The expansion and diversification of epigenetic regulatory networks underpins major transitions in the evolution of land plants

Romy Petroll*[1], Ranjith K. Papareddy*[2], Rafal Krela*[3], Alice Laigle[1], Quentin Riviere[3], Kateřina Bišová[4], Iva Mozgová[3]† & Michael Borg[1]†

[1] Department of Algal Development and Evolution, Max Planck Institute for Biology, Tübingen, Germany
[2] Gregor Mendel Institute for Molecular Plant Biology, Vienna Biocenter, Vienna, Austria
[3] Biology Centre CAS - Institute of Plant Molecular Biology, České Budějovice, Czech Republic
[4] Institute of Microbiology CAS – Centre Algatech, Třeboň, Czech Republic

* These authors contributed equally to this work.
† Correspondence: iva.mozgova@umbr.cas.cz; michael.borg@tuebingen.mpg.de

Abstract

Epigenetic silencing is essential for regulating gene expression and cellular diversity in eukaryotes. While transposable elements (TEs) are mostly silenced with DNA and H3K9 methylation, gene silencing is mediated by H3K27me3, an epigenetic mark deposited by the Polycomb repressive complex 2 (PRC2). Despite the major role epigenetic silencing plays in the development of multicellular eukaryotes, little is known about how epigenetically-controlled regulatory networks were shaped over evolutionary time. Here, we analyse epigenomes from a diverse group of species across the green lineage and infer the chronological epigenetic recruitment of genes that occurred during land plant evolution. We first reveal the nature of plant heterochromatin in the unicellular green microalga *Chlorella sorokiniana* and identify a substantial number of genes marked with H3K27me3, highlighting the deep origin of PRC2-regulated genes in the green lineage. By incorporating genomic phylostratigraphy, we show how genes of differing evolutionary age are partitioned into distinct epigenetic states in plants, with evolutionarily young genes incorporated into developmental programs controlled by H3K9 methylation in *Arabidopsis*. We further reveal a major wave of PRC2 recruitment to genes that emerged during land terrestrialisation and flowering plant evolution, and identify an ancestral PRC2 network with a shared functional topology in green algae through to land plants, providing a glimpse of the earliest types of genes regulated by PRC2 during the course of plant evolution. Finally, we analyse the potential regulation of these ancestral PRC2 target genes and find a strong enrichment of motifs bound by ancient AP2/ERF transcription factors (TFs) known to interact with PRC2, which we hypothesise were key determinants in shaping some of the first gene regulatory networks controlled by PRC2 in plants. Our data thus reveal pivotal epigenetic adaptations that occurred during a significant period in the evolutionary history of plants, which likely contributed to key regulatory innovations that influenced major morphological and developmental change into the modern-day. More broadly, our findings offer insight into the evolutionary dynamics and molecular triggers that drive the adaptation and elaboration of epigenetic regulation, laying the groundwork for its future consideration in other major eukaryotic lineages.

Introduction

A key aspect of complex multicellularity involves the differential regulation of gene expression in space and time during development. This spatiotemporal regulation is orchestrated by various regulatory genes, including transcription factors (TFs) and cell signalling pathways, but it also involves regulation at the level of the epigenome (Zeitlinger and Stark 2010; Sparks et al. 2013). Epigenetic regulation helps reinforce cell identity in multicellular eukaryotes by dictating which parts of the genome should remain transcriptionally silent (Zhu & Reinberg, 2011; Borg *et al.*, 2021a). As a consequence, epigenetic patterns are highly variable between different cell types and are reconfigured during multicellular development to specify cellular and tissue differentiation (Reik *et al.*, 2001; Kawashima & Berger, 2014). In most eukaryotes, these patterns are shaped by three major types of heritable epigenetic marks - the methylation of DNA cytosine bases or by methylation of lysine 9 and 27 on the tail of histone H3 (i.e., H3K9 and H3K27 methylation). H3K27me3 marks form facultative heterochromatin and are deposited by PRC2 to silence key development genes (Loubiere *et al.*, 2019; Baile *et al.*, 2021; Blackledge & Klose, 2021a). In contrast, DNA and H3K9 methylation often act together in a feedback loop to form constitutive heterochromatin, which interfaces with specific classes of short-interfering RNAs (siRNAs) to silence transposable elements (TEs) and repeat-rich sequences (Feng *et al.*, 2010; Du *et al.*, 2015). This division of labour between facultative and constitutive heterochromatin thus bestows complex eukaryotic genomes with a robust transcriptional control over genome stability and cell identity.

The conserved and ubiquitous nature of epigenetic silencing across the eukaryotic tree of life has raised important questions about its evolutionary origins. While some studies have hypothesised that epigenetic silencing evolved to control invading "parasitic" elements like TEs and exogenous retroviruses (Déléris *et al.*, 2021), others propose that TEs accumulate in eukaryotic genomes because of, not despite, epigenetic silencing mechanisms by suppressing homologous recombination of repetitive regions (Fedoroff, 2012). PRC2 is likely to have emerged prior to the diversification of eukaryotes and consists of a conserved functional core of three-to-four subunits - E(z), ESC, Su(z)12 and p55 - that are dynamic among distinct eukaryotic lineages (Sharaf *et al.*, 2022). In distant unicellular relatives of multicellular eukaryotes, PRC2 largely deposits H3K27me3 at TEs rather than genes, including in ciliates (*Paramecium;* Frapporti et al. 2019), in unicellular relatives of the red algae (*Cyanidioschyzon merolae;* (Mikulski *et al.*, 2017; Hisanaga *et al.*, 2023) and Stramenopiles (*Phaeodactylum tricornutum;* (Veluchamy *et al.*, 2015; Zhao *et al.*, 2021)). In the unicellular chlorophyte green alga *Chlamydomonas reinhardtii*, a member of an early diverging lineage of the Viridiplantae phylum that led to modern-day plants, H3K27me3 is not detectable despite presence of the H3K27 methyltransferase subunit E(z) (Shaver *et al.*, 2010; Huang *et al.*, 2017; Khan *et al.*, 2018). Perturbed PRC2 activity in *Chlamydomonas* also results in the derepression of some transposons (Shaver *et al.*, 2010), suggesting that the ancestral role of PRC2 in TE silencing might also be conserved at the root of the Viridiplantae. However, a lack of genomic level data for H3K27 methylation in unicellular green algae has prevented definitive conclusions.

Because PRC2 plays such a key role in specifying cell differentiation in multicellular eukaryotes, its adaptation towards regulating gene expression could have been a major factor in facilitating the transition to multicellular life from the ancestral unicellular state (Gombar *et al.*, 2014). The emergence of multicellularity in plants, coupled with the transition to land and subsequent emergence of the seed plants, were major events in plant evolution that all coincided with repeated bursts of *de novo* gene emergence,

whole genome duplication and neo-functionalisation (de Vries & Archibald, 2018; Bowman, 2022; Barrera-Redondo *et al.*, 2023). The emergent genetic novelty arising from these events have gained distinct spatial and temporal expression patterns over evolutionary time through increasingly complex layers of gene regulation. The adaptation and elaboration of epigenetic silencing mechanisms during the course of plant evolution might have facilitated this, and has resulted in epigenetic landscapes among land plants where H3K9me2 and H3K27me3 largely repress either TEs or genes, respectively (Vigneau & Borg, 2021). *Marchantia polymorpha* and *Anthoceros agrestis*, two extant representatives of the bryophyte lineage that share a common ancestor with land plants, still bear remnants of this progressive adaptation since PRC2 deposits H3K27me3 at both TEs and genes (Montgomery et al., 2020; Hisanaga et al., 2023). In flowering plants, H3K27me3 predominantly silences genes and is reprogrammed throughout the plant life cycle to mediate tissue-specific gene expression (Vigneau & Borg, 2021), although its impact on TE silencing is also evident albeit to a lesser degree than at genes (Hisanaga *et al.*, 2023; Hure *et al.*, 2024). Studies in *Arabidopsis* pollen suggest that H3K9me2 also silences a substantial number of pollen-specific genes, which are re-activated via epigenetic reprogramming in the vegetative cell (Borg *et al.*, 2021b). Similar reprogramming of DNA and H3K9me2 also occurs in the central cell of the female gametophyte, suggesting that H3K9me2-mediated silencing of genes may be a general feature of flowering plants (Pillot *et al.*, 2010; Ibarra *et al.*, 2012; Park *et al.*, 2016). The extent to which H3K9me2 methylation silences genes in other plant species, however, remains unclear. This raises fundamental questions about how adaptations in epigenetic regulation could have contributed to the emergence of complex multicellularity during land plant evolution. At what time point in evolution did these epigenetic pathways recruit genes? What were the dynamics of this evolutionary adaptation and did this favour particular groups or families of genes? And are the genes silenced by these epigenetic pathways common among all land plants or are they specific to each major clade? Such questions remain largely unresolved, not least through a systematic assessment across a major eukaryotic lineage.

Here, we perform an evolutionary analysis of epigenomes from six distantly related species of the Viridiplantae, including unicellular green algae, bryophytes and flowering plants. We first provide insight into the nature of heterochromatin in the unicellular alga *Chlorella sorokiniana* (Trebouxiophyceae), a member of the chlorophyte lineage that diverged from the streptophyte lineage over 1 billion years ago (Moczydlowska *et al.*, 2011; Leliaert *et al.*, 2011). We show heterochromatin in *Chlorella* is defined by H3K27me3 but is partitioned into two distinct forms by the co-occurrence or absence of H3K9me2. A substantial number of *Chlorella* genes marked solely with H3K27me3 in the absence of adjacent TEs highlight how PRC2-regulated genes predate the emergence of land plants. By aging genes across the plant kingdom using genomic phylostratigraphy, we reveal how evolutionarily young genes are preferentially silenced with H3K9me in flowering plants, some of which form part of the gene regulatory networks activated during reproductive development in *Arabidopsis*. In contrast, genes that emerged during Streptophyte evolution are preferentially silenced with H3K27me3, highlighting a major period of PRC2 recruitment to genes that emerged during this foundational period of plant evolution. We also show how PRC2 regulates a conserved network of gene families across land plants as well as in green microalgae, which includes ancient regulators of the plant life cycle and architecture. Finally, we analyse *cis*-regulatory elements at the ancestral PRC2-regulated genes and reveal prominent enrichment for known PRC2-interacting TFs belonging to the AP2/ERF family, which we postulate were key players that could have helped establish the first PRC2-controlled gene networks in plants.

Results

H3K27me3 marks both TEs and genes in the chlorophyte green microalga *Chlorella sorokiniana*

The Viridiplantae (or green lineage) represents the most dominant phylum in the Archaeplastida kingdom and is represented by modern-day chlorophyte and streptophyte algae together with their land plant relatives. Existing chromatin profiles from chlorophyte green algae are only available for the model species *Chlamydomonas reinhardtii* (Ngan *et al.*, 2015; Strenkert *et al.*, 2022), which has been shown to lack H3K27me3, likely due to the distinct amino acid residue changes surrounding K27 on histone H3 in this species (Shaver et al. 2010). This has made it difficult to assess the role of PRC2 in the unicellular representatives of the green lineage. To determine whether the lack of H3K27me3 is a feature common to all green algae, we performed a phylogenetic analysis of histone H3 variants from representatives of the Chlorophyta and Streptophyta, which diverged from each other *circa* 1250 mya (Evanovich *et al.*, 2020). Interestingly, the characteristic S28T mutation and deletion of residue A29 on the tail of *Chlamydomonas* histone H3 was only present in its close relative *Volvox carteri* (**Supplemental Fig. S1A**). These specific residue changes are likely to alter methylation states of K27 on *Chlamydomonas* histone H3, as is the case for atypical histone H3 variants in flowering plants (Borg *et al.*, 2021a). This suggests that the S28T mutation and A29 deletion are a feature specific to the volvocine green algae within the Chlamydomonadales order. A related species without these histone H3 mutations is the unicellular green alga *Chlorella sorokiniana* (hereon in *Chlorella*), which belongs to the Trebouxiophyceae, a sister clade to the Chlorophyceae (**Supplemental Fig. S1A**). We thus selected *Chlorella* as a representative species to clarify the role of H3K27me3 in chlorophytes (**Fig. 1A)**.

Using chromatin immunoprecipitation coupled with DNA sequencing (ChIP-seq), we obtained genomic profiles for the histone marks H3K4me3, H3K9me2 and H3K27me3 from *Chlorella* cells synchronised in G1 phase (**Fig. 1B-C; Supplemental Fig. S1B**). We observed that each histone mark was enriched at different genomic features as in several land plants, suggesting that chromatin functions similarly in this unicellular green alga (**Fig. 1D**). We next assessed the deposition profile of the histone marks across the *Chlorella* genome. Most of the contigs in *Chlorella* were characterised by one or more TE-rich hotspots that were strongly enriched for both H3K9me2 and H3K27me3 but depleted in active H3K4me3 marks (**Fig. 1B-C**), which likely represent the centromeres (Wang *et al.*, 2024). Consistently, H3K9me2 and H3K27me3 peaks were strongly enriched at TEs (**Fig. 1D**) and were deposited at many of the same TEs (cluster 1; 1038), while H3K27me3 was enriched at a second group independently of H3K9me2 (cluster 2; 346) (**Fig. 1E-F**). Closer inspection revealed that the class of repeats present in these clusters also differed (**Fig. 1E**). Co-deposition of H3K9me2 and H3K27me3 mostly occurred at LTR retrotransposons, whereas sole deposition of H3K27me3 mostly occurred at DNA transposons belonging to the CACTA and hAT terminal inverted repeat (TIR) element family, as well as Gypsy LTRs compared to the genome average (**Fig. 1E**). This suggests that two types of heterochromatin mediate TE silencing in *Chlorella*, which are both composed of H3K27me3 but differentiated by the presence or absence of H3K9me2. While H3K9me2 was largely restricted at the TE-rich hotspots in *Chlorella*, H3K27me3 was dispersed across the chromosomes (**Fig. 1B**). Clustering of the chromatin profiles over genes revealed that H3K27me3 was also deposited at multiple genes in *Chlorella* (12%; 1548). Functional classification of these genes revealed an over-representation of 24 functional terms in the GO, KEGG and KOG classifications. They highlighted the targeting of transposon-encoded proteins and related processes, including DNA

integration and RNA-dependent DNA replication, but were also related to protein phosphorylation, proteolysis and carbohydrate metabolism (**Supplemental Fig. S1C**). Thus, H3K27me3 does not appear to solely silence TEs in *Chlorella* but also marks several protein-coding genes, suggesting that the recruitment of genes by PRC2 likely predated the emergence of land plants.

Evolutionarily young and old genes are partitioned into distinct epigenetic states in plants

Land plant evolution is characterised by repeated bursts of *de novo* gene emergence and whole genome duplications that coincided with the transition to complex multicellularity, terrestrialisation and the emergence of flowering plants (Bowles *et al.*, 2020; Barrera-Redondo *et al.*, 2023). Many of the emergent genes arising from these events have gained distinct developmental expression patterns over evolutionary time, which is likely to involve epigenetic regulation. To gain insight into how epigenetic recruitment of these genes might have occurred during land plant evolution, we re-analysed ChIP-seq datasets from five Viridiplantae species **(Fig. 1A; Supplemental Table S1)** - the two bryophytes *Physcomitrium patens* and *Anthoceros agrestis* and the three flowering plants *Spirodela polyrhiza*, *Oryza sativa* and *Arabidopsis thaliana* (hereon in referred to by their respective genus). We then assigned phylostratigraphic ages to genes in each plant species and assessed the relative enrichment or depletion of genes marked with H3K9me1/2 and H3K27me3 in each phylogenetic rank (**Supplemental Fig. S2; Supplemental Table S2**). Strikingly, our analysis revealed a consistent and biased pattern that correlated with evolutionary age in divergent species of plants (**Fig. 2A-F**).

Among the land plants, genes marked with H3K9me2 (or H3K9me1 in *Anthoceros*) were biased towards those with limited or untraceable homology in other species, so-called taxonomically restricted genes (TRGs) (**Fig. 2B-F**). This bias was also evident at the level of DNA methylation, with TRGs having a higher rate of methylated cytosines in all contexts compared to older genes (**Supplemental Fig. S3**). In *Arabidopsis*, H3K27me1-marked genes also showed a pattern similar to that of H3K9me2-marked genes (**Fig. 2F**), further highlighting the increased tendency of TRGs to be silenced with constitutive heterochromatin. To confirm whether the TRGs marked with H3K9me2 have any functional relevance in *Arabidopsis*, we assessed their expression pattern across development. We further calculated *tau* scores for each gene to determine whether these had a narrow (*tau* > 0.8) or broad (*tau* < 0.5) expression patterns (Lüleci & Yılmaz, 2022). H3K9me2-marked genes had a narrow range of very high *tau* scores compared with all *Arabidopsis* genes of the equivalent phylogenetic rank, suggesting restricted expression patterns during development (**Supplemental Fig. S4A-B**). TRGs specific to the Brassicaceae (ranks 8-11) had the highest and narrowest range of *tau* scores, with those marked with H3K9me2 less likely to be transcribed than older genes (ranks 1-7) (**Supplemental Fig. S4B**). The few H3K9me2-marked TRGs that were transcribed tended to do so in reproductive stages like microspores, pollen and early zygotes (**Supplemental Fig. S4B**). No such bias was evident for H3K27me3-marked genes, which had a very narrow range of very high *tau* scores, consistent with the role of PRC2 in restricting the expression of genes during development (**Supplemental Fig. S4C**).

