Title

The chromatin landscape of bryophytes

Authors

Tetsuya Hisanaga[1*], Shuangyang Wu[1*], Elin Axelsson[1], Svetlana Akimcheva[1], Liam Dolan[1], and Frédéric Berger[1]

Affiliations

[1]Gregor Mendel Institute, Austrian Academy of Sciences, Vienna BioCenter; Dr. Bohr-Gasse 3, 1030 Vienna, Austria

Contact Info

*These authors contributed equally

Corresponding author: Email: frederic.berger@gmi.oeaw.ac.at

Keywords

up to ten keywords

Evolution, chromatin, *Marchantia polymorpha, Anthoceros agrestis,* bryophytes

Summary

In mammals and flowering plants chromatin is modified by DNA methylation (5mC) and methylation of lysine residues of the N-terminal tail of histone H3 that reflect transcriptional activity in three chromosomal domains: euchromatin comprising transcribed genes, facultative heterochromatin comprising repressed genes, and constitutive heterochromatin comprising transposons. Yet recent studies have shown that the correlation between chromatin modifications and

transcriptional regulation of transposons and genes vary among different lineages of eukaryotes including the bryophytes that diverged from vascular plants 510-480 mya. The monophyletic bryophytes comprise mosses, liverworts and hornworts. The chromatin landscapes of mosses and liverworts are different, leaving open the question regarding the identity of the chromatin landscape of all bryophytes.

Here we assembled the chromatin landscape of the model hornwort *Anthoceros agrestis*. By comparing chromatin landscapes between mosses, liverworts and hornworts we define the common chromatin landscape of the ancestor of extant bryophytes. The constitutive heterochromatin of bryophytes is characterized by the absence of gene body DNA methylation, low levels of non-CG DNA methylation and the absence of segregation of transposons in a few well defined chromosomal compartments. In addition, whereas histone modifications of euchromatin and facultative heterochromatin are primarily associated with genes in angiosperms and mosses, nearly half of the hornwort transposons are associated with facultative heterochromatin and euchromatin. Hence the ancestral genome of bryophytes was likely a patchwork of interspersed small units of euchromatin, facultative and constitutive heterochromatin. Each unit contained a few transposons and genes that share the same chromatin state. We propose that in bryophytes, this specific type of genome organization prevented the recruitment of transposons to form constitutive heterochromatin around point centromeres observed in the genomes of angiosperms.

Introduction

Nucleosomes efficiently package the genetic information in eukaryotes. A nucleosome is composed of 147 bp of DNA wrapped around an octamer composed of two heterodimers of histones H2A and H2B and a tetramer of histone H3 and H4 (Zhou et al., 2019). This packaging unit of chromatin is conserved across all eukaryotes (Sato et al., 2022; Talbert and Henikoff, 2021a; Malik and Henikoff, 2003). Nucleosomes enable eukaryotes not only to pack the genome inside the nucleus but also allow for post-translational modifications (PTMs) on N- and C- terminal tails of core histones to signal the position of transcriptional regulatory features such as transcriptional start sites,

gene bodies or to effect transcriptional silencing (Leng et al., 2020; Gardner et al., 2011; Talbert and Henikoff, 2021b; Kornberg and Lorch, 2020). PTMs of histone tails are involved in the regulation of cell cycle checkpoints, heterochromatic formation, centromere assembly, DNA replication, DNA repair and gene transcription among others (Schwämmle et al., 2016; Bannister and Kouzarides, 2011). Although a plethora of PTMs have been identified, the most conserved modifications are methylation, phosphorylation, and acetylation of H2A, H2B, and H3 (Allis and Jenuwein, 2016). Although some PTMs are restricted to specific lineages, many PTMs are conserved and are found in similar genomic contexts across eukaryotes (Grau-Bové et al., 2022; Sequeira-Mendes et al., 2014; Jamge et al., 2022).

In euchromatin, trimethylation of the lysine 4 on histone H3 (H3K4me3) is a PTM associated with the transcription start sites (TSS) of transcribed genes (Shilatifard, 2012; Leng et al., 2020) while H3K36me3 is associated with transcriptional elongation (Bannister et al., 2005). Facultative heterochromatin forms nuclear domains marked by H3K27me3 deposited by Polycomb Repressive Complexes 2 (PRC2) (Schuettengruber and Cavalli, 2009; Gan et al., 2015). H3K27me3 is crucial for silencing developmental genes (Del Prete et al., 2015; Müller and Verrijzer, 2009; Xiao and Wagner, 2015). By contrast, constitutive heterochromatin occupies transposons and is marked by H3K9 methylation in yeast, animals, and flowering plants (Allshire and Madhani, 2018). In yeast and animals H3K9me3 is bound by heterochromatin protein 1 (HP1) which promotes compaction of heterochromatin and could phase separate by recruiting other heterochromatin interacting proteins (Larson and Narlikar, 2018). In flowering plants, H3K9me2 does not bind to HP1, but instead recruits the plant-specific DNA methyltransferases CHROMOMETHYLASE (CMT) 2 and 3 (Jamge and Berger, 2022). CMT2/3 deposits methyl groups specifically on cytosines in CHG context (Du et al., 2012). H3K9me2 also recruit the *de novo* DNA methyltransferases DOMAINS REARRANGED METHYLASE 1(DRM1) and DRM2 (Li et al., 2018) and DNA methylation is bound by a domain present in the SU(VAR)3-9 HOMOLOG (SUVH) 4/5/6 that methylates H3K9 (Li et al., 2018). Hence feed-back loops involving H3K9 methylation and DNA methylation in flowering plants maintain heterochromatin. In addition, H3K27me1 is deposited by specific histone methyltransferases

ARABIDOPSIS TRITHORAX-RELATED 5 (ATXR5) and ATXR6 in constitutive heterochromatin of *Arabidopsis thaliana (Jacob et al., 2009)*.

Although PTMs and their function are conserved amongst angiosperms, recent investigations have begun to show a fairly distinct chromatin landscape in bryophytes that diverged ca. 500-480 MYA from the vascular plants. Bryophytes comprise three monophyletic groups – hornworts, liverworts and mosses (Harris et al., 2021). Five histone PTMs and DNA methylation have been profiled in the genomes of two model bryophytes, the moss *Physcomitrium patens* (Yaari et al., 2019; Widiez et al., 2014) and the model liverwort *Marchantia polymorpha* (Ikeda et al., 2018; Montgomery et al., 2020). DNA methylation at CGs is maintained by an ortholog of the flowering plant METHYLTRANSFERASE 1 (MET1), but other pathways controlling DNA methylation in angiosperms are not conserved in bryophytes. In *P. patens*, de novo DNA methylation depends on DNMT3, which is distinct from DRMs involved in the RNA-directed DNA methylation (RdDM) pathway in *A. thaliana (Yaari et al., 2019). M. polymorpha* has two CMT-like proteins which reside outside of the clade comprising CMT3 (Bewick et al., 2017). In addition, horizontal gene transfer of a N4 methyltransferase from prokaryote led to N4 methylation in *M. polymorpha* (Walker et al., 2021). The enzymes that deposit H3K9 methylation have not been characterized in bryophytes but orthologs of SUVH 4/5/6 responsible for the deposition of this PTMs in *A. thaliana* are present in bryophytes (Bowman et al., 2017). H3K27me1 is also present at heterochromatin of *M. polymorpha* (Montgomery et al., 2020). While H3K36me3 marks the body of actively transcribed genes in *M. polymorpha* and *P. patens* as in *A. thaliana*, H3K4me3 appears to be associated with repressed genes in *M. polymorpha* but not in *P. patens* (Montgomery et al., 2020; Widiez et al., 2014). While forty percent of TEs are covered by H3K27me3 in *M. polymorpha* (Montgomery et al., 2020) TEs in *P. patens* are covered mostly by H3K9me2 (Widiez et al., 2014). These studies show that flowering plants and bryophytes differ in the mechanisms of DNA methylation and association of histone PTMs with transcriptional states, genes and TEs.

