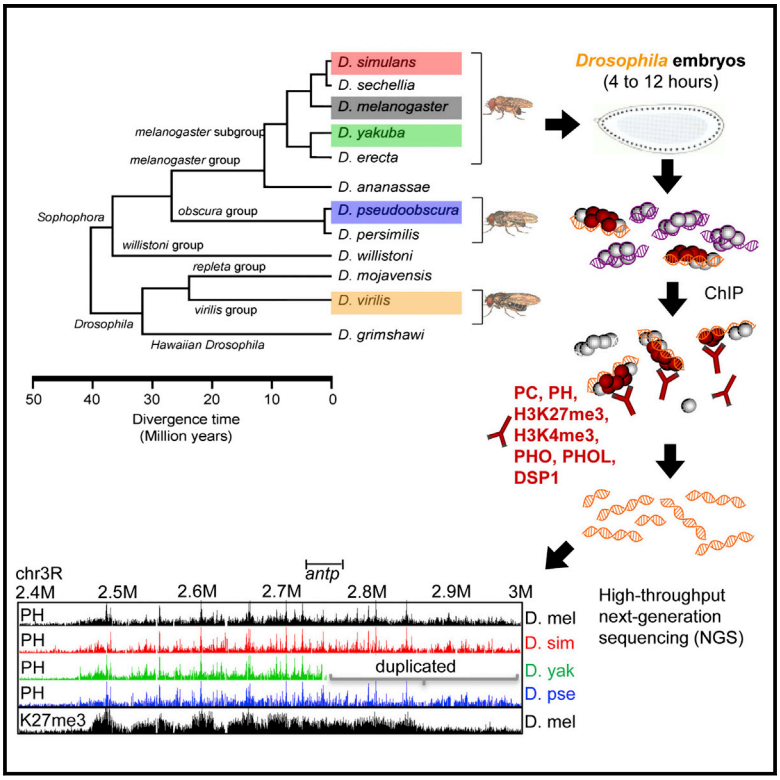
Article

Cell Reports

Cooperativity, Specificity, and Evolutionary Stability of Polycomb Targeting in *Drosophila*

Graphical Abstract



Highlights

Extremely high conservation of Polycomb repressive domains in five *Drosophila* species

cis-driven Polycomb response element (PRE) evolution is rare but possible

PHO recruitment to Polycomb domains is evolutionarily robust to motif changes

PRC1 stabilizes binding of PHO to PREs with low-quality motifs

Authors

Bernd Schuettengruber, Noa Oded Elkayam, ..., Amos Tanay, Giacomo Cavalli

Correspondence

amos.tanay@weizmann.ac.il (A.T.),
giacomo.cavalli@igh.cnrs.fr (G.C.)

In Brief

Schuettengruber et al. present an extensive comparative epigenomics data set, providing new insights into *cis*-driven versus buffered evolution of Polycomb recruitment and Polycomb domain specificity. Using chromatin immunoprecipitation sequencing and transgenic assays, they demonstrate an extremely high conservation of Polycomb repressive domains in five *Drosophila* species. Using Hi-C and knockout experiments, they challenge the standard hierarchical Polycomb recruitment model and demonstrate that cooperative rather than hierarchical interactions among DNA motifs, transcription factors, and Polycomb group complexes define Polycomb domains.

Accession Numbers

GSE60428
GSE61471

Schuettengruber et al., 2014, Cell Reports 9, 219–233
October 9, 2014 ©2014 The Authors
http://dx.doi.org/10.1016/j.celrep.2014.08.072

Cell Reports
Article

Cooperativity, Specificity, and Evolutionary Stability of Polycomb Targeting in *Drosophila*

Bernd Schuettengruber,[1,4] Noa Oded Elkayam,[2,4] Tom Sexton,[1,5] Marianne Entrevan,[1] Shani Stern,[2] Aubin Thomas,[1] Eitan Yaffe,[2] Hugues Parrinello,[3] Amos Tanay,[2,*] and Giacomo Cavalli[1,*]
[1]Institute of Human Genetics, UPR 1142, CNRS, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France
[2]Department of Computer Science and Applied Mathematics and Department of Biological Regulation, The Weizmann Institute of Science, Rehovot 7610001, Israel
[3]Montpellier GenomiX IBiSA, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France
[4]Co-first author
[5]Present address: Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, 67400 Illkirch, France
*Correspondence: amos.tanay@weizmann.ac.il (A.T.), giacomo.cavalli@igh.cnrs.fr (G.C.)
http://dx.doi.org/10.1016/j.celrep.2014.08.072

SUMMARY

Metazoan genomes are partitioned into modular chromosomal domains containing active or repressive chromatin. In flies, Polycomb group (PcG) response elements (PREs) recruit PHO and other DNA-binding factors and act as nucleation sites for the formation of Polycomb repressive domains. The sequence specificity of PREs is not well understood. Here, we use comparative epigenomics and transgenic assays to show that *Drosophila* domain organization and PRE specification are evolutionarily conserved despite significant *cis*-element divergence within Polycomb domains, whereas *cis*-element evolution is strongly correlated with transcription factor binding divergence outside of Polycomb domains. Cooperative interactions of PcG complexes and their recruiting factor PHO stabilize PHO recruitment to low-specificity sequences. Consistently, PHO recruitment to sites within Polycomb domains is stabilized by PRC1. These data suggest that cooperative rather than hierarchical interactions among low-affinity sequences, DNA-binding factors, and the Polycomb machinery are giving rise to specific and strongly conserved 3D structures in *Drosophila*.

INTRODUCTION

The regulation of complex genomes in multicellular organisms requires both flexibility and stability. Genomes must be flexible enough to accommodate multiple cell-type-specific transcriptional programs. Simultaneously, genome regulation must be sufficiently stable to avoid aberrant gene activation in committed or differentiated cells. Genomes have adapted to this challenge by evolving a sparse dispersion of genes within vast genomic territories, which are dotted with hundreds to thousands of small regulatory elements. Importantly, the local sequence specificity of metazoan regulatory elements within such vast territories is not higher than that observed in much more compact genomes, leading to the spontaneous evolution of millions of possible spurious binding sites for a typical sequence-specific transcription factor (TF). Nevertheless, extensive mapping of binding sites for hundreds of DNA-binding factors has shown that only many thousands of enhancer elements, rather than millions of potential spurious binding sequences, are specifically identified and engaged by combinations of TFs and chromatin regulators. It was suggested that such specificity is facilitated by cooperative binding of TFs (Junion et al., 2012) and by epigenetic mechanisms that selectively provide access to a small subset of the genome. The specificity of epigenetic regulation itself, however, remains poorly understood. For example, simple hierarchical models postulating that sequence-specific "pioneer" factors dictate specificity in genome regulation during cell-fate commitment cannot explain much of the experimental data (Rothenberg, 2014). Recently, the discovery of topologically associating domains (TADs) in flies and mammals showed that chromosomes can compartmentalize genomes into relatively isolated building blocks (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012), but how such compartmentalization affects regulatory specificity remains unclear.

Repressive Polycomb topological domains, characterized by the presence of the H3K27me3 mark, constitute a major subdivision of the eukaryotic genome and provide a paradigm for understanding epigenome regulation and chromosomal domain structure. Initial genetic studies in flies characterized the Polycomb group (PcG) system as being responsible for maintaining *HOX* gene repression following its initial setup during early embryonic development (Duncan, 1982; Lewis, 1985). Subsequent work demonstrated a more dynamic role of PcG proteins in defining cellular identities through the epigenetic repression of key developmental regulators (reviewed in Schuettengruber and Cavalli, 2009). Genomic analysis generated a multilayered view of PcG-mediated gene regulation involving a combination of Polycomb repressive complex 1 (PRC1) and 2 (PRC2) their associated histone marks, HA2 ubiquitylation of lysine 119 (H2AK119ub) or trimethylation of lysine 27 on histone H3 (H3K27me3), and chromosomal looping (Bantignies and Cavalli, 2011). In *Drosophila*, specific sequences termed Polycomb group response elements (PREs) are known to recruit PcG complexes to their target genes

A

Species	All Genome				H3K27me3 Sites			
	% of D. xxx not aligned to D. mel	% of D. mel not aligned to D. xxx	Not Mappable (within aligned)	Non-Exon Point Mutation (D. xxx - D.mel)	% of D. xxx not aligned to D. mel	% of D. mel not aligned to D. xxx	Not Mappable (within aligned)	Non-Exon Point Mutation (D. xxx - D.mel)
D. mel	NA	NA	10.70%	NA	NA	NA	NA	NA
D. sim	13.06%	11.02%	3.33%	9%	3.80%	0.01%	1.61%	5%
D. yak	10.27%	16.82%	7.45%	17%	0.92%	4.05%	8.11%	11%
D. pse	30.98%	29.74%	1.81%	45%	8.80%	16.35%	2.07%	36%
D. vir	32.66%	53.90%	4.22%	66%	18.62%	25.72%	0.30%	56%