To determine whether the presence of H3K9me2 is causal for silencing TRGs in *Arabidopsis*, we re-analysed gene expression in mutant lines that perturb H3K9me2 function, namely in mutants of the H3K9me2-binding protein AGDP1, which links DNA methylation with H3K9me2, and triple mutants of SUVH4/5/6, the SET domain proteins responsible for H3K9me2 deposition (Zhang *et al.*, 2018). TRGs were significantly over-represented among the genes derepressed in *agdp1* and *suvh456* mutants, which was not

the case for down-regulated genes (**Supplemental Fig. S4D-E**). The H3K9me2-marked TRGs derepressed in *suvh456* mutants had highly enriched expression in reproductive cell types like pollen, early zygotes and floral tissues, suggesting that H3K9me2 silencing is likely to be developmentally relevant and selectively reprogrammed throughout the life cycle (**Supplemental Fig. S4F**). Collectively, these results show that H3K9me is deposited at a substantial number of genes across land plants but is biased towards evolutionarily young TRGs, which have been incorporated into gene expression programs that are mainly activate during reproductive development in *Arabidopsis*.

In contrast to H3K9me2, genes marked with H3K27me3 in land plants were strongly biased towards those that arose after the emergence of the Streptophyta, whereas young TRGs showed a relative depletion for H3K27me3 (**Fig. 2B-F**). This coincides with major events in eukaryotic evolution when plants transitioned to complex multicellularity and conquered land (Umen, 2014). H3K27me3-marked genes were also enriched among genes that emerged in the Bryopsida (or true mosses) in *Physcomitrium* and to *Anthoceros*-specific genes (**Fig. 2B-C**). In the flowering plants *Spirodela*, *Oryza* and *Arabidopsis*, H3K27me3-marked genes were also biased towards those that arose during the evolution of flowering plants (Magnoliopsida) and their respective phylogenetic order (**Fig. 2D-F**). Neither H3K9me- nor H3K27me3-marked genes showed any bias in the oldest phylogenetic rank (i.e., ancient genes that emerged in the first cellular organisms), despite the fact that the majority of genes were classified within this rank in all five land plants species (**Fig. 2B-F)**. The unicellular green alga *Chlorella* showed a different pattern compared to the land plants as TRGs were instead enriched among H3K27me3-marked genes (**Fig. 2A**), which incidentally harbours a distinct form of constitutive heterochromatin where H3K9me co-localises together with H3K27me3 mostly at TEs (**Fig. 1F-G**). The bias of H3K27me3 at *Chlorella*-specific genes is suggestive of a recent wave of PRC2 recruitment to genes that occurred during evolution of the Trebouxiophyceae.

A gene ontology (GO) enrichment analysis revealed that the *Arabidopsis* genes marked with H3K9me2 and H3K27me3 were enriched for core metabolic processes within the oldest phylogenetic ranks, whereas the H3K27me3-marked genes that emerged in the Streptophytina and Magnoliopsida (ranks 5 and 6) contained several developmental processes (**Supplemental Fig. S5; Supplemental Table S3**). These foundational periods in plant evolution are associated with massive bursts in genomic novelty (Bowles *et al.*, 2020; Barrera-Redondo *et al.*, 2023), suggesting that the PRC2 recruitment to gene families that were founded during this evolutionary period could have been key in establishing novel regulatory programs that facilitated the evolution of increasingly complex plants. In summary, our analysis has revealed a striking correlation between the evolutionary age of plant genes and the type of heterochromatin used to silence their transcription. Evolutionarily younger genes are more likely to be silenced with heterochromatin marks normally associated with TE silencing, whereas developmental genes that emerged after the advent of complex multicellularity in plants are preferentially silenced with facultative H3K27me3 marks.

Epigenetic silencing of repeat elements is distinguished by the age and class of plant TEs

Since distinct forms of heterochromatin preferentially silences genes of differing evolutionary age, we wondered whether this phenomenon would also extend to TEs. Unlike genes however, TEs vary significantly in both sequence and copy number between and even within the same species, making it challenging to assign them to a distinct phylogenetic age. Instead, we estimated TE age in each species by computing their divergence from the consensus sequence of their assigned family, which serves both as a

proxy of evolutionary age and as a prediction of recent transpositional activity (Chalopin *et al.*, 2015). We then compared the relative age and distribution of TEs marked with H3K9me, H3K27me3 or both in the six different species.

In *Chlorella*, the vast majority of TEs were marked with both H3K9me2 and H3K27me3, consistent with their co-localisation at TE-rich hotspots in the genome (**Fig.1F; Fig. 3A**). The low sequence divergence of these TEs suggests that they are relatively young and likely to be recently active in *Chlorella*, which appears to retain very few old and degenerate TEs compared to land plants (**Fig. 3B**). However, the class of TEs marked solely with H3K27me3 in *Chlorella*, while largely young, did include a small proportion of relatively old DNA transposons (**Fig. 3B**). In *Anthoceros*, H3K9me1 and H3K27me3 co-localise at a distribution of TEs that are predominantly young and largely composed of LTR elements (**Fig. 3B**). In contrast, the TEs marked solely with either H3K9me1 or H3K27me3 in *Anthoceros* were relatively much older and shorter than those marked with both epigenetic marks, with the TEs marked solely with H3K27me3 predominantly being DNA transposons (**Fig. 3B-C**). *Physcomitrium* showed a different pattern since H3K9me2 solely marked the majority of TEs that were mostly old and presumably inactive, with very few being marked with both forms of histone methylation (**Fig. 3A-B**). TEs marked solely with H3K27me3 in *Physcomitrium* also included several DNA transposons that were much shorter in length when compared to TEs marked solely with H3K9me2 (**Fig. 3B-C**). The bias of H3K27me3 towards silencing DNA transposons was particularly prominent among the flowering plants, which increased from *Spirodela* to *Oryza* and *Arabidopsis* (**Fig. 3B**). In fact, the TEs marked with H3K27me3 in *Oryza* and *Arabidopsis* were overwhelmingly classified as DNA transposons that were highly divergent and very short in length, suggesting that they largely represent ancient domesticated TEs as has been reported in *Arabidopsis* (Hisanaga *et al.*, 2023; Hure *et al.*, 2024). In contrast, TEs marked with H3K9me2 alone in flowering plants included both LTRs and DNA transposons and appeared to be evolutionarily younger than H3K27me3-marked TEs (**Fig. 3B**). Moreover, unlike in *Chlorella* and *Anthoceros*, very few TEs were marked with both H3K9me and H3K27me3 in *Physcomitrium* and flowering plants, consistent with these two epigenetic pathways diverging in function during land plant evolution (**Fig. 3A**). Our analysis has thus revealed that, unlike for genes, the evolutionary age of TEs does not correlate with distinct forms of heterochromatin. Instead, the epigenetic silencing of TEs appears to be dictated by the age and mode of TE transposition, particularly among flowering plants, where short remnants of DNA transposons are primarily covered with facultative H3K27me3-marked heterochromatin.

Neighbouring transposons spread heterochromatin to evolutionarily young genes

Heterochromatic H3K9me domains typically span across large chromosomal domains to promote silencing of repetitive DNA and TEs, which can sometimes spread beyond their intended targets and repress neighbouring genes (Cutter DiPiazza *et al.*, 2021). We were thus curious whether the tendency of TRGs to be silenced with H3K9me might be caused by the spreading of heterochromatin from neighbouring TEs. To assess this, we compared the chromosomal distribution of the phylogenetically-ranked genes in each species (**Fig. 4**). In *Anthoceros* and *Physcomitrium*, old and young genes were evenly scattered across each chromosome and dispersed between the very high frequency of TEs present in these genomes (**Fig. 4A,D**). With the exception of TRGs, the density of surrounding TEs was comparable among genes of varying evolutionary age (**Fig. 4B,E**). A notably high proportion of *Anthoceros*-specific genes (rank 6) lied in the vicinity of H3K9me1-marked TEs (**Fig. 4B-C**). In comparison, the proportion of genes found within 1kb of

a TE was generally much lower in *Physcomitrium* (**Fig. 4E-F**), which is likely attributed to its significantly larger genome size when compared with *Anthoceros* (Zhang *et al.*, 2020; Li *et al.*, 2020; Bi *et al.*, 2024). Nevertheless, H3K9me2-associated TEs were also more frequently adjacent to *Physcomitrium*-specific genes (rank 7), although at a much-reduced proportion when compared with *Anthoceros* (**Fig. 4E-F**). Thus, TRGs are more inclined to lie adjacent to heterochromatic TEs in both bryophyte species, explaining their increased propensity for silencing with H3K9me1/2 (**Fig. 2C-F**).

In *Spirodela* and *Arabidopsis*, TEs are known to be enriched within pericentromeric chromosomal regions that also happen to be relatively depleted of genes (Kaul *et al.*, 2000; Dombey *et al.*, 2024). While this was indeed the case for genes classified in the oldest phylogenetic ranks, TRGs appeared to have a more equal dispersion across the chromosomes in both *Spirodela* (ranks 7-8) and *Arabidopsis* (ranks 8-11), with no clear depletion within the TE-rich pericentromeric regions (**Fig. 4G,J**). Consistently, the relative proportion of TRGs lying in close proximity of TEs and H3K9me2 peaks was higher when compared with older genes, with a higher frequency of TE insertions within the body of TRGs also more evident (**Fig. 4H-I,K-L**). The proportion of *Arabidopsis* genes of all ages that overlapped with H3K9me2 peaks was also increased when compared to the other land plant species, which likely reflects the compact and gene-dense nature of the *Arabidopsis* genome (**Fig. 4L**). Although the pericentromeric regions of the *Oryza* genome are less obvious than in *Spirodela* and *Arabidopsis*, they nonetheless manifest as distinct chromosomal regions that are relatively enriched with TEs (**Fig. 4M**). While these TE-enriched hotspots were once again depleted of old genes compared to neighbouring chromosomal regions, no obvious depletion was observed for TRGs (ranks 9-11) (**Fig. 4M**). Consistent with the high complement of TEs in *Oryza,* around two-thirds of all genes lied within 1 kb of a TE (**Fig. 4N**). However, TRGs were once again more frequently associated with H3K9me2-marked TEs and had more TE insertions within their gene body compared with older genes (**Fig. 4N-O**). Our results demonstrate that TRGs persist within highly heterochromatic TE-rich regions of flowering plant genomes and are consequently more likely to be recruited for silencing by the spreading of H3K9me2 domains from neighbouring TEs.

Interestingly, despite the obvious TE-rich hotpots present in the *Chlorella* genome, we did not observe a depletion of old genes as seen within the pericentromeres of the flowering plant genomes (**Fig. 4P**). Nevertheless, TRGs in *Chlorella* (ranks 7-8) were found more frequently in the vicinity of an H3K27me3-marked TE, although only subtly more so than older genes (**Fig. 4Q-R**). In fact, most TRGs in *Chlorella* that overlapped with an H3K27me3 domain had no adjacent TEs within 1 kb (**Fig. 4R**). This suggests that PRC2 recruitment of TRGs is not simply caused by H3K27me3 spreading from neighbouring TEs but might rather involve a TE-independent PRC2 targeting mechanism.

PRC2 regulates a conserved group of gene families across the green lineage

Our epigenetic analysis of gene ages has revealed that H3K27me3 preferentially marks genes that emerged during early Streptophyte evolution. PRC2 thus appears to have been recruited to a large swathe of genes during this evolutionary period, which might have occurred early on in the common ancestor of all land plants. We thus wondered whether signatures of this epigenetic adaptation might have persisted as a shared network of PRC2-regulated gene families into the modern-day. To investigate this, we performed an orthology inference analysis to group gene families from the six representative species into 18,415 orthogroups, then calculated the relative proportion of genes marked with H3K27me3 (**Supplemental Table S4; Supplemental Table S5**). Hierarchical clustering of 1,520 orthogroups that were deeply-

conserved in all six species revealed that the gene families regulated by PRC2 are largely unique to each species, consistent with the distinct evolutionary history of each organism (**Fig. 5A**). However, our analysis did reveal a substantial number of orthogroups that were commonly marked with H3K27me3 between the different species, including a core of 92 gene families common to the five land plants we analysed (**Fig. 5B**).

We divided this core set of PRC2-regulated orthogroups into three main groups and further classified their biological function based on gene annotations in *Arabidopsis* (**Fig. 5C-D; Supplemental Table S6**). This revealed a collection of biological processes that would have served as useful traits for terrestrial life. The largest group of gene families encoded catabolic enzymes involved in various metabolic functions, including lipid biogenesis, cell wall metabolism and proteins that modulate responses to osmotic and oxidative stress (**Fig. 5B-C**). Some signalling processes were also present and included a Phosphatidylethanolamine binding protein (PEBP) gene family related to *FLOWERING LOCUS T* and *TERMINAL FLOWER 1* (Wickland & Hanzawa, 2015). Gene families involved in transport activity of inorganic ions, sugars and hormones were also abundant, as were a defined set of TF families (**Fig. 5B-C**). While the majority of PRC2-regulated gene families were only conserved in the five land plants (52 of 92 orthogroups; 56.0 %) (**Fig. 5B-C**), around a third were also conserved in *Chlorella* but only marked with H3K27me3 in land plants (29 of 92 orthogroups; 31.5 %) (**Fig. 5C**), suggesting that PRC2 was recruited to these gene families upon or after the emergence of the Streptophyta (**Fig. 5C**). The final group was composed of a small subset of gene families marked with H3K27me3 in all six species (11 of 92 orthogroups; 0.5%), highlighting a core of deeply-conserved PRC2-regulated gene families involved in core biological processes, including ion and sugar transport, lipid and carbohydrate metabolism, and stress tolerance (**Fig. 5B-C; Supplemental Table S7)**. This deeply-conserved group of PRC2 targets also included Squamosa promoter binding protein–like (SPL) genes (**Fig. 5B**), which belong to an ancient family of plant-specific TFs that play diverse roles in land plants (Preston & Hileman, 2013). Our analysis has thus revealed how PRC2 regulates a core of orthologous genes across land plants as well as in their unicellular green algal relatives, providing a glimpse of the first types of genes regulated by PRC2 during early phases of plant evolution.

Among the 103 transcription-associated proteins (TAPs) regulated by PRC2 across the six species we analysed, only 16 were marked with H3K27me3 in all five land plants species, which included members of the AP2/ERF, bHLH, C2H2, R2R3-MYB and WRKY family of TFs (**Fig. 5E**; **Supplemental Table S8**). Argonaute and SET domain proteins were also marked with H3K27me3 in the land plants, highlighting how PRC2 control also evolved to modulate the expression of major components in the small RNA and chromatin pathways (**Fig. 5E**; **Supplemental Table S8**). Although only thirteen TAPs were marked with H3K27me3 in *Chlorella*, seven of these were shared with the five land plants and included AP2/ERF, PHD and SPL type TFs and SWI/SNF chromatin remodelling proteins. (**Fig. 5E**). Most notable among the TFs with conserved PRC2 regulation were two orthogroups descended from the TALE-HD TFs KNOX (KNOTTED-like homeobox) and BELL (BEL-Like) (**Fig. 5B,E**). KNOX/BELL have deep evolutionary origins in unicellular green algae (Lee *et al.*, 2008; Nishimura *et al.*, 2012; Kariyawasam *et al.*, 2019) and play a conserved role in sexual reproduction by activating the zygotic program after fertilisation in *Chlamydomonas* and *Marchantia* (Sakakibara *et al.*, 2013; Horst *et al.*, 2016; Dierschke *et al.*, 2021; Hisanaga *et al.*, 2021). Although we were unable to identify any obvious homologs of KNOX/BELL in *Chlorella*, which is assumed to reproduce asexually (Hovde *et al.*, 2018), the KNOX/BELL orthologs present in *Anthoceros* and *Physcomitirium* were all marked with H3K27me3 (**Fig. 5B; Supplemental Fig. S6**). Analysis of

CUT&RUN data in *Marchantia* also confirmed that most KNOX/BELL orthologs are also silenced with H3K27me3 in liverworts (**Supplemental Fig. S6**). As the KNOX/BELL gene families expanded during flowering plant evolution, PRC2 regulation appears to have also diversified by targeting a subset of KNOX/BELL homologs (**Fig. 5B**). Other noteworthy examples of TF orthogroups with deeply conserved PRC2 regulation included members of the WOX, WRKY, bHLH and DREB-A2 subfamily of AP2/ERF TFs, which regulate a wide range of vital processes in *Arabidopsis* ranging from seed dormancy, embryogenesis, stomatal patterning through to abiotic stress response (**Fig. 5B**). Key transcriptional regulators of the plant life cycle and cellular patterning were thus among the first types of genes recruited by PRC2 during early land plant evolution, which likely helped diversify the regulatory programs underlying the increased morphological and developmental complexity observed in this group.

Interestingly, the TAP families regulated by PRC2 in flowering plants were a larger and more diverse group compared with those in *Chlorella* and *Anthoceros* (**Fig. 5E**), emphasising the broader control PRC2 evolved to exert over gene regulatory networks in flowering plants. *Physcomitrium* stood as an exception since the number of TAPs marked with H3K27me3 were similar in number to those in flowering plants, although at least seven of these were exclusive to this species, such as E2F/DP and HMG-type TFs (**Fig. 5E**). The expansion of PRC2-regulated TAPs among the flowering plants raised the question as to whether they also share a distinct network of PRC2-regulated gene families. Within the H3K27me3-marked orthogroups were a group of 52 gene families that were specifically regulated by PRC2 in *Spirodela*, *Oryza* and *Arabidopsis* but not in bryophytes (**Fig. 5F**). While many of the functions were broadly similar to those observed in the ancestral PRC2-regulated gene network, the flowering plant-specific network was notably more diverse in gene families involved in signalling processes (**Fig. 5F-G**). Among these were gene families encoding at least two distinct types of protein kinases, a group of ABA-induced Type 2C protein phosphatases (PP2C) as well as AUX/IAA genes involved in auxin responses (**Fig. 5F-G**). Several TF families were also present and notably included the floral meristem identity regulator LEAFY and a CCT MOTIF FAMILY (CMF) group of TFs related to ASML2, a TF with high expression in reproductive organs that regulates sugar-inducible genes (**Fig. 5E-F**) (Weigel *et al.*, 1992; Masaki *et al.*, 2005). Transcriptomic analysis of the flowering plant-specific PRC2 network of genes in *Arabidopsis* revealed strong developmentally-regulated expression in reproductive and vegetative organs and tissues that distinguish the flowering plant lineage, including flowers, pollen, seeds and roots (**Fig. 5H**). These data highlight how PRC2 was recruited to a network of genes specifically during the emergence of flowering plants, which includes key transcriptional regulators and signalling proteins that specify major organs and tissues characteristic of flowering plants.

