Uncoupled evolution of the Polycomb system and deep origin of non-canonical PRC1

Bastiaan de Potter[1,†], Maximilian W. D. Raas[1,2], Michael F. Seidl[1], C. Peter Verrijzer[3#], Berend Snel[1#*]

[1]Theoretical Biology and Bioinformatics, Department of Biology, Science Faculty, Utrecht University, Utrecht, Netherlands

[2,†]Hubrecht institute, Royal Netherlands Academy of Arts and Sciences, Utrecht, Netherlands

[3]Department of Biochemistry, Erasmus University Medical Center, Rotterdam, Netherlands

[†]Present address

#Equal contribution

*Corresponding author

Email: b.snel@uu.nl

Keywords: Polycomb, Evolution, PRC1, PRC2, Comparative genomics

Abstract

Polycomb group (PcG) proteins modulate chromatin states to silence gene transcription in plants and animals. Most PcG proteins function as part of distinct multi-subunit Polycomb repressive complexes (PRCs). Gene repression by the Polycomb system involves chromatin compaction by canonical PRC1 (cPRC1), mono-ubiquitylation of histone H2A (H2Aub1) by non-canonical PRC1 (ncPRC1) and tri-methylation of histone H3K27 (H3K27me3) by PRC2. Prevalent models for Polycomb repression emphasize a tight functional coupling between PRC1 and PRC2. However, whether this paradigm indeed reflects the evolution and functioning of the Polycomb system remains unclear. Here, we examined the relationship between cPRC1, ncPRC1 and PRC2 through a comprehensive analysis of their presence and evolution across the entire eukaryotic tree of life. We show that both PRC1 and PRC2 were present in the Last Eukaryotic Common Ancestor (LECA), but that their subsequent evolution is uncoupled. The identification of orthologs for ncPRC1-defining subunits in unicellular relatives of animals and of fungi suggests that the origin of ncPRC1 predates that of cPRC1, and we develop a scenario for the evolution of cPRC1 from ncPRC1. Our results demonstrate the independent evolution and function of PRC1 and PRC2 and show that crosstalk between these complexes is a secondary development in evolution.

Main Text

Introduction

Packaging of the eukaryotic genome into chromatin is crucial for its maintenance, replication and expression. The nucleosome, comprising 147 bp stretches of DNA wrapped in a left-handed superhelical turn around a histone octamer, is the fundamental repeating unit of chromatin. Histones and additional structural chromatin proteins impede access of DNA-binding proteins, such as transcription factors, thereby generating a fundamental level of gene regulation. Consequently, modulation of chromatin structure through post-translational histone modifications, nucleosome re-positioning and higher-order folding are central to eukaryotic gene regulation (Bracken et al. 2019). Polycomb group (PcG) proteins are conserved chromatin regulators that repress gene transcription to maintain cellular identity in plants and animals (Schuettengruber et al. 2017; Kassis et al. 2017; Baile et al. 2022). The *PcG* genes were first identified in *Drosophila* as repressors that are crucial to prevent mis-expression of homeotic (*Hox*) genes in inappropriate parts of the body (Schuettengruber et al. 2017; Kassis et al. 2017; Baile et al. 2022). This function is reflected by the name Polycomb, which refers to the extra sex combs appearing on the second and third pair of legs of male flies, due to de-repression of the *sex combs reduced* gene. Subsequent research showed that, in addition to silencing of *Hox* genes, PcG proteins are involved in the regulation of cell proliferation, stem cell pluripotency and oncogenesis (Bracken et al. 2019; Hennig and Derkacheva 2009; Kim and Kingston 2022; Piunti and Shilatifard 2021).

PcG proteins function as part of two main classes of multi-protein complexes, named Polycomb Repressive Complex (PRC) 1 and 2, which change chromatin structure across multiple scales (Baile et al. 2022; Kang et al. 2022; Kim and Kingston 2022; Piunti and Shilatifard 2021; Schuettengruber et al. 2017; Loubiere et al. 2019). The PRC2 methyltransferases mediate mono-, di- and tri-methylation of H3K27 (H3K27me3), whereas the PRC1s mediate chromatin compaction and mono-ubiquitylation of H2A (H2Aub1) on Lys 118, 119 or 120 in *Drosophila*, mammals or *Arabidopsis*, respectively (**Fig.1A**). The catalytic core of PRC2 comprises the methyltransferase EZH associated with EED, SUZ12 and RBBP (or their paralogs) (Yu et al. 2019; Laugesen et al.

2019). EZH proteins contain a catalytic SET domain that is responsible for methylating H3K27, while EED is a WD40 repeat protein that allosterically activates EZH upon binding to H3K27me3, thus creating a positive feedback loop (Yu et al. 2019; Laugesen et al. 2019; Glancy et al. 2021). The C-terminal VEFS-box of SUZ12 further stabilizes the interaction between EED and EZH. The WD40 repeat protein RBBP, enhances chromatin association through direct binding to histones. Selective association of the PRC2 core with accessory proteins generates unique subassemblies with different functional properties (Glancy et al. 2021; Baile et al. 2022; Laugesen et al. 2019; Yu et al. 2019; Bieluszewski et al. 2021).

The PRC1 class can be subdivided in canonical (cPRC1) and non-canonical PRC1 (ncPRC1) (Baile et al. 2022; Kang et al. 2022; Loubiere et al. 2019; Piunti and Shilatifard 2021). Both are assembled around a RING-RING heterodimer, which in animals comprise a RING1 protein paired with a PcG Finger (PCGF) protein. Notably, plant RING1 and PCGF orthologs can form homo- as well as heterodimers (Baile et al. 2022). RING1 and PCGF proteins share a similar structural organization characterized by an N-terminal zf-RING domain and a C-terminal RAWUL domain (**Fig.1B**). cPRC1 is defined by the association of a Chromobox protein (CBX) and Polyhomeotic homolog (PHC) subunit, which are absent in ncPRC1 (Gao et al. 2012). Additionally, SCM is a PcG protein that is more loosely associated with cPRC1. The presence of a RYBP/YAF subunit defines the ncPRC1s, which through association with additional partner proteins form multiple variant complexes. Pertinently, most or all of H2Aub1 is deposited by ncPRC1 rather than cPRC1 (Tamburri et al. 2022; Barbour et al. 2020; Lagarou et al. 2008).

The relative importance of the different PRC activities and their order of recruitment to chromatin remains controversial (Bracken et al. 2019; Kim and Kingston 2022; Loubiere et al. 2019; Piunti and Shilatifard 2021; Tamburri et al. 2022). While there is compelling evidence for the essential function of H3K27me3 for Polycomb repression (Sankar et al. 2022; Pengelly et al. 2013), the role of H2Aub1 is more ambiguous (Barbour et al. 2020). Although H2Aub1 is essential for animal viability, mutational studies established that PRC1 rather than its enzymatic activity is essential for transcriptional repression of canonical Polycomb target genes (Bonnet et al. 2022;

Illingworth et al. 2015; Pengelly et al. 2015; Tsuboi et al. 2018). Moreover, cPRC1 mediates long-range chromatin interactions via polymerization of the SAM domains within PH, which is independent of H2Aub1 (Isono et al. 2013; Boyle et al. 2020; Seif et al. 2020). Thus, it appears that chromatin compaction is essential for Polycomb repression in animals, while H2Aub1 stimulates this process but is not absolutely essential. Intriguingly, while the core subunits RING1 and PCGF of PRC1 are well-conserved in plants, the nc- and cPRC1 accessory subunits appear to be absent. This implies an alternate scenario for the potential formation of higher-order Polycomb structures in plants (Baile et al. 2022). The binding of CBX proteins in cPRC1 to H3K27me3, and conversely, the binding of PRC2-accessory proteins to H2Aub1, provides cross-talk between PRC1 and PRC2 (Glancy et al. 2021; Tamburri et al. 2022; Blackledge et al. 2020; Tamburri et al. 2020). These observations prompted the formulation of hierarchical recruitment models through “writing” and “reading” of histone modifications by PRC1 and PRC2 **(Fig.1A)** (Bracken et al. 2019; Kim and Kingston 2022; Loubiere et al. 2019; Piunti and Shilatifard 2021; Tamburri et al. 2022). At odds with such scenarios, there are accumulating examples of largely uncoupled gene repression by PRC1 and PRC2 (Bonnet et al. 2022; Cohen et al. 2021; Leeb et al. 2010; Pengelly et al. 2015; Sijm et al. 2022; Zepeda-Martinez et al. 2020; Chiacchiera et al. 2016; Rossi et al. 2016). Moreover, it is becoming increasingly clear that in mammals, as previously established in *Drosophila*, DNA-binding plays a major role in the genomic targeting of PRCs (Bracken et al. 2019; Glancy et al. 2021; Loubiere et al. 2019; Piunti and Shilatifard 2021; Kassis et al. 2017). In summary, PRC recruitment involves a combination of cooperative- and redundant mechanisms that may differ depending on the cellular- or genomic context. However, the functional- and evolutionary coupling between PRC1 and PRC2 remains an unresolved question.