The difference of the mechanisms of DNA methylation and histone modification in mosses and liverworts and the absence of data in the sister group of hornworts have made it impossible to

reconstruct the ancestral state (Jamge and Berger, 2022). To reconstruct the ancestral chromatin landscape of bryophytes, we investigated DNA methylation and a set of five PTMs in the chromatin of the model hornwort *Anthoceros agrestis* and compared it to *P. patens* and *M. polymorpha*. The chromatin synapomorphies shared by *A. agrestis* and *M. polymorpha* or *P. patens* are thus considered to represent the ancestral features of chromatin in bryophytes.

Results

Chromatin profiling of *Anthoceros agrestis*

To address the conservation of the association of PTMs of chromatin and DNA methylation among bryophytes, we used the model hornwort *Anthoceros agrestis* that diverged from bryophyte ancestors earlier than liverworts and mosses (Puttick et al., 2018; Harris et al., 2021). We used Enzymatic Methyl-seq (Vaisvila et al., 2021) to obtain a genome wide profile of 5-methylated cytosine from four week old vegetative tissue of *A. agrestis*. Chromatin immunoprecipitation coupled with DNA sequencing (ChIP-seq) was applied to the same tissue to obtain genomic profiles of five histone PTMs (H3K4me3, H3K36me3, H3K9me1, H3K27me1 and H3K27me3) and H3 (Figure 1A). Although H3K9me2 is often used as a mark for constitutive heterochromatin, ChIP procedure for this PTM did not work for *A. agrestis*. Instead, we used H3K9me1 as a mark for constitutive heterochromatin as this PTM shows a similar coverage to H3K9me2 in *M. polymorpha*. Biological replicates clustered together in a Spearman correlation matrix (Figure S1A) and the profiles of PTMs typically considered as repressive (H3K9me1 and H3K27me1) or active (H3K4me3 and H3K36me3) grouped among themselves (Figure S1B), showing the robustness of our data.

To gain insights into the chromosome level organization of chromatin landscape in *A. agrestis*, we used Hi-C data from *A. agrestis* Bonn strain to assemble genome scaffolds of *A. agrestis* Oxford strain into chromosome level assembly. 140 scaffolds out of 153 were assembled into 6 large scaffolds which corresponds to the 6 chromosomes described in *A. agrestis* (Li et al., 2020). In addition, we re-annotated TEs in *A. agrestis* Oxford strain and identified 90,809 TEs including 990 intact TEs belonging to various TE families (see Methods for the new annotation). A large majority of

the annotated TEs were relatively short (<500bp) and to the exception of MITEs, represented primarily fragments of intact TEs. Hence, we mainly focused on long TEs (≥500 bp, 26,070 TEs including 923 intact TEs). In this annotation, we found 4,955 TEs overlapped with protein coding gene (PCG) annotations. Using this assembly and TE annotation, we plotted densities of genes, TEs and chromatin marks to see the distribution of these features over chromosomes, as exemplified by their general profiles over chromosome 1 (Figure 1A). Overall, genes and TEs are distributed evenly over the chromosome of *A. agrestis* Oxford strain as observed in *A. agrestis* Bonn strain. Corresponding to this pattern, all the chromatin marks were distributed evenly over the chromosome (Figure 1A). Such a broadly even distribution of chromatin marks, genes and TEs was comparable to chromatin organization in *M. polymorpha* and *P. patens*.

Peaks of H3K9me1 and H3K27me1 were enriched on various types of TE sequences and showed high levels of DNA methylation in both CG and CHG context (Figures 1B-D). Peaks of H3K4me3 were enriched on coding sequences and had very low levels of DNA methylation (Figures 1B-D). While these associations were reported in other model plants, peaks of H3K36me3 and H3K27me3 were present on both TEs and coding sequences with modest levels of cytosine methylation that were not reported on other model plant species (Figures 1B, C and E). Hence while we observed some of the associations between genomic features and PTMs reported in other model plants, this overview suggested additional associations present in hornworts.

Association of histones PTMs and DNA methylation on protein coding genes

We explored preferential associations between chromatin marks and the transcriptional status of PCGs based on their expression in the vegetative tissues (thallus) publicly available (Li et al., 2020). H3K36me3 and H3K4me3 were strongly associated with expressed PCGs (Figure 2A), while H3K9me1, H3K27me1 and H3K27me3 were associated with repressed PCGs (Figure 2A). To untangle relationships between chromatin profiles and PCGs in *A. agrestis*, we performed unsupervised k-means clustering of chromatin profiles over PCGs. This analysis led to defining five major clusters of PCGs (named cluster P1 to P5) exhibiting different chromatin environments (Figure 2B). We calculated the average expression level in each cluster (Figure 2C). Cluster P3, P4 and P5

comprised 24%, 21% and 25% of all PCGs, respectively and accounted for expressed PCGs (Figures 2B and 2C). In these three clusters H3K36me3 was enriched over gene bodies and H3K4me3 was enriched in the transcription start sites (TSS) as described in other land plants. The presence of H3K4me3 in gene bodies observed in the cluster P3 (Figure 2B) was not reported in other bryophytes. In these three clusters, the gene bodies showed low levels of CG methylation and no cytosine methylation in non-CG context (Figure 2D) as described in *P. patens (Zemach et al., 2010)* and *M. polymorpha (Takuno et al., 2016; Ikeda et al., 2018)*. Non-expressed PCGs formed clusters P1 and P2 (Figures 2B and 2C). Cluster P1 contained 11% of all PCGs and was characterized by enrichment of H3K27me3 and H3K4me3 and depletion of cytosine methylation in all contexts, which was similar to the cluster 1 observed over 8% of *M. polymorpha* PCGs (Montgomery et al., 2020). Cluster P2 contained 18% of all PCGs with enrichment of H3K9me1, H3K27me1 and a high level of cytosine methylation in CG and CHG contexts. No comparable chromatin state was observed over PCGs in *M. polymorpha* (Montgomery et al., 2020). PCGs in cluster P2 were shorter than PCGs in other clusters (Figure S2B). About 40% of PCGs from cluster 2 overlapped with TEs (Figure S2A), suggesting that these PCGs are coding regions of intact TEs or PCGs that have co-opted coding regions of TEs.

To test whether repressed PCGs from clusters P1 and P2 were expressed in non-vegetative tissue or in different physiological conditions, we explored transcriptome data sets available (Li et al., 2020). We found that 15% of PCGs in cluster P1 are repressed in gametophytes but expressed in sporophytes, suggesting H3K27me3 functions as a facultative heterochromatic mark to regulate expression of developmental genes (Figure S2C). Surprisingly, 9.3% of PCGs in cluster P2 are repressed in gametophytes but expressed in sporophyte stage, suggesting that H3K9me1 and/or H3K27me1 also behave as a facultative heterochromatic mark in *A. agrestis* (Figure S2D). These PCGs did not overlap with annotated TEs and encoded proteins such as GDSL lipase, transporters and pectin lyase which might have physiological function during the sporophyte stage.

We conclude that in euchromatin, the patterns of enrichment of H3K4me3 and H3K36me3 over expressed genes are broadly conserved in land plants but bryophytes exhibit a much lower enrichment in DNA methylation on gene bodies than in angiosperms. In addition to the conserved

association of H3K27me3 with facultative heterochromatin, H3K9me1 and H3K27me1 also repress gene expression in a context dependent manner in hornworts.

Association of histones PTMs and DNA methylation on TEs

To address the relationships between chromatin marks and TEs, we performed k-means clustering of chromatin marks over long TEs, and defined five major clusters of TEs showing different chromatin environments (cluster T1 to T5 in Figure 3A). Clusters T2 and T3 contained 39 and 25% of all TEs and were covered with H3K9me1, H3K27me1 and CG and CHG methylation (Figures 3A and 3B). TEs in cluster T2 showed weaker enrichment in DNA methylation, H3K9me1 and H3K27me1 compared to cluster T3. These clusters contained longer TEs compared to those in other clusters (Figure S2E) and were more enriched in LTR families (Figure 3B). TEs from clusters T1, T4 and T5 were occupied by chromatin states distinct from the typical constitutive heterochromatin state present in TEs from clusters T2 and T3. Cluster T1 comprised 3% of all TEs that were covered with H3K4me3 and H3K27me3 but not with DNA methylation (Figures 3A and 3D). This association of PTMs was also observed on 6.5% of short TEs (Figure S3A). Clusters T4 and T5 contained 11% and 18 % of all long TEs that were occupied by chromatin with euchromatic marks (H3K36me3 and H3K4me3) and low level of CG methylation (Figures 3A and D). These TEs consisted of relatively short TEs (less than 1.0 kb) (Figure S2E), and a majority of them (60% of TEs in cluster T4 and 55% of TEs in cluster T5) overlapped with PCGs (Figure S2F), in contrast to long TEs from clusters T1 – T3. This suggested that TEs covered with euchromatic PTMs are fragments of TEs located near PCGs or were mis-annotated as PCGs .