Ancestral PRC2-regulated genes share common *cis*-regulatory motifs

In animals and plants, H3K27me3 deposition is targeted to specific loci by particular TFs that tether PRC2 to *cis*-regulatory regions called Polycomb response elements (PREs) (Blackledge & Klose, 2021b). We thus pondered what type of TFs might recruit PRC2 to the ancestral set of H3K7me3-marked genes and whether these would be similar among the five land plants. To address this, we performed motif enrichment analysis to determine which TFs are likely to bind within the H3K27me3 domains marking the ancestral PRC2 target genes. Because the binding specificity of orthologous TFs is known to be highly conserved deep across evolutionary time (Nitta *et al.*, 2015), we reasoned that the binding preference of *Arabidopsis* TFs could be used to identify potential PRE elements. For this we used the set of experimentally-derived

DNA binding motifs for *Arabidopsis* TFs generated with DAP-seq (O'Malley *et al.*, 2016), but focused only on DNA binding motifs from TF families that had at least one ortholog present in all five species.

In total we identified 210 enriched motifs at the ancestral PRC2 target genes across all five species (**Fig. 6A; Supplemental Table S9**). The number of enriched motifs understandably varied, with *Anthoceros* and *Arabidopsis* having the lowest and highest number of enriched motifs, respectively (**Fig. 6A**). Among these were 15 from six major TF families that were commonly enriched across all five land plants, with another 27 motifs that were also commonly enriched in all species except for *Anthoceros* (**Fig. 6A; Supplemental Table S9**). Importantly, just over a quarter of these common motifs (28.6%; 12/42) are bound by AP2/ERF TFs that have been previously shown to function as PRE-binding TFs in *Arabidopsis,* many of which also can interact with PRC2 (**Fig. 6B)** (Xiao *et al.*, 2017; Xie *et al.*, 2022). To further probe the relationship of these TFs in PRC2 recruitment, we plotted the occurrence of DNA binding motifs over the ancestral PRC2 target genes alongside the profile of H3K27me3 (**Fig. 6C-D**). In each of the five land plants, the motifs occurred in a pattern that strongly mirrored the profile of H3K27me3, with the flowering plants having a higher density of motifs compared to the bryophyte species (**Fig. 6D**). Although some motifs were also present in the upstream flanking regions of the genes, most were positioned downstream of the TSS along the gene body (**Fig. 6D**). This is consistent with recent reports of several plant TFs binding within the body of genes, including PRE-binding TFs in *Arabidopsis* (Xiao *et al.*, 2017; Voichek *et al.*, 2024).

Because the DAP-seq dataset assayed TF binding to naked *Arabidopsis* genomic DNA, DAP-seq peaks also give an indication of *in vivo* binding potential of a given TF. We thus further plotted the coverage of *Arabidopsis* DAP-seq peaks along the ancestral PRC2-regulated genes to further validate our motif enrichment profiles. This analysis revealed that, where available, the DAP-seq peaks of TF binding were also enriched over gene bodies, particularly for AP2/ERF TFs, demonstrating that the pattern of motif predictions we observe strongly mirror TF binding in a genomic context (**Fig. 6C**). In contrast, a control set of non-enriched motifs showed no distinct pattern and had much lower motif occurrences compared to the enriched motifs (**Supplemental Fig. S7; Supplemental Table S9**). Thus, the common set of genes regulated by PRC2 across land plants also share similar TF-binding *cis*-elements, including those of known AP2/ERF PRE-binding TFs, alluding to ancestral PRC2-TF relationships that could have recruited PRC2 to genes during early land plant evolution.

Discussion

The genomic era has heralded essential insights into the genetic bases of land plant evolution and points to the expansion and diversification of a genetic toolkit with deep origins in their aquatic algal ancestors (Bowman, 2022; Dhabalia Ashok *et al.*, 2024). This includes most of the TF families found in modern-day plants, many of which play conserved roles in environmental and developmental regulation across large evolutionary distances (Romani & Moreno, 2021). Although these TFs would have undoubtedly shaped gene regulatory networks in the last common Streptophyte ancestor, added layers of gene regulation would have been necessary to modulate TF expression and define new cell types and developmental traits. Here, we have described the evolutionary dynamics of epigenetic regulation across the Viridiplantae and postulate that the progressive expansion and elaboration of epigenetically-regulated gene networks was a key step in diversifying gene expression and cellular differentiation during land plant evolution.

Our analysis has revealed that evolutionarily young or TRGs are highly over-represented among genes that are silenced by H3K9me1/2-marked constitutive heterochromatin in plants (**Fig. 4**). Closer inspection of the chromosomal distribution of genes grouped by their evolutionary age revealed that evolutionarily older genes were depleted from heterochromatic TE-rich regions like pericentromeres. In contrast, TRGs had an equal distribution along chromosome in several plant species, including TE-rich heterochromatic regions, increasing their likelihood for silencing with constitutive heterochromatin. In *Drosophila*, TRGs tend to emerge within pre-existing repressive domains that are enriched with H3K27me3 (Zhang & Zhou, 2019). In the nematode *Pristionchus pacificus*, TRGs are also strongly enriched for heterochromatic states marked with H3K27me3, H3K9me3 or both (Werner *et al.*, 2018). In the brown alga *Ectocarpus*, TRGs are highly enriched along UV sex chromosomes that are distinguished by a brown algal-specific repressive chromatin state enriched with H3K79me2 (Gueno *et al.*, 2013; Luthringer *et al.*, 2015). TRGs thus appear to be largely partitioned into repressive chromatin domains in a diverse range of eukaryotes and tend to be transcribed in a restricted number of cell types and tissues (Luthringer *et al.*, 2015; Werner *et al.*, 2018; Zhang & Zhou, 2019). Our analysis has revealed a similar pattern in *Arabidopsis* where Brassicaceae-specific H3K9me2-marked genes are transcribed in a restricted range of mostly reproductive cell types (**Supplemental Fig. S4**). These included at least three pollen tube-related genes that are targeted for epigenetic reprogramming in pollen, a known a hotspot for the origin of new genes in flowering plants (Wu *et al.*, 2014; Borg *et al.*, 2021b; Rutley *et al.*, 2021), with similar reprogramming also occurring during central cell specification in the female gametophyte (Pillot *et al.*, 2010; Ibarra *et al.*, 2012; Park *et al.*, 2016). H3K9me2 has also been shown to be key in repressing key regulators of leaf development in *Arabidopsis* (Hung *et al.*, 2023). Thus, although PRC2 appears to dominate gene silencing across land plants, lineage-specific cases of H3K9me2-silenced gene networks have emerged during flowering plant evolution that help specify reproductive development in *Arabidopsis*.

Our analysis of heterochromatin in the unicellular green microalga *Chlorella sorokiniana* revealed a substantial proportion of genes marked with H3K27me3 (**Fig. 1**), suggesting that PRC2 regulation of genes has deep origins in the green lineage. H3K9me2 and H3K27me3 always co-occurred together at TEs in *Chlorella*, which is strikingly reminiscent of their co-deposition at TEs in *Paramecium* (Frapporti *et al.*, 2019). This once again highlights the common evolutionary history shared by these two epigenetic pathways in unicellular eukaryotes prior to their functional bifurcation later during evolution. In contrast, H3K27me3-marked genes in *Chlorella* were devoid of H3K9me2 and generally did not lie in the vicinity of LTR or DNA

transposons, indicating a separation between the chromatin states that define TEs and genes in this microalga. This suggests that *cis* recruitment of PRC2 to genes in *Chlorella* occurs by an unknown mechanism, potentially via PRE-like elements commonly found in more complex multicellular eukaryotes (Blackledge & Klose, 2021b). The evolution of PRE elements is not well understood but studies in *Drosophila* suggest that they are extraordinarily dynamic, even among closely-related species (Hauenschild *et al.*, 2008). Unicellular eukaryotes like green microalgae could thus make interesting model systems to study the origin and dynamics of PRC2 recruitment in the future.

By stratifying genes by their evolutionary age and epigenetic state in multiple species of land plants, we observed that genes specific to the Streptophytina lineage were consistently over-represented among the genes targeted by PRC2 (**Fig. 2**), which interestingly coincides with the point in evolution when multicellularity first arose in the charophyceaen macroalgae (Umen, 2014). An over-representation of PRC2-target genes was also evident among genes that emerged during the rise of the embryophytes and subsequent evolution of the monocots and dicots. This highlights a major period of regulatory adaptation during Viridiplantae evolution where hundreds of genes were preferentially incorporated into epigenetic networks controlled by PRC2. These landmark events are associated with considerable *de novo* gene birth and gene family diversification (Bowles *et al.*, 2020; Barrera-Redondo *et al.*, 2023), which would have provided ripe context for new PRC2 gene regulatory networks to emerge and fuel evolutionary change. Based on our knowledge of PRC2 function in the modern-day, this is likely to have been driven by genetic changes through the emergence of motifs bound by TFs that recruit PRC2, so-called PRE-binding TFs (Xiao *et al.*, 2017; Blackledge & Klose, 2021b). Recent studies have suggested that H3K27me3-marked TE fragments, particularly those originating from DNA transposons, were co-opted as PRE elements during Archaeplastida evolution (Hisanaga *et al.*, 2023). Our analysis of the epigenetic state and evolutionary age of TEs suggests that DNA transposons are the preferred target of H3K27me3 across distantly-related Viridiplantae species, including unicellular green microalgae, suggesting an ancient evolutionary relationship (**Fig. 3)**. H3K27me3-marked DNA transposons were particularly prevalent among the flowering plants we analysed and appear to be particularly old given their short length and high sequence divergence. Our work thus provides further credence to the hypothesis that DNA transposons were domesticated to impact gene regulation during Archaeplastida evolution, and we hypothesise that embryophyte evolution in particular was a major period of epigenetic adaptation given the considerable surge in the recruitment of PRC2 to genes that emerged during this key period in plant evolution.

This raised the question as to whether the functional topology of the gene networks regulated by PRC2 are common or unique among distantly-related plants. Overall, the gene families regulated by PRC2 were largely unique to the species we analysed, which is unsurprising given the distinct evolutionary history of each lineage (**Fig. 5A**). Nevertheless, our analysis revealed a group of ancient gene families with conserved PRC2 regulation across large evolutionary distances (**Fig. 5B**). Given their maintenance across land plants for over 475 million years since the emergence of bryophytes, we speculate that this could represent an ancestral signature of the earliest PRC2-regulated gene networks that were established during land plant evolution. This ancestral network includes several biological processes whose spatial, temporal and environmental regulation would have undoubtedly facilitated land terrestrialisation. For example, the evolution of a sturdy cell wall that is adaptable across development and during environmental stress is proposed to have been essential for land adaptation (de Vries & Archibald, 2018). Several glycosyl hydrolase gene families stood out in this ancestral network, which are crucial for carbohydrate and

glycoconjugate metabolism during various cellular processes, including the mobilisation of starch reserves, pathogen defence and the assembly and disassembly of the cell wall (Kfoury *et al.*, 2024). In addition to a host of gene families with diverse metabolic functions, several others involved in sugar, amino acid and organic ion transport also form part of this network, whose fine-tuned epigenetic regulation would have also facilitated nutrient relocation within specific tissues and organs of an increasingly complex plant body plan (Martinoia *et al.*, 2000; Nagata *et al.*, 2008). Interestingly, phytohormone signalling pathways were not part of the ancestral PRC2-regulated network, despite the fact that auxin signalling pathways are known to be prominent targets of PRC2 in flowering plants (Lafos *et al.*, 2011; Wójcikowska *et al.*, 2020; Wu *et al.*, 2023). Our results show that PRC2 regulation of auxin signalling, specifically AUX/IAA proteins, emerged later during flowering plant evolution, likely in response to the expansion of the minimal auxin signalling systems seen in bryophytes (**Fig. 5E-F**) (Flores-Sandoval *et al.*, 2015; Suzuki *et al.*, 2023).

The most prominent and well-studied PRC2 target genes in complex multicellular eukaryotes tend to encode for TFs. Our analysis has confirmed this in several land plants but has provided further insight into breadth of TAPs regulated by PRC2 across distantly-related species in the green lineage (**Fig. 5E**). The number and diversity of TAPs regulated by PRC2 was markedly increased in flowering plants compared to bryophytes and included not only DNA binding TFs but also key players of small RNA silencing (Argonautes), histone methylation (SET domain proteins) and chromatin remodelling (SWI/SNF proteins). We speculate that the progressive recruitment of diverse TAP families by PRC2 would have been key in creating new gene regulatory networks over the course of land plant evolution and likely facilitated the emergence and specification of new cell types and tissues. Interestingly, the number of TAP families regulated by PRC2 in *Chlorella* was fairly low in number compared to the land plants, despite the fact that many of the same plant TAPs have orthologs in *Chlorella* (**Fig. 5E**). This nevertheless included seven TAP families with deeply-conserved PRC2 regulation across the green lineage, including AP2/ERF, PHD and SPL type TFs as well as SWI/SNF chromatin remodellers. This emphasises how an increased diversity of PRC2-regulated TAPs underpins developmental complexity in the green lineage, and unveils an ancient regulatory relationship between PRC2 and seven TAP families that has persisted into modern-day bryophytes and flowering plants. The master regulator of floral meristem identity LEAFY was also identified as a TF family that came under control of PRC2 during flowering plant evolution and forms part of a broader PRC2-regulated gene network associated with the emergence of this dominant group (**Fig. 5E,F)**. Interestingly, our analysis of DNA binding motifs at the ancestral target genes of PRC2 revealed a commonly enriched set of motifs, which also happened to occur in a pattern that mirrored H3K27me3 enrichment in each of the land plants we analysed (**Fig. 6**). Of particular interest were those bound by AP2/ERF family TFs, which have been reported to physically interact with PRC2 to recruit it to its target loci in *Arabidopsis* in a manner similar to that reported in animals (Xiao *et al.*, 2017). Coupled with the fact that the PRC2 regulation of AP2/ERF TFs can be traced all the way back to green microalgae, we speculate that AP2/ERF TFs could have been some of the first TFs to interact with PRC2 and could have played a key role in the emergence and evolution of PRC2-regulated gene networks during land plant evolution. Future validation of AP2/ERF-PRC2 interactions in green microalgae and bryophytes will hope to shed more light on this ancestral regulatory relationship.

Among the ancestral PRC2-regulated TF families we identified were the heterodimerisation partners KNOX and BELL (Bellaoui *et al.*, 2001). Given the ancestral role of these TALE-HD TFs in activating the diploid zygotic/sporophytic program in *Chlamydomonas reinhardtii* and *Marchantia polymorpha*, it is

hypothesised that they were key genetic determinants that sparked the evolution of haploid-diploid life cycles in plants (Lee *et al.*, 2008; Sakakibara *et al.*, 2013; Hisanaga *et al.*, 2021). Our analysis shows how BELL and KNOX are both regulated by PRC2 in the haploid gametophyte of all three major bryophyte groups, suggesting that their recruitment by PRC2 is an ancestral feature that could have occurred prior to the emergence of land plants. This regulation is likely to be highly relevant since *BELL* is derepressed in PRC2 mutants of *Physcomitirium* and results in the gametophyte displaying sporophyte-like features (Pereman *et al.*, 2016). Transcriptional silencing of BELL and KNOX by PRC2 in the haploid phase of the life cycle would have been key for repressing diploid identity, further fuelling the idea that their upstream regulation could have facilitated the evolution of haploid-diploid transitions in plants (Vigneau & Borg, 2021). KNOX/BELL homologs appear to have been lost in *Chlorella* so we were unable to assess whether they are also regulated by PRC2 in this asexual microalga. Broader profiling of chromatin landscapes in other green algal lineages and land plants, including the more complex charophycean macroalgae, is required to better understand when PRC2 silencing of KNOX/BELL first emerged, and also help trace when and how PRC2-regulated gene networks were shaped during the earliest phases of Viridiplantae evolution.

Materials and Methods

Chlorella ChIP-seq profiling

The unicellular green alga *Chlorella sorokiniana* CCALA 259 (equivalent to UTEX 1230) was obtained from the Culture Collection of Autotrophic Organisms at the Institute of Botany of the Czech Academy of Sciences (Trebon, Czech Republic). The cultures were grown on ½ strength Šetlík-Sluková (SS) medium (Hlavová *et al.*, 2016). For routine subculture, cultures were streaked every three weeks on ½ SS medium solidified with agar (1.5 %, w/v) and grown at an incident light intensity of 100 $\mu mol.m^{-2}.s^{-1}$ of photosynthetically active radiation (PAR). For the experiment, 300 ml of the liquid ½ SS medium was inoculated directly from the plates and the cultures were placed in glass cylinders (inner diameter 30 mm, height 500 mm) at 30 °C and "aerated" with a mixture of air and CO2 (2%, v/v) at a flow rate of 15 $l.h^{-1}$. The cylinders were illuminated from one side with dimmable fluorescent lamps (OSRAM DULUX L55W/950 Daylight, Milano, Italy), the light intensity of which was adjusted so that 500 $\mu mol.m^{-2}.s^{-1}$ of PAR hit the surface of the cylinders. The cultures were grown under continuous light until a cell density of approximately 1×10^8 $cells.ml^{-1}$ was reached. They were then shifted to 40 °C and a light intensity of 700 $\mu mol.m^{-2}.s^{-1}$ for 24 hours to allow the cells to grow but block cell division. The cultures were then transferred to the dark at 30 °C and the cells were allowed to divide for 24 hours. These partially synchronized cultures were diluted to a cell density of approximately $1x10^7$ $cells.ml^{-1}$ either in ½ SS medium. The cultures were grown at a light intensity of 500 $\mu mol.m^{-2}.s^{-1}$ at 30 °C. The cultures were sampled in the early G1 phase, shortly after one cell cycle was completed.