Our knowledge of the function and evolution of the Polycomb system is primarily based on studies on animals and plants. This has limited our understanding of its role in developmental gene regulation in other eukaryotic lineages (Gahan et al. 2020; Huang et al. 2017, 2019; Berke and Snel 2015). Recent research has started to expand the diversity of eukaryotes in which Polycomb repression is analyzed. For example, studies in ciliates have shown that PRC2-mediated

trimethylation of histones is involved in the silencing of transposons (Xu et al. 2021; Wang et al. 2022; Déléris et al. 2021). In addition, a phylogenetic screening has shown that PRC2 is conserved across major eukaryotic groups and is likely present in the last eukaryotic common ancestor (LECA) (Sharaf et al. 2022). However, PRC1 has not been studied as extensively in non-animal lineages, and therefore, whether PRC1 and PRC2 co-evolved remains unresolved.

In the present study we used highly sensitive and comprehensive phylogenetic profiling to analyze the evolutionary history of PRC1 and PRC2 across the available scope of genomes covering the eukaryotic tree of life. Our analysis suggests that both PRC1 and PRC2 were present in LECA. Importantly, we found that while their intra-complex subunits evolved cohesively, PRC1 and PRC2 evolve independently. Furthermore, we discovered previously unreported ncPRC1-defining subunits in the relatives of animals and fungi, indicating that the differentiation of ncPRC1 occurred before cPRC1. These findings provide a foundation for future research on the biological functions of these proteins in a wider range of eukaryotes and in areas where PRC1 is understudied.

Results

Conservation of PRC1 subunits throughout eukaryotes traces its presence to LECA

To study the evolution of PRC1 beyond animals and plants, we traced the presence of PRC1 core subunits, RING1 and PCGF, in a diverse collection of 178 eukaryotes that covers all known major eukaryotic groups **(SI Data and Methods and External Dataset S1)**. Both RING1 and PCGF consist of an N-terminal zf-RING domain and a C-terminal RAWUL domain **(Fig.1B)**. These protein domains are the structural and functional units that are responsible for PRC1 functionality. As such, the conservation of specific protein domains across different species can be a strong indication that the protein is performing the same or similar functions in those species. To study the domain conservation of RING1 and PCGF proteins across our major eukaryotic groups,

we utilized Pfam (v.35) domain profile hidden-Markov models (HMMs) to annotate the sequences included in the phylogenetic trees (Mistry et al. 2021). These profiles, however sometimes failed to detect zf-RING or RAWUL domains in protein sequences that clustered in potential orthologous groups, while their presence was evident after manual inspection of sequence alignments combined with AlphaFold2 structural predictions (Varadi et al. 2022; Mirdita et al. 2022; Jumper et al. 2021). To overcome the lack of sensitivity in default Pfam profiles (v.35), we manually curated HMM profiles for the zf-RING, and RAWUL domain, by combining sequence alignments with structural predictions by AlphaFold2 (Varadi et al. 2022; Jumper et al. 2021). A detailed description of our approach and curated profiles are presented in **SI Data and Methods and External Dataset S2**.

This approach enabled us to identify 69 and 67 species with high-confidence orthologs for RING1 and PCGF, respectively. We considered an ortholog of high-confidence if it clustered monophyletically with the other sequences and contained both the zf-RING and RAWUL domain as predicted by our sequence models and AlphaFold2 **(Fig.2, SI Fig.S1A,B, Fig.S2)**. Additionally, we detected 15 species with a putative RING1- and ten species with putative PCGF orthologs, containing a zf-RING-, but lacking a recognizable RAWUL domain (**SI Fig.S1A,B, Fig.S2)**. The phylogenetic trees of RING1 and PCGF contained multiple groups of sequences from species that are part of all major eukaryotic groups in which the deep duplication nodes displayed large species overlap, which is a defining feature of orthologous groups in a gene phylogeny. The subsequent speciation node can thus be identified with high reliability as a LECA presence **(Fig.2, 3A, S1A,B)**. The increased sensitivity of our approach is illustrated by the detection of a previously unknown RING1 ortholog in the diatom (Stramenopile) *Phaeodactylum tricornutum*. In addition, we were able to detect a PCGF ortholog in *Arabidopsis thaliana* that was missed in a recent study (Grau-Bové et al. 2022), most likely because the canonical Pfam model (v.35) is not sensitive enough to detect the RAWUL domain. Importantly, we identified RING1 and PCGF orthologs in all unicellular eukaryotes that are currently placed proximal to the root of the eukaryotic tree of life, such as

Discoba, Metamonada, Malawimonadida, CRuMs, Hemimastigophora, and Provora **(Fig.2, S1A,B)**.

Although the eukaryotic group of fungi is generally believed to have lost PRC1 (Schuettengruber et al. 2017), we uncovered PRC1 orthologs in multiple deep branching fungal species. For instance, we observed PRC1 (either RING1 or PCGF) in the deep branching fungi *Rozella allomycis* and *Gonapodya prolifera* **(Fig.2,3B)**. Moreover, we confirmed the presence of PRC1 orthologs in the fungus *Spizellomyces punctatus* (Grau-Bové et al. 2022). Collectively, our phylogenetic analysis reveals that the PRC1 core was already present in LECA and remained highly conserved throughout eukaryotic life **(Fig.2, SI Fig.S2)**.

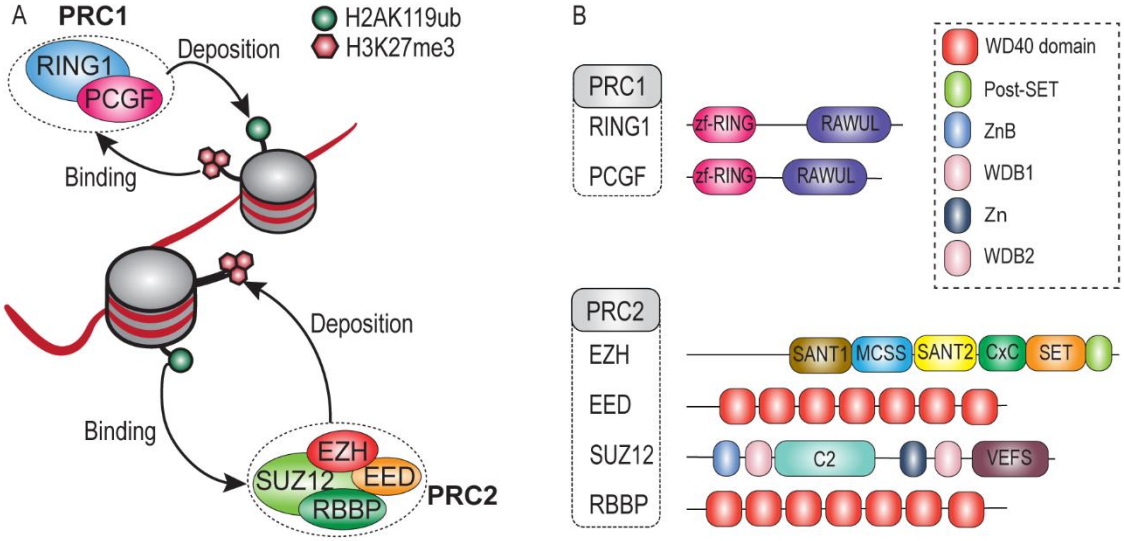

Figure 1. Schematic representation of the core subunits of PRC1 and PRC2 and their enzymatic activities. (A) Overview of histone modifications by PRC1 and PRC2, and cross-talk via recognition of each other's histone marks (B) Schematic representation of the core subunits of PRC1 and PRC2. Shown are the major conserved domains for each protein based on structural information in humans. In EZH, we did not depict the additional SBD, EBD, BAM, SAL, and SRM domain. RAWUL, RING finger- and WD40-associated ubiquitin-like; MCSS, motif connecting SANT1 and SANT2; SANT, SWI3, ADA2, N-CoR and TFIIIB DNA-binding domain; CxC, cysteine-rich domain; SET, Su(var) 3–9, enhancer of zeste, trithorax domain; WD, WD-40 domain; WDB, WD-40 binding domain; Zn, Zn-finger region; VEFS, conserved among VRN2-EMF2-FIS2-SU(Z)12.

Evidence for the presence of PRC2 in LECA

To reconstruct the evolution of PRC2 we traced the functional core of PRC2, consisting of EZH, EED, SUZ12 and RBBP (Glancy et al. 2021; Bieluszewski et al. 2021), across our diverse set of eukaryotes. Similar to PRC1, we studied the domain conservation of these subunits. Because Pfam domain profiles (v.35) again failed to detect key domains in a substantial number of orthologs, we curated our HMM profiles by combining sequence alignments with AlphaFold2 structural predictions **(Fig.1B, SI Data and Methods and External Dataset S2)**. Within our selection of 178 eukaryotes, we detected 94 species with orthologs for EZH, 91 with EED, 87 with SUZ12, and 178 with RBBP **(Fig.2, SI Fig.S1C-E)**. In accordance with our phylogenetic analysis of PRC1, the phylogenetic trees of PRC2 subunits also exhibited numerous orthologous groups with extensive species overlap. This observation thus similarly enables the reliable inference of the speciation node and the presence in LECA for the PRC2 subunits. EZH serves as the key enzymatic subunit and is a crucial indicator of the presence of a functional PRC2 (Kasinath et al. 2018). In our phylogenetic profile, EZH, EED, and SUZ12 exhibited strong co-occurrence. However, the co-occurrence of RBBP with other PRC2 was less prominent, owing to its involvement in various other protein assemblies (Suganuma et al. 2008). The phylogenetic trees of EZH and RBBP had the highest support values throughout all major eukaryotic groups **(SI Fig.S1C-E)**. In contrast, the phylogenomic delineation of EED orthologs was more ambiguous, due to their low sequence conservation which results in trees with lower resolution as well as difficulties in findings homologs and annotating the tree with protein domains **(SI Fig.S1D)**. For instance, PRC2 was recently experimentally identified in the ciliates *Paramecium tetraulia* and *Tetrahymena thermophila* (Xu et al. 2021; Wang et al. 2022), but the EED sequences in these organisms were undetectable by standard sequence-based homology methods. By performing a HMM search with a query based on the profile of the sequence alignment of these two sequences however, we retrieved EED orthologs in most ciliates, but not in the Alveolates *Stentor coeruleus* and *Colponemida vietnamica* **(Fig.3C)**.

