA Res 7(1):9–17

Okano Y, Aono N, Hiwatashi Y, Murata T, Nishiyama T, Ishikawa T, Kubo M, Hasebe M (2009) A polycomb repressive complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution. Proc Natl Acad Sci USA 106(38):16321–16326

Papa CM, Springer NM, Muszynski MG, Meeley R, Kaeppler SM (2001) Maize chromomethylase Zea methyltransferase2 is required for CpNpG methylation. Plant Cell 13(8):1919–1928

Pavlopoulou A, Kossida S (2007) Plant cytosine-5 DNA methyltransferases: structure, function, and molecular evolution. Genomics 90(4):530–541

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita T, Oishi K, Shin IT, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu SH, Stueber K, Theodoulou FL, Tu H, Van de Peer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319(5859):64–69

Sakakibara K, Nishiyama T, Deguchi H, Hasebe M (2008) Class 1 KNOX genes are not involved in shoot development in the moss Physcomitrella patens but do function in sporophyte development. Evol Dev 10(5):555–566

Sakakibara K, Nishiyama T, Kato M, Hasebe M (2001) Isolation of homeodomain-leucine zipper genes from the moss Physcomitrella patens and the evolution of homeodomain-leucine zipper genes in land plants. Mol Biol Evol 18(4):491–502

Saze H, Kakutani T (2011) Differentiation of epigenetic modifications between transposons and genes. Curr Opin Plant Biol 14(1):81–87

Schaefer M, Lyko F (2010) Solving the Dnmt2 enigma. Chromosoma 119(1):35–40

Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152(1–2):352–364

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153(1):193–205

Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328(5980):916–919

Zemach A, Zilberman D (2010) Evolution of eukaryotic DNA methylation and the pursuit of safer sex. Curr Biol 20(17):R780–R785

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 126(6):1189–1201

Zimmer AD, Lang D, Buchta K, Rombauts S, Nishiyama T, Hasebe M, Van de Peer Y, Rensing SA, Reski R (2013) Reannotation and extended community resources for the genome of the non-seed plant Physcomitrella patens provide insights into the evolution of plant gene structures and functions. BMC Genom 14:498