ChIP-seq experiments were performed with two biological replicates using the photoautotrophically-grown cells described above. ChIP was performed as described previously (Strenkert *et al.*, 2011) with modifications based on (Mozgová *et al.*, 2015). *Chlorella sorokiniana* UTEX1230 cells (9 x 10^8 cells) were crosslinked at room temperature using 0.37% formaldehyde for 10 minutes, after which crosslinking was quenched by 0.125 M glycine for 10 minutes at room temperature. Cells were washed twice with water and centrifuged at 3000 g for 5 min at 4°C, subsequently resuspended in 200µl of Lysis Buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8, 1x cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche #04693132001)) and frozen with liquid nitrogen. Next, cells were sonicated using a Bioruptor Plus with 20 cycles of 30 sec ON/OFF at 4°C to achieve DNA fragments size of approximately 200 bp and centrifuged for 5 min at 4500 g at 4°C to remove cell debris. The supernatant was diluted ten times with ChIP dilution buffer (1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8, 167 mM NaCl, cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche #04693132001)), 200µl aliquot was used for each IP incubated overnight at 4°C with 1.5µg of antibody. Antibodies specific for the following epitopes were used: H3 (Millipore/Merck, 07-690, lot# 3683182), H3K4me3 (Millipore/Merck, 07-473, lot# 3660317), H3K9me2 (Abcam, ab1220, lot #GR3247768-1), H3K27me3 (Diagenode, C15410069, lot # A1818P). The next day, the samples were mixed with 20 µl Dynabeads Protein A (Thermo) and incubated for another 2 hours at 4°C. Next, beads were successively washed three times (5 min/wash) with Low Salt Wash Buffer (150 mM NaCl, 0.1% SDS, 1% Triton X 100, 2 mM EDTA, 20 mM Tris-HCl pH 8), High Salt Buffer (500 mM NaCl, 0.1% SDS, 1% Triton X 100, 2 mM EDTA, 20 mM Tris-HCl pH 8), LiCl Wash Buffer (0.25 M LiCl, 1% NP40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8) and finally with TE Buffer (10 mM Tris-HCl pH 8, 1 mM EDTA). DNA was decrosslinked, purified, and eluted using IPureKit (Diagenode). ChIP was

performed in three biological replicates, of which two were subject to Illumina sequencing. Sequencing libraries were made with 5 ng DNA using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (NEB #E7103S/L) following the manufacturer's instructions. Paired-end 150bp Illumina Sequencing was performed using the NovaSeq 6000 System, with an average output of 20 Mio reads/sample.

ChIP-seq data analysis

After a survey of existing ChIP-seq data from multiple members of the Viridiplantae, we chose a series of H3K4me3, H3K9me1/2 and H3K27me3 datasets from five Viridiplantae species - the two bryophytes *Physcomitrium patens* and *Anthoceros agrestis* and the three flowering plants *Spirodela polyrhiza*, *Oryza sativa* and *Arabidopsis thaliana* **(Fig. 1A; Supplemental Table S1)**. These were all generated from vegetative tissue of the dominant life cycle phase of each species i.e., haploid gametophytic tissue for the bryophyte species and sporophytic leaf tissue for the flowering plants, and whole plants in case of the duckweed *Spirodela*. For consistency in cross-comparing between different species, we chose not to include *Marchantia polymorpha* given that it was generated with CUT&RUN rather than traditional ChIP-seq (Montgomery *et al.*, 2020). Both the *Chlorella* and public ChIP-Seq data were processed using the Nextflow nf-core/chipseq v2.0.0 pipeline (https://nf-co.re/chipseq/2.0.0/) (Ewels *et al.*, 2020). The ChIP-seq data for each species was mapped to the latest genome assembly and annotation available at the time of our analysis, which are included in an online data repository (https://doi.org/10.17617/3.PJSEUL), together with a MultiQC reports with quality control metrics such as trimming, mapping, coverage and complexity metrics, as well as the version of each tool used in the pipeline. Because the *Physcomitrium* H3K9me2 ChIP-seq data was generated using the legacy SOLiD sequencing format (Widiez *et al.*, 2014), we developed a custom pipeline to process these reads and re-mapped it to the latest *Physcomitrium* patterns genome assembly available at the time of our analysis (https://genomevolution.org/coge/GenomeInfo.pl?gid=33928; see online data repository) using BLAT-like Fast Accurate Search Tool (BFAST v0.7.0a) (https://github.com/nh13/BFAST). First, the reference genome was prepared in color space using the command *bfast fasta2brg* with a *-A 1* flag, then indexed with the command *bfast index* using the *-A 1 -m 1111111111111111111111 -w 14* flag. The color space reads were matched with *bfast match* using a *-A 1 -z -r* flag, then locally aligned with *bfast localalign* using the flag *-A 1 -n 8 -U*. The resulting alignment was then filtered using *bfast postprocess* with the flag *-A 1 -a 2 -O* 1. The downstream processing of the resulting aligned sam file followed that as in the nf-core/chipseq v2.0.0 pipeline. For data visualisation and plotting, normalised $\log_2$ bigwig coverage files of each histone mark relative to H3 or input were generated using deepTools version 3.5.1 bamCompare with a bin size of 10 bp (Ramírez *et al.*, 2014). Biological replicates were merged where available (**Supplemental Table S1**). A cross-correlation matrix of Spearman's correlation coefficient of the *Chlorella* samples was generated using deepTools v3.5.1 *multiBamSummary* (Ramírez *et al.*, 2014). Bigwig coverage files were visualized along each genome assembly using IGV v2.16.2 (Robinson *et al.*, 2011).

Functional analysis of H3K27me3-marked genes in *Chlorella* was based on the gene annotations provided by Phycocosm (Chloso_1 assembly). Five different classifications were included: Gene ontology (GO) - biological process (GO_BP), molecular function (GO_MF), and cellular component (GO_CC), Kyoto encyclopedia of genes and genomes (KEGG) - pathway (KEGG_pathway) and pathway class (KEGG_pathway_class) and Eukaryotic orthologous groups (KOG_definition). Hypergeometric tests were performed using the enricher function in the R package clusterProfiler (Yu *et al.*, 2012), with default

parameters, considering the five classifications separately. A functional term was considered enriched among the Chlorella H3K27me3-marked genes if its false-discovery rate (FDR) was lower than 5%.

RNA-seq analysis

The analysis of RNA-seq data from across *Arabidopsis* development was performed as described previously (Borg *et al.*, 2020). In brief, raw FASTQ files for each dataset were downloaded from the Gene Expression Omnibus (GEO) database (**Supplemental Table S1**). Adapters were trimmed using TrimGalore v0.4.1 (https://github.com/FelixKrueger/TrimGalore) and the resulting reads aligned to the *Arabidopsis* genome (TAIR10) using the STAR aligner v2.5.2a (Dobin *et al.*, 2013). Transcripts per million (TPM) values were generated using Kallisto v0.43.1 (Bray *et al.*, 2016) with an index built on TAIR10 cDNA sequences (see online data repository). The *agdp1* and *suvh456* RNA-seq datasets (Zhang *et al.*, 2018) were processed similarly except that read count quantification per gene was summarised using Salmon v1.10.1 in alignment-based mode (Patro *et al.*, 2017).The resulting count table was used for differential gene expression analysis using DESeq2 v1.40.2 (Love *et al.*, 2014). Differentially expressed genes (DEGs) were defined as having a $\log_2$ fold-change >1 or < -1 and an adjusted *p*-value < 0.05 compared to the wild-type control. Overlap of enrichment of DEGs with *Arabidopsis* TRGs (ranks 8-11) was determined with the R package GeneOverlap v1.36.0 function newGOM (https://github.com/shenlab-sinai/GeneOverlap). Expression specificity *tau* scores of each *Arabidopsis* gene were calculated in R across the RNA-seq datasets shown in **Supplemental Figure S4** as described previously (Lüleci & Yılmaz, 2022).

DNA methylation analysis

Raw FASTQ files for methylomes previously generated from *Physcomitrium patens* (Yaari *et al.*, 2019), *Spirodela polyrhiza* (Harkess *et al.*, 2024), *Oryza sativa* (Xu *et al.*, 2020) and *Arabidopsis thaliana* (Gallego-Bartolomé *et al.*, 2019) were first quality and adapter trimmed with using TrimGalore v0.6.10 with default settings. Bisulfite-converted reads were then aligned against each respective genome in directional mode using Bismark v0.23.0 (*bismark -q --score-min L,0,–0.4*) (Krueger & Andrews, 2011). Deduplicated and uniquely mapped alignments were then used to call weighted methylation rates at every converted cytosine using Methylpy v1.2.9 (Schultz *et al.*, 2015). Cytosines with a coverage ≥ 4 were used to calculate the average methylation rate per genomic feature and calculated using the BEDTools v2.30.0 *map* function. For methylation analysis genes or TEs with ≥ 4 mapped cytosines were retained.

Chromatin state of evolutionarily aged genes

Genomic phylostratigraphy was performed to age protein-coding genes in each species using GenERA with the flag *-u "--query-cover 50"* to increase the threshold of protein homology to at least 50% (Barrera-Redondo *et al.*, 2023). H3K9me1/2- and H3K27me1/3-marked genes were identified using the R packages *ChIPpeakAnno* and *GenomicRanges* (Zhu *et al.*, 2010; Lawrence *et al.*, 2013) and defined as having a minimum overlap of 100bp of coding sequence with the respective ChIP-seq peaks. The $\log_2$ ratio of observed to expected values in **Figure 2** represents the proportion of genes in each phylogenetic rank marked with a given histone mark (observed) divided by the proportion of genes in each phylogenetic rank found across the whole genome (expected). Circos plots showing the distribution of the phylogenetically-ranked genes in each species were generated using the R package circlize (Gu *et al.*, 2014). Meta profiles of TE enrichment were generated using the EnrichedHeatmap function *normalizeToMatrix* and plotted using a custom script

in R. Taxonomically-restricted genes (TRGs) were defined as those that arose within the family and in later ranks for each species (e.g,. those in the Brassicaceae rank onwards in the case of *Arabidopsis*). GO enrichment computation of H3K9me2- and H3K27me3-marked genes in each phylogenetic rank in *Arabidopsis* was performed using Panther 19.0 (DOI: 10.5281/zenodo.12173881).

TE annotation and divergence landscapes

De novo annotation of TEs and TE library preparation, including the consensus sequences of all identified TE families, was performed for all six Viridiplantae species using EDTA v2.0.0 (Ou *et al.*, 2019a). Each TE library was used to mask TEs across the respective genomes using RepeatMasker vopen-4.0.9 (Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-4.0*. 2013-2015 http://www.repeatmasker.org). To identify TEs marked with H3K9me1/2, H3K27me3 or both, ChIP-seq peaks were intersected with each TE using the R packages *ChIPpeakAnno* and *GenomicRanges* (Zhu *et al.*, 2010; Lawrence *et al.*, 2013) . TE landscapes of DNA transposons, LTR retrotransposons and MITEs were generated from the RepeatMasker output file (.out file) using the percent diversity parameter, which represents the percentage of substitutions of a given TE relative to the consensus sequence of its assigned family. Given that EDTA is unable to successfully annotate LINE elements (Ou *et al.*, 2019b), this class of TEs were not considered in our study and thus warrant further investigation in the future.

Analysis of conserved PRC2-regulated gene families

Orthology inference analysis was first performed to identify orthologous gene families (or orthogroups) across the proteome of each species using OrthoFinder v2.5.2 (Emms & Kelly, 2019). The relative proportion of genes marked with H3K27me3 within each orthogroup was then computed for each species to identify conserved PRC2 target genes. The ancestral land plant PRC2 network of genes was defined as having at least one gene within an orthogroup marked with H3K27me3 in each of the five land plant species. The deeply-conserved PRC2-regulated network of genes was similarly defined but with the inclusion of orthogroups that were also conserved in *Chlorella*, whereas the flowering-plant specific PRC2-regulated network was restricted to orthogroups that has at least one gene marked with H3K27me3 exclusively in *Spirodela*, *Oryza* or *Arabidopsis*. Functional annotation of the orthogroups was performed with *Arabidopsis* gene identifiers using the online database GenFAM (https://www.mandadilab.com/genfam/) (Bedre & Mandadi, 2019). Further manual curation was performed to assign the orthogroups to one of the nine functional classes defined in **Figure 5C**. The annotation of transcription-associated proteins (TAPs) was performed with the proteome of each species using TAPscan v4 (https://tapscan.plantcode.cup.uni-freiburg.de) (Petroll *et al.*, 2024).

Motif enrichment analysis

Motif enrichment was performed using motifs in the DAP-seq data set (O'Malley *et al.*, 2016) with the *AME* function in the MEME suite version 5.5.4 (Bailey *et al.*, 2009). The H3K27me3 peaks overlapping genes that form part of the ancestral land plant PRC2-regulated network in each species were provided as input, with the respective set of complete H3K27me3 peaks used as the background model. Commonly enriched motifs were defined as having an adjusted *p*-value less than 0.05 in each land plant species. Heatmaps of DAP-seq peaks, motif occurrences and H3K27me3 enrichment were generated using the R package EnrichedHeatmap (Gu *et al.*, 2018).

Data availability

Chlorella ChIP-seq data generated in this study have been deposited in the European Nucleotide Archive (ENA) under accession code PRJEB79967. A repository containing all the genome assemblies, annotation and processed data generated in this study can be accessed at https://doi.org/10.17617/3.PJSEUL, which includes re-analysis of the public datasets listed in **Supplemental Table S1**. All scripts are available upon request.

Contributions

MB conceived and designed the study. MB performed the bioinformatic analyses with support from RP, RKP and AL. Green algal histone H3 sequences were collated and analysed by MB and IM. Synchronised *Chlorella* cells in G1 were prepared by KB. *Chlorella* ChIP-seq data was generated by RK under supervision of IM. QR and IM performed the functional enrichment analysis of H3K27me3 target genes. RP generated the TE landscapes for Figure 4. RKP processed public DNA methylation datasets and the *adgp1* and *suvh456* RNA-seq data for Supplemental Figures 2 and 3. AL surveyed public ChIP-seq datasets and developed scripts for downstream analysis under supervision of MB. MB interpreted the data, assembled the figures and wrote the manuscript.

Acknowledgements

The authors thank Josue Barrera-Redondo and Sodai Lotharukpong for advice on running GenERA. MB, RP and AL were supported by the Max-Planck-Gesellschaft. RK was supported by an ERA Fellowship from the European Research Executive Agency (101090308 - PAMFGAL). IM was supported by an ERC-CZ grant from the Czech Academy of Sciences (ERC200961901).

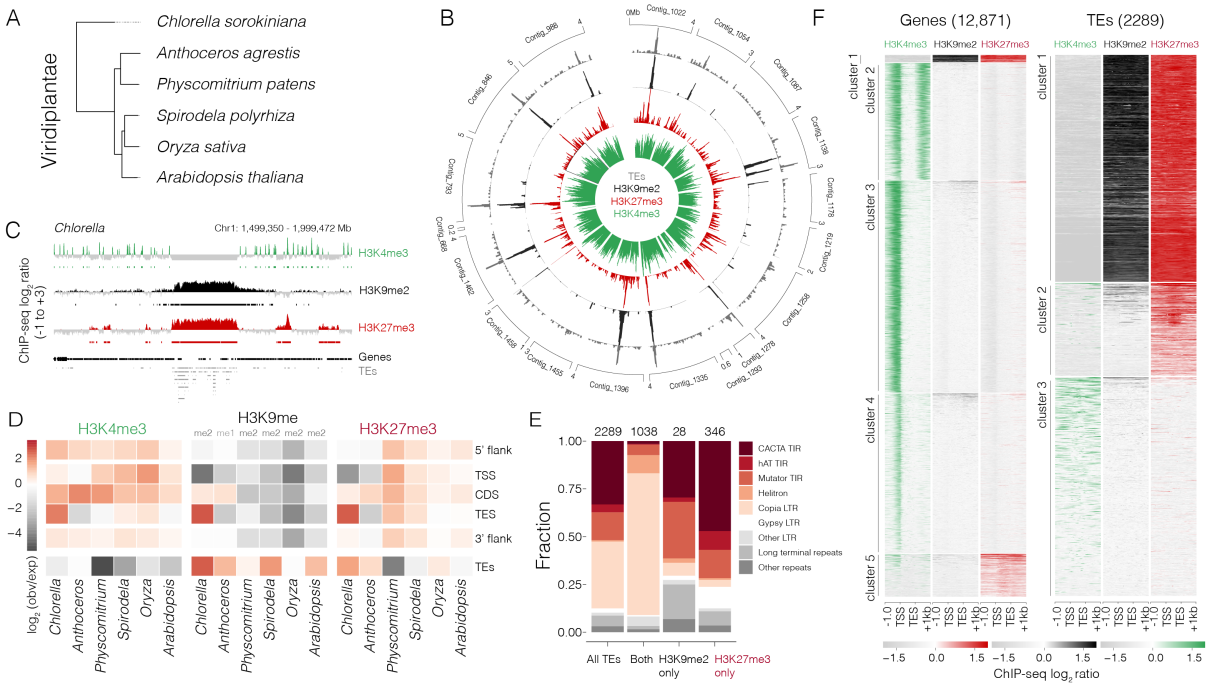


Figure 1. The heterochromatin landscape of the unicellular chlorophyte alga *Chlorella sorokiniana*

(**A**) Phylogenetic relationships of the six species analysed in this study. (**B**) Circos plot representing the density of TEs (grey) and ChIP-seq peaks of H3K9me2 (black), H3K27me3 (red) and H3K4me3 (green) along each chromosome in the *Chlorella* genome. (**C**) ChIP-Seq tracks of the TE-rich hotspot on chromosome 1. Coloured and grey shading indicate enriched or depleted signals, respectively. Coverage is represented as the ChIP-seq $\log_2$ ratio relative to H3. (**D**) Distribution of H3K4me3, H3K9me2 and H3K27me3 peaks at genomic features across the green lineage. The $\log_2$ enrichment or depletion is calculated relative to the frequency of genomic features in each species relative to 10,000 random permutations. TSS = translation start site, CDS = coding sequence, TES = translation end site, TEs = transposable elements. (**E**) Distribution of TE classes marked with either H3K9me2, H3K27me3 or both alongside their relative frequency of all TE classes in the Chlorella genome. (**F**) Heatmaps of H3K4me3, H3K9me3 and H3K27me3 enrichment centred on genes and TEs in *Chlorella*. Plotted is the ChIP-seq $\log_2$ ratio relative to H3. ChIP-seq was performed with two biological replicates.