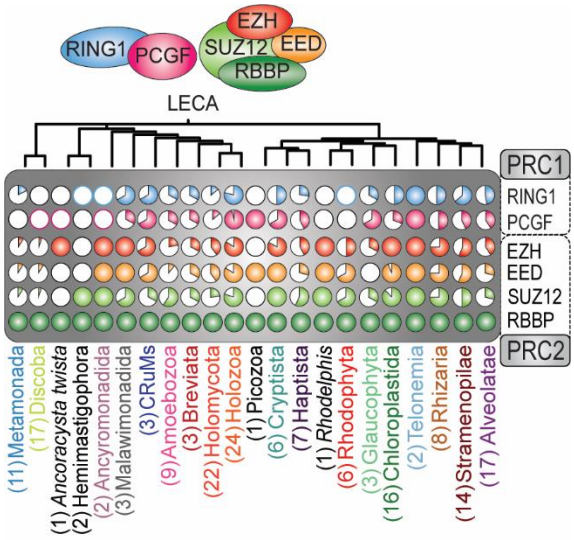

Figure 2. The evolutionary history of the core subunits of PRC1 and PRC2 can be traced back to LECA ***(Fig.S1A-E)****. Shown is a summary of the phylogenetic profiles of PRC1 and PRC2 in all major eukaryotic groups included in our study* ***(SI Fig.S2)****. Detected presences are presented as fractions of the species included in that group illustrated as a pie chart. For groups in which we only detected putative orthologs, only the outer line of the pie chart was colored.*

Like EED, SUZ12 was also found to be highly divergent in various species, e.g., a recent study demonstrated that MES-3 in *Caenorhabditis elegans* is a diverged SUZ12 ortholog rather than a *Caenorhabditis elegans* specific addition to PRC2 (Snel et al. 2022). The N-terminus of SUZ12 in animals contains five motifs: a zinc-finger binding (ZnB), WD-domain binding 1 (WDB1), C2 domain, zinc finger (Zn), and WD-domain binding 2 (WDB2) **(Fig.1B)** (Chen et al. 2018; Glancy et al. 2021; Kasinath et al. 2018). All these motifs are highly conserved throughout animals and plants. Pertinently, the C-terminal VEFS box of SUZ12 is crucial for the formation of the PRC2 catalytic lobe (Chen et al. 2018; Glancy et al. 2021; Kasinath et al. 2018). Therefore, we used the VEFS-box as the basis for our inference analysis to systematically detect SUZ12 orthologs. This approach uncovered a strong correlation between SUZ12 orthologs and the phylogenetic profiles of the other PRC2 subunits **(SI Fig.S2)**. Note that examples of SUZ12-like orthologs without a VEFS-box were recently described, e.g., *Paramecium tetraulia* has an adapted SUZ12-like ortholog

without a VEFS-box (Wang et al. 2022). Consequently, we could not detect these SUZ12-like orthologs in our phylogenetic inference **(Fig.3C)**. In addition, a previous study reported that in the yeast *Cryptococcus neoformans,* SUZ12 is substituted by a structurally unrelated protein, Bnd1 (Dumesic et al. 2015). Indeed, we initially missed a *Cryptococcus neoformans* SUZ12 ortholog in the yeast *Cryptococcus neoformans*. However, by combining HHpred searches with AlphaFold2 structural predictions, we found that, like MES-3 in *Caenorhabditis elegans,* Bnd1 is most likely a highly diverged SUZ12 ortholog **(SI Fig.S3)**. Further inferred absences of subunits predicted by our analysis might thus have been substituted by other, currently unknown, subunits, or were not detected due to high sequence divergence (e.g., EED, SUZ12). Consistent with the notion that EZH, EED and SUZ12 comprise the crucial functional core of PRC2, these subunits typically co-occur across the eukaryotic tree of life, reflecting the close and ancient biochemical interdependence of EZH, EED and SUZ12.

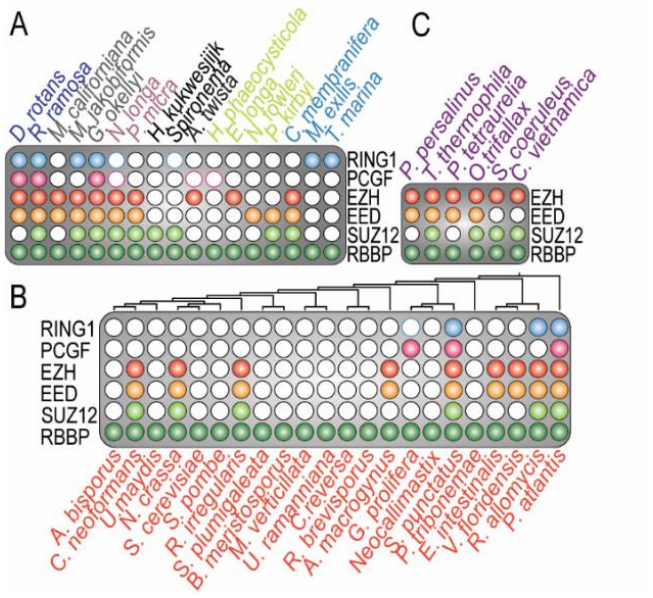

Figure 3. Phylogenetic profiles of the core subunits of PRC1 and PRC2. (A) Phylogenetic profiles of species in major eukaryotic groups proximal to the eukaryotic root, solidifying their presence in LECA. (B) Representation of the presence of PRC1 and PRC2 core subunits in fungi, a group that was previously believed to have lost PRC1 completely. (C) Phylogenetic profile of ciliates, highlighting the detection of highly diverged EED sequences.

Intra-complex evolution of PRC1 and PRC2 is cohesive while inter-complex evolution is uncoupled

The evolution of RING1 and PCGF is largely coupled, i.e. these core PRC1 subunits were mostly lost or retained together during eukaryotic evolution **(Fig.4A)**. Likewise, the genes encoding the core of PRC2 were typically lost or retained together. This suggests that both PRC1 and PRC2 function and evolve as cohesive units. To obtain a quantitative measure of their cohesiveness throughout eukaryotic evolution, we calculated the average Pearson's correlation coefficient between the phylogenetic profiles of the subunits of PRC1 and PRC2 **(Fig.4B)**. We found that the subunits of the PRC2 core (EZH, EED, and SUZ12) had a high average correlation (r = 0.73), as did the core subunits of PRC1 (RING1 and PCGF, r = 0.7) **(SI Fig.S4)**. This confirms that these subunits tend to evolve together within their resident complexes. As expected given its participation in multiple additional chromatin regulating pathways, RBBP only correlated weakly with other PRC2 subunits.

These observations raise the question whether the evolution of PRC1 and PRC2 might be coupled too. To address this question, we scored the presence of PRC1 or PRC2 only if all defining core subunits were present (i.e., both RING1 and PCGF for PRC1, and EZH, EED and SUZ12 for PRC2). Our analysis of 178 eukaryotes, selected to cover all known major eukaryotic groups, revealed that 40 species retained both PRC1 and PRC2, 12 lost PRC2 but retained PRC1, 17 retained only PRC2, whereas 52 lost both complexes. The Polycomb status of the remaining 57 eukaryotes remains unclear because they contained sequences without a recognizable RAWUL domain or harbored apparently partial PRCs. Thus, ~33% of species we analyzed contains remnants of the Polycomb system but lacks an unambiguous complete PRC1 or PRC2. About 22% of eukaryotes retained both PRC1 and PRC2, ~16% of eukaryotes retained either PRC1 or PRC2, without a clear preference for one over the other, while ~29% lost both. Thus, a substantial portion (~42%) of organisms that contain (part of) the Polycomb system use only PRC1 or only PRC2. We expect that with increasingly sensitive detection, the degree of uncoupling between PRC1 and

PRC2 might turn-out to be even higher than presented in this study. The presence of either only PRC1 or only PRC2 was particularly noticeable in Amoebozoans, fungi, Cryptista, Haptista, Rhodophyta, Glaucophyta, Chloroplastida, and SAR, but we were unable to determine a consistent trend of either complex being more likely to be lost or retained **(Fig.4A)**. In conclusion, while PRC1 and PRC2 were both present in LECA, their evolution throughout eukaryotic life has been largely uncoupled. Concurrent conservation of both complexes was most frequent in land plants and animals, while other major taxa displayed a more diverse pattern of conservation for PRC1 and PRC2 (**Fig.4A**). In contrast to their high intra-complex correlations, PRC1 and PRC2 subunits only displayed a low degree of inter-PRC correlation (r = 0.4). Collectively, these results reflect great evolutionary flexibility of the Polycomb system.