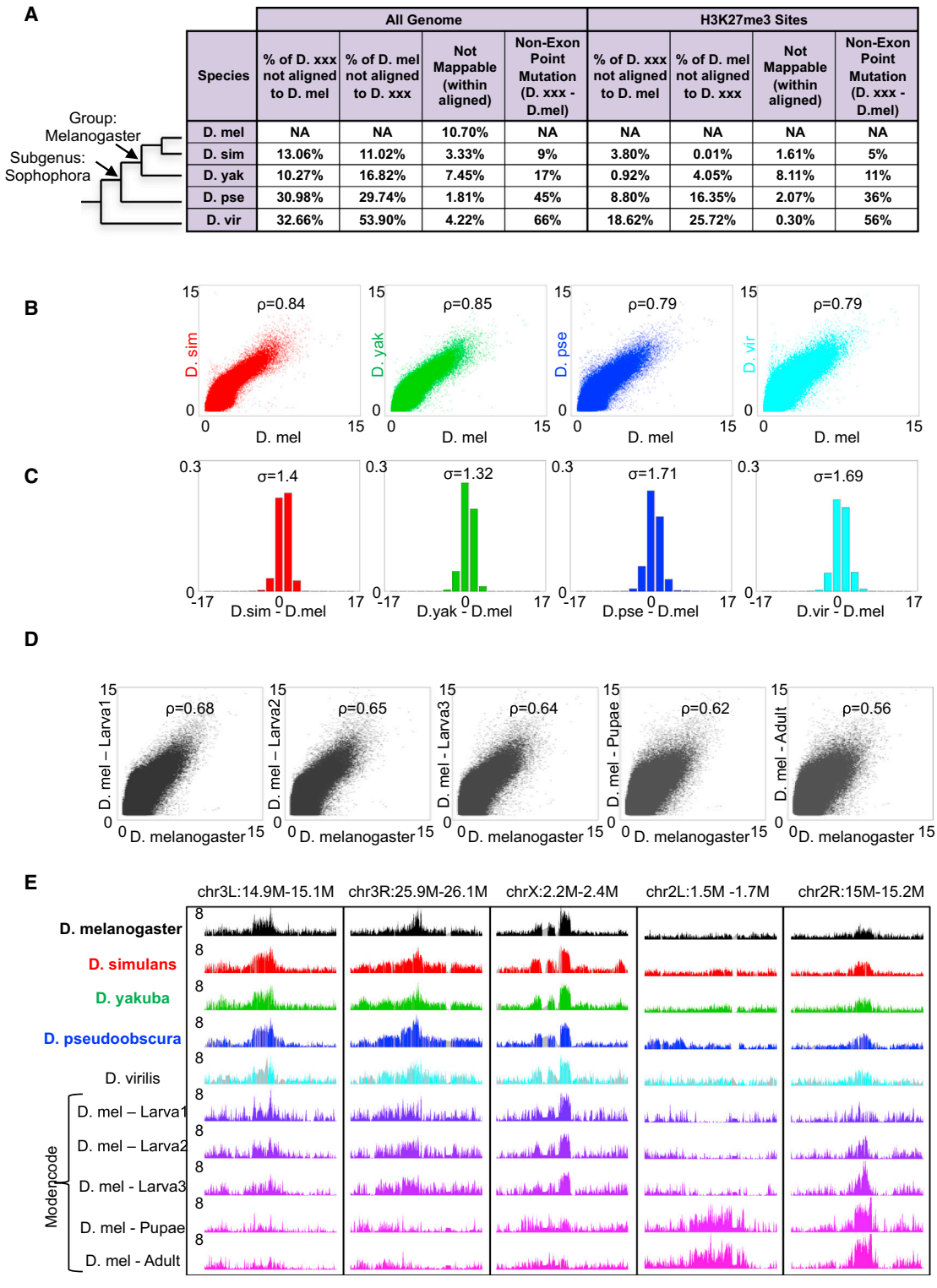


Figure 1. Comparative Epigenomics of Polycomb Domains

(A) Phylogenetic statistics of the *Drosophila* species used in this study. Data on divergence are provided with respect to the entire genome or regions that are annotated as H3K27me3-marked domains in D.mel.

(B) Scatterplots depicting the correlation between H3K27me3 ChIP-seq enrichment in syntenic loci for the indicated *Drosophila* species.

(legend continued on next page)

via their interaction with sequence-specific DNA-binding proteins defined as PcG recruiters (Müller and Kassis, 2006). Wang et al. (2004) proposed a recruitment model that suggests a hierarchy of binding events to PREs: first, sequence-specific binding of DNA-binding factors lead to the subsequent targeting of PRC2, which then helps to recruit the later-acting PRC1 complex via its interaction with the PRC2-specific histone mark H3K27me3. However, in mammals it was recently demonstrated that PRC1 variant complexes can be recruited to DNA independently of PRC2, but in turn the PRC1-associated mark H2AK119ub helps to recruit PRC2 (Blackledge et al., 2014; Cooper et al., 2014). The sequence requirements for targeting of mammalian PRC complexes are still unclear, but CpG islands seem to play a major role.

Isolation and perturbation analyses of *Drosophila* PREs have uncovered several sequence motifs that are required for the assembly of the PcG machinery on reporter constructs (Brown et al., 1998; Déjardin et al., 2005). Among these, the binding sites of Pleiohomeotic (PHO) and Dorsal Switch Protein 1 (DSP1) were particularly enriched in genomic catalogs of putative PREs (Schuettengruber et al., 2009). However, none of the motifs or their associated DNA-binding proteins are sufficient to recruit PcG proteins to their targets (Müller and Kassis, 2006). Whereas mutation of PHO induces homeotic phenotypes similar to those observed in loss-of-function PcG mutants, mutations of DSP1 exhibit a variety of phenotypes—some similar to PcG mutants and others more typical of mutants for a Trithorax group gene, a factor known to counteract PcG function. Hence, PcG-recruiting factors are associated with transcriptional activation as well as repression (Fujioka et al., 2008; Kwong et al., 2008; Schuettengruber et al., 2009). PHO, in particular, binds numerous putative promoter and enhancer elements outside a PRE context (Schuettengruber et al., 2009), and it was shown to interact with the INO80 chromatin-remodeling complex in addition to PcG complexes (Klymenko et al., 2006). Of note, the mammalian homolog of PHO, YY1, is able to rescue a *pho* zygotic mutant, but does not seem to play a major role in PcG recruitment in mammals (Atchison et al., 2003; Mendenhall et al., 2010). DSP1 was also shown to bind to many non-PRE loci in *Drosophila* embryos, which are marked strongly by GAGA motifs (Schuettengruber et al., 2009). Thus, despite the unambiguous genetic and genomic evidence for PREs, and for the involvement of PHO and DSP1 in their function, the specificity of the process remains difficult to understand based on studies of individual factors or loci. Thus far, any attempts to predict PREs from genomic sequence alone have only been partly successful (Fiedler and Rehmsmeier, 2006; Kassis and Brown, 2013; Ringrose et al., 2003; Schuettengruber et al., 2009; Zeng et al., 2012).

Here, we study the function and evolution of PRE sequences within the broader context of multigenic Polycomb topological domains (hereafter referred to as Polycomb domains). Comparative epigenomics shows that during the evolution of *Drosophila* species, and despite extensive sequence divergence, the structure of the Polycomb domain in syntenic genomic regions remained perfectly conserved. This remarkable stability was facilitated by the high conservation of PRC1 binding at putative PREs. In the relatively few cases in which PRC1 binding diverged, transgenic PRE assays indicate that *cis*-element divergence was the likely cause of functional divergence. In other cases, however, divergence in *cis* is not linked to functional PRE divergence in Polycomb domains. We explain this effect by showing that targeting of the PcG recruiter PHO to PREs within Polycomb domains can be driven by PHO-DNA interactions that occur over a wide spectrum of noncanonical, low-affinity binding sites. Such sites are hypothesized to buffer pronounced evolutionary divergence without significant loss of PRE function. Instead of high sequence affinity, our data indicate that PHO recruitment relies on cooperative effects with other TFs (e.g., DSP1) and on a positive feedback loop induced by PRC1 binding. Therefore, the sequence specificity of Polycomb domains in flies is established through cooperative rather than hierarchical interactions between sequence-specific factors and the Polycomb machinery, in a way that potentially also involves the formation of 3D chromatin hubs associating with several PREs. A similar interplay among sequence-specific genomic signals, epigenetic factors, and large-scale chromosomal structure may have general implications for regulatory genomics and epigenomics in metazoans.

RESULTS

Evolutionary Conservation of H3K27me3 Repressive Domains in *Drosophila* Embryos

To characterize the evolution of Polycomb domains in *Drosophila*, we performed chromatin immunoprecipitation sequencing (ChIP-seq) for the H3K27me3 mark on 4- to 12-hr-old embryos from *D. melanogaster* (D.mel), *D. simulans* (D.sim), *D. yakuba* (D.yak), *D. pseudoobscura* (D.pse), and *D. virilis* (D.vir) (see Supplemental Experimental Procedures). These species represent variable levels of evolutionary distance from D.mel, with overall local identity on alignable sequences varying between 92% (D.sim) and 39% (D.vir) of the D.mel genome. We note that sequences marked by H3K27me3 are evolving more slowly than other genomic regions, as indicated by both the point mutation level and the overall retained aligned sequences (Figure 1A). Nonetheless, the rate of sequence divergence is substantial and in principle could support changes in the definition of genomic Polycomb domains. Surprisingly, however, a comparative analysis of the species ChIP-seq that was projected onto the D.mel genome indicated a remarkable pointwise conservation of H3K27me3 occupancies (Figure 1B). The overall correlation between the orthologous profiles was higher than 0.79 even for the remote species D.vir. Moreover, the root-mean-square of pairwise differential ChIP-seq ranged between 1.32 and 1.7 (Figure 1C), maintaining levels comparable to those

(C) Summary of H3K27me3 divergence in each species compared with D.mel.
(D) Comparison of H3K27me3 enrichment levels in D.mel embryos and other developmental stages.
(E) Examples of evolutionarily conserved H3K27me3 domains that are developmentally plastic. The y axis represents normalized ChIP-seq values.
See also Figure S1.