We investigated whether PCGs and surrounding TEs share the same chromatin environment by measuring the distance between each PCG and the closest TE per cluster (Figure 3C) and vice versa (Figure S2G). While TEs in cluster T1 were mostly surrounded by PCGs from cluster P1, PCGs from cluster P1 did not always associate with TEs in cluster T1. Active PCGs from clusters P3-P5 were surrounded not only by TEs in clusters T4 and T5 which are marked by active marks, but also by TEs covered by H3K9me1 and H3K27me1(Figures 1D and S2G). PCGs from cluster P2 were mostly surrounded by TEs in clusters T2 and T3 (Figure S2G) and vice versa (Figure 3D). Altogether we

observed a strong association between PCGs and TEs sharing the same location, and forming small domains dominated by a single type of chromatin.

Discussion

In flowering plants, TEs are enriched in pericentromeric heterochromatin while PCGs are enriched along chromosomal arms (Jamge and Berger, 2022; Sequeira-Mendes et al., 2014). By contrast, both TEs and PCGs are distributed evenly over the entire chromosomes of bryophytes (Montgomery et al., 2020; Lang et al., 2018; Zhang et al., 2020; Li et al., 2020). Corresponding to these distributions, it was shown that the chromatin PTMs associated with TEs are also distributed evenly over the chromosomes of *M. polymorpha* and *P. patens (Montgomery et al., 2020; Widiez et al., 2014)*. In this study, we demonstrated that a comparable global and even distribution of the chromatin modifications shapes the global architecture of the genome of the hornwort, *A. agrestis*. The conservation of genome size and number of chromosomes amongst hornworts (generally 6; (Villarreal and Renner, 2013) and liverworts (always 9 ; (Bowman et al., 2022) is also in support of a general conservation of architecture of genomes across bryophytes with deviation of this state possible in mosses with much larger genomes.

H3K9 methylation forms constitutive heterochromatin that represses the expression of TEs in many eukaryotic species (Allshire and Madhani, 2018). In addition, this mark strongly co-occurs with H3K27me1 and DNA methylation in angiosperms (Jacob et al., 2009; Li et al., 2018). In all model bryophytes a majority of TEs is also associated with H3K9me1 and H3K27me1 in agreement with the presence of orthologs of *A. thaliana* SUVH4/5/6 and ATXR5/6 in bryophytes. In *M. polymorpha* the orthologs of *A. thaliana* MET1 maintains CG methylation and is involved in silencing TEs (Ikeda et al., 2018). In *P. patens*, CHG methylation is maintained by an ortholog of *A .thaliana* CHROMOMETHYLASE 3 (Domb et al., 2020) and CHH methylation is deposited primarily by the *de novo* methyltransferase DNMT3 but not by an ortholog of the methyltransferases DRM1 and DRM2 (Yaari et al., 2019) nor by an ortholog of CMT2 which are responsible for CHH deposition in *A. thaliana (Stroud et al., 2014)*. Together these pathways contribute to TE silencing (Domb et al., 2020). In *A. agrestis* we identified orthologs of MET1 and CMT in agreement with the presence of

CG and CHG methylation and DRM (Table 1). However, the near complete absence of CHH methylation suggests that the ortholog of DRM is inactive in vegetative haploid tissues. DNA methylation in CHG and CG is primarily observed in TEs, suggesting the absence of gene body methylation on PCGs as described in *A. thaliana.* Such reduction or absence of gene body methylation was also reported in *M. polymorpha* (Bewick et al., 2017) and *P. patens* (Domb et al., 2020) and it is thus a likely common feature to all bryophytes.

Despite the overall conservation of chromatin environments in bryophytes, several features appear to be specific to *A. agrestis*. First, a significant fraction of PCGs is covered by chromatin marks of constitutive heterochromatin (H3K9me1 and H3K27me1). These PCGs overlap with or are surrounded by TEs in clusters T2 and T3, indicating clusters of constitutive heterochromatin containing both PCGs and TEs dispersed over the entire genome. Strikingly, half of TEs show very low levels of DNA methylation, suggesting that TEs play an important role outside of constitutive heterochromatin in *A. agrestis*. Similarly, 5% of TEs are covered by H3K27me3 and most of them are neighbors of PCGs covered by H3K27me3. This is reminiscent of the association between TEs and H3K27me3 in *M. polymorpha*, supporting further a shared role of H3K27me3 in silencing TEs across a broad range of eukaryotes (Déléris et al., 2021). On the other hand, short TE fragments covered by euchromatic marks (H3K4me3 and H3K36me3) are located near or inside PCGs. Such strong association between TEs and euchromatin is not observed in *M. polymorpha* and *A. thaliana* and may be the result of the size reduction of the genome of *A. agrestis* genome (120 Mb and 25,000 PCGs) compared with *M. polymorpha* (230Mb and 20,000 PCGs), and the absence of cumulation of TEs in a specialized domain of heterochromatin as in *A. thaliana*.

We observed 18% of PCGs are covered by H3K9me1 together with H3K27me1 which is also known to form constitutive heterochromatin in plants. A fraction of these PCGs is specifically expressed in sporophytes, suggesting that in addition to H3K27me3, H3K9me1 and/or H3K27me1 function as facultative heterochromatin during life cycle of *A. agrestis*. Almost half of TEs in Anthoceros are present in euchromatin or facultative heterochromatin indicating a large degree of co-optation of TEs to PCGs and/or their regulatory sequences. Such a role has been

reported mammals (Ninova et al., 2019) and in the pollen of *A. thaliana*, where reprogramming of H3K9me2 covered loci in vegetative nuclei promote expression of genes required for pollen tube growth (Borg et al., 2021).

The location of constitutive heterochromatin in bryophytes contrasts with the position of heterochromatin around point centromeres in flowering plants with small genomes (Sigman and Slotkin, 2016; Naish et al., 2021). It has been proposed that heterochromatin participates in the recruitment of centromeric proteins in fission yeast and *A. thaliana* (Scott and Sullivan, 2014) but this is not the case in bryophytes. Hence bryophytes might have evolved mechanisms of centromere recruitment distinct from that of flowering plants. This mechanism could be related to the pervasive co-optation of TEs in bryophytes that prevented their recruitment in discrete domains surrounding centromeres. Whether bryophytes have preserved a more ancestral status of chromatin organization inherited from charophytes awaits chromatin profiling in these algal members of streptophytes.

Acknowledgements

We thank Sean A. Montgomery for suggestions and critical reading of the manuscript. F.B. acknowledges support from the PlantS and Next Generation Sequencing at the Vienna BioCenter Core Facilities (VBCF), and the Molecular Biology Services from the Institute for Molecular Pathology (IMP), members of the Vienna BioCenter (VBC), Austria. This work was funded by FWF grants P32054, and P36231 to F.B., funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Marie Curie Skłodowska Grant Agreement Nr. 847548 (VIP2) to T.H. and S.W..

Author contributions

T.H. and F.B. conceived and designed the experiments. T.H. performed ChIP-seq and EM-seq in *A. agrestis* with help from S. A.. S. W. annotated TEs in the genomes of *A. agrestis.* T.H. and S. W.

performed the analysis of the data with help from E. A who also performed statistical analyses and curated data. L.D. provided funding of S.W. and F.B supervised the study. T.H. and F.B. wrote the manuscript draft with input from all authors.

Declaration of interests

The authors declare no conflict of interest.