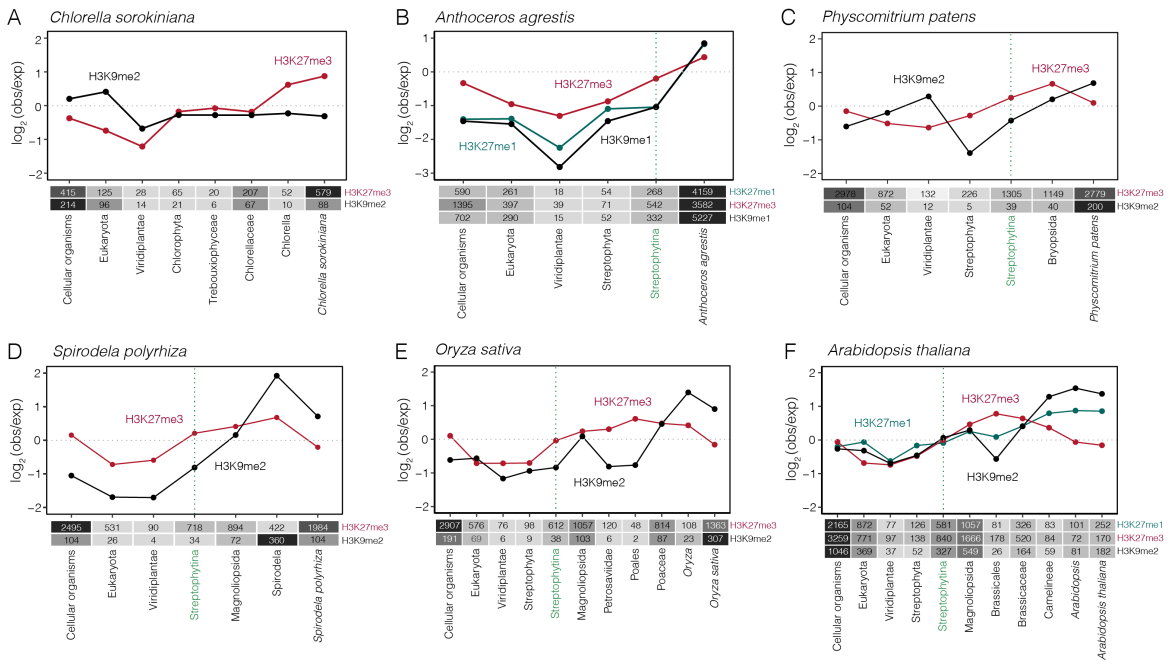


Figure 2. Epigenetic state correlates with the evolutionary age of genes across the Viridiplantae.

Line plots showing the relative enrichment of H3K9me1/2 (black), H3K27me1 (green) and H3K27me3 (red) at genes grouped by their phylostratigraphic (or evolutionary) age in *Chlorella* (**A**), *Anthoceros* (**B**), *Physcomitrium* (**C**), *Spirodela* (**D**), *Oryza* (**E**) and *Arabidopsis* (**F**). Genes are ordered left-to-right from the oldest genes (i.e., those with homology in all cellular organisms) to the youngest (i.e., those restricted to each respective species). Plotted is the $\log_2$ ratio of marked genes relative to all genes in each phylogenetic rank. The green dotted line represents the Streptophytina rank when complex multicelluarity arose during streptophyte evolution. The heatmap below each plot summarises the total number of genes marked with either epigenetic mark in each rank.

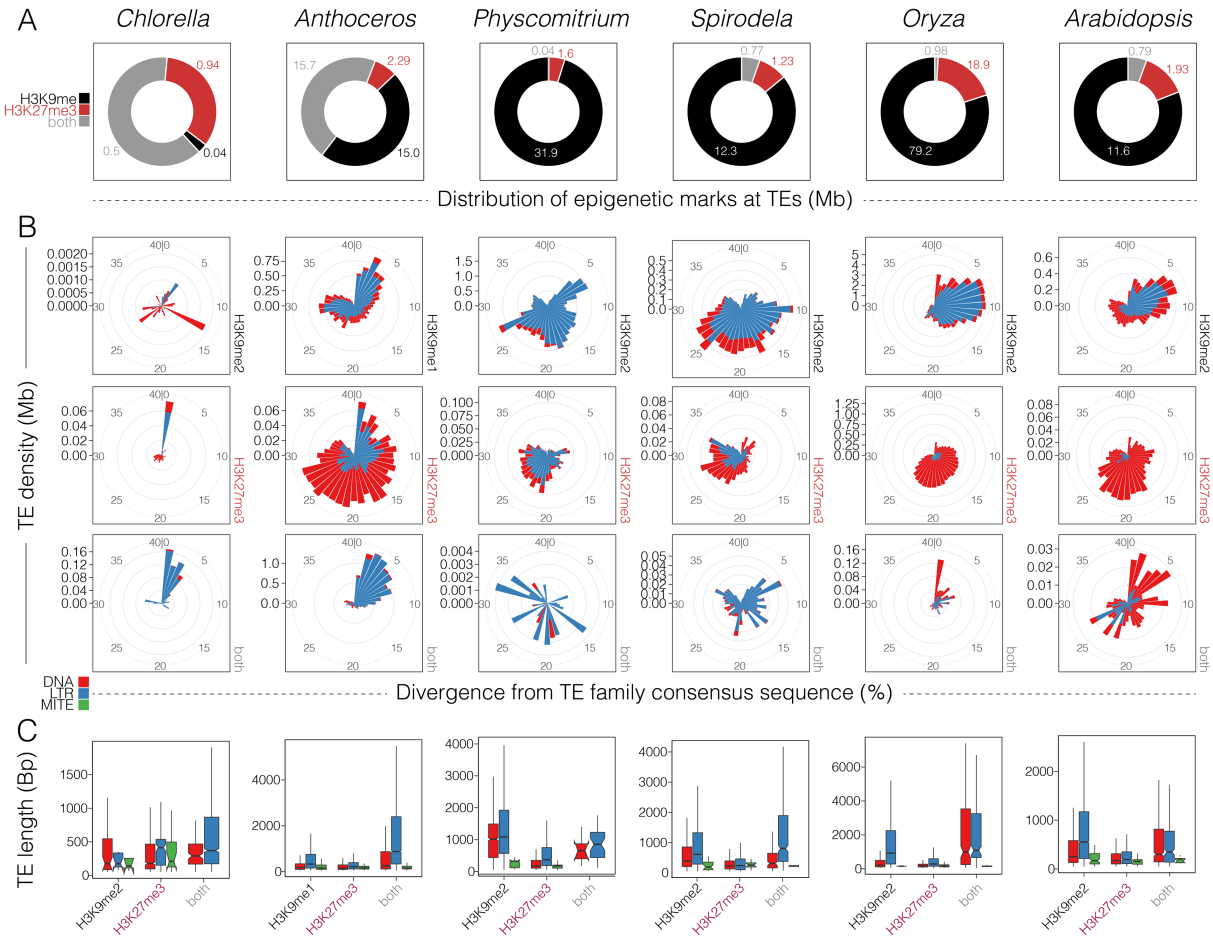


Figure 3. Evolutionary age and epigenetic state of TEs across the green lineage.

(**A**) Each pie chart summarises the relative distribution of all DNA, LTR and MITE transposons marked with H3K9me1/2 only (black), H3K27me3 only (red) or both (grey) in each species. The total genomic space taken up by each group is indicated in Mb. (**B**) Circular histograms summarising the relative age and distribution of DNA (red), LTR (blue) and MITE (green) transposons marked with H3K9me1/2 only (top row), H3K27me3 only (middle row) or both (bottom row) in each species (grouped by column, species indicated above). Increasing sequence divergence of TEs from the consensus sequence of their assigned family is plotted clockwise, with higher divergence indicative of an older age and reduced potential for transposition. (**C**) The length distribution of DNA, LTR and MITE transposons marked with H3K9me1/2 only, H3K27me3 only or both in each species. Each boxplot indicates minimum and maximum values as well as 25th, 50th, and 75th quartiles.

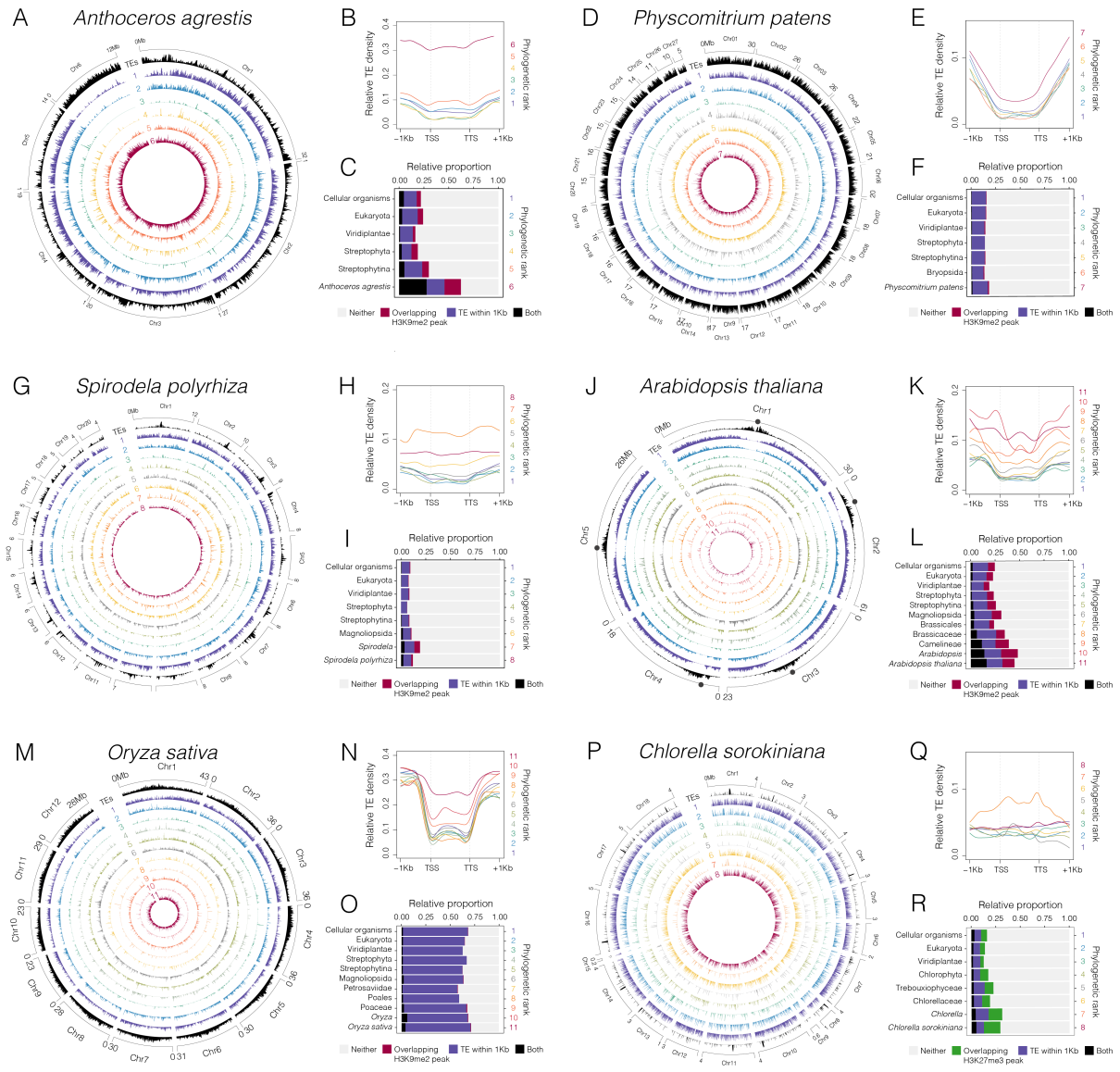


Figure 4. Neighbouring TEs spread constitutive heterochromatin to evolutionarily young genes

A trio of panels are grouped for *Anthoceros* (**A-C**), *Physcomitrium* (**D-F**), *Spirodela* (**G-I**), *Arabidopsis* (**J-L**), *Oryza* (**M-O**) and *Chlorella* (**P-R**) under their respective species name. The first element of the trio is a circos plot (A,D,G,J,M,P) showing the chromosomal density of phylogenetically-ranked genes in each species. Phylogenetically-ranked genes are colour-coded from old (rank 1) to young (highest rank) according to the colour scheme indicated in the circos plot tracks. The second element is a profile plot (B,E,H,K,N,Q) of relative TE density centred on each group of phylogenetically-ranked genes. The third element is a barplot showing the relative proportion of genes in each phylogenetic rank that lie within 1 kb of a TE (purple), that overlap with an H3K9me1/2 peak (black) (or H3K27me3 peak in *Chlorella* in green), or both (maroon).

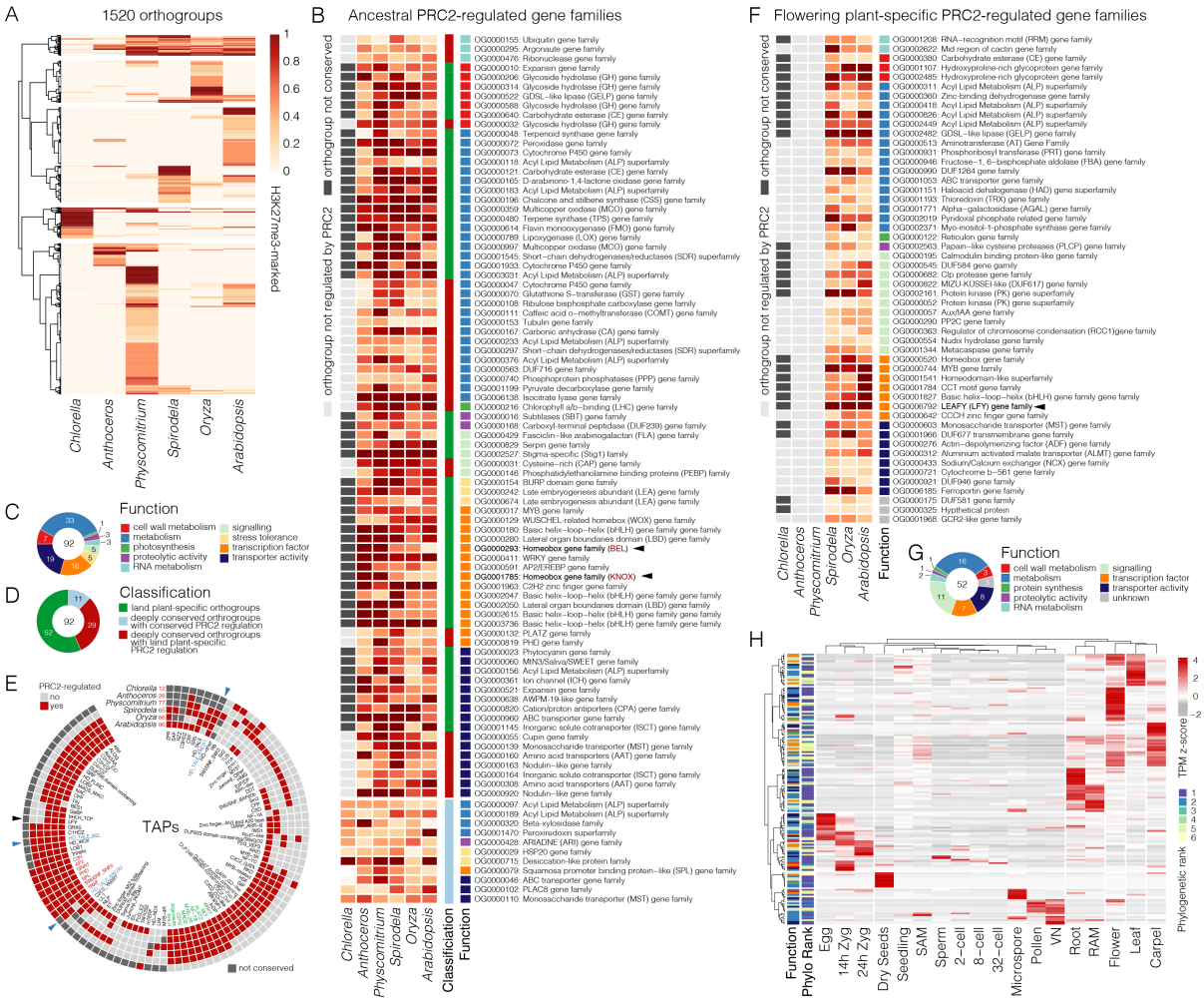


Figure 5. The conserved topology of PRC2-regulated gene networks across the green lineage

(**A**) Heatmap summarising the 1520 orthogroups (or gene families) conserved across the six species analysed in this study. The relative proportion of genes in each orthogroup marked with H3K27me3 in each species are indicated and coloured according to the inset scale. (**B**) Heatmap showing the cluster of gene families marked with H3K27me3 in all five land plants using the colour scale in panel B. Conserved orthogroups containing no genes marked with H3K27me3 in *Chlorella* are shown in light grey, whereas orthogroups not conserved in *Chlorella* are shown in dark grey (see classification column and panel D). Assigned functions are coloured according to panel C. (**C**) Pie chart summarising the different functions of the gene families in panel B. (**D**) Pie chart summarising the number of orthogroups in panel B. This was composed of PRC2-regulated gene families only conserved in the five land plants (green), gene families conserved in *Chlorella* but only marked with H3K27me3 in land plants (red), and PRC2-regulated gene families deeply-conserved in all six species (light blue). (**E**) Circular heatmap summarising the 102 transcription-associated protein (TAP) families regulated by PRC2 across each species. TAPs with deeply-conserved regulation are indicated in red, land plant-specific regulation in green, and BELL-KNOX homologs in blue. (**F**) Heatmap showing PRC2-regulated gene families specific to flowering plants. (**G**) Pie chart summarising the different functions of the gene families in panel F. (**H**) Heatmap illustrating the developmentally regulated expression of the *Arabidopsis* PRC2-regulated gene families specific to flowering plants in panel F. Expression represents the z-score of normalised RNA-seq TPM values.