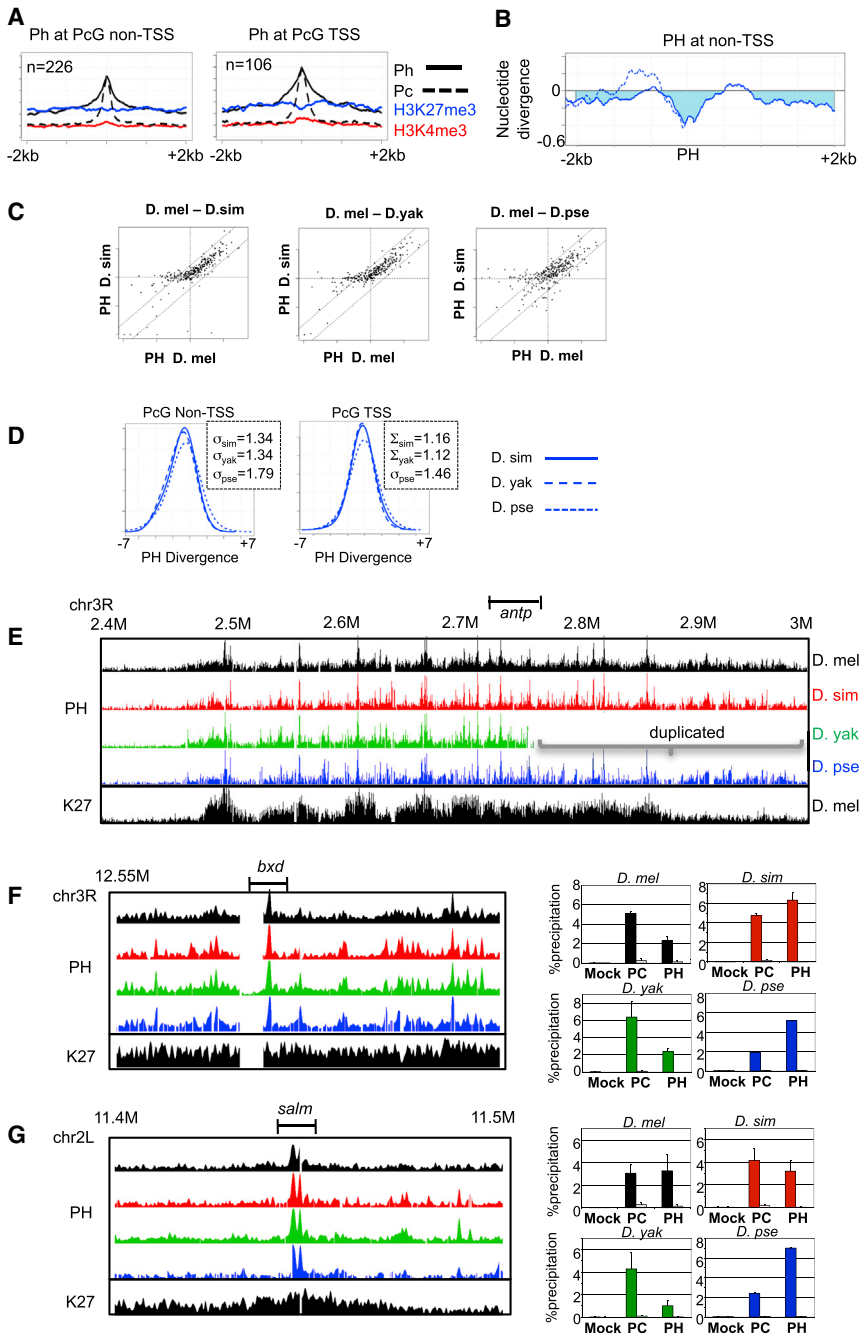


Figure 2. High Conservation of PRC1-Binding Sites

(A) Spatial distributions of D.mel ChIP-seq data for PH, PC, H3K27me3, and H3K4me3, pooling statistics around identified PH sites that are classified according to their proximity to a TSS (TSS: <500 bp from a known gene start site; non-TSS: >500 bp distant from TSSs).

(B) Spatial distributions of divergence statistics (log of ratios between observed and expected nucleotide substitutions) around PH sites within Polycomb domains (defined by Hi-C data). Sites in a non-TSS context (N, >500 bp from a TSS) are shown by solid lines, and sites in a TSS context are shown by dashed lines. All sites are oriented according to the strand of the nearest TSS.

(C) Genome-wide comparative analysis of PH ChIP-seq enrichment, showing pairwise data for all PH sites that were mapped on a syntenic, mappable locus.

(D) Summary of the divergence in PH ChIP-seq enrichment between D.mel and three other species, stratified according to TSS distance. The reported σ values are computed as the SD of the differential ChIP-seq data over all PH sites.

(E) Comparative ChIP-seq for PH in four *Drosophila* species, reflecting conservation of the epigenetic structure in the *Antp* complex. A region in D.yak that is duplicated (and therefore is not uniquely identifiable in ChIP-seq) is marked.

(F and G) Comparative ChIP-seq (left) and validation data by qChIP (right) for the conserved bxd and salm loci. Error bars represent the SD of the means of three independent experiments.

See also Figure S2 and Tables S1 and S2.

observed between replicate experiments (Figures S1A and S1B). The extreme evolutionary conservation of H3K27me3 domains is nontrivial, since H3K27me3 domains are developmentally plastic (Nègre et al., 2011; Figures 1D, 1E, and S1C). Thus, although reprogramming of Polycomb organization is not observed evolutionarily, it occurs throughout *Drosophila*'s life cycle. In conclusion, nonduplicated and nondeleted Polycomb domains in the embryonic stage conserved their H3K27me3 association to near completeness despite substantial underlying sequence divergence.

PRC1-Binding Sites Are Conserved amidst Dynamic Sequence Evolution

We expanded our ChIP-seq data set to study the evolution of PRC1-binding sites (via its core components Polyhomeotic [PH] and Polycomb [PC]). We also generated H3K4me3 ChIP-seq profiles to map transcription start sites (TSSs) in four *Drosophila* species. Using our previously constructed *Drosophila* Hi-C maps (Sexton et al., 2012), we identified Polycomb topological domains and studied the different localization of PRC1, H3K27me3, and H3K4me3 distributions around PH peaks within these Polycomb domains, classing sites according to their proximity to a TSS (defining TSS versus non-TSS sites; Figure 2A). There is some sequence conservation (mean 30% decrease in convergence rate) in the −200 to +200 bp range around non-TSS PRC1 sites, with accelerated evolution in the 400–800 bp around the site (Figure 2B), consistent with the broad evolutionary patterns observed around *Drosophila* enhancers (Kenigsberg and Tanay, 2013). Despite this relatively mild sequence conservation, a comparative PH ChIP-seq analysis (Figures S2A and S2B) indicated that

PRC1-binding sites are very highly conserved and rarely diverge at all. Systematically, we estimated a high degree of conservation of PH-binding levels in all species, with conservation slightly higher in TSSs (Figures 2C and 2D). These data indicated that PRC1 occupancy tolerates significant *cis*-element divergence, but also provided us with specific cases of divergence for analysis at higher resolution.

Linking *cis*-Element and PRC1 Occupancy Divergence

When we screened the ChIP-seq profiles, we identified 379 sites within PcG domains that are conserved across the *Drosophila* species (Figures 2E–2G and S3; Table S1) and 32 potentially diverged elements (Figure S2C). For example, a putative diverged PRC1-binding site, located within the Antp locus, shows increased PH-binding levels in D.pse that correlates with the gain of one strong GAGA repeat and two PHO (GCCATTT) boxes (Figure 3A). Both of these motifs were previously suggested to be important for PRC1 recruitment in D.mel (Schuettengruber et al., 2009). In another case, an element within the *Sox21b* region loses PRC1 binding in D.pse concomitantly with divergence of the GAGA repeat sequence, but not of a PHO box (Figure 3B). To validate that these elements are indeed functionally divergent due to local sequence perturbation, we generated transgenic reporter D.mel flies carrying either the D.mel sequence or the orthologous D.pse sequence upstream of a mini-white reporter gene. As controls, fly lines carrying the empty vector, a promoter region (*Zif*) that is not associated with PcG proteins at any developmental stage, or the well-characterized D.mel bxd PRE were generated. The reporter constructs were integrated at the same genomic position to avoid position effects, and PRE activity was determined by analyzing repression of the reporter gene using the eye color as a readout (Figures 3C–3H) or by determining the ability of the transgene to recruit the repressive histone mark H3K27me3 by quantitative ChIP (qChIP) assays (Figures 3I–3K). For the analyzed *Antp* locus (3R282), the D.mel sequence did not show any PRE characteristics (i.e., no significant repression of the reporter gene compared with the control "vector only," and no association of the transgene with repressive histone marks), whereas the orthologous sequence derived from D.pse significantly repressed the reporter gene in a dosage-dependent manner (pairing-sensitive silencing [PSS]) and was associated with increased levels of the H3K27me3 mark (Figures 3C, 3F, and 3I). In contrast, the putative diverged PRE sequences from the *Sox21b* gene locus (3L141) only showed functional PRE features when derived from D.mel (Figures 3E, 3H, and 3K). In summary, our transgenic analysis shows that the divergence of specific sequence elements (PHO and GAGA) could underlie *cis*-driven divergence of PRC1 binding in a few cases.

Extension of the comparative sequence analysis to 12 *Drosophila* species suggested that an association between GAGA/PHO motifs and the orthologous PREs defined by the conserved recruitment of PRC1 in four species is more universally conserved (Figures S4A and S4B). In fact, PHO and GAGA motifs are somewhat more conserved in the context of these orthologous PRE sites than in the rest of the genome (Figures S4C and S4D), even when stratifying for regional sequence conservation (Figure S4E). However, this degree of conservation is far from sufficient to predict the very highly conserved recruitment of PRC1 alone (Figure S4F). A possible explanation for this discrepancy is that some buffering mechanism contributes to the stabilization of PRC1 recruitment in a way that can tolerate significant divergence of local sequence elements.

Combinatorial PHO and DSP1 Occupancy Marks PRC1 Sites

To explore the relationship between PHO/DSP1 binding conservation and PRC1 recruitment conservation in more detail, we performed ultradeep ChIP-seq for the PcG recruiter factors (Figure S5A), focusing first on the D.mel genome. We compared the sites for one or both factors with PRC1 recruitment and H3K27me3 labeling, classing the sites based on whether or not they are present at TSSs and/or within Polycomb domains. Figure 4A shows that joint PHO/DSP1 sites within Polycomb domains co-occur perfectly with PRC1-binding sites (n = 159 and n = 103 for non-TSS and TSS contexts, respectively). PHO-only binding sites can also be observed at non-TSSs within Polycomb domains (n = 206), although in these cases PRC1 enrichment is observed at weaker levels. In contrast to these patterns, outside of Polycomb domains, we observe all combinations of PHO- and DSP1-binding sites (joint binding and PHO- or DSP1-only sites), at non-TSSs or TSSs. Taken together, these results show that PHO and DSP1 cobinding distributions are globally correlated with PRC1 occupancy, whereas PHO binding and DSP1 binding are less correlated with each other in other genomic contexts where PRC1 binding is generally lacking.