Inclusion and diversity statement

Figure legends

Figure 1. Association of chromatin marks with genomic features

(A) Circos plot showing epigenomics features over the chromosome 1 of *A. agrestis*. From top to bottom, the following information was depicted with a window size of 100 kb. Gene density, TE density, level of DNA methylation, densities of H3K9me1, H3K27me1, H3K27me3, H3K4me3 and H3K36me3 peaks

(B) Distribution of chromatin marks over genomic features. The total length of chromatin mark peaks overlapping specified genomic features was divided by the total length of peaks of chromatin marks to determine each proportion. Features which cover less than 0.5% of chromatin mark peaks are not shown.

(C) and (D) Integrative Genomics Viewer (IGV) browser screenshot demonstrating vicinity of TEs covered by H3K9me1 and expressed genes (C) and TEs covered by H3K27me3 (D). The region shown is 55 kb (C) or 34 kb (D) in length and from the scaffolds utg000063l and utg000028l, respectively. Chromatin mark tracks are bigwig files scaled by H3 coverage in 10-bp windows. ‘‘TE’’ and ‘‘Gene’’ tracks are annotation files for TEs and genes, respectively. ‘‘RNA-seq’’ track is a bigwig of mapped RNA-seq reads from vegetative tissue (ref). Scales are noted in square brackets beside each track.

(E) Profile plot of CG, CHG and CHH methylation levels over chromatin mark peaks. Each chromatin mark peak is scaled to 1 kb and sequences 1 kb upstream and downstream are included. Average methylation over 10 bp bins is plotted.

Figure 2. Association of chromatin marks on genes

(A) Violin plot showing expression level of PCGs associated with chromatin marks. Width is relative to the density of genes. Red dots indicate median expression values.

(B) Heatmap of k-means clustering of genes based on chromatin marks. Prevalence of each mark (columns) based on its score normalized against H3 signals per 10 bp bins. Gene body of each gene is scaled to 2 kb and sequences 1 kb upstream and downstream are included. Each color (orange, green and blue) stands for enrichment and white for depletion. Each row corresponds to one gene, with multiple genes grouped into blocks that have been defined as clusters P1 through P5.

(C) Violin plot showing expression level of PCGs per PCG cluster. Width is relative to the density of genes. Red dots indicate median expression values.

(D) Profile plot of CG, CHG and CHH methylation levels over PCGs per PCG cluster. Gene body of each PCG is scaled to 2 kb and sequences 1 kb upstream and downstream are included. Average methylation over 10 bp bins is plotted.

Figure 3. Association of chromatin marks on TEs

(A) Heatmap of k-means clustering of TEs based on chromatin marks. Prevalence of each mark (columns) based on its score normalized against H3 signals per 10 bp bins. Each TE annotation is scaled to 1 kb and sequences 1 kb upstream and downstream are included. Each color (orange, green and blue) stands for enrichment and white for depletion. Each row corresponds to one TE, with multiple TEs grouped into blocks that have been defined as clusters T1 through T5.

(B) Stacked bar chart indicating proportion of TE families in each TE cluster.

(C) Boxplot of distances between each TE and the nearest gene per gene cluster. Briefly each TE is compared to all PCGs belonging to a PCG cluster to find its nearest neighbor. TEs are divided based

on the TE cluster they belong to. Distances are in kilobases (kbp). Colored boxes represent interquartile range, and lines represent median values. Outliers are not shown.

(D) Profile plot of CG, CHG and CHH methylation levels over TEs per TE cluster. Each TE annotation is scaled to 1 kb and sequences 1 kb upstream and downstream are included. Average methylation over 10 bp bins is plotted.

Figure S1. Spearman Correlation matrix of ChIP-seq reads from *A. agrestis*

(A) Spearman correlation matrix showing that biological replicates of each mark cluster together.

(B) Spearman correlation matrix showing that active and repressive marks cluster together.

Figure S2. Characterization of *A. agrestis* gene clusters and long TE clusters

(A) Stacked bar chart indicating number of PCGs overlapped by TEs per PCG cluster.

(B) Violin plot showing length of PCGs per PCG cluster. Width is relative to the density of PCGs.

(C) and (D) Heatmap showing expression level of PCGs in PCG cluster 1 (C) or Cluster 2 (D). Each row corresponds to a cluster based on expression pattern and each column corresponds to each tissue or condition.

(E) Stacked bar chart indicating number of TEs overlapped by PCGs per TE cluster.

(F) Violin plot showing length of TEs per TE cluster. Width is relative to the density of TEs.

(G) Boxplot of distances between each gene and the nearest TE per TE cluster. Briefly each gene is compared to all TEs belonging to a TE cluster to find its nearest neighbor. Genes are divided based on the gene cluster they belong to. Distances are in kilobases (kbp). Colored boxes represent interquartile range, and lines represent median values. Outliers are not shown.

Figure S3. Characterization of *A. agrestis* short TE clusters

(A) Heatmap of k-means clustering of short TEs based on chromatin marks. Prevalence of each mark (columns) based on its score normalized against H3 signals per 10 bp bins. Each TE annotation is scaled to 1 kb and sequences 1 kb upstream and downstream are included. Each color (orange, green

and blue) stands for enrichment and white for depletion. Each row corresponds to one TE, with multiple TEs grouped into blocks that have been defined as clusters sT1 through sT5.

(B) Stacked bar chart indicating proportion of TE families in each TE cluster.

(C) Stacked bar chart showing proportion of PCG clusters of nearby genes of TEs in cluster sT1.

Materials and Methods

Plant materials

Anthoceros agrestis strain Oxford were cultured on 0.5 Gamborg's B5 medium solidified with 1% agar under continuous white light at 22°C.

Chromosomal assembly of *A. agrestis* Oxford strain genome

Since the Oxford strain genome is close to the Bonn strain which is already assembled in chromosomal level (Li et al., 2020), we were using RagTag (v2.0.1) (Alonge et al., 2021) to perform homology-based assembly scaffolding and patching. To be briefly, RagTag uses the Bonn genome as reference to scaffold contig joins and fill gaps in the Oxford genome, resulting in an assembly comprising 6 chromosomes and 13 super-scaffolds.

Annotation of TEs in *A. agrestis*

Transposable elements of *A. agrestis* were annotated using EDTA 1.9.9 (Ou et al., 2019), which incorporates a bunch of tools including LTRharvest, LTR_FINDER, LTR_retriever, Generic Repeat Finder , TIR-Learner, MITE-Hunter, HelitronScanner, and RepeatMasker. All softwares were adjusted to EDTA with proper filters and parameters. Final non-redundant TE libraries were produced by removing nested insertions and protein-coding genes by EDTA customized scripts.

Chromatin profiling of *A. agrestis* using ChIP-seq

ChIP experiments were performed following a previously described protocol with some modifications (Yelagandula et al., 2017). Four weeks old gametophyte tissue of *A. agrestis* were collected and crosslinked using 1% paraformaldehyde in 1x PBS with vacuum on ice for 10 minutes.

Cross-linking reaction was quenched by adding 2 M glycine with vacuum on ice for 10 minutes. Excess solution was removed from cross-linked tissue by paper towels. Cross-linked tissue was then snap- frozen in liquid nitrogen and grinded to fine powder using mortar and pestle. The powder was transferred into a 50 ml plastic tube and suspended in 40 ml of MP1 buffer (10 mM MES-KOH buffer pH 5.3, 10 mM NaCl, 10 mM KCl, 0.4 M sucrose, 2% (w/v) PVP-10, 10 mM MgCl2, 10 mM 2-mercaptoethanol, 6 mM EGTA, 1× cOmplete protease inhibitor cocktail). Suspended samples were then filtered through one layer of Miracloth twice, 40 μm nylon mesh once and 10 μm nylon mesh twice. Filtered samples were centrifuged at 3000 xg at 4 °C for 10 min and then supernatant was discarded. The pellet was washed using 15 ml of MP2 buffer (10 mM MES-KOH buffer pH 5.3, 10 mM NaCl, 10 mM KCl, 0.25 M sucrose, 10 mM MgCl2, 10 mM 2-mercaptoethanol, 0.2% Triton-X 100, 1× cOmplete protease inhibitor cocktail) three times. The final pellet was then resuspended in 5 ml of MP3 buffer (10 mM MES-KOH buffer pH 5.3, 10 mM NaCl, 10 mM KCl, 1.7 M sucrose, 2 mM MgCl2, 10 mM 2-mercaptoethanol, 1× cOmplete protease inhibitor cocktail) and centrifuged at 16000 xg at 4C for 1 h. After centrifuge, the supernatant was discarded and the pellet was resuspended in 900 μl of covaris buffer (0.1% SDS, 1 mM EDTA, 10 mM Tris-HCl pH 8.0, 1× cOmplete protease inhibitor cocktail). Resuspended pellet containing chromatin fraction was fragmented using a Covaris E220 High Performance Focused Ultrasonicator for 15 min at 4 °C (duty factor, 5.0; peak incident power, 140.0; cycles per burst, 200) in a 1 ml Covaris milliTUBE. Sheared chromatin was centrifuged at 15000 rpm at 4C for 10 min and the supernatant was transferred into a new 5 ml tube and diluted by adding xx ml of ChIP dilution buffer. Diluted chromatin was precleared by incubating with proteinA/G beads (Thermo Fisher Scientific) beads at 20 rpm spinning at 4C for 1 h. Beads were removed by magnetic racks and precleared chromatin was separated into 6 tubes and incubated with 1ug of specific antibodies for histone modifications (listed in xx) at 20 rpm spinning at 4C for overnight. Chromatin bound by antibodies was collected by incubating with protein A/G beads for 3 h. The beads were collected by magnetic racks and washed by low salt wash buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100 and 0.1% SDS) twice, high salt wash buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100 and 0.1% SDS) once,