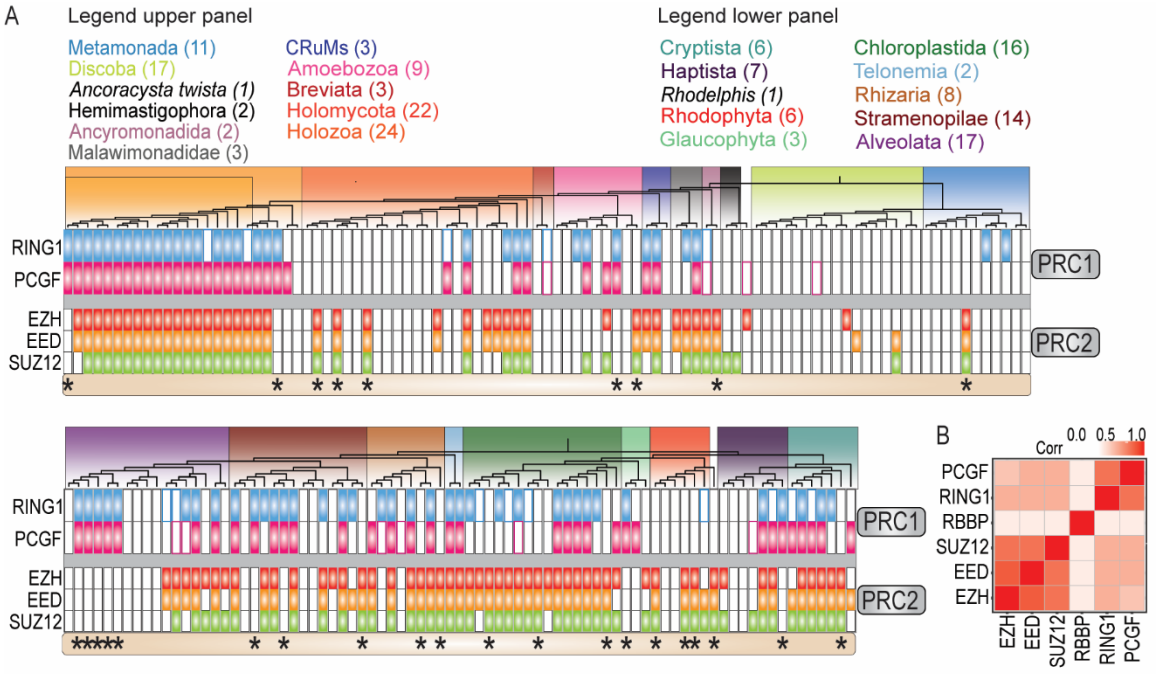

*Figure 4. Intra-complex evolution of PRC1 and PRC2 is cohesive, while inter-complex evolution is largely uncoupled. (A) Phylogenetic profiles of PRC1 and PRC2 core subunits in all independent species included in our analysis. The profiles show the overall cohesive loss- or retention of intra-complex subunits, and the largely uncoupled evolution of PRC1 and PRC2. RBBP was not included because it does not represent the presence or absence of the PRC2. Data was split into two panels and subunits are colored as follows: RING (blue), PCGF (Pink), EZH (Red), Orange (EED), SUZ12 (Green). Asterisks highlight species where either PRC1 or PRC2 is uncoupled. (B) Heatmap of the correlations between the subunits. Intra-complex subunits have a higher correlation than inter-complex subunits. Full names of all species are provided in **SI Fig.S2**.*

Structural similarity between RYBP and CBX suggests a common ancestor

Both PRC1 and PRC2 form subcomplexes through the incorporation of accessory subunits or through the combinatorial assembly of paralogous components (Baile et al. 2022; Kang et al. 2022; Piunti and Shilatifard 2021; Kim and Kingston 2022). ncPRC1 and cPRC1 are distinguished by the mutually exclusive binding of either CBX (cPRC1) or RYBP (ncPRC1) (Gao et al. 2012). CBX harbors a chromodomain at its N-terminus and a CBX7_C domain at its C-terminus, while RYBP contains a zf-RanBP domain at its N-terminus and a YAF2 domain at its C-terminus (**Fig.5A**). Interestingly, both the CBX7_C- and YAF2 domain form a ß-hairpin of around 16 amino acids. From an evolutionary perspective, mutual exclusivity (such as between RYBP and CBX) is often found between paralogs that define variant complexes (Szklarczyk et al. 2008; Ori et al. 2016). Therefore, we sought to investigate if CBX and RYBP might actually be distant paralogs by determining whether there is structural or highly diverged sequence homology between them. Whereas we did not detect full length structural- or sequence homology between CBX and RYBP, a structural alignment of their C-terminal ß-hairpins using the TM-align tool (Zhang 2005) revealed that they superimpose almost perfectly (RMSD = 2.35 Å) **(Fig.5A, SI External Dataset S3)**. Moreover, a sensitive sequence profile-vs-profile (HHpred) (Hildebrand et al. 2009) search with the C-terminal motifs of RYBP and CBX8 revealed that they were reciprocal best hits **(Fig.5B, SI External Dataset S3)**. Based on this evidence, we conclude that the domains CBX7_C and YAF are likely to be distant paralogs. Thus, to avoid any confusion resulting from the use of multiple names referring to the same domain, we will use ß-HP throughout the remainder of this article. Combined, the structural similarity and the reciprocal best hits of their ß-hairpin domains, provide evidence that CBX and RYBP might be switching paralogs in PRC1.

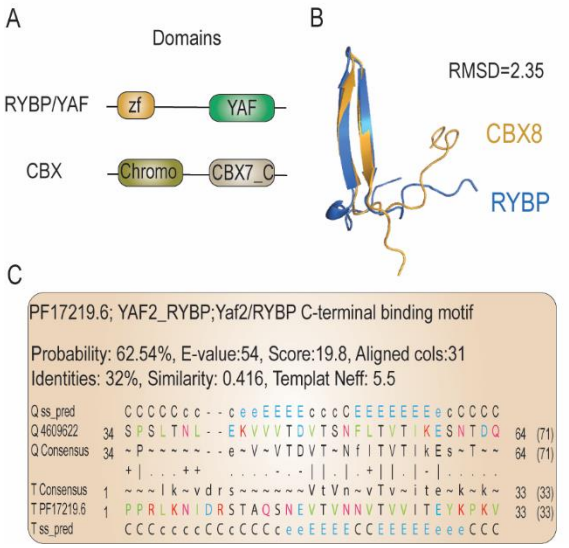

Figure 5. Structural homology and reciprocal best hits suggests that the ß-hairpin of RYBP and CBX are orthologous. (A) Schematic representation of RYBP/YAF and CBX. Both the C-terminal YAF domain and CBX7_C domain consist of a ß-hairpin. (B) Structural alignment of the ß-hairpin of RYBP, and CBX. The structures superimpose almost perfectly (RMSD = 2.35). (C) Output of the YAF2 C-terminal ß-hairpin as the best hit of a sensitive sequence profile-vs-profile (HHpred) search with the ß-hairpin of CBX.

Evolutionary evidence for the occurrence of ncPRC1 prior to cPRC1

To further study the emergence of ncPRC1 and cPRC1, we investigated the evolutionary origin of their signature subunits RYBP, CBX, and PHC. Recently, a putative ortholog of RYBP, containing only the N-terminal zf-RanBP motif, was identified in the Choanoflagellate *Salpingoeca rosetta*, which is a close unicellular relative of animals (Gahan et al. 2020). Additionally, we identified full length orthologs of RYBP in two Choanoflagellata, *Salpingoeca kvevrii* and *Diaphanoeca grandis.* These orthologs contain both the N-terminal zf-RanBP motif as well as the C-terminal ß-hairpin, suggesting that these are *bona fide* RYBP orthologs **(Fig.6A, SI External Dataset S4,5)**. We also detected a putative ortholog containing only the N-terminal zf-RanBP motif in the nucleariid *Parvularia atlantis*, which is a close relative of fungi **(Fig.6A)**. In contrast, we were not able to detect orthologs of CBX outside of animals (**SI External Dataset S4)**. Thus, it appears that RYBP originated before CBX. Next, we performed a phylogenetic analysis of SAM domains which are present in the PRC1-associated proteins PHC, SCML, SCMH, L3MBTL, SFMBT and

MBTD (**SI Data and Methods External Dataset S4).** The SAM domains clustered together in a monophyletic group, reflecting a shared ancestry **(SI Fig.S1F)**, but could be separated further into two distinctive groups; one that consisted mostly of proteins containing an MBT domain, and one without such a domain **(Fig.6B)**. We therefore next performed a phylogenetic analysis of MBT domains with the aim to find orthologs outside of animals that would imply an earlier origin for these proteins (**SI Data and Methods and External Dataset S4)**. Our analysis revealed that the MBT domains of the PRC1 accessory subunits were part of an orthologous group. Strikingly, we identified two sequences in the relatives of animals and fungi that contained both a potential SAM- and a MBT domain: one filasterean, *Capsaspora owczarzaki*, and the nucleariid *Parvularia atlantis* **(SI Fig.S1G)**. Although these sequences most likely encode orthologs of the L3MBTL/SCML PRC1 accessory subunits, the SAM- and MBT domains are in reverse order compared to their animal counterparts **(Fig.7)**. Adding the *Capsaspora owczarzaki*, and *Parvularia atlantis* SAM domains to our earlier inference analysis did not yield a monophyletic clustering with the SAM domains of SCML, SCMH, L3MBTL, and PHC **(SI Fig.S1H)**. This negative result might be due the limited phylogenetic signal of the relatively small SAM domain (65-70 amino acids). Thus, the orthology of the filasterean, *Capsaspora owczarzaki*, and the nucleariid *Parvularia atlantis* SAM domains remains unclear.