PHO Binds Weak *cis* Elements in a Polycomb-Domain-Dependent Fashion

A motif enrichment analysis in PHO- and DSP1-binding sites gave the expected binding sequences for the two factors. Further analysis revealed that combined PHO/DSP1-binding sites are characterized by strong GAGA motif enrichment (up to 85-fold higher than background level) but little or no PHO motif enrichment (Figures 4B and 4C). On the other hand, PHO sites lacking DSP1 enrichment are enriched for PHO motifs (45- to 60-fold over background levels), but not GAGA motifs, and DSP1-only binding sites are enriched for GAGA motifs, but not PHO motifs. To better understand the corecruitment of PHO with DSP1 in the absence of canonical PHO motifs, we derived ChIP-seq distributions at genomic sequences based on their fit to the PHO consensus (from a completely randomized sequence to a perfect consensus match), further breaking down the sequences to their presence at TSSs and the epigenetic identity (Polycomb, active, or null; Sexton et al., 2012) of their topological domains (Figures 4D and 4E). Strikingly, PHO recruitment to non-TSS sequences within Polycomb domains is effective even when the sequence is far from the optimal consensus. PHO recruitment to such low-specificity sites is far less efficient within null domains and is only efficient at active domains when in a TSS context. DSP1 recruitment to weak GAGA motifs is generally weaker outside of a TSS context, and is equally sensitive in active and Polycomb domains, with very poor recruitment in null domains. From these analyses, we conclude that sequence information is interpreted in a context-dependent manner to

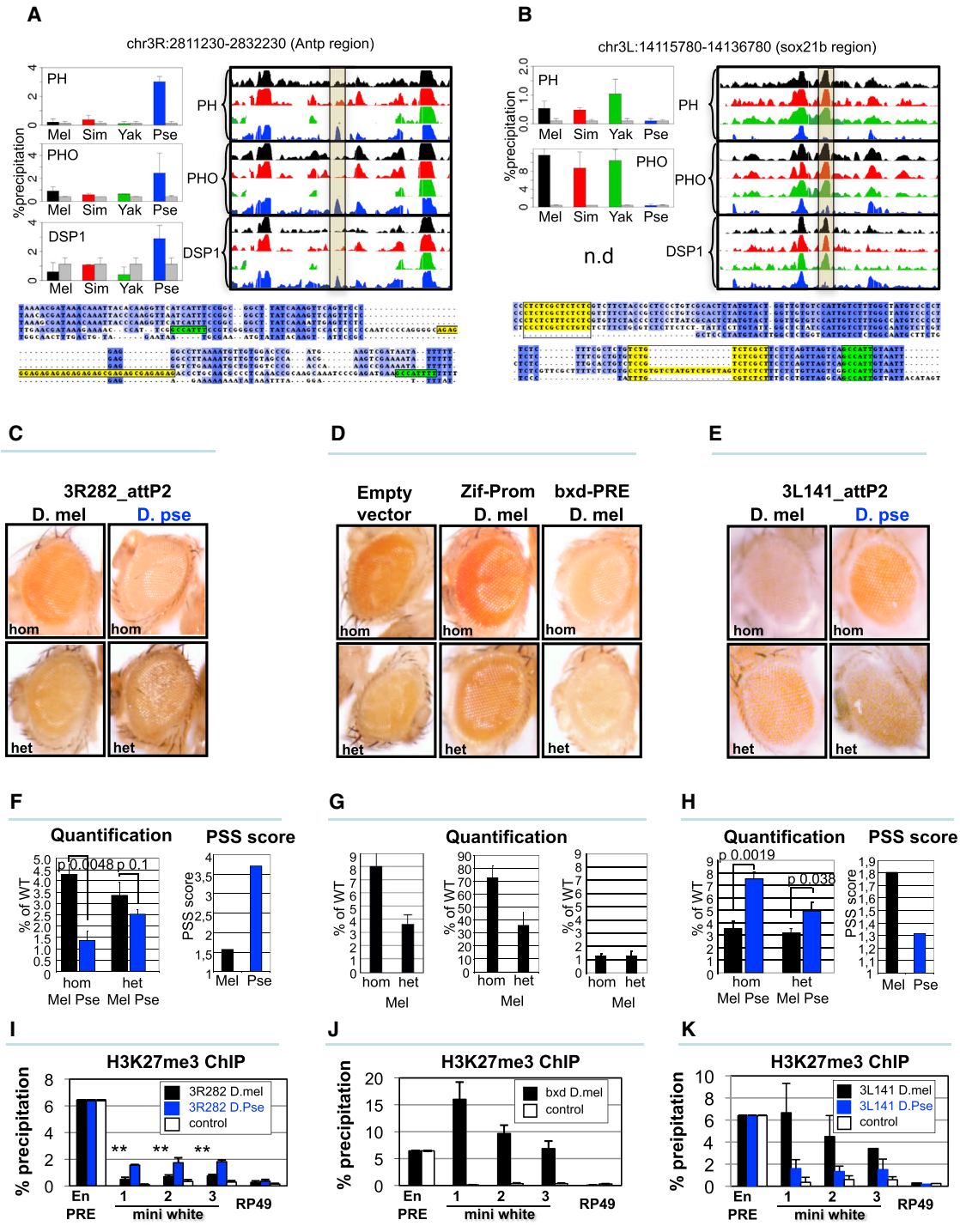


Figure 3. Scenarios for *cis*-Driven PRE Evolution

(A) qChIP (left), regional ChIP-seq profile (right), and annotated sequence alignment (bottom) for a region in the Antp locus in which a D.pse-specific PH site is detected. GAGA repeats are highlighted in yellow and PHO motifs are highlighted in green.

(B) Similar to (A), but showing data for PH binding that is specifically lost in D.pse at the *Sox21b* gene region compared with the other species.

(C–K) Results of transgenic reporter assays.

(C–E) Eye phenotype of transgenic fly lines. Four-day-old male flies, either heterozygous (bottom) or homozygous (top), are shown.

(F–H) Quantification of eye pigment in the heads of transgenic flies. Pigment levels are expressed as the percentage of WT pigment. Heads of 4-day-old male flies were used for each assay. The SD from three independent experiments is shown. The PSS score was calculated from mean pigment levels as heterozygote/homozygote for each diverged PRE region, normalized to the same ratio calculated for the "vector only" control line.

(legend continued on next page)

determine PHO or DSP1 binding, or their joint binding. Sequences within Polycomb domains, in particular, are capable of recruiting PHO through weak binding sites, possibly through cooperative mechanisms.

PHO Binding Evolutionary Divergence Is Correlated with *cis*-Element Divergence Only outside of Polycomb Domains

Analysis of overall sequence conservation around strong D.mel PHO and GAGA motifs (Figure 5A and 5B) shows that these motifs are moderately well conserved themselves (showing only a 30% decrease in divergence rate compared with the background). On the other hand, these motifs are typically located in a region of 200–400 bp, showing significant overall conservation. Regional conservation is observed, to some extent, even around motifs that lack PHO or DSP1 binding in embryos. Therefore, the evolutionary conservation of PHO-and DSP1-binding sites may involve not only the known *cis* elements analyzed above but also additional sequences that together define an element on a scale of a few hundred bases. We performed comparative ChIP-seq of PHO and DSP1 in four *Drosophila* species and showed that indeed, despite the limited conservation of the motifs, most of the factors' binding landscapes are well conserved (Figures 5C, 5D, and S5B–S5G), showing quantitatively higher stability than previously described for other factors that are not part of the Polycomb recruitment machinery (Figure S5H). Breaking down sites according to their context, we observed generally high PHO binding conservation ($\sigma = \sim 1.5$) with some preference for TSSs within Polycomb domains ($\sigma = 1.1$), and higher divergence of Dsp1 sites ($\sigma = \sim 2$), with more constrained evolution at TSSs within active domains ($\sigma = 1.6$). Importantly, the divergence in PHO binding within Polycomb domains is uncorrelated with the sequence divergence of PHO motifs (Figure 5E), whereas we observe a significant correlation between the two in active domains or null domains (spearman rho coefficient up to 0.35, $p << 10^{-10}$). On the other hand, the divergence of DSP1 binding is significantly correlated with GAGA motif divergence in all contexts (rho 0.19–0.42, $p << 10^{-10}$). Combined with the observations of low-affinity motif enrichment described above, our evolutionary analysis suggests that within Polycomb domains, PHO binding is buffered by cooperative factors and thus is capable of engaging low-affinity sites and being less sensitive to the evolutionary divergence of high-affinity binding sites. This buffering might contribute to a highly conserved evolutionary signature and in turn may underlie the conservation of H3K27me3 domains and PRC1 binding landscapes.

Recruiting the Recruiter: Predicting PHO Binding Intensity Given PHO Motifs and PRC1 Occupancy

According to the hierarchical PcG recruitment model (Wang et al., 2004), PHO should be a pioneer factor, binding its cognate DNA motifs specifically to promote the recruitment of PRC2 and PRC1. As noted above, however, sequence specificity alone cannot predict PHO and DSP1 binding accurately, so the high degree of correlation and predictability of PRC1 binding levels from the factors' binding profiles in all species (Figures S6A–S6C) reflects a strong, but not necessarily causative, correlation. Using classification of PHO-binding sites into Polycomb, active, and null domains, we observed a remarkably high correlation between the quantitative binding intensity of PH and PHO within the Polycomb domain context. A large range of PHO binding levels in the Polycomb domain context is also observed (Figure 6A). However, an analysis of the correlation between factor binding and sequence affinity (as predicted from the pho motif; Figure 6B) suggested that sequence-driven recruitment of PHO is strongly supported in active or null domains, but not in non-TSS contexts within Polycomb domains. To resolve this apparent contradiction, which goes against the hierarchical Polycomb recruitment model (Wang et al., 2004), we normalized PHO binding intensity by PH binding intensity to the same site, and recomputed the degree of sequence to binding correlation (Figure 6C). Modeling PHO as a function of PH binding and PHO motifs in this manner generated accurate quantitative predictions, explaining 73%–80% of the variance and significantly exceeding other hypothetical models that aim to predict PHO binding from combinations of its motifs and other factors (Figures S6D and S6E). The data therefore suggest that instead of a well-separated hierarchy of mechanisms, PRE specificity may involve a bidirectional interaction between recruiters and PRCs. An initial modest sequence specificity for PHO recruitment within Polycomb domains may be amplified through a PRC1-dependent positive feedback loop, making the ultimate PHO binding landscapes within Polycomb domains a complex function of sequence, DNA-binding factors, and PRC1 interactions.

Topological Clustering of PHO Sites within Polycomb Domains

In flies, Polycomb domains cluster in the nucleus to form specific compartments called Polycomb bodies (Cheutin and Cavalli, 2014). Within these nuclear compartments, PREs form specific spatial contacts (Lanzuolo et al., 2007; Lo Sardo et al., 2013), and it was suggested that PcG-mediated repression works in part by forming chromosomal loops that bring PREs into contact and antagonize transcriptional activation on repressed TSSs (Cheutin and Cavalli, 2014). A possible scenario for more robust binding of PHO within Polycomb domains is that looping brings PHO-binding sites into close spatial proximity, thereby increasing the local concentration of PHO as well as Polycomb proteins and facilitating PHO binding even at suboptimal DNA motifs. Indeed, analysis of the degree of PHO clustering and its conservation between D.mel and D.pse strongly suggested that the spatial organization of PHO-binding sites within Polycomb domains is more conserved compared with other genomic regions (Figure 6D). In order to better characterize the 3D environment of PHO-binding sites, we generated an

(I–K) qChIP assays performed on embryos (0–12 hr old) of the indicated transgenic fly lines using H3K27me3 antibodies. Primers 1 and 2 amplify the TSS region, whereas primer 3 amplifies the *white* coding region. Data were normalized to the positive control (*engrailed* PRE, "En"). As a negative control, the housekeeping gene *rp49* was used (**$p < 0.01$, *$p < 0.05$ as calculated from a two-tailed t test). Error bars represent the SD of the means of three independent experiments. See also Figure S4 and Table S2.