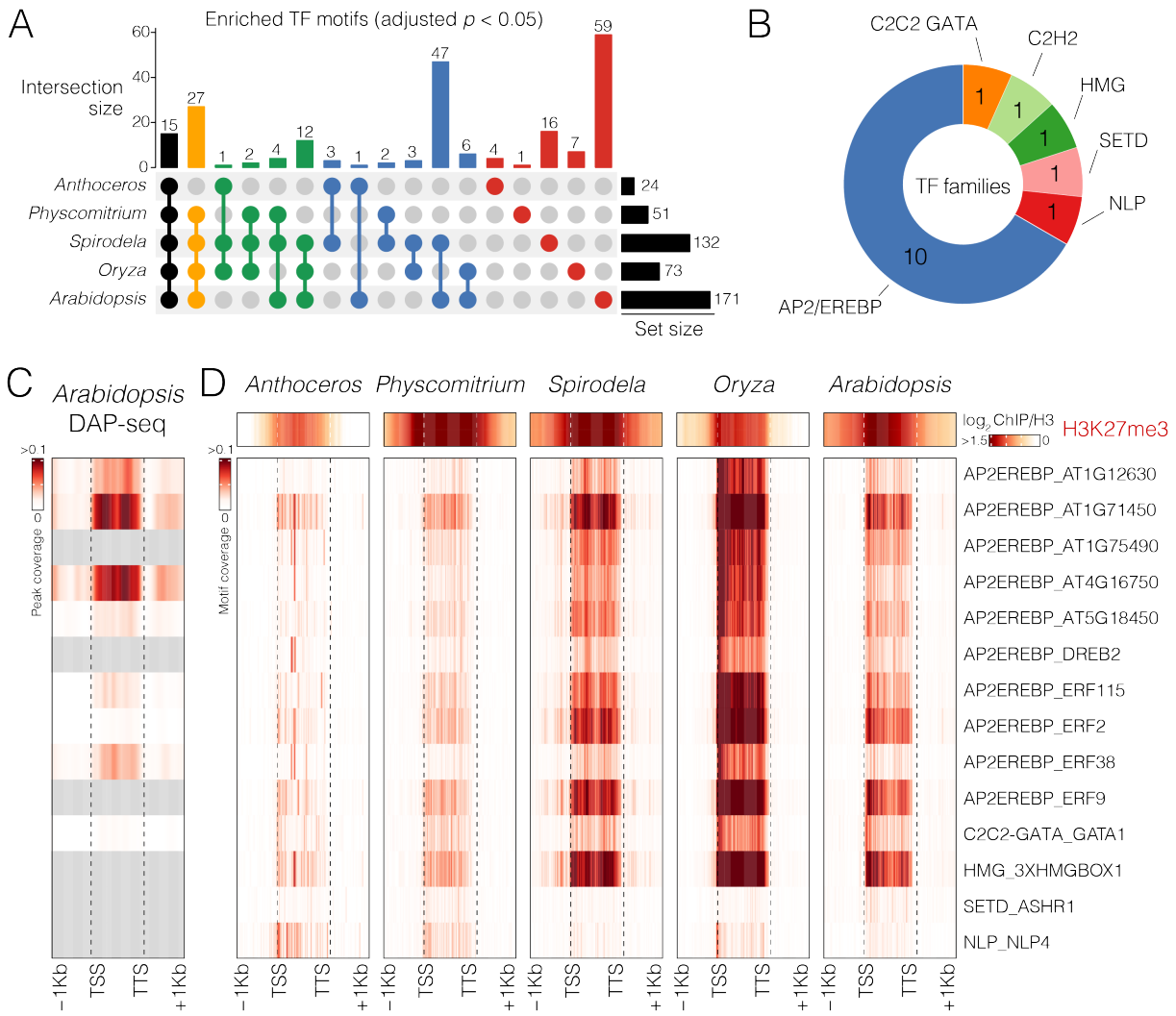


Figure 6. Ancestral PRC2 target genes share DNA binding motifs across large evolutionary distances

(**A**) Upset plot summarising the overlap and number of DNA binding motifs significantly enriched (adjusted *p*-value < 0.05) within H3K27me3 domains marking the ancestral land plant network of PRC2-regulated genes in each land plant species (see Fig. 4B; Supplemental Table S6). (**B**) The number and type of DNA binding motifs that are commonly enriched across all land plants in panel A (black bar). (**C**) Averaged signal of DAP-seq peaks centred on the PRC2-regulated *Arabidopsis* genes that form part of the ancestral land plant network. (**B**) Occurrence of the commonly enriched DNA binding motifs (panel B) within H3K27me3 domains marking the ancestral PRC2-regulated genes in each land plant species. ChIP-seq $\log_2$ ratio of H3K27me3 enrichment relative to H3 over the same genes is indicated in the heatmaps above.

References

Baile F, Gómez-Zambrano Á, Calonje M. **2021**. Roles of Polycomb complexes in regulating gene expression and chromatin structure in plants. *Plant communications* **3**.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. **2009**. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research* **37**.

Barrera-Redondo J, Lotharukpong JS, Drost HG, Coelho SM. **2023**. Uncovering gene-family founder events during major evolutionary transitions in animals, plants and fungi using GenEra. *Genome Biology* **24**: 1–21.

Bedre R, Mandadi K. **2019**. GenFam: A web application and database for gene family-based classification and functional enrichment analysis. *Plant Direct* **3**: e00191.

Bellaoui M, Pidkowich MS, Samach A, Kushalappa K, Kohalmi SE, Modrusan Z, Crosby WL, Haughn GW. **2001**. The Arabidopsis BELL1 and KNOX TALE Homeodomain Proteins Interact through a Domain Conserved between Plants and Animals. *The Plant Cell* **13**: 2455.

Bi G, Zhao S, Yao J, Wang H, Zhao M, Sun Y, Hou X, Haas FB, Varshney D, Prigge M, *et al.* 2024. Near telomere-to-telomere genome of the model plant Physcomitrium patens. *Nature plants* **10**: 327–343.

Blackledge NP, Klose RJ. **2021a**. The molecular principles of gene regulation by Polycomb repressive complexes. *Nature Reviews Molecular Cell Biology 2021 22:12* **22**: 815–833.

Blackledge NP, Klose RJ. **2021b**. The molecular principles of gene regulation by Polycomb repressive complexes. *Nature Reviews Molecular Cell Biology 2021 22:12* **22**: 815–833.

Borg M, Jacob Y, Susaki D, LeBlanc C, Buendía D, Axelsson E, Kawashima T, Voigt P, Boavida L, Becker J, *et al.* 2020. Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. *Nature cell biology*: 1–9.

Borg M, Jiang D, Berger F. **2021a**. Histone variants take center stage in shaping the epigenome. *Current Opinion in Plant Biology* **61**.

Borg M, Papareddy RK, Dombey R, Axelsson E, Nodine MD, Twell D, Berger F. **2021b**. Epigenetic reprogramming rewires transcription during the alternation of generations in arabidopsis. *eLife* **10**.

Bowles AMC, Bechtold U, Paps J. **2020**. The Origin of Land Plants Is Rooted in Two Bursts of Genomic Novelty. *Current Biology* **30**: 530-536.e2.

Bowman JL. **2022**. The origin of a land flora. *Nature Plants 2022 8:12* **8**: 1352–1369.

Bray NL, Pimentel H, Melsted P, Pachter L. **2016**. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**: 525–527.

Chalopin D, Naville M, Plard F, Galiana D, Volff JN. **2015**. Comparative Analysis of Transposable Elements Highlights Mobilome Diversity and Evolution in Vertebrates. *Genome Biology and Evolution* **7**: 567.

Cutter DiPiazza AR, Taneja N, Dhakshnamoorthy J, Wheeler D, Holla S, Grewal SIS. **2021**. Spreading and epigenetic inheritance of heterochromatin require a critical density of histone H3 lysine 9 tri-methylation. *Proceedings of the National Academy of Sciences of the United States of America* **118**.

Déléris A, Berger F, Duharcourt S. **2021**. Role of Polycomb in the control of transposable elements. *Trends in genetics : TIG* **37**: 882–889.

Dhabalia Ashok A, de Vries S, Darienko T, Irisarri I, de Vries J. **2024**. Evolutionary assembly of the plant terrestrialization toolkit from protein domains. *Proceedings B* **291**: 20240985.

Dierschke T, Flores-Sandoval E, Rast-Somssich MI, Althoff F, Zachgo S, Bowman JL. **2021**. Gamete-specific expression of TALE class HD genes activates the diploid sporophyte program in Marchantia polymorpha. *eLife* **10**.

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

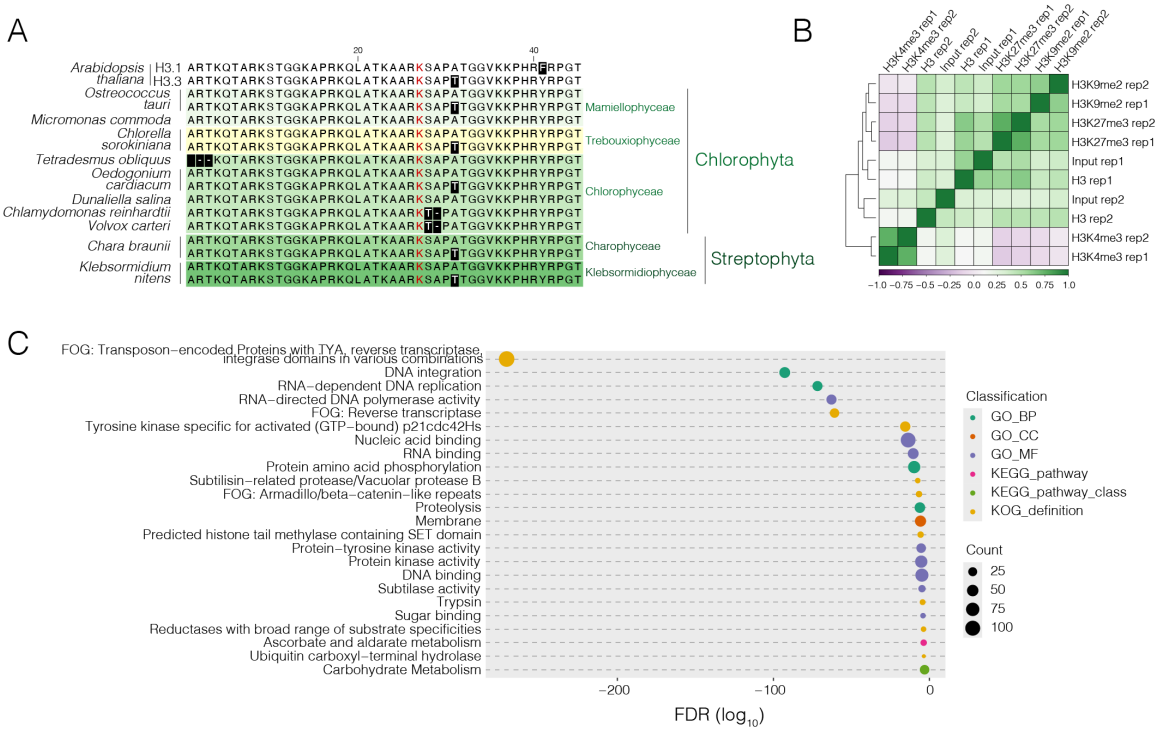
Dombey R, Barragán-Borrero V, Buendía-Ávila D, Ponce-Mañe A, Vargas-Guerrero JM, Elias R, Marí-Ordóñez A. **2024**. Atypical epigenetic and small RNA control of transposons in clonally reproducing Spirodela polyrhiza. *bioRxiv*: 2024.04.03.587901.
Du J, Johnson LM, Jacobsen SE, Patel DJ. **2015**. DNA methylation pathways and their crosstalk with histone methylation. *Nature Reviews Molecular Cell Biology 2015 16:9* **16**: 519–532.
Emms DM, Kelly S. **2019**. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 1–14.
Evanovich E, Mendonça-Mattos PJ de S, Guerreiro JF. **2020**. A Timescale for the Radiation of Photosynthetic Eukaryotes. *bioRxiv*: 2020.04.18.047969.
Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. **2020**. The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology 2020 38:3* **38**: 276–278.
Fedoroff N V. 2012. Transposable Elements, Epigenetics, and Genome Evolution. *Science*.
Feng S, Jacobsen SE, Reik W. **2010**. Epigenetic reprogramming in plant and animal development. *Science (New York, N.Y.)* **330**: 622–7.
Flores-Sandoval E, Eklund DM, Bowman JL. **2015**. A Simple Auxin Transcriptional Response System Regulates Multiple Morphogenetic Processes in the Liverwort Marchantia polymorpha. *PLoS Genetics* **11**.
Frapporti A, Miró Pina C, Arnaiz O, Holoch D, Kawaguchi T, Humbert A, Eleftheriou E, Lombard B, Loew D, Sperling L, *et al.* 2019. The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in Paramecium. *Nature Communications* **10**.
Gallego-Bartolomé J, Liu W, Kuo PH, Feng S, Ghoshal B, Gardiner J, Zhao JMC, Park SY, Chory J, Jacobsen SE. **2019**. Co-targeting RNA Polymerases IV and V Promotes Efficient De Novo DNA Methylation in Arabidopsis. *Cell* **176**: 1068-1082.e19.
Gombar S, MacCarthy T, Bergman A. **2014**. Epigenetics Decouples Mutational from Environmental Robustness. Did It Also Facilitate Multicellularity? *PLoS Computational Biology* **10**.
Gu Z, Eils R, Schlesner M, Ishaque N. **2018**. EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. *BMC Genomics* **19**: 234.
Gu Z, Gu L, Eils R, Schlesner M, Brors B. **2014**. circlize Implements and enhances circular visualization in R. *Bioinformatics (Oxford, England)* **30**: 2811–2812.
Gueno J, Borg M, Bourdareau S, Cossard G, Godfroy O, Lipinska A, Tirichine L, Cock JM, Coelho SM. **2013**. Chromatin landscape associated with sexual differentiation in a UV sex determination system. *Nucleic Acids Research* **1**: 13–14.
Harkess A, Bewick AJ, Lu Z, Fourounjian P, Michael TP, Schmitz RJ, Meyers BC. **2024**. The unusual predominance of maintenance DNA methylation in Spirodela polyrhiza. *G3 Genes|Genomes|Genetics* **14**.
Hauenschild A, Ringrose L, Altmutter C, Paro R, Rehmsmeier M. **2008**. Evolutionary plasticity of polycomb/trithorax response elements in Drosophila species. *PLoS Biology* **6**.
Hisanaga T, Fujimoto S, Cui Y, Sato K, Sano R, Yamaoka S, Kohchi T, Berger F, Nakajima K. **2021**. Deep evolutionary origin of gamete-directed zygote activation by KNOX/BELL transcription factors in green plants. *eLife* **10**.
Hisanaga T, Romani F, Wu S, Kowar T, Lintermann R, Jamge B, Montgomery SA, Axelsson E, Akimcheva S, Dierschke T, *et al.* 2023. The Polycomb repressive complex 2 deposits H3K27me3 and represses transposable elements in a broad range of eukaryotes. *Current Biology*: 4367–4380.
Hisanaga T, Wu S, Schafran P, Axelsson E, Akimcheva S, Dolan L, Li FW, Berger F. **2023**. The ancestral chromatin landscape of land plants. *New Phytologist* **240**: 2085–2101.
Hlavová M, Vítová M, Bišová K. **2016**. Synchronization of Green Algae by Light and Dark Regimes for Cell Cycle and Cell Division Studies. *Methods in molecular biology (Clifton, N.J.)* **1370**: 3–16.
Horst NA, Katz A, Pereman I, Decker EL, Ohad N, Reski R. **2016**. A single homeobox gene triggers phase transition, embryogenesis and asexual reproduction. *Nature Plants* **2**.