In summary, our collective results suggest that the ancestor of animals and fungi already harbored RYBP and a L3MBTL-associated ortholog **(Fig.7)**. We propose a scenario in which RYBP was lost in fungi and Filasterea, but retained in Nuclearidae, Choanoflagellata and animals. Furthermore, we hypothesize that CBX arose via a gene duplication of RYBP at the root of animals, after which the zf-RanBP was replaced by a chromo-domain **(Fig.7)**. Similarly, our data suggests that the SAM domain containing accessory subunits of PRC1 most likely originated from a singular gene rearrangement event of an ancestral L3MBTL-associated ortholog. The occurrence of RYBP and L3MBTL-associated proteins in these early branching lineages suggests that ncPRC1 defining subunits, as we know them in animals, originated before cPRC1 accessory subunits.

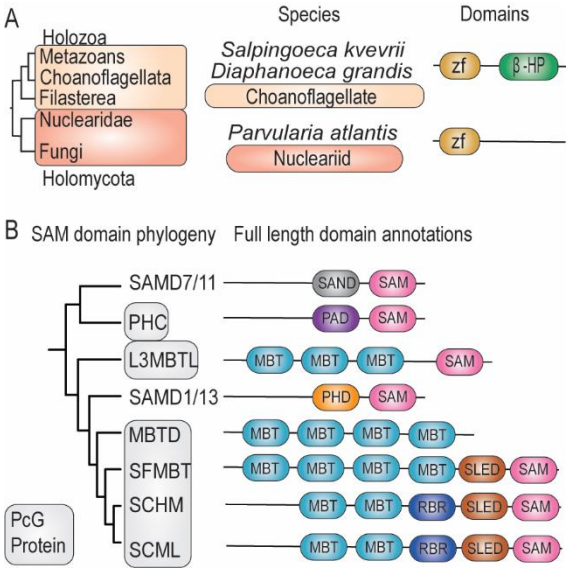

Figure 6. Phylogenetic history and domain organization of RYBP and SAM-domain proteins. (A) Domain organizations of the identified RYBP orthologs in the relatives of animals and fungi. (B) Schematic representation of the two clusters in the SAM domain phylogeny ***(SI Fig.S1F)*** *and the full-length domain annotations of the identified proteins.*

Discussion

Here, we present a systematic characterization of the Polycomb system throughout the eukaryotic tree of life. Phylogenetic and structural analysis of 178 eukaryotes representing all known major eukaryotic groups revealed broad conservation of both PRC1 and PRC2. Our results imply that PRC1 and PRC2 were both present in LECA, but that their subsequent evolution is uncoupled. Sensitive analyses allowed us to identify highly diverged orthologs, which previously escaped detection. Contrary to the general believe that fungi have lost PRC1, we uncovered PRC1 orthologs in multiple deep branching fungal species. Furthermore, we identified orthologs that had previously gone undetected because of their significant sequence divergence, such as EED in ciliates and SUZ12 orthologs in C*ryptococcus neoformans* and *Caenorhabditis elegans* (Snel et al. 2022). We expect that increasingly sensitive techniques will uncover additional species harboring divergent Polycomb orthologs or other, currently unknown, subunits. Thus, the evolutionary conservation and flexible development of the Polycomb system might be even more prevalent than

presented here. Finally, we identified orthologs of ncPRC1-signature domains, but not cPRC1-specific domains, in unicellular relatives of animals and fungi. Based on the distant homology between the C-terminal domains of RYBP and CBX, we propose that these subunits might be switching paralogs. Moreover, we conclude that the origin of ncPRC1 predates that of cPRC1, strengthening earlier hypotheses that ncPRC1 accessory subunits originated before cPRC1 (Schuettengruber et al. 2017).

Ever since their discovery, the relationship between PRC1 and PRC2 has been debated. The binding of CBX subunits of cPRC1 to H3K27me3, and the recognition of H2AK119ub1 by PRC2.2-associated factors JARID2 or AEBP2, has inspired multiple hierarchical or interdependent recruitment models (Baile et al. 2022; Kang et al. 2022; Kim and Kingston 2022; Piunti and Shilatifard 2021; Schuettengruber et al. 2017; Loubiere et al. 2019; Blackledge et al. 2020; Tamburri et al. 2020). However, in spite of cooperative cross-talk, *in vivo* studies showed that there is no crucial dependency between PRC1 and PRC2 that is essential for repression of canonical Polycomb targets (Bonnet et al. 2022; Cohen et al. 2021; Leeb et al. 2010; Pengelly et al. 2015; Sijm et al. 2022; Zepeda-Martinez et al. 2020). Rather, PRC1 and PRC2 function in largely redundant pathways, thereby ensuring robust gene silencing. Our analysis showed that a substantial portion (~42%) of organisms that contain (part of) the Polycomb system use only PRC1 or only PRC2. These results provide compelling evolutionary support for the predominantly independent function of PRC1 and PRC2.

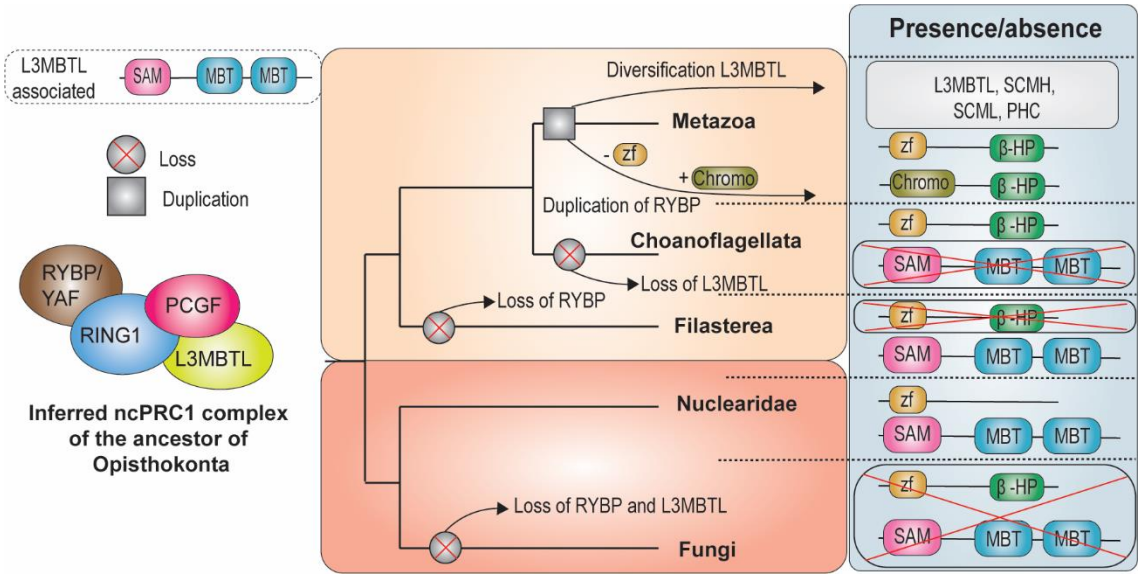

Figure 7. ncPRC1 evolved prior to cPRC1. The ancestor of animals and fungi most likely already had a distinct ncPRC1 functionality, while cPRC1 originated later in the root of animals. The upper left panel presents a schematic overview of the identified L3MBTL-associated ortholog. Shown is furthermore an evolutionary scenario of the origin and evolution of ncPRC1 and cPRC1 accessory subunits. Our scenario suggests that RYBP was lost in fungi and Filasterea, and retained in Nuclearidae, Choanoflagellata and animals, while CBX arose via a gene duplication of RYBP at the root of animals, after which the zf-RanBP got replaced by a chromo-domain. SAM domain containing accessory subunits of PRC1 likely originated from diversification of the ancestral L3MBTL-associated ortholog, after duplications.

While the central function of H3K27me3 for Polycomb repression in animals is unequivocal (Sankar et al. 2022; Pengelly et al. 2015), the role of H2Aub1 turned-out to be more enigmatic. During mouse and *Drosophila* development, H2Aub1 contributes, but is not absolutely required for the silencing of Polycomb targets (Bonnet et al. 2022; Boyle et al. 2020; Illingworth et al. 2015; Pengelly et al. 2015; Tsuboi et al. 2018). Rather, it appears that higher order chromatin organization by cPRC1, which is independent of H2Aub1, is crucial for Polycomb function in animals (Bonnet et al. 2022; Illingworth et al. 2015; Pengelly et al. 2015; Tsuboi et al. 2018). Supporting this notion, phylogenetic profiling of histone modifications revealed broad conservation of H3K27me3, while H2Aub1 is mainly restricted to animals and plants (Grau-Bové et al. 2022). Thus, while PRC1 is highly conserved and can be traced back to LECA, H2Aub1 is not. This implies that the primary function of PRC1 does not depend on H2Aub1. Collectively, these observations argue that

biochemical cross-talk between PRC1 and PRC2 is a secondary development during the evolution of animals.