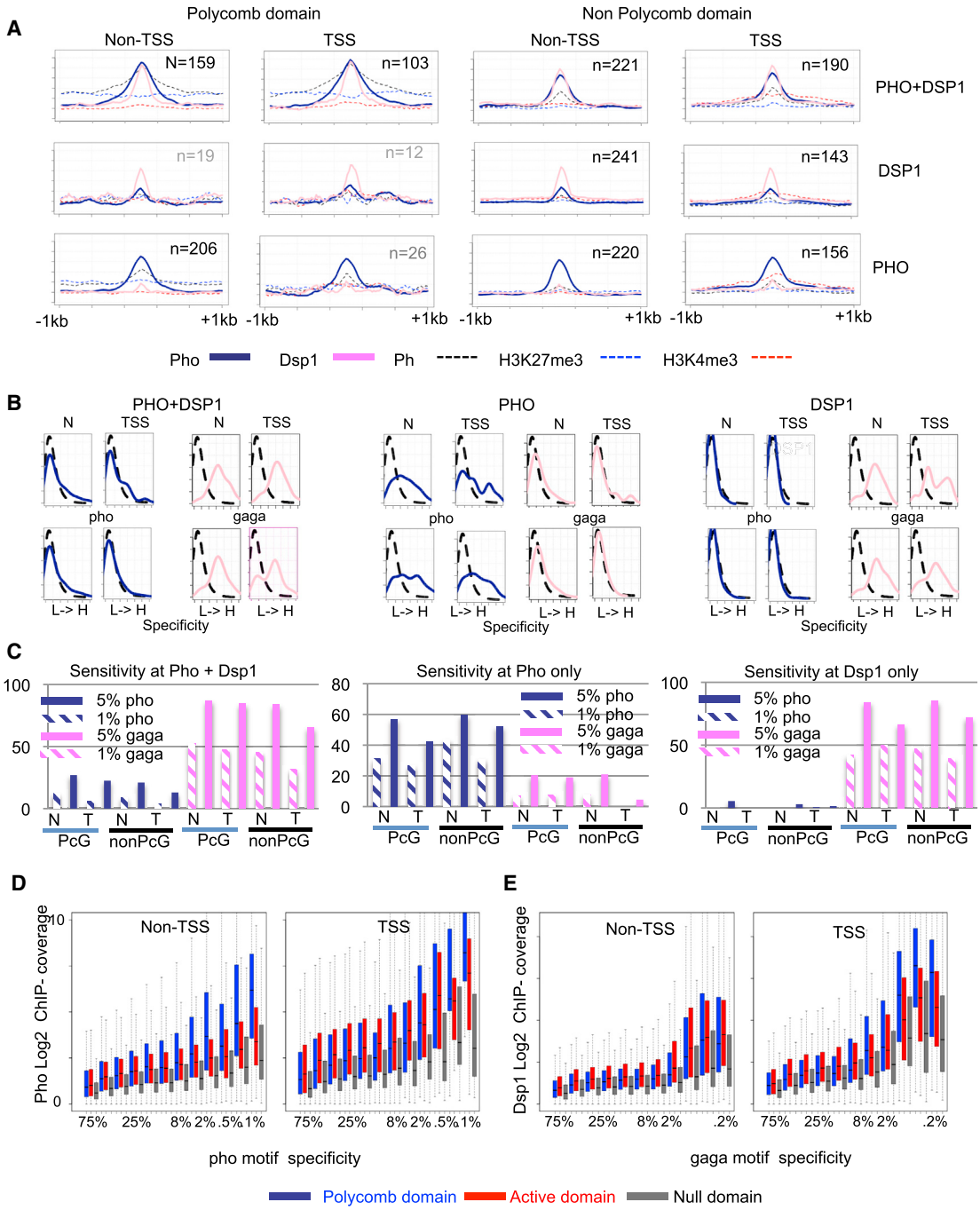


Figure 4. PcG Recruiter Sequence Specificity Is Dependent on Context

(A) Mean spatial enrichment statistics for PHO and DSP1 for groups of joint (top) and factor-specific (middle and bottom) sites. Sites are further stratified according to their association with a Hi-C Polycomb domain and their TSS proximity (TSS, within 500 bp of a TSS; N, >500 bp away). Data for H3K27me3 and H3K4me3, as well as for the PRC1 component PH, are added in dashed lines.

(B) Enrichment of sequence motifs is depicted for groups of joint and factor-specific sites as defined in (A). For each group, the distribution of motif scores is compared with the background distribution (black dashes) that is estimated from TSS-linked or unlinked sequences as appropriate.

(C) Summary of motif enrichment. Fold change is estimated from a comparison with the top first and fifth percentiles of the background distribution.

(D) Boxplot showing the genome-wide distribution of PHO ChIP-seq enrichments at sites with increasingly more specific motif scores (x axis, lower quartile on the left, upper 0.1 percentile on the right). The analysis is stratified to TSS and non-TSS contexts (right and left panels) and to the type of Hi-C domain in which each locus is contained (color-coded boxplots).

(E) Similar to (D), but for DSP1 ChIP-seq enrichments with respect to GAGA motif scores.

See also Figure S4.

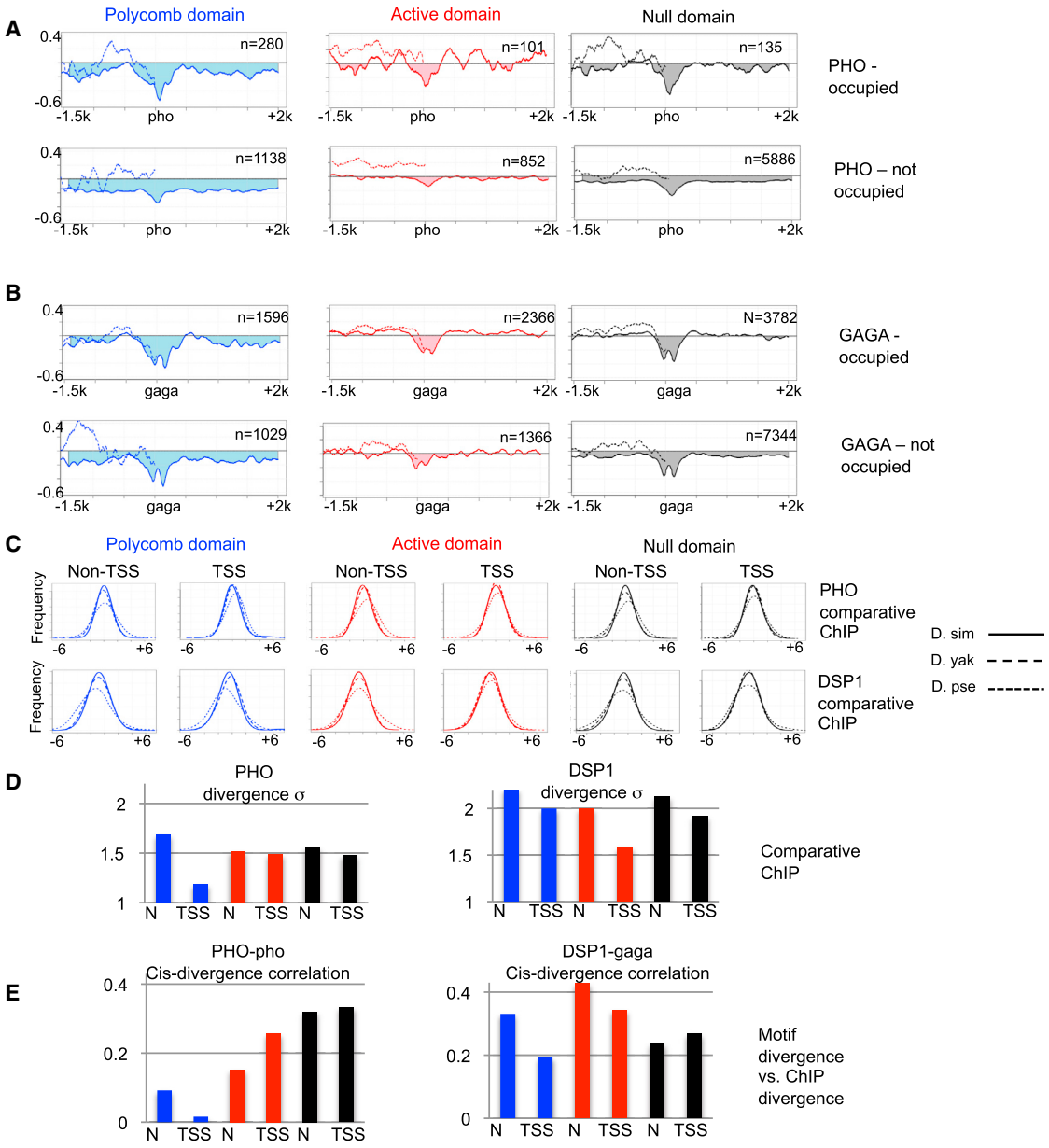


Figure 5. *cis*-Driven Evolution of PHO Is Buffered within Polycomb Domains

(A) Sequence divergence statistics around PHO motifs, estimated by the log of the ratio between the observed and expected numbers of substitutions in a 12-species *Drosophila* phylogeny. All occurrences of a PHO motif above some threshold were stratified according to their domain context (blue, Polycomb; red, active; black, null), their TSS proximity (>500 bp for a TSS [solid line], <500 bp from a TSS [dashed line in the left half of the graph]), and their PHO ChIP-seq occupancy.

(B) Similar to (A), but analyzing GAGA motifs and DSP1 sites.

(C) Divergence of PHO and DSP1 ChIP-seq on syntenic *Drosophila* sequences. Data for a pairwise comparison of D.mel and three other species are shown. For each pair, divergence is estimated on sites that are occupied by a factor in either of the species. Stratification according to domain type and TSS proximity was done as described above.

(D) Summary of divergence statistics shown in (C).