Hovde BT, Hanschen ER, Steadman Tyler CR, Lo CC, Kunde Y, Davenport K, Daligault H, Msanne J, Canny S, Eyun S il, *et al.* 2018. Genomic characterization reveals significant divergence within Chlorella sorokiniana (Chlorellales, Trebouxiophyceae). *Algal Research* **35**: 449–461.
Huang Y, Chen DH, Liu BY, Shen WH, Ruan Y. **2017**. Conservation and diversification of polycomb repressive complex 2 (PRC2) proteins in the green lineage. *Briefings in Functional Genomics* **16**.
Huang Y, Jiang L, Liu BY, Tan CF, Chen DH, Shen WH, Ruan Y. **2019**. Evolution and conservation of polycomb repressive complex 1 core components and putative associated factors in the green lineage. *BMC Genomics* **20**.
Hung FY, Feng YR, Hsin KT, Shih YH, Chang CH, Zhong W, Lai YC, Xu Y, Yang S, Sugimoto K, *et al.* 2023. Arabidopsis histone H3 lysine 9 methyltransferases KYP/SUVH5/6 are involved in leaf development by interacting with AS1-AS2 to repress KNAT1 and KNAT2. *Communications Biology 2023 6:1* **6**: 1–13.
Hure V, Piron-Prunier F, Yehouessi T, Vitte C, Kornienko AE, Adam G, Nordborg M, Déléris A. **2024**. Alternative silencing states of Transposable Elements in Arabidopsis. *bioRxiv*: 2024.03.16.585326.
Ibarra CA, Feng X, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T, *et al.* 2012. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337**: 1360–4.
Kariyawasam T, Joo S, Lee J, Toor D, Gao AF, Noh KC, Lee JH. **2019**. TALE homeobox heterodimer GSM1/GSP1 is a molecular switch that prevents unwarranted genetic recombination in Chlamydomonas. *The Plant Journal* **100**: 938–953.
Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin X, *et al.* 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature 2000 408:6814* **408**: 796–815.
Kawashima T, Berger F. **2014**. Epigenetic reprogramming in plant sexual reproduction. *Nature Reviews Genetics*: 1–12.
Kfoury B, Felipe W, Rodrigues C, Kim S-J, Brandizzi F, Del-Bem L-E. **2024**. Multiple horizontal gene transfer events have shaped plant glycosyl hydrolase diversity and function. *New Phytologist* **242**: 809–824.
Khan A, Eikani CK, Khan H, Iavarone AT, Pesavento JJ. **2018**. Characterization of Chlamydomonas reinhardtii Core Histones by Top-Down Mass Spectrometry Reveals Unique Algae-Specific Variants and Post-Translational Modifications. *Journal of Proteome Research* **17**.
Krueger F, Andrews SR. **2011**. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**: 1571–1572.
Lafos M, Kroll P, Hohenstatt ML, Thorpe FL, Clarenz O, Schubert D. **2011**. Dynamic Regulation of H3K27 Trimethylation during Arabidopsis Differentiation. *PLOS Genetics* **7**: e1002040.
Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. **2013**. Software for Computing and Annotating Genomic Ranges. *PLOS Computational Biology* **9**: e1003118.
Lee JH, Lin H, Joo S, Goodenough U. **2008**. Early Sexual Origins of Homeoprotein Heterodimerization and Evolution of the Plant KNOX/BELL Family. *Cell* **133**: 829–840.
Leliaert F, Verbruggen H, Zechman FW. **2011**. Into the deep: new discoveries at the base of the green plant phylogeny. *BioEssays : news and reviews in molecular, cellular and developmental biology* **33**: 683–692.
Li FW, Nishiyama T, Waller M, Frangedakis E, Keller J, Li Z, Fernandez-Pozo N, Barker MS, Bennett T, Blázquez MA, *et al.* 2020. Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nature Plants 2020 6:3* **6**: 259–272.
Loubiere V, Martinez AM, Cavalli G. **2019**. Cell Fate and Developmental Regulation Dynamics by Polycomb Proteins and 3D Genome Architecture. *BioEssays : news and reviews in molecular, cellular and developmental biology* **41**.
Love MI, Huber W, Anders S. **2014**. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**.

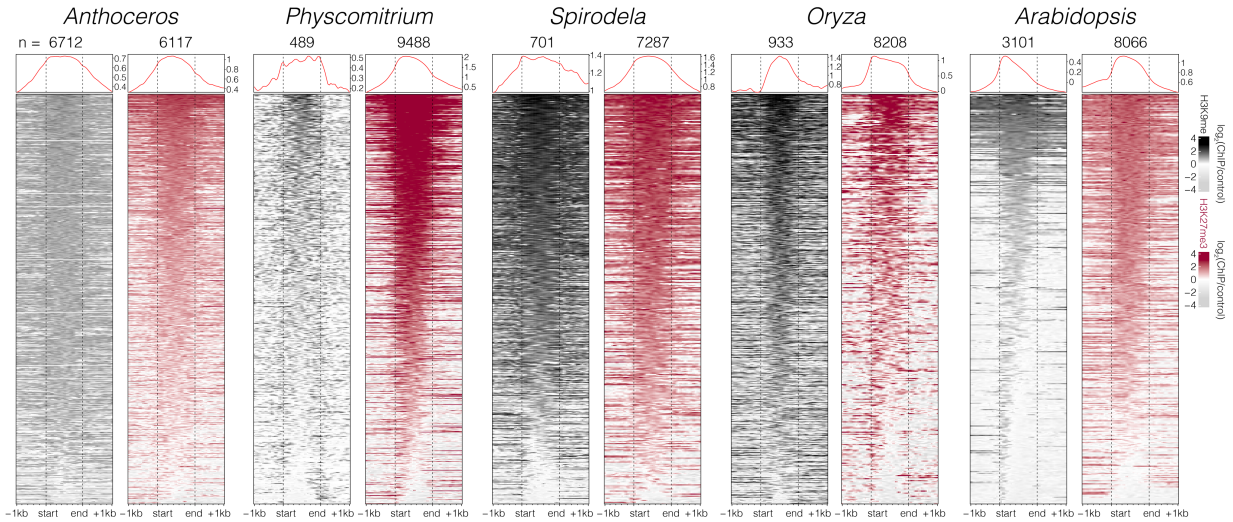
Lüleci HB, Yılmaz A. **2022**. Robust and rigorous identification of tissue-specific genes by statistically extending tau score. *BioData Mining* **15**: 1–14.
Luthringer R, Lipinska AP, Roze D, Cormier A, Macaisne N, Peters AF, Cock JM, Coelho SM. **2015**. The Pseudoautosomal Regions of the U/V Sex Chromosomes of the Brown Alga Ectocarpus Exhibit Unusual Features. *Molecular Biology and Evolution* **32**: 2973–2985.
Martinoia E, Massonneau A, Frangne N. **2000**. Transport Processes of Solutes across the Vacuolar Membrane of Higher Plants. *Plant and Cell Physiology* **41**: 1175–1186.
Masaki T, Tsukagoshi H, Mitsui N, Nishii T, Hattori T, Morikami A, Nakamura K. **2005**. Activation tagging of a gene for a protein with novel class of CCT-domain activates expression of a subset of sugar-inducible genes in Arabidopsis thaliana. *The Plant Journal* **43**: 142–152.
Mikulski P, Komarynets O, Fachinelli F, Weber APM, Schubert D. **2017**. Characterization of the polycomb-group mark H3K27me3 in unicellular algae. *Frontiers in Plant Science* **8**.
Moczydlowska M, Landing E, Zang W, Palacios T. **2011**. Proterozoic phytoplankton and timing of Chlorophyte algae origins. *Palaeontology* **54**: 721–733.
Montgomery SA, Tanizawa Y, Galik B, Wang N, Ito T, Mochizuki T, Akimcheva S, Bowman JL, Cognat V, Maréchal-Drouard L, *et al.* 2020. Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. *Current Biology* **30**.
Mozgová I, Wildhaber T, Liu Q, Abou-Mansour E, L'Haridon F, Métraux JP, Gruissem W, Hofius D, Hennig L. **2015**. Chromatin assembly factor CAF-1 represses priming of plant defence response genes. *Nature Plants 2015 1:9* **1**: 1–8.
Nagata T, Iizumi S, Satoh K, Kikuchi S. **2008**. Comparative molecular biological analysis of membrane transport genes in organisms. *Plant Molecular Biology* **66**.
Ngan CY, Wong CH, Choi C, Yoshinaga Y, Louie K, Jia J, Chen C, Bowen B, Cheng H, Leonelli L, *et al.* 2015. Lineage-specific chromatin signatures reveal a regulator of lipid metabolism in microalgae. *Nature plants* **1**.
Nishimura Y, Shikanai T, Nakamura S, Kawai-Yamada M, Uchimiya H. **2012**. Gsp1 Triggers the Sexual Developmental Program Including Inheritance of Chloroplast DNA and Mitochondrial DNA in Chlamydomonas reinhardtii. *The Plant Cell* **24**: 2401–2414.
Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, Hens K, Toivonen J, Deplancke B, Furlong EEM, *et al.* 2015. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *eLife* **2015**.
O'Malley RC, Huang SSC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. **2016**. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell*.
Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, *et al.* 2019a. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* **20**: 1–18.
Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, *et al.* 2019b. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* **20**: 1–18.
Park K, Kim MY, Vickers M, Park JS, Hyun Y, Okamoto T, Zilberman D, Fischer RL, Feng X, Choi Y, *et al.* 2016. DNA demethylation is initiated in the central cells of Arabidopsis and rice. *Proceedings of the National Academy of Sciences of the United States of America* **113**.
Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. **2017**. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods 2017 14:4* **14**: 417–419.
Pereman I, Mosquna A, Katz A, Wiedemann G, Lang D, Decker EL, Tamada Y, Ishikawa T, Nishiyama T, Hasebe M, *et al.* 2016. The Polycomb group protein CLF emerges as a specific tri-methylase of H3K27 regulating gene expression and development in Physcomitrella patens. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1859**.

Petroll R, Varshney D, Hiltemann S, Finke H, Schreiber M, Vries J de, Rensing SA. **2024**. Enhanced sensitivity of TAPscan v4 enables comprehensive analysis of streptophyte transcription factor evolution. *bioRxiv*: 2024.07.13.602682.
Pillot M, Baroux C, Vazquez MA, Autran D, Leblanc O, Vielle-Calzada JP, Grossniklaus U, Grimanelli D. **2010**. Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in Arabidopsis. *Plant Cell* **22**.
Preston JC, Hileman LC. **2013**. Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. *Frontiers in Plant Science* **4**: 46893.
Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. **2014**. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research* **42**: W187-91.
Reik W, Dean W, Walter J. **2001**. Epigenetic reprogramming in mammalian development. *Science* **293**: 1089–1093.
Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. **2011**. Integrative genomics viewer. *Nature biotechnology* **29**: 24–26.
Romani F, Moreno JE. **2021**. Molecular mechanisms involved in functional macroevolution of plant transcription factors. *New Phytologist* **230**: 1345–1353.
Rutley N, Poidevin L, Doniger T, Tillett RL, Rath A, Forment J, Luria G, Schlauch KA, Ferrando A, Harper JF, *et al.* 2021. Characterization of novel pollen-expressed transcripts reveals their potential roles in pollen heat stress response in Arabidopsis thaliana. *Plant Reproduction* **34**: 61–78.
Sakakibara K, Ando S, Yip HK, Tamada Y, Hiwatashi Y, Murata T, Deguchi H, Hasebe M, Bowman JL. **2013**. KNOX2 genes regulate the haploid-to-diploid morphological transition in land plants. *Science* **339**.
Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H, *et al.* 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature 2015 523:7559* **523**: 212–216.
Sharaf A, Vijayanathan M, Oborník M, Mozgová I. **2022**. Phylogenetic profiling resolves early emergence of PRC2 and illuminates its functional core. *Life Science Alliance* **5**.
Shaver S, Casas-Mollano JA, Cerny RL, Cerutti H. **2010**. Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. *Epigenetics* **5**.
Strenkert D, Schmollinger S, Schroda M. **2011**. Protocol: Methodology for chromatin immunoprecipitation (ChIP) in Chlamydomonas reinhardtii. *Plant Methods* **7**: 1–9.
Strenkert D, Yildirim A, Yan J, Yoshinaga Y, Pellegrini M, O'Malley RC, Merchant SS, Umen JG. **2022**. The landscape of Chlamydomonas histone H3 lysine 4 methylation reveals both constant features and dynamic changes during the diurnal cycle. *The Plant Journal* **112**: 352–368.
Suzuki H, Kato H, Iwano M, Nishihama R, Kohchi T. **2023**. Auxin signaling is essential for organogenesis but not for cell survival in the liverwort Marchantia polymorpha. *Plant Cell* **35**.
Umen JG. **2014**. Green Algae and the Origins of Multicellularity in the Plant Kingdom. *Cold Spring Harbor Perspectives in Biology* **6**.
Veluchamy A, Rastogi A, Lin X, Lombard B, Murik O, Thomas Y, Dingli F, Rivarola M, Ott S, Liu X, *et al.* 2015. An integrative analysis of post-translational histone modifications in the marine diatom Phaeodactylum tricornutum. *Genome Biology* **16**.
Vigneau J, Borg M. **2021**. The epigenetic origin of life history transitions in plants and algae. *Plant Reproduction 2021 34:4* **34**: 267–285.
Voichek Y, Hristova G, Mollá-Morales A, Weigel D, Nordborg M. **2024**. Widespread position-dependent transcriptional regulatory sequences in plants. *bioRxiv*: 2023.09.15.557872.
de Vries J, Archibald JM. **2018**. Plant evolution: landmarks on the path to terrestrial life. *New Phytologist* **217**: 1428–1434.

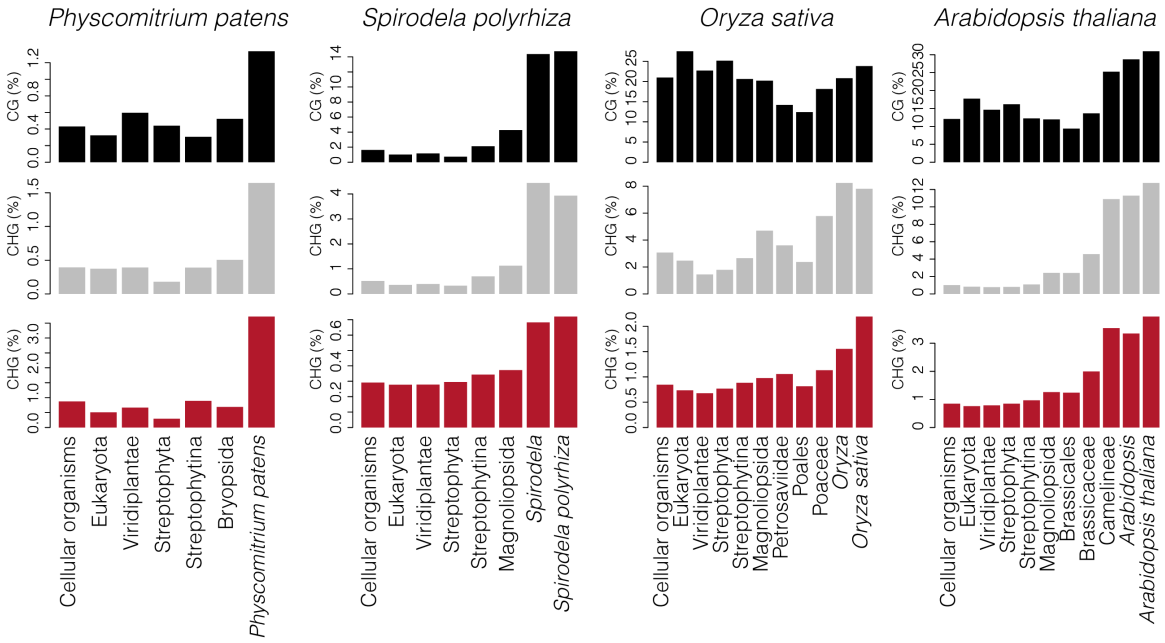
Wang B, Jia Y, Dang N, Yu J, Bush SJ, Gao S, He W, Wang S, Guo H, Yang X, *et al.* 2024. Near telomere-to-telomere genome assemblies of two Chlorella species unveil the composition and evolution of centromeres in green algae. *BMC Genomics* **25**: 1–14.
Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM. **1992**. LEAFY controls floral meristem identity in Arabidopsis. *Cell* **69**: 843–859.
Werner MS, Sieriebriennikov B, Prabh N, Loschko T, Lanz C, Sommer RJ. **2018**. Young genes have distinct gene structure, epigenetic profiles, and transcriptional regulation. *Genome Research* **28**: gr.234872.118.
Wickland DP, Hanzawa Y. **2015**. The FLOWERING LOCUS T/TERMINAL FLOWER 1 Gene Family: Functional Evolution and Molecular Mechanisms. *Molecular Plant* **8**: 983–997.
Widiez T, Symeonidi A, Luo C, Lam E, Lawton M, Rensing SA. **2014**. The chromatin landscape of the moss Physcomitrella patens and its dynamics during development and drought stress. *Plant Journal* **79**.
Wójcikowska B, Wójcik AM, Gaj MD. **2020**. Epigenetic Regulation of Auxin-Induced Somatic Embryogenesis in Plants. *International Journal of Molecular Sciences* **21**.
Wu DD, Wang X, Li Y, Zeng L, Irwin DM, Zhang YP. **2014**. "Out of Pollen" Hypothesis for Origin of New Genes in Flowering Plants: Study from Arabidopsis thaliana. *Genome Biology and Evolution* **6**: 2822.
Wu X, Xie L, Sun X, Wang N, Finnegan EJ, Helliwell C, Yao J, Zhang H, Wu X, Hands P, *et al.* 2023. Mutation in Polycomb repressive complex 2 gene OsFIE2 promotes asexual embryo formation in rice. *Nature Plants 2023 9:11* **9**: 1848–1861.
Xiao J, Jin R, Yu X, Shen M, Wagner JD, Pai A, Song C, Zhuang M, Klasfeld S, He C, *et al.* 2017. Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nature Genetics 2017 49:10* **49**: 1546–1552.
Xie W, Ding C, Hu H, Dong G, Zhang G, Qian Q, Ren D. **2022**. Molecular Events of Rice AP2/ERF Transcription Factors. *International Journal of Molecular Sciences* **23**.
Xu L, Yuan K, Yuan M, Meng X, Chen M, Wu J, Li J, Qi Y. **2020**. Regulation of Rice Tillering by RNA-Directed DNA Methylation at Miniature Inverted-Repeat Transposable Elements. *Molecular Plant* **13**: 851–863.
Yaari R, Katz A, Domb K, Harris KD, Zemach A, Ohad N. **2019**. RdDM-independent de novo and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs. *Nature Communications 2019 10:1* **10**: 1–10.
Yu G, Wang LG, Han Y, He QY. **2012**. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS : a Journal of Integrative Biology* **16**: 284.
Zhang C, Du X, Tang K, Yang Z, Pan L, Zhu P, Luo J, Jiang Y, Zhang H, Wan H, *et al.* 2018. Arabidopsis AGDP1 links H3K9me2 to DNA methylation in heterochromatin. *Nature communications* **9**.
Zhang J, Fu XX, Li RQ, Zhao X, Liu Y, Li MH, Zwaenepoel A, Ma H, Goffinet B, Guan YL, *et al.* 2020. The hornwort genome and early land plant evolution. *Nature Plants 2020 6:2* **6**: 107–118.
Zhang JY, Zhou Q. **2019**. On the Regulatory Evolution of New Genes Throughout Their Life History. *Molecular Biology and Evolution* **36**: 15–27.
Zhao X, Rastogi A, Deton Cabanillas AF, Ait Mohamed O, Cantrel C, Lombard B, Murik O, Genovesio A, Bowler C, Bouyer D, *et al.* 2021. Genome wide natural variation of H3K27me3 selectively marks genes predicted to be important for cell differentiation in Phaeodactylum tricornutum. *The New phytologist* **229**: 3208–3220.
Zhu LJ, Gazin C, Lawson ND, Pagès H, Lin SM, Lapointe DS, Green MR. **2010**. ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**: 1–10.
Zhu B, Reinberg D. **2011**. Epigenetic inheritance: Uncontested? *Cell Research* **21**.