LiCl wash buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.25 M LiCl, 1% IGEPAL CA-630 and 0.1% sodium deoxycholate) once and TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA) twice. Immunoprecipitated DNA was eluted using 500 ul elution buffer (1% SDS and 0.1 M NaHCO3) at 65C for 15 min. To reverse cross-link eluted DNA was mixed with 51 µl of reverse cross-link buffer (40 mM Tris-HCl pH 8.0, 0.2 M NaCl, 10 mM EDTA, 0.04 mg ml−1 proteinase K (Thermo Fisher Scientific)), and then incubated at 45 °C for 3 h and subsequently at 65 °C for 16 h. After cross-link reversal, DNA was treated with 10 µg of RNase A (Thermo Fisher Scientific), incubated at room temperature for 30 min and purified using the MinElute PCR purification kit (Qiagen). ChIP-seq library was generated from ChIPed DNAs using Ovation® Ultralow Library Systems V2 (TECAN). The ChIP-seq libraries were sequenced on a illumina Hiseq v4 to generate 50 bp single end reads.

ChIP-seq data analyses

The bam files of ChIP-seq reads were sorted with SAMtools v1.9 (Li et al., 2009) and converted to fastq format using bamtofastq function of BEDTools v2.27.1 (Quinlan and Hall, 2010), trimmed with Cutadapt v1.18 (Martin, 2011) and aligned to *A. agrestis* Oxford strain genome (Li et al., 2020) using Bowtie2 v2.3.4.2 (Langmead and Salzberg, 2012). Resulting bam files were sorted and indexed with SAMtools v1.9. Reads with MAPQ less than ten were removed with Samtools v1.9 and duplicates were removed with Picard v2.18.27 (http://broadinstitute.github.io/picard/). Spearman correlation matrices were generated using multiBamSummary and plotCorrelation functions in deepTools v3.3.1 (Ramírez et al., 2016). Deduplicated reads from 2 biological replicates were merged. The read coverage of each chromatin mark was normalized against the read coverage of H3 in 10 bp windows with the bamCompare function in deepTools v3.3.1, generating bigwig files. Broad peaks of each chromatin mark were called by using macs2 v2.2.5 with default settings (Zhang et al., 2008). Overlaps between genomic features and each chromatin mark were calculated by using the intersect function of BEDTools. Ratio of each genomic feature was calculated by dividing total length of overlaps by total length of each chromatin mark.

Clustering analyses of ChIPseq data

K-means clustering of chromatin marks was performed using deepTools v3.3.1. Matrices were computed using computeMatrix for either PCGs or TEs using bigwig files as input and each region is scaled to 2 kb for PCGs and 1 kb for TEs with 1 kb upstream and downstream. Heatmaps of matrices were plotted with plotHeatmap with k-means clustering. Cluster assignments can be found in Data S3. Overlaps between PCG annotations and TE annotations were calculated using the intersect function of BEDTools v2.27.1. PCGs are considered as overlapped by TEs when more than 50% of the regions of each PCG are overlapped by each TE and vice versa. Numbers of PCGs overlapped by TEs and TEs overlapped by PCGs per cluster were plotted using the ggplot2 package in R (Wickham, 2016). Distance between closest TE and PCG pairs per cluster was calculated by using the closest function of BEDTools.

Gene Expression analysis of *A. agrestis*

Gene expression data from (Li et al., 2020) were downloaded from the SRA (PRJNA574453) or ENA (PRJEB34743) and processed with RSEM v1.2.31 (Li and Dewey, 2011) and STAR v2.5.2a (Dobin et al., 2013). Transcript Per Million (TPM) values were averaged from two or three biological replicates each condition and used for further analyses. Association of each peak over PCGs are calculated using the intersect function of BEDtools v2.27.1. PCGs were considered as overlapped by each chromatin mark when more than 50% of the regions of each PCG were overlapped by each peak. Average expression levels per chromatin peak were plotted using ggplot2 package in R. Heatmaps of expression levels of PCGs in cluster 1 and 2 over various conditions were plotted using pheatmap function in R.

Genome wide profiling of 5mC in A. agrestis

Genomic DNA was extracted from 100 mg of 4 weeks old gametophyte tissue of *A. agrestis* using Nucleon PhytoPure (cytiva). Sequencing libraries for genome wide DNA methylation profiles were generated from 200 ng of genomic DNA using NEBNext® Enzymatic Methyl-seq Kit (New England Biolabs). These libraries were sequenced on a Illumina NextSeq 2000 to generate 100 bp paired end reads.

EMseq data analysis

The bam files of EM-seq reads were sorted with SAMtools v1.9 and converted to fastq format using bamtofastq function of BEDTools v2.27.1, then trimmed with Trim Galore (https://github.com/FelixKrueger/TrimGalore). Bisulfite converted reference genome was prepared from *A. agrestis* Oxford strain genome sequence using Bismark (Krueger and Andrews, 2011). Trimmed reads were mapped to the bisulfite genome using the Bowtie2 option of Bismark. Duplicates were removed using the deduplicate function in Bismark. Cytosine methylation reports were created from deduplicated reads using the bismark_methylation_extractor function in Bismark. Each cytosine which is covered by at least four reads was used for further analyses. Methylation ratio of each cytosine was calculated and summarized to a bed file. These bed files were converted to bigwig files using bedGraphToBigWig (Kent et al., 2010) and used as inputs for the computMatrix function in deepTools v3.3.1. Aggregate profile plots of matrices were plotted with the plotProfile function in deepTools v3.3.1.

Tables

Table 1. Expression levels of homologs of DNA methyltransferase in *A. agrestis*.

Gene name, gene ID and averaged TPM value in each developmental stage are shown.

Gene name	Gene ID	Gametophyte_2w	Gmetophyte_1m	Gametophyte_2m	Sporophyte<5mm	Sporophyte5-10mm	Sporophyte>10mm
AaMET	AagrOXF_evm.model.utg000084l.8	9.78	11.78	7.64	6.2	5.81	5.315
AaCMTa	AagrOXF_evm.model.utg000009l.436	14.365	15.36	11.22	10.005	11.035	8.405
AaCMTb	AagrOXF_evm.model.utg000045l.50	15.61	17.87	12.44	12.19	12.765	8.64
AaDRMa	AagrOXF_evm.model.utg000006l.710	7.11	6.435	7.015	3.39	3.275	3.6
AaDRMb	AagrOXF_evm.model.utg000063l.237	7.345	10.29	8	12.885	15.985	13.985
AaDNMT3	AagrOXF_evm.model.utg000006l.43	54.61	47.54	43.73	32.81	34.205	41.145

References

Allis, C.D. and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat. Rev. Genet. **17**: 487–500.

Allshire, R.C. and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. Nat. Rev. Mol. Cell Biol. **19**: 229–244.