(E) Spearman correlation values testing the linkage between divergence in motif scores (following sequence divergence) and divergence in factor binding as estimated by comparative ChIP-seq. All values above 0.1 are statistically highly significant ($p << 10^{10}$).

See also Figure S5.

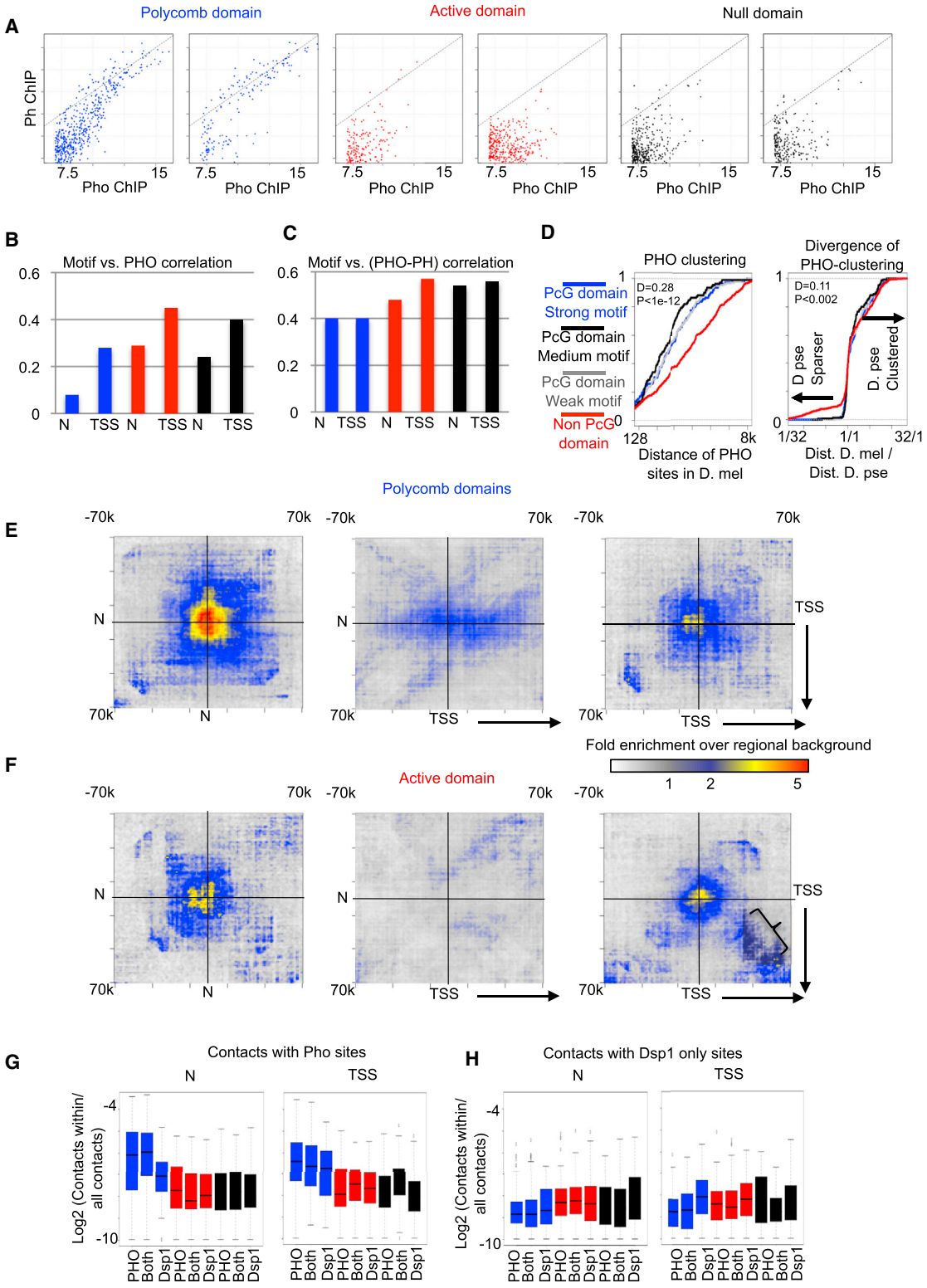


Figure 6. Potential Cooperative Factors underlying PHO Genomic Specificity

(A) Scatterplots comparing ChIP-seq enrichment values for PHO and PH in different epigenomic contexts.

(B) Spearman correlations between PHO motif scores and PHO ChIP-seq for different epigenomic contexts.

(C) Spearman correlations between pho motif scores and the PH-normalized PHO ChIP-seq binding intensity for different epigenomic contexts.

(legend continued on next page)

ultrahigh-coverage Hi-C map from D.mel embryos and explored the internal Polycomb domain structure at a higher resolution than was previously attainable. Hi-C was performed as previously described (Sexton et al., 2012) and sequencing was extended in order to obtain ~281 million mapped and filtered contacts. When we looked at the interaction maps around pairs of PHO non-TSS sites within Polycomb domains, we observed a high degree of enrichment of interactions within 10 kb of the sites compared with regions more distal (~50 kb) to the sites (Figure 6E). Contact enrichment was significantly weaker for non-TSS sites in active domains (Figure 6F). Furthermore, we observed enrichment of contacts between non-TSS and TSSs sites within Polycomb domains, but not in active domains. For TSSs within Polycomb domains, the contact enrichment was localized, whereas for TSSs in active domains, we observed preferential contacts on the TSSs but also some potential contacts over the gene body (the diagonal of increased contacts in the lower-right quadrant of the Hi-C submatrix). We also quantified the absolute (rather than regionally normalized) fraction of contacts that PHO-binding sites form with other PHO sites (with or without DSP1 cobinding; Figure 6G). We found that PHO sites preferentially contact each other in Polycomb domains, but not in active or null domains, showing that in addition to the spatial preferences identified in Figure 6E, genomic clustering of PHO sites and intradomain compaction give rise to a distinct topological environment around PHO sites in Polycomb domains. These data raise some questions about the role of PHO- and PRC1-bound chromatin hubs in combining the low specificities of several unlinked genomic loci into a more stable folded structure.

PRC1 Knockout Results in a Polycomb-Domain-Specific Reduction in PHO Binding

To further test the cooperative nature of PHO and PRC1 recruitment within Polycomb domains, we analyzed PHO binding within or outside of Polycomb domains in PH mutant embryos. Notably, in these mutants, PC recruitment was also strongly reduced (Figure 7D), suggesting that the stability of the whole PRC1 complex was affected. qChIP experiments revealed a reduced binding of PHO to PREs within Polycomb domains in PH mutant embryos, whereas binding of PHO to active promoters outside the Polycomb context was not significantly affected (Figures 7A and S7A). In order to extend this analysis to a genome-wide scale, we performed PHO ChIP-seq in PH mutant embryos. Analysis of differential PHO binding in different contexts showed a highly significant decrease in PHO recruitment specifically in Polycomb domains (Figures 7B and 7C). This was concomitant with a significant reduction in PRC1 binding, as expected (Figure 7D). Interestingly, in PH mutant embryos, we detected a significant correlation between pho motifs and PHO binding even within Polycomb domains (rho = 0.17, $p < 6 \times 10^{-5}$ in non-TSS loci, rho = 0.44 $p < 10^{-8}$ in TSSs), in contrast to the lack of such dependency in wild-type (WT) (rho = 0.08, $p < 0.07$ in non TSSs, rho = 0.28, $p < 3 \times 10^{-4}$ in TSSs). We next wished to test whether outside of Polycomb domains and PRC1 presence, PHO-binding sites are colocalized with some alternative chromosomal factor. Analysis of domain-type-dependent colocalization of PHO and DSP1 with insulator proteins (Nègre et al., 2010) showed that in non-PcG contexts, TSS PHO sites were strongly colocalized with CP190 and BEAF32 (Figure S7A), and possibly linked to active transcription (Figure S7B) and enrichment of long-range contacts (Figure S7C). We did not detect significant insulator enrichment on DSP1 sites, or on PHO sites in a non-TSS context. Together, these data indicate that PHO recruitment to DNA within Polycomb domains involves a feedback interaction with PRC1. This feedback is likely facilitated through local cooperativity, but also occurs within the context of long-range contacts formed in PcG hubs as described above. Even outside of Polycomb domains, PHO binding may be promoted by and/or promote the formation of long-range contacts via colocalization with the general, non-sequence-specific CP190 protein.

DISCUSSION

Multilayer Organization and the Evolutionary Buffering of Polycomb Domains

We used comparative epigenomics to demonstrate that Polycomb domains are an extremely well conserved feature of the genome during fly evolution. In fact, the evolutionary profile of epigenomic domain organization in embryos of five *Drosophila* species indicates a complete lack of divergence of H3K27me3-marked Polycomb domains in syntenic regions. A similar high conservation of the H3K27me3 pattern across *Drosophila* species was recently described (Arthur et al., 2014). Polycomb domains typically harbor several PH-marked PREs, and a comparative analysis showed that these are also highly conserved and the few loci that show a divergence of PRC1 occupancy patterns are not correlated with overall domain divergence. Likewise, the binding of PHO and DSP1 is highly conserved (to a degree at

(D) Cumulative distributions (left) of the distances between PHO sites and the nearest adjacent PHO site outside of PcG domains (red) and within PcG domains on three levels of motif energy (blue, high; gray, medium; black, low). Right: cumulative distribution of differential PHO-to-PHO distances for syntenic loci between D.mel and D.pse, again stratifying according to domain context. Kolmogorov-Smirnov statistics for the PcG versus non-PcG distributions are indicated.
(E) 2D submatrices derived from Hi-C data, centered according to pairs of PHO-binding sites within PcG domains. The submatrices are oriented according to the strand of the closest TSS. Site pairs are classified as non-TSS (left), TSS (right), and mixed contexts (middle). The total number of observed contacts in each spatial bin was divided by the number expected from a technical correction model. Ratios were then further normalized according to the mean ratio in the outer 7 kb frame of each matrix. A fixed color scale was then applied to visualize the matrices.
(F) Similar to (E), but using PHO-binding sites in active domains.
(G) The total (marginal) number of contacts observed for 2 kb elements centered on PHO sites were counted and normalized to the total number of contacts between PHO sites and other sites, classed according to sites binding only PHO, only DSP1, or both PHO and DSP1. This analysis is distinct from that shown in (E) because it omits the regional normalization and thus reflects the absolute frequency of recruiter-recruiter interactions within Polycomb domains.
(H) Similar to (G), but based on control sites that show DSP1 enrichment without PHO.
See also Figure S6.