In agreement with our phylogenetic analysis, functional studies in organisms other than animals revealed that the Polycomb system is highly plastic and employs diverse mechanisms of action. The Polycomb system in plants is the result of an independent evolutionary trajectory, yielding unique PRCs, which are absent in animals. E.g., plants lack cPRC1 but contain a wide variety of plant-specific PRC1 and PRC2 accessory subunits. One of the plant-specific PRC2-associated proteins, like-heterochromatin protein 1 (LHP1) contains a chromo domain and chromo-shadow domain that bind to H3K27me3 (Libault et al. 2005; Gaudin et al. 2001). Thus, LHP1 might mediate a positive feedback loop that stimulates the formation of H3K27me3 domains (Baile et al. 2022). The EZH ortholog in ciliates, Ezl1, is a dual-specificity methyltrasferase that mediates both H3K27me3 and H3K9me3 (Frapporti et al. 2019). Ezl1 binds a PRC2 core associated with the RNA interference effector Ptiwi09, and RING-finger proteins Rnf1 and Rnf2 (Wang et al. 2022; Xu et al. 2021; Miró-Pina et al. 2022). Thus, the Ezl1 complex appears to resemble a combination of PRC2 and PRC1. However, Rnf1 and Rnf2 are not orthologous to the canonical RING1 proteins found in PRC1 and are not required for H2Aub1, but they do contribute to histone methylation by the Ezl1 complex. Ptiwi09 mediates small-RNA-guided recruitment of the Ezl1-PRC2 complex to transposons, where it deposits H3K27me3 and H3K9me3, triggering transposon elimination (Wang et al. 2022; Xu et al. 2021). Examples of adaptation in the absence of PRC1 (Fig.3B) are provided by the filamentous fungus *Neurospora crassa* and the yeast *Cryptococcus neoformans*. In *Neurospora crassa*, the BAH- and PHD-domain protein EPR-1 acts as a H3K27me2/3 reader. EPR1 is a crucial protein involved in H3K27me3-directed silencing and mediates the formation of nuclear foci resembling Polycomb bodies (Wiles et al. 2020). The *Cryptococcus neoformans* PRC2-associated chromo-domain protein Ccc1 recognizes H3K27me3, which is crucial for its localization to repressive subtelomeric domains (Dumesic et al. 2015). Thus, in the absence of PRC1, other proteins recognize H3K27me3 to effectuate transcriptional silencing. Finally, it should be noted that some Polycomb factors have been repurposed for additional, chromatin-independent functions.

E.g., *Drosophila* PSC uses a diptera-specific domain to bind and regulate cyclin B, thereby controlling cell cycle progression (Fig.S5) (Mohd-Sarip et al. 2012). Mammalian SCML2B regulates the G1/S checkpoint by binding and modulating the CDK/Cyclin/p21 complex (Lecona et al. 2013). These examples from diverse eukaryotes emphasize the remarkable evolutionary flexibility of the Polycomb system.

What was the original function of the Polycomb system? There is accumulating evidence that the primary purpose of PRC2 in unicellular eukaryotes is the silencing of transposable elements to protect genome integrity (Shaver et al. 2010; Xu et al. 2021; Montgomery et al. 2020; Hisanaga et al. 2022). Similarly, in addition to gene control, plant PcG proteins ensure transcriptional repression of repetitive genomic elements and transposable elements (Fuchs et al. 2008; Montgomery et al. 2020; Shaver et al. 2010; Widiez et al. 2014; Deleris et al. 2012; Moreno-Romero et al. 2016; Schubert 2019). Collectively, these observations give rise to the hypothesis that gene-selective transcriptional repression by the Polycomb system evolved from an initial function in silencing of transposal elements (Déléris et al. 2021). Additional experimental studies will be necessary to establish the ancestral functions of PRC1 and PRC2, and delineate how these evolved towards their role in gene regulation in animals and plants. The findings presented here provide unique insights into the evolutionary development of the Polycomb system and its function in genome regulation. Additionally, it provides a framework for experimental analysis of the Polycomb system in diverse eukaryotic organisms. We anticipate that these future studies will shed light on the connection between genome evolution and functional adaptation of PRC1 and PRC2.

Materials and Methods

Proteome database

To study the occurrences of PRC1 and PRC2 genes across the eukaryotic tree of life, a reference dataset of predicted eukaryote proteomes comprising 178 diverse eukaryotic species

was assembled. A detailed description of the included species and how the data was collected is described in **SI Data and Methods and External Dataset S1**. Briefly, species were selected based on multiple criteria; i) we attempted to include representatives of all major eukaryotic groups encompassing most currently known eukaryotic diversity; ii) we weighed the species against genome quality statistics (BUSCO); if multiple proteomes or different representatives of a single species were available, the most complete one was selected; and iii), in clades with many closely related, high quality genome assemblies, we selected important (model) organisms.

Homolog detection and phylogenetic inferences of core subunits PRC1 and PRC2

To study the evolutionary history of the PRC1 and PRC2 subunits, a general homolog detection and phylogenetic inference method was initially applied and iterated upon if needed. Briefly, eukaryotic homologs were identified using pairwise searches with BLASTP v.2.12.0+ (Altschul et al. 1990) against our local proteome database with full length *Homo sapiens* sequences as queries (**SI External Dataset S6)**. At most 500 significant hits (default: 1e-5) were aligned using MAFFT v.7.490 (settings: genafpair, maxiterate = 1000) and processed with trimalAl v1.4.rev15 [gt = 0.1]) (Capella-Gutierrez et al. 2009; Katoh and Standley 2013) **(SI External Dataset S7)**. Phylogenetic trees were inferred with IQ-TREE v.2.1.4 (ModelFinder, ultrafast bootstrap = 1,000) (Nguyen et al. 2015), and subsequently visualized and annotated using iTOL **(SI Fig.S1A-H, External Dataset S7)** (Letunic and Bork 2021). To manually delineate orthologous groups from the phylogenetic trees, all sequences in the trees were annotated with independent function predictions (eggnog) and predicted domains based on HMM searches with the 'hmmsearch' tool from the HMMER package (http://hmmer.org/, HMMER 3.3.2), with profiles from Pfam (v.35) (Huerta-Cepas et al. 2019; Mistry et al. 2021). Delineation of the orthologous group was based on i) support values (>90); ii) consistency of domain annotations; and iii) consistency of independent function predictions. If all three sufficed, all sequences in the group were considered to constitute an orthologous group.

In some cases, interpretation of the phylogenetic trees was hampered by highly diverged sequences, incomplete genomes, low support values, or low sensitivity of Pfam profiles (v.35) that were unable to detect protein domains. Subsequent manual inspection of homologues by AlphaFold2 structures through ColabFold and multiple sequence alignments revealed the likely presence of known, yet highly diverged, proteins domains (Varadi et al. 2022; Mirdita et al. 2022; Jumper et al. 2021). In these cases, custom profiles for domains were created with the 'hmmbuild' command from the HMMER package (http://hmmer.org/, HMMER 3.3.2) **(SI External Dataset S2)**. These custom profiles were used to annotate sequences included in the respective phylogenetic trees, which enabled the detection of proteins domains that were previously reported to be absent, and therefore supported our orthologous inference. For some analyses, we also performed HMM searches with full length profiles. The specific analyses for these subunits, and supporting data, are described in **SI Data and Methods and External Dataset S4**. Fasta sequence files of our orthologous groups are provided in **SI External Dataset S8**. AlphaFold2 predicted structures used in this study are provided in **SI External Dataset S5**.

Structural analysis of CBX and RYBP

To detect structural similarity between the C-terminal regions of CBX and RYBP, the pdb_selres.py script from pdb-tools (https://wenmr.science.uu.nl/pdbtools/) was used to cut out the ß-sheet motifs. Accordingly, we used TMalign tool (https://zhanggroup.org/TM-align/) to perform a structural alignment (Zhang 2005) **(SI External Dataset S3)**. Additionally, a single blast was performed with the ß-sheet motifs using the MPI bioinformatics toolkit (https://toolkit.tuebingen.mpg.de/) and the MSA was used as a query to perform a HH search (https://toolkit.tuebingen.mpg.de/) against the Euk_Homo_sapiens_04_Jul_2017 proteome, with PDB_mmCIF70_12_Aug, Pfam-A_v35, and SMART_v6.0 as structural/domain databases (Hildebrand et al. 2009; Zimmermann et al. 2018; Gabler et al. 2020) **(SI External Dataset S3)**.

Correlation plot quantification and visualization

To study whether PRC1 and PRC2 co-evolved, a correlation plot of the phylogenetic profiles was created. For this correlation plot, putative orthologs were excluded, i.e. they did not contribute to the presences nor the absences. Phylogenetic profiles were created by setting a '1' for presence and '0' for absence. To then evaluate the co-occurrence of PRC1 and PRC2, the similarity of their phylogenetic profiles was quantified by using the Pearson correlation coefficient which for binary data is equivalent to the phi coefficient. Genes were clustered based on their phylogenetic profiles using average linkage and the resulting heatmap was visualized with ggplot in R version 4.1.2 (Wickham 2009).

Competing Interest Statement: The authors declare no conflict of interest.

Acknowledgments

We thank Eelco Tromer, Jolien van Hooff and Carlos Santana-Molina for advice in compiling the eukaryotic proteome database. We thank Carlos Santana-Molina for sharing scripts to remove splice variants from the proteome database. This work was supported by the Netherlands Organisation for Scientific Research (NWO-Vici 016.160.638, to B.S.).