Alonge, M., Lebeigle, L., Kirsche, M., Aganezov, S., Wang, X., Lippman, Z.B., Schatz, M.C., and Soyk, S. (2021). Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. bioRxiv: 2021.11.18.469135.

Bannister, A.J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. **21**: 381–395.

Bannister, A.J., Schneider, R., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. J. Biol. Chem. **280**: 17732–17736.

Bewick, A.J., Niederhuth, C.E., Ji, L., Rohr, N.A., Griffin, P.T., Leebens-Mack, J., and Schmitz, R.J. (2017). The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. Genome Biol. **18**: 65.

Borg, M., Papareddy, R.K., Dombey, R., Axelsson, E., Nodine, M.D., Twell, D., and Berger, F. (2021). Epigenetic reprogramming rewires transcription during the alternation of generations in Arabidopsis. Elife **10**.

Bowman, J.L. et al. (2017). Insights into Land Plant Evolution Garnered from the

Marchantia polymorpha Genome. Cell **171**: 287–304.e15.

Bowman, J.L. et al. (2022). The Renaissance and Enlightenment of Marchantia as a model system. Plant Cell.

Déléris, A., Berger, F., and Duharcourt, S. (2021). Role of Polycomb in the control of transposable elements. Trends Genet.

Del Prete, S., Mikulski, P., Schubert, D., and Gaudin, V. (2015). One, Two, Three: Polycomb Proteins Hit All Dimensions of Gene Regulation. Genes **6**: 520–542.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**: 15–21.

Domb, K., Katz, A., Harris, K.D., Yaari, R., Kaisler, E., Nguyen, V.H., Hong, U.V.T., Griess, O., Heskiau, K.G., Ohad, N., and Zemach, A. (2020). DNA methylation mutants in Physcomitrella patens elucidate individual roles of CG and non-CG methylation in genome regulation. Proc. Natl. Acad. Sci. U. S. A. **117**: 33700–33710.

Du, J. et al. (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell **151**: 167–180.

Gan, E.-S., Xu, Y., and Ito, T. (2015). Dynamics of H3K27me3 methylation and demethylation in plant development. Plant Signal. Behav. **10**: e1027851.

Gardner, K.E., Allis, C.D., and Strahl, B.D. (2011). Operating on chromatin, a colorful language where context matters. J. Mol. Biol. **409**: 36–46.

Grau-Bové, X. et al. (2022). A phylogenetic and proteomic reconstruction of eukaryotic chromatin evolution. Nat Ecol Evol **6**: 1007–1023.

Harris, B.J., Clark, J.W., Schrempf, D., Szöllősi, G.J., Donoghue, P.C.J., Hetherington, A.M., and Williams, T.A. (2021). Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. bioRxiv: 2021.10.28.466308.

Ikeda, Y., Nishihama, R., Yamaoka, S., Arteaga-Vazquez, M.A., Aguilar-Cruz, A., Grimanelli, D., Pogorelcnik, R., Martienssen, R.A., Yamato, K.T., Kohchi, T., Hirayama, T., and Mathieu, O. (2018). Loss of CG Methylation in Marchantia polymorpha Causes Disorganization of Cell Division and Reveals Unique DNA Methylation Regulatory Mechanisms of Non-CG Methylation. Plant Cell Physiol. **59**: 2421–2431.

Jacob, Y., Feng, S., LeBlanc, C.A., Bernatavichute, Y.V., Stroud, H., Cokus, S., Johnson, L.M., Pellegrini, M., Jacobsen, S.E., and Michaels, S.D. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat. Struct. Mol. Biol. **16**: 763–768.

Jamge, B. and Berger, F. (2022). Diversification of chromatin organization in eukaryotes. Curr. Opin. Cell Biol. **74**: 1–6.

Jamge, B., Lorković, Z.J., Axelsson, E., Yelagandula, R., Akimcheva, S., and Berger, F.

(2022). Transcriptional activity is shaped by the chromatin landscapes in Arabidopsis. bioRxiv: 2022.06.02.494419.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics **26**: 2204–2207.

Kornberg, R.D. and Lorch, Y. (2020). Primary Role of the Nucleosome. Mol. Cell **79**: 371–375.

Krueger, F. and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics **27**: 1571–1572.

Lang, D. et al. (2018). The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. Plant J. **93**: 515–533.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**: 357–359.

Larson, A.G. and Narlikar, G.J. (2018). The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. Biochemistry **57**: 2540–2548.

Leng, X., Thomas, Q., Rasmussen, S.H., and Marquardt, S. (2020). A G(enomic)P(ositioning)S(ystem) for Plant RNAPII Transcription. Trends Plant Sci. **25**: 744–764.

Li, B. and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12**: 323.

Li, F.-W. et al. (2020). Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. Nat Plants **6**: 259–272.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**: 2078–2079.

Li, X., Harris, C.J., Zhong, Z., Chen, W., Liu, R., Jia, B., Wang, Z., Li, S., Jacobsen, S.E., and Du, J. (2018). Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. Proc. Natl. Acad. Sci. U. S. A. **115**: E8793–E8802.

Malik, H.S. and Henikoff, S. (2003). Phylogenomics of the nucleosome. Nat. Struct. Biol. **10**: 882–891.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17**: 10–12.

Montgomery, S.A. et al. (2020). Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. Curr. Biol. **30**: 573–588.e7.

Müller, J. and Verrijzer, P. (2009). Biochemical mechanisms of gene regulation by

polycomb group protein complexes. Curr. Opin. Genet. Dev. **19**: 150–158.

Naish, M. et al. (2021). The genetic and epigenetic landscape of the Arabidopsis centromeres. Science **374**: eabi7489.

Ninova, M., Fejes Tóth, K., and Aravin, A.A. (2019). The control of gene expression and cell identity by H3K9 trimethylation. Development **146**.

Ou, S. et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. **20**: 275.

Puttick, M.N., Morris, J.L., Williams, T.A., Cox, C.J., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Schneider, H., Pisani, D., and Donoghue, P.C.J. (2018). The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte. Curr. Biol. **28**: 733–745.e2.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**: 841–842.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. **44**: W160–5.

Sato, S., Dacher, M., and Kurumizaka, H. (2022). Nucleosome Structures Built from Highly Divergent Histones: Parasites and Giant DNA Viruses. Epigenomes **6**.

Schuettengruber, B. and Cavalli, G. (2009). Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development **136**: 3531–3542.

Schwämmle, V., Sidoli, S., Ruminowicz, C., Wu, X., Lee, C.-F., Helin, K., and Jensen, O.N. (2016). Systems Level Analysis of Histone H3 Post-translational Modifications (PTMs) Reveals Features of PTM Crosstalk in Chromatin Regulation. Mol. Cell. Proteomics **15**: 2715–2729.

Scott, K.C. and Sullivan, B.A. (2014). Neocentromeres: a place for everything and everything in its place. Trends Genet. **30**: 66–74.

Sequeira-Mendes, J., Aragüez, I., Peiró, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S.E., Bastolla, U., and Gutierrez, C. (2014). The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. Plant Cell **26**: 2351–2366.

Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu. Rev. Biochem. **81**: 65–95.

Sigman, M.J. and Slotkin, R.K. (2016). The First Rule of Plant Transposable Element Silencing: Location, Location, Location. Plant Cell **28**: 304–313.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat. Struct. Mol. Biol. **21**: 64–72.

Takuno, S., Ran, J.-H., and Gaut, B.S. (2016). Evolutionary patterns of genic DNA methylation vary across land plants. Nat Plants **2**: 15222.

Talbert, P.B. and Henikoff, S. (2021a). Histone variants at a glance. J. Cell Sci. **134**.

Talbert, P.B. and Henikoff, S. (2021b). The Yin and Yang of Histone Marks in Transcription. Annu. Rev. Genomics Hum. Genet. **22**: 147–170.

Vaisvila, R. et al. (2021). Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. Genome Res.

Villarreal, J.C. and Renner, S.S. (2013). Correlates of monoicy and dioicy in hornworts, the apparent sister group to vascular plants. BMC Evol. Biol. **13**: 239.

Walker, J., Zhang, J., Liu, Y., Vickers, M., Dolan, L., Nakajima, K., and Feng, X. (2021). Extensive N4 cytosine methylation is essential for Marchantia sperm function. Cold Spring Harbor Laboratory: 2021.02.12.428880.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis.