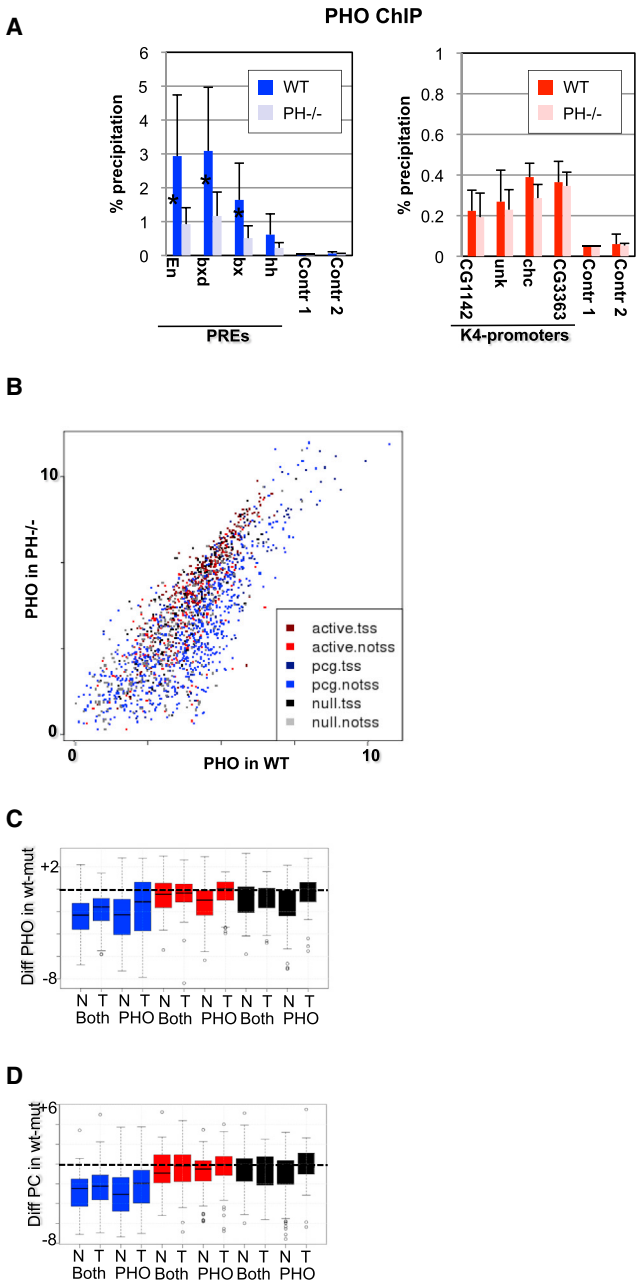


Figure 7. PRC1 Knockout Selectively Impairs PHO Recruitment to Polycomb Domains

(A) qChIP experiments of WT or PH mutant embryos ($PH^{-/-}$) using PHO antibodies. Primers specific for PREs (left) or active promoters bound by PHO outside the PcG context (right) were used. Results are represented as the percentage of input chromatin precipitated. The SD was calculated from at least four independent replicate experiments (*$p < 0.05$ as calculated from a two-tailed t test).

(B) PHO ChIP-seq intensities of WT and PH mutant embryos, color-coded according to context.

(C) Boxplot depicting differential PHO ChIP-seq binding in WT and PH mutant embryos, classified according to domain type (color), co-occurrence of DSP1 (both/PHO), and TSS context (N, non-TSS; T, TSS).

(D) Similar to (C), but showing differential PC ChIP-seq intensities.

See also Figure S7 and Table S2.

least as strongly, and possibly more strongly, than binding of individual factors; Bradley et al., 2010; He et al., 2011), but even cases of diverged factor occupancies are usually not correlated with overall PRE divergence. In marked contrast, the sequences underlying PREs and Polycomb domains are diverging extensively, and sequence-based prediction of PREs across *Drosophila* species suggested that divergence of PREs could occur frequently (Hauenschild et al., 2008). However, neither our ChIP-seq experiments nor our transgenic reporter assays support this dynamic behavior (Figure S3). Instead, we show that such sequence divergence is buffered by the epigenetic targeting mechanisms to maintain Polycomb domains. We suggest that the multilayered organization uses redundancy and cooperativity to facilitate the remarkable Polycomb domain conservation. This is occurring both in *cis*, where several TFs collaborate to define a regulatory element even when the underlying sequence is imperfect (see, e.g., Stefflova et al., 2013), and at the domain level, where several PREs participate to define the PcG domain structure and possibly stabilize each other.

The Sequence Specificities of PREs Are Not Fully Encoded in *cis*

Although PREs are associated with several known sequence features (such as GAGA- and PHO-binding motifs) in a statistically significant way, these features are not sufficient to distinguish many PREs from the genomic background and from other PHO- or DSP1-bound active chromatin elements (Schuettengruber et al., 2009). There are many possible explanations for this lack of specificity, including the existence of additional, yet-to-be-characterized sequence-specific recruiting factors; the involvement of nucleosome positioning; transcription of noncoding RNAs; or imperfect modeling of the sequence specificity of the known factors. The data presented here, however, introduce a new perspective that can help resolve this conundrum. In contrast to previous hypotheses, the data show that even when strong binding sites are lacking, PHO and DSP1 may bind PREs directly through weak (but highly nonrandom) motifs. Remarkably, sequence affinities that are completely nonspecific on a genomic scale (possibly defining millions of spurious sites) are still highly informative for predicting the binding intensity within the context of a PRE. The strong correlation of PHO binding with weak but nonrandom motifs makes it unlikely that binding to these sites represents indirect binding via interaction/looping with strong binding sites. The data show that in order to understand PRE sequence specificity, we must take into account multiple potential binding sites with variable affinities and fidelities, and consider their cooperative interaction in the context of the PRE chromosomal landscape. This idea is compatible with the evolutionary constraints on PRE sequences, which we have demonstrated here to affect a spectrum of binding affinities rather than to conserve classical binding sites alone.

Cooperative Rather than Hierarchical Interactions of PHO and PRC1 Contribute to PRE and Domain Specification

What might be the molecular mechanism that allows the specific binding of weak sites in the context of Polycomb domains? One possibility is that cooperative binding of TFs at PREs supports

their occupancy of weak motifs. Indeed, we found that PHO and DSP1 are bound jointly at PREs (with weak underlying sequence motifs), whereas at other regions of the genome where the factors bind alone, they are usually associated with strong sequence motifs. This observation is in agreement with the recently proposed "TF collective model," according to which combinatorial TF binding occurs with little or no apparent sequence motifs for at least a subset of the bound factors (Junion et al., 2012).

In addition, we show that transient interactions of DNA-binding proteins with weak affinity sites are stabilized by the presence of the PcG proteins themselves. A similar observation of a positive feedback of PRC1 on PHO binding was recently reported (Kahn et al., 2014) and is further supported by the fact that cooperative binding of PHO and Polycomb to PREs can occur even in vitro (Mohd-Sarip et al., 2005). In vivo, long-range contacts involving remote PREs within the same (or even a different) Polycomb domain may contribute to this process (Bantignies et al., 2011; Lanzuolo et al., 2007; Sexton et al., 2012). Clustering of multiple flanking PREs in the 3D space of the nucleus might generate Polycomb compartments characterized by high concentrations of PcG proteins as well as their recruiting DNA-binding proteins. In this scenario, loss of occupancy following the dissociation of any of these factors from DNA may be more easily replenished by the concentrated stock of factor within a Polycomb compartment compared with individual binding sites present elsewhere in the genome. This may push the equilibrium toward increased PHO and DSP1 binding to low-affinity sites and partially reduce the evolutionary pressure to maintain the nucleotidic sequence of recruiter motifs at PREs. Structural long-range effects may also inhibit PcG recruitment in cases where active enhancers and TSSs are in proximity to a candidate PRE sequence. Our analysis suggests that H3K4me3-marked loci are also highly conserved, but the low-affinity PHO- or DSP1-binding sites in them are completely uncorrelated with occupancy of these factors, further supporting a model of highly organized and cooperative epigenomic organization.

In conclusion, the data presented here indicate that sequence conservation collaborates with 3D chromatin architecture to maintain an exceptional evolutionary stability of Polycomb-regulated loci in fly genomes. This phenomenon highlights the contribution of chromosome domains and their particular looping structures to epigenomic specificity and genome evolution. Hi-C analysis in mammals has revealed that topological domains are a strikingly conserved feature between the mouse and human genomes (Dixon et al., 2012). Our data raise the possibility that, beyond combinatorial contributions by TF-binding sites in close proximity, the confinement of regulatory elements within TADs and their frequent DNA contacts constitute significant driving forces that also affect DNA sequence evolution in these and possibly many other species.

EXPERIMENTAL PROCEDURES

Fly Strains

Drosophila species were obtained from the *Drosophila* Stock Center (http://cis.arl.arizona.edu/stock.htm; stock numbers: D.sim 14021-0251.195, D.yak 14021-0261.01, D.pse 14011-0121.94, and D.vir 15010-1051.17). In addition, the Oregon-R w1118 line of D.mel (R. Paro, Center of Biosystems, Science and Engineering, Basel, Switzerland) was used.

Generation of Transgenic Fly Lines

Approximately 1.6 kb DNA fragments corresponding to conserved or diverged PRE regions from both D.mel and D.pse were obtained by PCR using specific primers. Each PCR fragment was cloned into a slightly modified attB-P[acman]-CmR-BW vector (BACPAC Resources Center) at the BamHI restriction site. Plasmids were sent to BestGene for site-specific integration using the PhiC31 integrase system into attp2 (chromosome 3L, cytological position 68A4) or VK7 sites (chromosome 3R, cytological position 82A1). Site-specific integration was verified by standard PCR analyses.

Eye Color Pigmentation Assay and Eye Imaging

For the eye pigmentation assay, 10 or 20 heads (depending on the eye color) of 4-day-old male flies were collected and homogenized in EPE buffer (30% EtOH-HCL [pH 2]) and incubated for 1 hr at 25°C in the dark. After centrifugation, eye pigmentation was quantified by spectrometry at 480 nm. The PSS score was calculated from mean pigment levels as heterozygote/homozygote for each diverged PRE region, normalized to the same ratio calculated for the "vector only" control line. Transgenic flies were photographed with a standard light microscope, always using the same time of exposure for each diverged PRE and controls (*$p < 0.01$ as calculated from a two-tailed t test).