Author Contributions: Author contributions: B.d.P., M.F.S., P.V., and B.S. designed research; B.d.P. performed research; B.d.P., analyzed data; M.W.D.R., developed the methods for, and created, the proteome database; and B.d.P, M.W.D.R., M.F.S., P.V., and B.S. wrote the paper. All authors approved the final version of the manuscript.

Data Availability

The data supporting the conclusions made in this article are available via its online Supplementary information, or deposited at the figshare depository as described in the Supplementary information. Scripts to remove splice variants from the proteome database are available upon reasonable request to the corresponding author.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Baile F, Gómez-Zambrano Á, Calonje M. 2022. Roles of Polycomb complexes in regulating gene expression and chromatin structure in plants. *Plant Commun* **3**: 100267. https://www.sciencedirect.com/science/article/pii/S2590346221001814.

Barbour H, Daou S, Hendzel M, Affar EB. 2020. Polycomb group-mediated histone H2A monoubiquitination in epigenome regulation and nuclear processes. *Nat Commun* **11**: 5947.

Berke L, Snel B. 2015. The plant Polycomb repressive complex 1 (PRC1) existed in the ancestor of seed plants and has a complex duplication history. *BMC Evol Biol* **15**: 44.

Bieluszewski T, Xiao J, Yang Y, Wagner D. 2021. PRC2 activity, recruitment, and silencing: a comparative perspective. *Trends Plant Sci* **26**: 1186–1198.

Blackledge NP, Fursova NA, Kelley JR, Huseyin MK, Feldmann A, Klose RJ. 2020. PRC1 Catalytic Activity Is Central to Polycomb System Function. *Mol Cell* **77**: 857-874.e9.

Bonnet J, Boichenko I, Kalb R, le Jeune M, Maltseva S, Pieropan M, Finkl K, Fierz B, Müller J. 2022. PR-DUB preserves Polycomb repression by preventing excessive accumulation of H2Aub1, an antagonist of chromatin compaction. *Genes Dev* **36**: 1046–1061.

Boyle S, Flyamer IM, Williamson I, Sengupta D, Bickmore WA, Illingworth RS. 2020. A central role for canonical PRC1 in shaping the 3D nuclear landscape. *Genes Dev* **34**: 931–949.

Bracken AP, Brien GL, Verrijzer CP. 2019. Dangerous liaisons: interplay between SWI/SNF, NuRD, and Polycomb in chromatin regulation and cancer. *Genes Dev* **33**: 936–959.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Chen S, Jiao L, Shubbar M, Yang X, Liu X. 2018. Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding. *Mol Cell* **69**: 840-852.e5.

Chiacchiera F, Rossi A, Jammula S, Piunti A, Scelfo A, Ordóñez-Morán P, Huelsken J, Koseki H, Pasini D. 2016. Polycomb Complex PRC1 Preserves Intestinal Stem Cell Identity by Sustaining Wnt/β-Catenin Transcriptional Activity. *Cell Stem Cell* **18**: 91–103.

Cohen I, Bar C, Liu H, Valdes VJ, Zhao D, Galbo PM, Silva JM, Koseki H, Zheng D, Ezhkova E. 2021. Polycomb complexes redundantly maintain epidermal stem cell identity during development. *Genes Dev* **35**: 354–366.

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

594 Deleris A, Stroud H, Bernatavichute Y, Johnson E, Klein G, Schubert D, Jacobsen SE. 2012. Loss of
595 the DNA Methyltransferase MET1 Induces H3K9 Hypermethylation at PcG Target Genes
596 and Redistribution of H3K27 Trimethylation to Transposons in Arabidopsis thaliana. *PLoS*
597 *Genet* **8**: e1003062.

598 Dumesic PA, Homer CM, Moresco JJ, Pack LR, Shanle EK, Coyle SM, Strahl BD, Fujimori DG, Yates
599 JR, Madhani HD. 2015. Product Binding Enforces the Genomic Specificity of a Yeast
600 Polycomb Repressive Complex. *Cell* **160**: 204–218.

601 Frapporti A, Miró Pina C, Arnaiz O, Holoch D, Kawaguchi T, Humbert A, Eleftheriou E, Lombard B,
602 Loew D, Sperling L, et al. 2019. The Polycomb protein Ezl1 mediates H3K9 and H3K27
603 methylation to repress transposable elements in Paramecium. *Nat Commun* **10**: 2710.

604 Fuchs J, Jovtchev G, Schubert I. 2008. The chromosomal distribution of histone methylation
605 marks in gymnosperms differs from that of angiosperms. *Chromosome Res* **16**: 891–8.

606 Gabler F, Nam S, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. 2020. Protein
607 Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics* **72**.

608 Gahan JM, Rentzsch F, Schnitzler CE. 2020. The genetic basis for PRC1 complex diversity
609 emerged early in animal evolution. *Proceedings of the National Academy of Sciences* **117**:
610 22880–22889.

611 Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, Kluger Y, Reinberg D. 2012. PCGF homologs,
612 CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* **45**:
613 344–56.

614 Gaudin V, Libault M, Pouteau S, Juul T, Zhao G, Lefebvre D, Grandjean O. 2001. Mutations in LIKE
615 HETEROCHROMATIN PROTEIN 1 affect flowering time and plant architecture in
616 Arabidopsis. *Development* **128**: 4847–58.

617 Glancy E, Ciferri C, Bracken AP. 2021. Structural basis for PRC2 engagement with chromatin. *Curr*
618 *Opin Struct Biol* **67**: 135–144.

619 Grau-Bové X, Navarrete C, Chiva C, Pribasnig T, Antó M, Torruella G, Galindo LJ, Lang BF, Moreira
620 D, López-Garcia P, et al. 2022. A phylogenetic and proteomic reconstruction of eukaryotic
621 chromatin evolution. *Nat Ecol Evol* **6**: 1007–1023.

622 Hennig L, Derkacheva M. 2009. Diversity of Polycomb group complexes in plants: same rules,
623 different players? *Trends Genet* **25**: 414–23.

624 Hildebrand A, Remmert M, Biegert A, Söding J. 2009. Fast and accurate automatic structure
625 prediction with HHpred. *Proteins: Structure, Function, and Bioinformatics* **77**: 128–132.

626 Hisanaga T, Romani F, Wu S, Kowar T, Lintermann R, Jamge B, Montgomery SA, Axelsson E,
627 Dierschke T, Bowman JL, et al. 2022. Transposons repressed by H3K27me3 were co-opted
628 as cis-regulatory elements of H3K27me3 controlled protein coding genes during evolution
629 of plants. *bioRxiv* 2022.10.24.513474.
630 http://biorxiv.org/content/early/2022/10/24/2022.10.24.513474.abstract.

631 Huang Y, Chen D-H, Liu B-Y, Shen W-H, Ruan Y. 2017. Conservation and diversification of
632 polycomb repressive complex 2 (PRC2) proteins in the green lineage. *Brief Funct Genomics*
633 **16**: 106–119.

634 Huang Y, Jiang L, Liu B-Y, Tan C-F, Chen D-H, Shen W-H, Ruan Y. 2019. Evolution and
635 conservation of polycomb repressive complex 1 core components and putative associated
636 factors in the green lineage. *BMC Genomics* **20**: 533.

637 Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR,
638 Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and
639 phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses.
640 *Nucleic Acids Res* **47**: D309–D314.

641 Illingworth RS, Moffat M, Mann AR, Read D, Hunter CJ, Pradeepa MM, Adams IR, Bickmore WA.
642 2015. The E3 ubiquitin ligase activity of RING1B is not essential for early mouse
643 development. *Genes Dev* **29**: 1897–1902.

644 Isono K, Endo TA, Ku M, Yamada D, Suzuki R, Sharif J, Ishikura T, Toyoda T, Bernstein BE, Koseki
645 H. 2013. SAM Domain Polymerization Links Subnuclear Clustering of PRC1 to Gene
646 Silencing. *Dev Cell* **26**: 565–577.

647 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R,
648 Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with
649 AlphaFold. *Nature* **596**: 583–589.

650 Kang H, Cabrera JR, Zee BM, Kang HA, Jobe JM, Hegarty MB, Barry AE, Glotov A, Schwartz YB,
651 Kuroda MI. 2022. Variant Polycomb complexes in Drosophila consistent with ancient
652 functional diversity. *Sci Adv* **8**: eadd0103.

653 Kasinath V, Faini M, Poepsel S, Reif D, Feng XA, Stjepanovic G, Aebersold R, Nogales E. 2018.
654 Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science (1979)* **359**: 940–
655 944.

656 Kassis JA, Kennison JA, Tamkun JW. 2017. Polycomb and Trithorax Group Genes in Drosophila.
657 *Genetics* **206**: 1699–1725.

658 Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
659 Improvements in Performance and Usability. *Mol Biol Evol* **30**: 772–780.

660 Kim JJ, Kingston RE. 2022. Context-specific Polycomb mechanisms in development. *Nat Rev*
661 *Genet* **23**: 680–695. https://doi.org/10.1038/s41576-022-00499-0.

662 Lagarou A, Mohd-Sarip A, Moshkin YM, Chalkley GE, Bezstarosti K, Demmers JAA, Verrijzer CP.
663 2008. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during
664 Polycomb group silencing. *Genes Dev* **22**: 2799–2810.

665 Laugesen A, Højfeldt JW, Helin K. 2019. Molecular Mechanisms Directing PRC2 Recruitment and
666 H3K27 Methylation. *Mol Cell* **74**: 8–18.