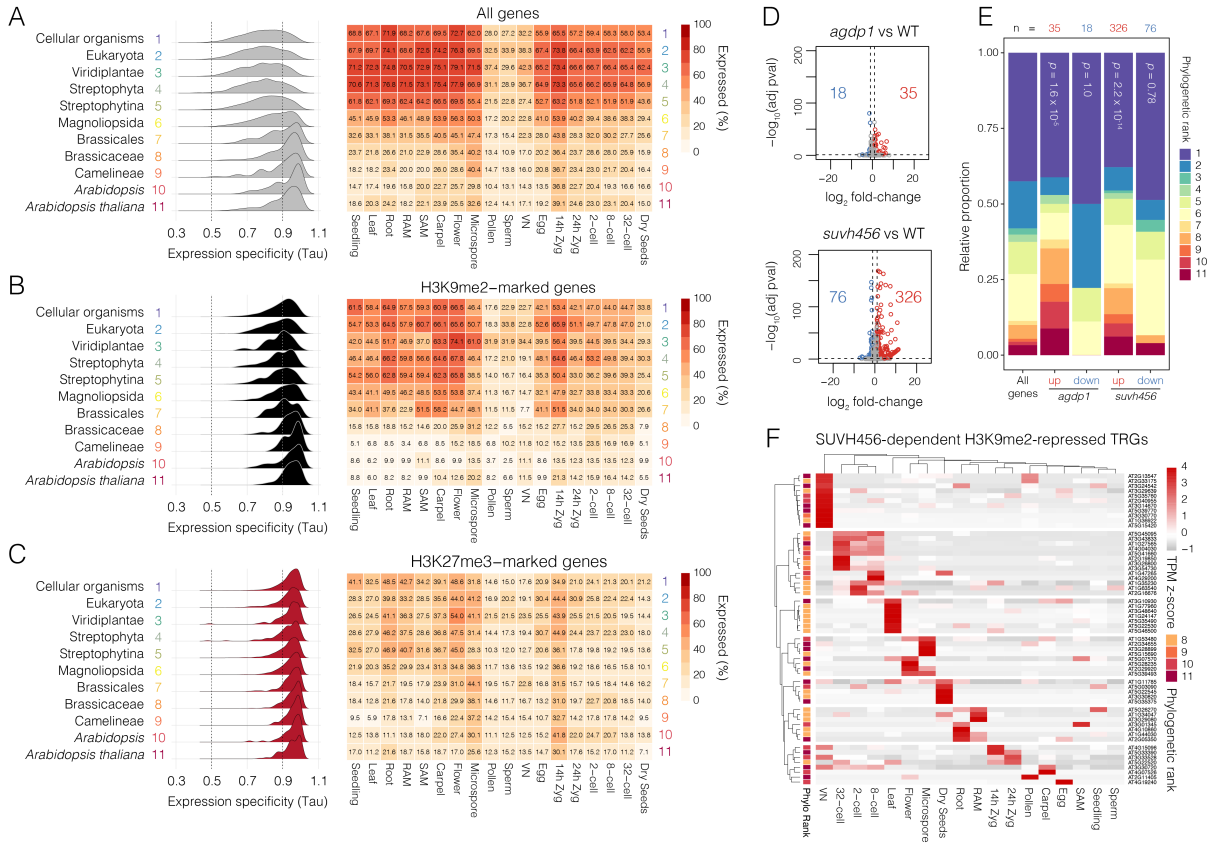
Supplemental Figure S1. (**A**) ClustalW alignment of the N-terminal tail of histone H3 variants among various clades of chlorophyte and streptophyte green algae. *Arabidopsis* histone H3.1 and H3.3 are included as an outgroup. Lysine 27 (K27) is highlighted in red. (**B**) Spearman cross-correlation matrix of the *Chlorella* ChIP-seq data sets generated in this study. (**C**) Enriched functional terms among the H3K27me3-marked genes in *Chlorella* (FDR < 0.05, hypergeometric test). Terms are plotted with their FDR (in log10) with dots coloured according to their classification and sized according to their count (number of genes among the H3K27me3-marked genes). GO_BP: Gene ontology, biological process; GO_MF: Gene ontology, molecular function; GO_CC: Gene ontology, cellular component; KEGG_pathway: Kyoto encyclopedia of genes and genomes, pathway; KEGG_pathway_class: Kyoto encyclopedia of genes and genomes, pathway class; KOG_definition: Eukaryotic orthologous groups, definition.



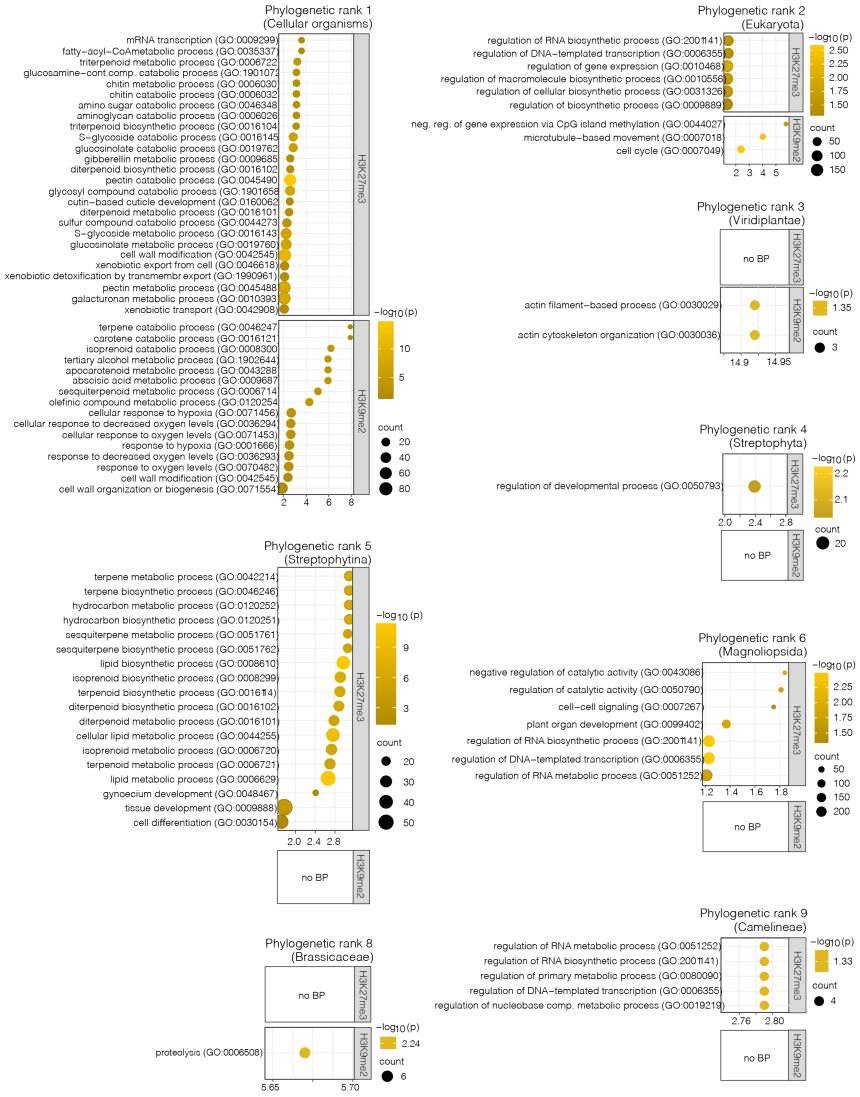
Supplemental Figure S2. Levels of H3K9me2 (or H3K9me1 in *Anthoceros*) and H3K27me3 marks centred over genes overlapping ChIP-seq peaks of each respective form of histone methylation in each land plant species (see Supplemental Table S2). Plotted is the $\log_2$ ChIP-seq enrichment relative H3 or input in the case of H3K9me2 in *Spirodela* and *Oryza*. Numbers of genes overlapping either form of histone methylation are indicated above each heatmap.



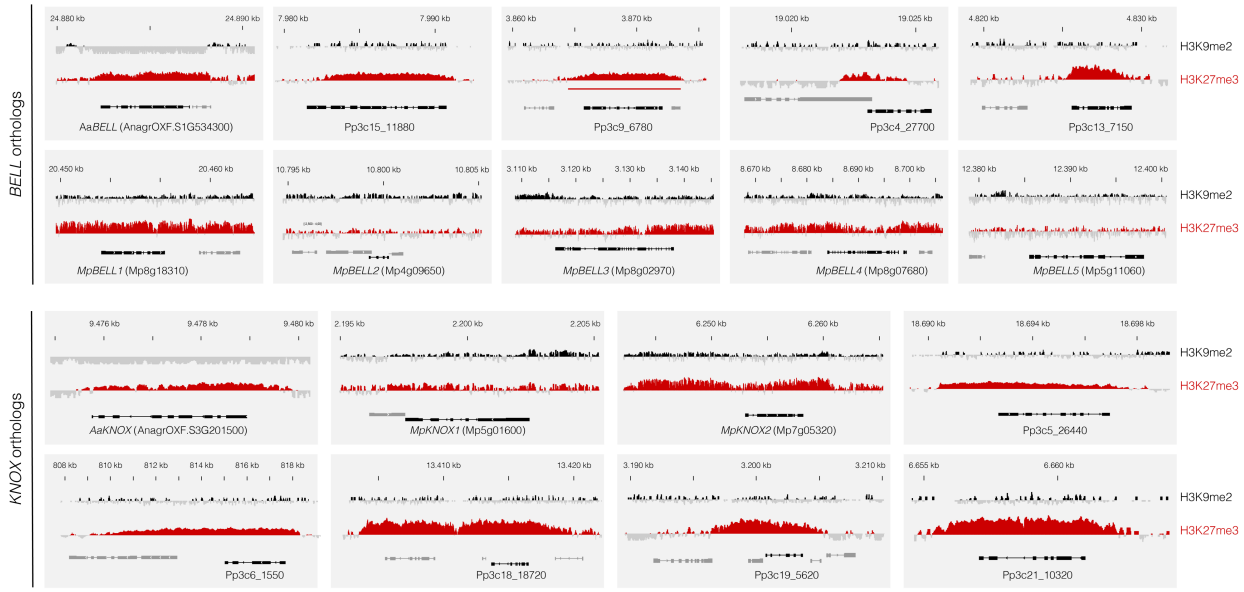
Supplemental Figure S3. Levels of CG and non-CG methylation over each group of phylogenetically-ranked genes in different land plant species (see Supplemental Table S2). Plotted is the percentage methylated cytosines overlapping gene bodies.



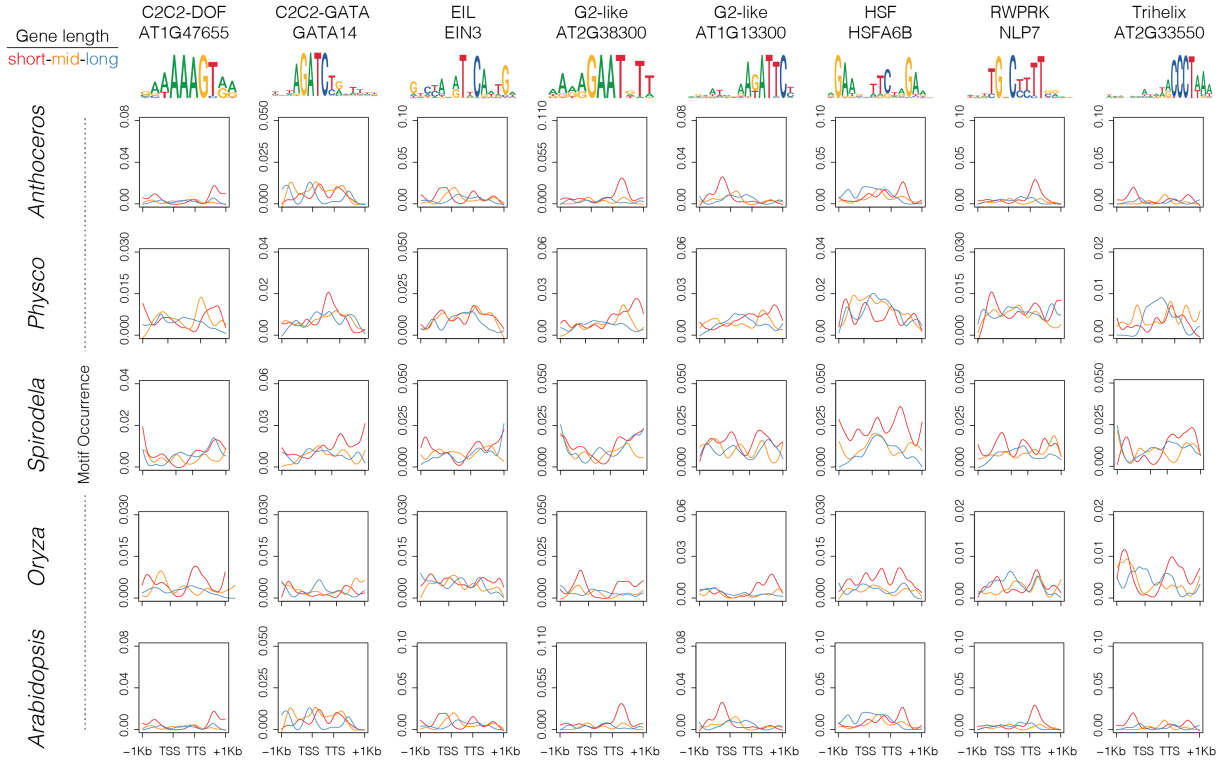
Supplemental Figure S4. (**A-C**) Ridgeline plots (left) showing the distribution of *tau* values for all phylogenetically-ranked genes in *Arabidopsis* (A) and those marked with H3K9me2 (B) and H3K27me3 (C). High *tau* > 0.9 (dashed line) indicates genes with highly restricted expression in the different tissues shown in the heatmaps to the right. The heatmaps summarise the percentage of genes in each phylogenetic rank that are expressed in a given tissue or cell type (TPM > 5). (**D**) Volcano plot of differentially expressed genes in *agdp1* or *suvh456* mutants compared to wild type (WT) (*Zhang* et al., *2018)*. Upregulated genes are shown in red while downregulated genes are shown in blue. DEGs were defined as having a $\log_2$ fold-change > 1 and FDR-adjusted *p*-value < 0.05. (**E**) Relative proportion of phylogenetically-ranks among all *Arabidopsis* genes and among the DEGs in panel D. The *p*-value indicates the statistical enrichment of taxonomically-restricted genes (TRGs; ranks 8-11) among each group of DEGs as determined using pairwise two-sided Fisher's exact tests. (**F**) Heatmap illustrating the developmentally-regulated expression of TRGs upregulated in *suvh456* mutants compared to WT. Expression represents the z-score of normalised RNA-seq TPM values. The left panel indicates the phylogenetic rank assigned to each TRG.



Supplemental Figure S5. Gene ontology (GO) enrichment analysis of H3K27me3-marked and H3K9me2-marked genes in each of the 11 phylogenetic ranks in *Arabidopsis*. All genes assigned to the corresponding phylogenetic rank were used as the reference gene set for the computation of *p*-values. Enriched GO terms belonging to the Biological Processes (BP) category are shown, with only those enriched more than 2-fold displayed for H3K27me3 targets in Phylogenetic rank 1 (See Supplemental Table S3 for more terms). Phylogenetic ranks 7, 10 and 11 with no significant enrichment of BP categories were not plotted.



Supplemental Figure S6. ChIP-Seq tracks showing H3K9me2 (black; H3K9me1 in *Anthoceros*) and H3K27me3 (red) levels over *BELL* and *KNOX* orthologs in *Anthoceros* (*Aa*), *Marchantia* (*Mp*) and *Physcomitrium* (*Pp*). Coloured and grey shading indicate enriched or depleted signals, respectively. Coverage is represented as the ChIP-seq $\log_2$ ratio relative to H3.



Supplemental Figure S7. Occurrence of non-enriched DNA binding motifs within H3K27me3 domains marking the ancestral PRC2-regulated genes in each land plant species. Genes were divided into three groups based on their average length.

Supplemental Tables

Supplemental Table S1. Publicly available genomic datasets used in this study.

Supplemental Table S2. Phylogenetic rank and chromatin state of the genes from each species.

Supplemental Table S3. Enriched GO terms among the H3K9me2- and H3K27me3-marked genes of differing evolutionary age in *Arabidopsis*.

Supplemental Table S4. Orthogroups identified in this study.

Supplemental Table S5. H3K27me3-marked orthogroups.

Supplemental Table S6. Ancestral land plant PRC2-regulated network.

Supplemental Table S7. Deeply-conserved PRC2-regulated network.

Supplemental Table S8. Transcription-associated proteins regulated by PRC2 across the green lineage.

Supplemental Table S9. DAP-seq motifs enriched within the H3K37me3 domains marking genes in the ancestral land plant PRC2-regulated network.