Widiez, T., Symeonidi, A., Luo, C., Lam, E., Lawton, M., and Rensing, S.A. (2014). The chromatin landscape of the moss Physcomitrella patens and its dynamics during development and drought stress. Plant J. **79**: 67–81.

Xiao, J. and Wagner, D. (2015). Polycomb repression in the regulation of growth and development in Arabidopsis. Curr. Opin. Plant Biol. **23**: 15–24.

Yaari, R., Katz, A., Domb, K., Harris, K.D., Zemach, A., and Ohad, N. (2019). RdDM-independent de novo and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs. Nat. Commun. **10**: 1613.

Yelagandula, R., Osakabe, A., Axelsson, E., Berger, F., and Kawashima, T. (2017). Genome-Wide Profiling of Histone Modifications and Histone Variants in Arabidopsis thaliana and Marchantia polymorpha. Methods Mol. Biol. **1610**: 93–106.

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

Zhang, J. et al. (2020). The hornwort genome and early land plant evolution. Nat Plants **6**: 107–118.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. **9**: R137.

Zhou, K., Gaullier, G., and Luger, K. (2019). Nucleosome structure and dynamics are coming of age. Nat. Struct. Mol. Biol. **26**: 3–13.

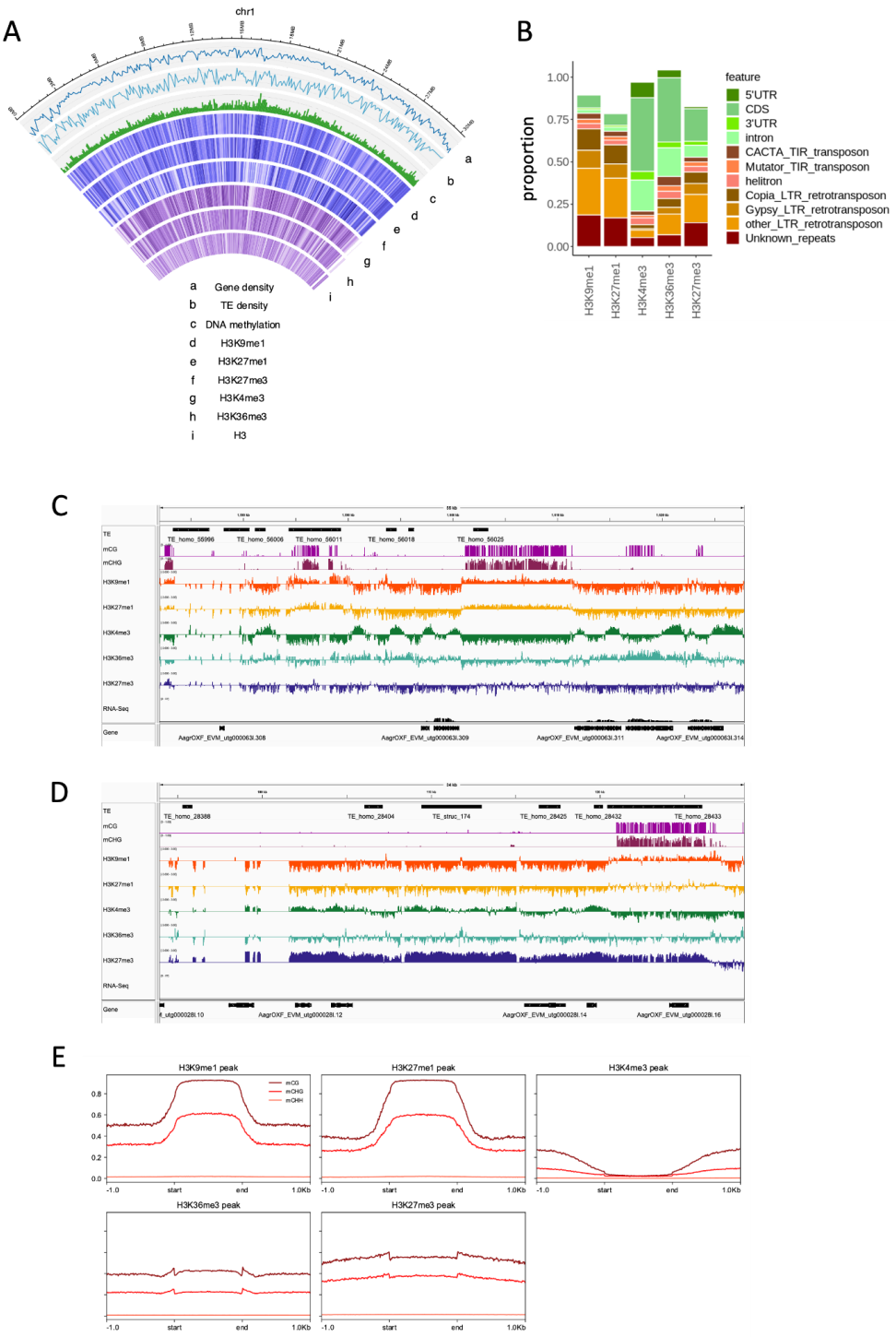

Figure 1. Association of chromatin marks with genomic features

(A) Circos plot showing epigenomics features over the chromosome 1 of *A. agrestis*. From top to bottom, the following information was depicted with a window size of 100 kb. Gene density, TE density, level of DNA methylation, densities of H3K9me1, H3K27me1, H3K27me3, H3K4me3 and H3K36me3 peaks

(B) Distribution of chromatin marks over genomic features. The total length of chromatin mark peaks overlapping specified genomic features was divided by the total length of peaks of chromatin marks to determine each proportion. Features which cover less than 0.5% of chromatin mark peaks are not shown.

(C) and (D) Integrative Genomics Viewer (IGV) browser screenshot demonstrating vicinity of TEs covered by H3K9me1 and expressed genes (C) and TEs covered by H3K27me3 (D). The region shown is 55 kb (C) or 34 kb (D) in length and from the scaffolds utg000063l and utg000028l, respectively. Chromatin mark tracks are bigwig files scaled by H3 coverage in 10-bp windows. "TE" and "Gene" tracks are annotation files for TEs and genes, respectively. "RNA-seq" track is a bigwig of mapped RNA-seq reads from vegetative tissue (ref). Scales are noted in square brackets beside each track.

(E) Profile plot of CG, CHG and CHH methylation levels over chromatin mark peaks. Each chromatin mark peak is scaled to 1 kb and sequences 1 kb upstream and downstream are included. Average methylation over 10 bp bins is plotted.