667 Lecona E, Rojas LA, Bonasio R, Johnston A, Fernández-Capetillo O, Reinberg D. 2013. Polycomb
668 Protein SCML2 Regulates the Cell Cycle by Binding and Modulating CDK/CYCLIN/p21
669 Complexes. *PLoS Biol* **11**: e1001737.

670 Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, Wutz A. 2010. Polycomb complexes act
671 redundantly to repress genomic repeats and genes. *Genes Dev* **24**: 265–276.

672 Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
673 display and annotation. *Nucleic Acids Res* **49**: W293–W296.

674 Libault M, Tessadori F, Germann S, Snijder B, Fransz P, Gaudin V. 2005. The Arabidopsis LHP1
675 protein is a component of euchromatin. *Planta* **222**: 910–25.

676 Loubiere V, Martinez A-M, Cavalli G. 2019. Cell Fate and Developmental Regulation Dynamics by
677 Polycomb Proteins and 3D Genome Architecture. *BioEssays* **41**: 1800222.

678 Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making
679 protein folding accessible to all. *Nat Methods* **19**: 679–682.

680 Miró-Pina C, Charmant O, Kawaguchi T, Holoch D, Michaud A, Cohen I, Humbert A, Jaszczyszyn
681 Y, Chevreux G, del Maestro L, et al. 2022. Paramecium Polycomb repressive complex 2
682 physically interacts with the small RNA-binding PIWI protein to repress transposable
683 elements. *Dev Cell* **57**: 1037-1052.e8.

684 Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE,
685 Paladin L, Raj S, Richardson LJ, et al. 2021. Pfam: The protein families database in 2021.
686 *Nucleic Acids Res* **49**: D412–D419.

687 Mohd-Sarip A, Lagarou A, Doyen CM, van der Knaap JA, Aslan Ü, Bezstarosti K, Yassin Y, Brock
688 HW, Demmers JAA, Verrijzer CP. 2012. Transcription-independent function of Polycomb
689 group protein PSC in cell cycle control. *Science* **336**: 744–7.

690 Montgomery SA, Tanizawa Y, Galik B, Wang N, Ito T, Mochizuki T, Akimcheva S, Bowman JL,
691 Cognat V, Maréchal-Drouard L, et al. 2020. Chromatin Organization in Early Land Plants
692 Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive
693 Heterochromatin. *Current Biology* **30**: 573-588.e7.

694 Moreno-Romero J, Jiang H, Santos-González J, Köhler C. 2016. Parental epigenetic asymmetry of
695 <scp>PRC</scp> 2-mediated histone modifications in the *Arabidopsis* endosperm. *EMBO J*
696 **35**: 1298–1311.

697 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective
698 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* **32**:
699 268–274.

700 Ori A, Iskar M, Buczak K, Kastritis P, Parca L, Andrés-Pons A, Singer S, Bork P, Beck M. 2016.
701 Spatiotemporal variation of mammalian protein complex stoichiometries. *Genome Biol* **17**:
702 47.

703 Pengelly AR, Copur Ö, Jäckle H, Herzig A, Müller J. 2013. A Histone Mutant Reproduces the
704 Phenotype Caused by Loss of Histone-Modifying Factor Polycomb. *Science (1979)* **339**:
705 698–699.

706 Pengelly AR, Kalb R, Finkl K, Müller J. 2015. Transcriptional repression by PRC1 in the absence of
707 H2A monoubiquitylation. *Genes Dev* **29**: 1487–1492.

708 Piunti A, Shilatifard A. 2021. The roles of Polycomb repressive complexes in mammalian
709 development and cancer. *Nat Rev Mol Cell Biol* **22**: 326–345.
710 https://doi.org/10.1038/s41580-021-00341-1.

711 Rossi A, Ferrari KJ, Piunti A, Jammula S, Chiacchiera F, Mazzarella L, Scelfo A, Pelicci PG, Pasini D.
712 2016. Maintenance of leukemic cell identity by the activity of the Polycomb complex PRC1
713 in mice. *Sci Adv* **2**: e1600972.

714 Sankar A, Mohammad F, Sundaramurthy AK, Wang H, Lerdrup M, Tatar T, Helin K. 2022. Histone
715 editing elucidates the functional roles of H3K27 methylation and acetylation in mammals.
716 *Nat Genet* **54**: 754–760.

717 Schubert D. 2019. Evolution of Polycomb-group function in the green lineage. *F1000Res* **8**: 268.

718 Schuettengruber B, Bourbon H-M, di Croce L, Cavalli G. 2017. Genome Regulation by Polycomb
719 and Trithorax: 70 Years and Counting. *Cell* **171**: 34–57.

720 Seif E, Kang JJ, Sasseville C, Senkovich O, Kaltashov A, Boulier EL, Kapur I, Kim CA, Francis NJ.
721 2020. Phase separation by the polyhomeotic sterile alpha motif compartmentalizes
722 Polycomb Group proteins and enhances their activity. *Nat Commun* **11**: 5609.

723 Sharaf A, Vijayanathan M, Oborník M, Mozgová I. 2022. Phylogenetic profiling resolves early
724 emergence of PRC2 and illuminates its functional core. *Life Sci Alliance* **5**: e202101271.

725 Shaver S, Casas-Mollano JA, Cerny RL, Cerutti H. 2010. Origin of the polycomb repressive
726 complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas.
727 *Epigenetics* **5**: 301–12.

728 Sijm A, Atlasi Y, van der Knaap JA, Wolf van der Meer J, Chalkley GE, Bezstarosti K, Dekkers
729 DHW, Doff WAS, Ozgur Z, van IJcken WFJ, et al. 2022. USP7 regulates the ncPRC1 Polycomb
730 axis to stimulate genomic H2AK119ub1 deposition uncoupled from H3K27me3. *Sci Adv* **8**.

731 Simon JA, Kingston RE. 2009. Mechanisms of Polycomb gene silencing: knowns and unknowns.
732 *Nat Rev Mol Cell Biol* **10**: 697–708.

733 Snel B, van den Heuvel S, Seidl MF. 2022. Caenorhabditis elegans MES-3 is a highly divergent
734 ortholog of the canonical PRC2 component SUZ12. *iScience* **25**: 104633.

735 Suganuma T, Pattenden SG, Workman JL. 2008. Diverse functions of WD40 repeat proteins in
736 histone recognition: Figure 1. *Genes Dev* **22**: 1265–1268.

737 Szklarczyk R, Huynen MA, Snel B. 2008. Complex fate of paralogs. *BMC Evol Biol* **8**: 337.

738 Tamburri S, Conway E, Pasini D. 2022. Polycomb-dependent histone H2A ubiquitination links
739 developmental disorders with cancer. *Trends in Genetics* **38**: 333–352.

740 Tamburri S, Lavarone E, Fernández-Pérez D, Conway E, Zanotti M, Manganaro D, Pasini D. 2020.
741 Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional
742 Repression. *Mol Cell* **77**: 840-856.e5.

743 Tsuboi M, Kishi Y, Yokozeki W, Koseki H, Hirabayashi Y, Gotoh Y. 2018. Ubiquitination-
744 Independent Repression of PRC1 Targets during Neuronal Fate Restriction in the
745 Developing Mouse Neocortex. *Dev Cell* **47**: 758-772.e5.

746 Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G,
747 Laydon A, et al. 2022. AlphaFold Protein Structure Database: massively expanding the
748 structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids*
749 *Res* **50**: D439–D444.

750 Wang C, Solberg T, Maurer-Alcalá XX, Swart EC, Gao F, Nowacki M. 2022. A small RNA-guided
751 PRC2 complex eliminates DNA as an extreme form of transposon silencing. *Cell Rep* **40**:
752 111263.

753 Wickham H. 2009. *ggplot2*. Springer New York, New York, NY.

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

757 Wiles ET, McNaught KJ, Kaur G, Selker JML, Ormsby T, Aravind L, Selker EU. 2020. Evolutionarily
758 ancient BAH–PHD protein mediates Polycomb silencing. *Proceedings of the National*
759 *Academy of Sciences* **117**: 11614–11623.

760 Xu J, Zhao X, Mao F, Basrur V, Ueberheide B, Chait BT, Allis CD, Taverna SD, Gao S, Wang W, et
761 al. 2021. A Polycomb repressive complex is required for RNAi-mediated heterochromatin
762 formation and dynamic distribution of nuclear bodies. *Nucleic Acids Res* **49**: 5407–5425.

763 Yu J-R, Lee C-H, Oksuz O, Stafford JM, Reinberg D. 2019. PRC2 is high maintenance. *Genes Dev*
764 **33**: 903–935.

765 Zepeda-Martinez JA, Pribitzer C, Wang J, Bsteh D, Golumbeanu S, Zhao Q, Burkard TR, Reichholf
766 B, Rhie SK, Jude J, et al. 2020. Parallel PRC2/cPRC1 and vPRC1 pathways silence lineage-
767 specific genes and maintain self-renewal in mouse embryonic stem cells. *Sci Adv* **6**.

768 Zhang Y. 2005. TM-align: a protein structure alignment algorithm based on the TM-score.
769 *Nucleic Acids Res* **33**: 2302–2309.

770 Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN,
771 Alva V. 2018. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred
772 Server at its Core. *J Mol Biol* **430**: 2237–2243.

773