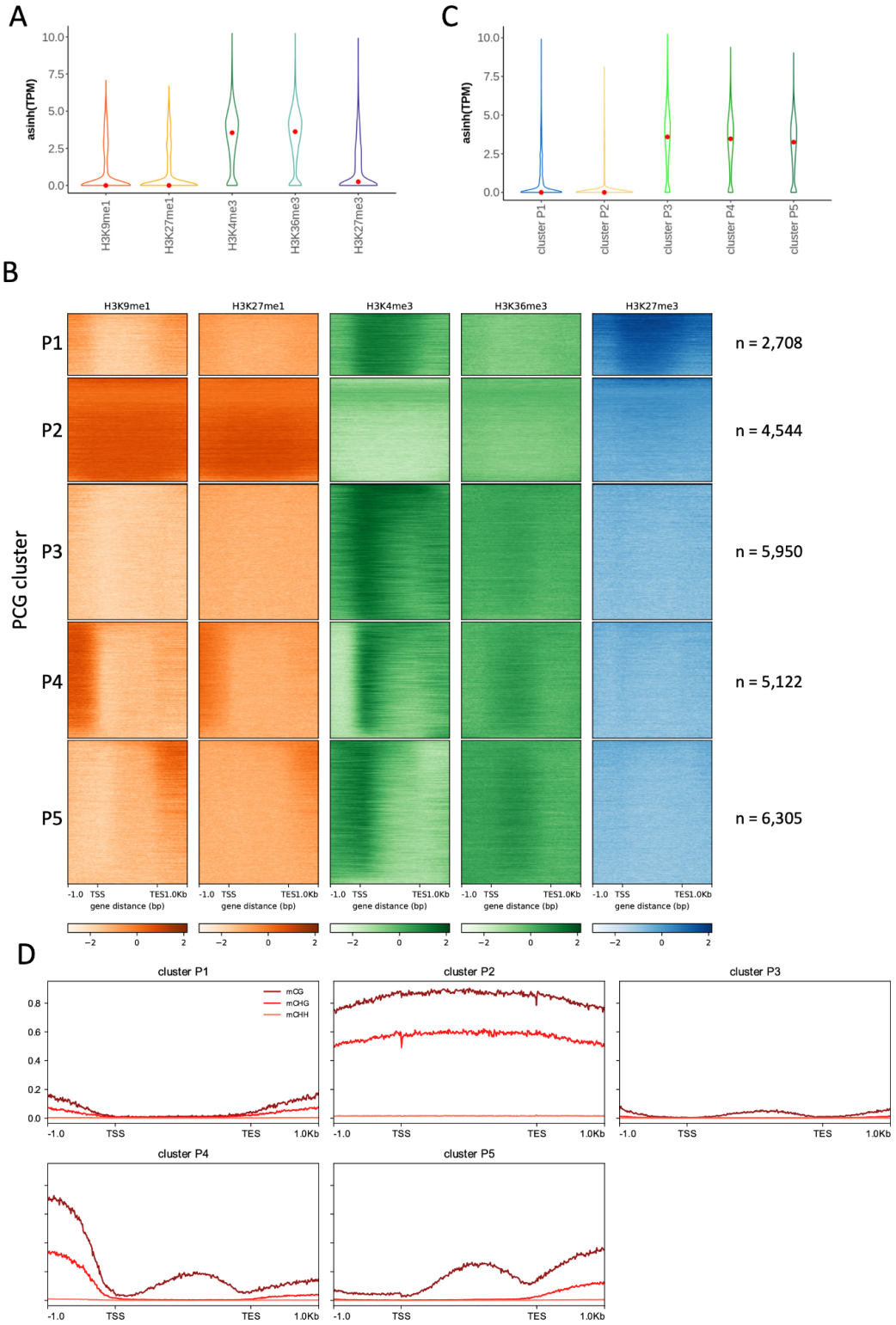

Figure 2. Association of chromatin marks on genes

(A) Violin plot showing expression level of PCGs associated with chromatin marks. Width is relative to the density of genes. Red dots indicate median expression values.

(B) Heatmap of k-means clustering of genes based on chromatin marks. Prevalence of each mark (columns) based on its score normalized against H3 signals per 10 bp bins. Gene body of each gene is scaled to 2 kb and sequences 1 kb upstream and downstream are included. Each color (orange, green and blue) stands for enrichment and white for depletion. Each row corresponds to one gene, with multiple genes grouped into blocks that have been defined as clusters P1 through P5.

(C) Violin plot showing expression level of PCGs per PCG cluster. Width is relative to the density of genes. Red dots indicate median expression values.

(D) Profile plot of CG, CHG and CHH methylation levels over PCGs per PCG cluster. Gene body of each PCG is scaled to 2 kb and sequences 1 kb upstream and downstream are included. Average methylation over 10 bp bins is plotted.

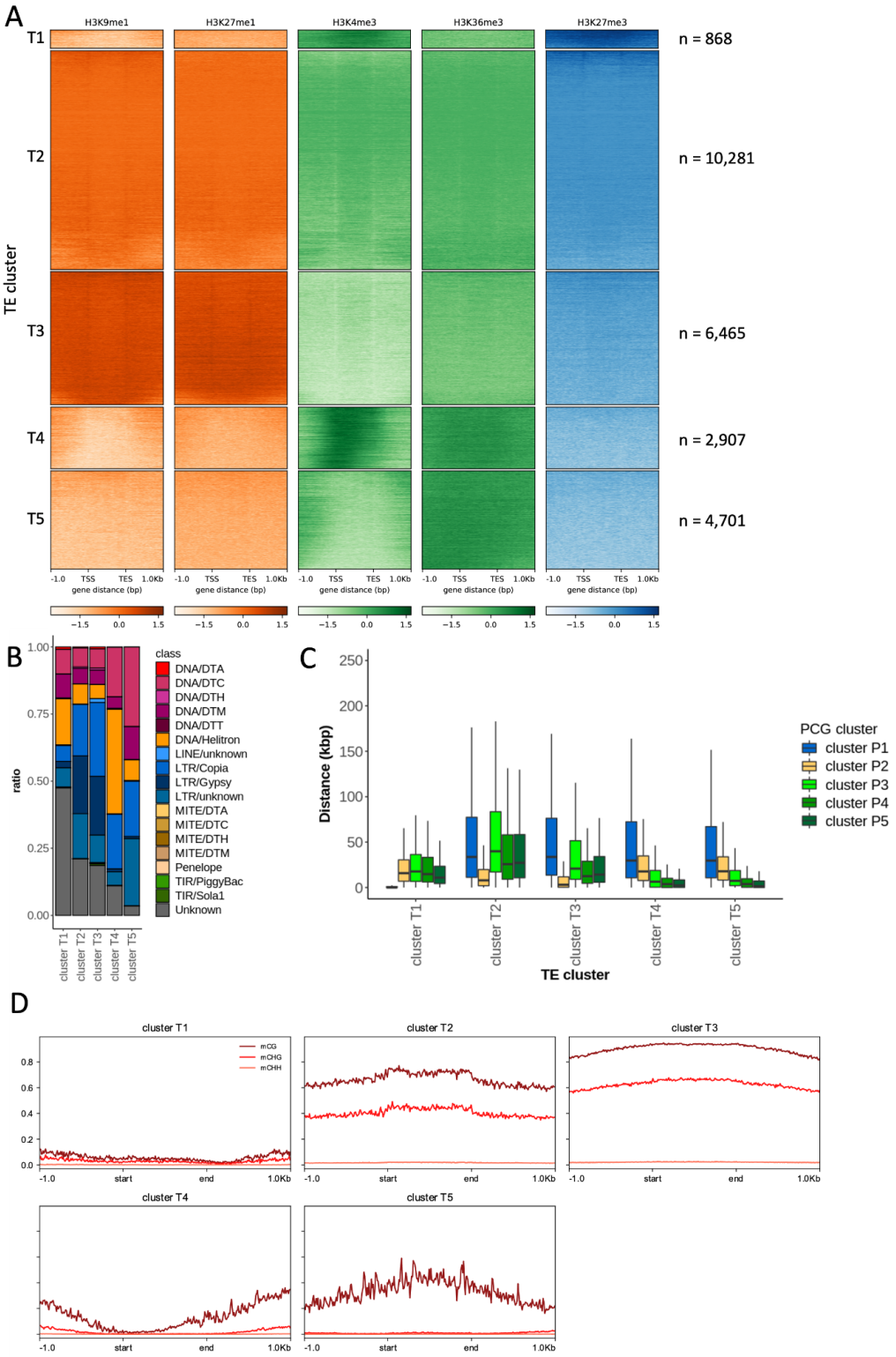

Figure 3. Association of chromatin marks on TEs

(A) Heatmap of k-means clustering of TEs based on chromatin marks. Prevalence of each mark (columns) based on its score normalized against H3 signals per 10 bp bins. Each TE annotation is scaled to 1 kb and sequences 1 kb upstream and downstream are included. Each color (orange, green and blue) stands for enrichment and white for depletion. Each row corresponds to one TE, with multiple TEs grouped into blocks that have been defined as clusters T1 through T5.

(B) Stacked bar chart indicating proportion of TE families in each TE cluster.

(C) Boxplot of distances between each TE and the nearest gene per gene cluster. Briefly each TE is compared to all PCGs belonging to a PCG cluster to find its nearest neighbor. TEs are divided based on the TE cluster they belong to. Distances are in kilobases (kbp). Colored boxes represent interquartile range, and lines represent median values. Outliers are not shown.

(D) Profile plot of CG, CHG and CHH methylation levels over TEs per TE cluster. Each TE annotation is scaled to 1 kb and sequences 1 kb upstream and downstream are included. Average methylation over 10 bp bins is plotted.