ChIP Experiments

ChIP experiments on whole *Drosophila* embryos 4–12 hr after egg laying were performed essentially as described in Schuettengruber et al. (2009). Antibodies were diluted 1:100 for IP. For qChIP, after immunoprecipitation and DNA purification, enrichment of specific DNA fragments was analyzed by real-time PCR using Roche Light Cycler equipment and accessories as described in Comet et al. (2006). Data are expressed as the percentage of input chromatin precipitated for each region examined. As a negative control, *Rp49* was included in the PCR experiments. For primer sequences, see Table S2.

For ChIP-seq, 20 ng of the ChIP reaction and input DNA were used for the library preparation. ChIP-seq was performed by the Montpellier GenomiX facility (MGX, http://www.mgx.cnrs.fr/index.php).

Antibodies

Antibodies against modified histones were obtained from Millipore (H3K4me3 [#04-745] and H3K27me3 [#07-449]). Antibodies against the proteins PC, PH, and DSP1 are described in Schuettengruber et al. (2009). The antibody against PHO used for ChIP-seq in the different *Drosophila* species is described in Klymenko et al. (2006).

Sorting of PH Mutant Flies

Ph^{del} mutant flies (Feng et al., 2011) were crossed to a balancer line expressing GFP from the Krüppel promoter. Homozygous Ph^{del}/Ph^{del} embryos were collected from the heterozygous stock by selecting for GFP-negative embryos using an embryo sorter (COPAS SELECT; Union Biometrica). ChIP on WT and $Ph^{del/del}$ embryos was performed as described above. The PHO antibodies used are described in Schuettengruber et al. (2009). Note that this antibody gives essentially the same ChIP-seq profiles as the PHO antibody used in the species ChIP-seq (Klymenko et al., 2006), but seems to have reduced affinity in some non-melanogaster strains (data not shown).

Low-Level ChIP-Seq Analysis and Phylogenetic Projection

ChIP-seq mapping, normalization, and phylogenetic projection were performed as described in Supplemental Experimental Procedures. Briefly, we created binding profiles and identified nonmappable and nonalignable regions for each of the species independently. We then projected all syntenic and mappable regions onto the D.mel coordinate space to facilitate comparison.

Motif Finding

Discriminative motif finding was performed as previously described (Schuettengruber et al., 2009). The positive set in each species consisted of 400 bp elements around peaks of PH or H3K4me3. The background set was defined by shifting the positive set by 1,000 bp.

Sequence Affinities

For each motif and each species, we generated the sequence affinity in 20 bp bins with respect to the D.mel genome by summing over the position weight matrix (PWM) likelihoods across the orthologous sequence in the appropriate species. Each 20 bp bin was extended to the summation of the surrounded 400 bp window. Again, in order to control for variable sequence content in the different species, we transformed the sum of PWM likelihoods value to minus log2 of its (1 − quantile) value.

Phylogenetic Motif Tracing

To analyze the enrichment of the inferred motifs in PH or H3K4me3 sites along the *Drosophila* phylogeny, we used multiple alignment of 12 *Drosophila* species and computed the enrichments as defined above in sequences that are orthologous to the D.mel and D.pse elements. To ensure that the orthologous elements were of a uniform size, we always used 400 bp around the center of the projected locus.

Sequence Conservation Analysis

We used our previously described context-aware inferred substitution statistics (Chachick and Tanay, 2012; Kenigsberg and Tanay, 2013) to extract statistics on the observed and expected numbers of point substitutions for each alignable D.mel locus. We then pooled together loci around given landmarks (e.g., PH-binding sites or PHO/GAGA motifs) and summed up the total number of such substitutions at each relative offset to the landmark. The conservation statistics were then derived as log2(observed/expected).

PHO Syntenic Clustering

To create Figure 6D, we defined the PHO clustering score for each PHO site as the distance to the nearest adjacent PHO site (in logarithmic scale). To determine the extent to which this clustering property is conserved, we projected all D.mel PHO sites to the D.pse genome and computed the projected clustering score for adjacent sites that maintain their co-occurrence on the same chromosome or contig. We note that one must use caution when interpreting these data, since the analysis disregards all sites that were mapped to different chromosomes in D.pse, and the analysis is by definition asymmetric (e.g., performing a similar analysis starting from D.pse sites would not generate precisely the same distribution).

Hi-C Analysis

Hi-C raw sequence filtering, mapping, and normalization were done as previously described (Sexton et al., 2012). This provided statistics on the observed number of contacts for each pair of restriction fragments and the number expected from a technical background model. Given a set of sites, we characterized the spatial contact structure around interactions between them (Figures 6E and 6F) by extracting and pooling observed and expected statistics from 100 kb × 100 kb submatrices centered on the contact point of each pair. In cases of overlaps between such submatrices, we always assigned the contact to the pair of sites more proximal to it. We performed our analysis at restriction fragment resolution and then pooled data for 5 kb × 5 kb bins defined by distance to the contact point. Since the matrices were extracted around contacts with variable distances (and therefore with variable background contact intensity), we normalized the pooled matrix statistics using the average log(observed/expected) ratio of bins in the outer 5 kb frame. We also used an alternative approach (Figure 6G) in which we computed for windows of 2 kb around each site the total number of observed contacts with other 2 kb windows around sites of the same family, and divided this value by the total (marginal) number of contacts observed for restriction fragments within this window. When we used this normalization approach, we did not further normalize by the technical correction model.

ACCESSION NUMBERS

The raw sequencing data reported in this work have been deposited in the NCBI GEO under accession numbers GSE60428 and GSE61471.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.08.072.

AUTHOR CONTRIBUTIONS

B.S. and G.C. designed the experiments. N.O. and A.T. designed evolutionary and computational analyses. B.S., T.S., and M.E. performed ChIP experiments, Hi-C assays, and fly work. N.O. and T.A analyzed ChIP-seq data. N.O. implemented modeling and regression approaches with help from S.S. E.Y. helped with processing and normalizing raw Hi-C data. H.P. performed library preparation and sequencing. A.T., G.C., N.O., and B.S analyzed the data and wrote the manuscript.

ACKNOWLEDGMENTS

We are grateful to Jurg Müller and Judy Kassis for providing PHO antibodies, and D. Locker for providing DSP1 antibodies. We thank Mythily Ganapathi for helpful discussions. Research at G.C.'s lab was supported by grants from the European Research Council (ERC-2008-AdG No 232947), the CNRS, the INSERM, the European Network of Excellence EpiGeneSys, the Agence Nationale de la Recherche, and the Association pour la Recherche sur le Cancer. Research at A.T.'s lab was supported by the European Research Council, the European Network of Excellence EpiGeneSys, and the Israel Science Foundation.

Received: May 13, 2014
Revised: July 8, 2014
Accepted: August 27, 2014
Published: October 2, 2014

REFERENCES

Arthur, R.K., Ma, L., Slattery, M., Spokony, R.F., Ostapenko, A., Nègre, N., and White, K.P. (2014). Evolution of H3K27me3-marked chromatin is linked to gene expression evolution and to patterns of gene duplication and diversification. Genome Res. *24*, 1115–1124.

Atchison, L., Ghias, A., Wilkinson, F., Bonini, N., and Atchison, M.L. (2003). Transcription factor YY1 functions as a PcG protein in vivo. EMBO J. *22*, 1347–1358.

Bantignies, F., and Cavalli, G. (2011). Polycomb group proteins: repression in 3D. Trends Genet. *27*, 454–464.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell *144*, 214–226.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell *157*, 1445–1459.

Bradley, R.K., Li, X.Y., Trapnell, C., Davidson, S., Pachter, L., Chu, H.C., Tonkin, L.A., Biggin, M.D., and Eisen, M.B. (2010). Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related Drosophila species. PLoS Biol. *8*, e1000343.

Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.L., and Kassis, J.A. (1998). The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. Mol. Cell *1*, 1057–1064.

Chachick, R., and Tanay, A. (2012). Inferring divergence of context-dependent substitution rates in Drosophila genomes with applications to comparative genomics. Mol. Biol. Evol. *29*, 1769–1780.

Cheutin, T., and Cavalli, G. (2014). Polycomb silencing: from linear chromatin domains to 3D chromosome folding. Curr. Opin. Genet. Dev. *25*, 30–37.

Comet, I., Savitskaya, E., Schuettengruber, B., Nègre, N., Lavrov, S., Parshikov, A., Juge, F., Gracheva, E., Georgiev, P., and Cavalli, G. (2006). PRE-mediated bypass of two Su(Hw) insulators targets PcG proteins to a downstream promoter. Dev. Cell *11*, 117–124.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. Cell Reports *7*, 1456–1470.

Déjardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., and Cavalli, G. (2005). Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. Nature *434*, 533–538.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Duncan, I.M. (1982). Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster. Genetics *102*, 49–70.

Feng, S., Huang, J., and Wang, J. (2011). Loss of the Polycomb group gene polyhomeotic induces non-autonomous cell overproliferation. EMBO Rep. *12*, 157–163.

Fiedler, T., and Rehmsmeier, M. (2006). jPREdictor: a versatile tool for the prediction of cis-regulatory elements. Nucleic Acids Res. *34*, W546–W550.

Fujioka, M., Yusibova, G.L., Zhou, J., and Jaynes, J.B. (2008). The DNA-binding Polycomb-group protein Pleiohomeotic maintains both active and repressed transcriptional states through a single site. Development *135*, 4131–4139.

Hauenschild, A., Ringrose, L., Altmutter, C., Paro, R., and Rehmsmeier, M. (2008). Evolutionary plasticity of polycomb/trithorax response elements in Drosophila species. PLoS Biol. *6*, e261.

He, Q., Bardet, A.F., Patton, B., Purvis, J., Johnston, J., Paulson, A., Gogol, M., Stark, A., and Zeitlinger, J. (2011). High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. Nat. Genet. *43*, 414–420.

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E., and Furlong, E.E. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. Cell *148*, 473–486.

Kahn, T.G., Stenberg, P., Pirrotta, V., and Schwartz, Y.B. (2014). Combinatorial interactions are required for the efficient recruitment of pho repressive complex (PhoRC) to polycomb response elements. PLoS Genet. *10*, e1004495.

Kassis, J.A., and Brown, J.L. (2013). Polycomb group response elements in Drosophila and vertebrates. Adv. Genet. *81*, 83–118.

Kenigsberg, E., and Tanay, A. (2013). Drosophila functional elements are embedded in structurally constrained sequences. PLoS Genet. *9*, e1003512.

Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev. *20*, 1110–1122.

Kwong, C., Adryan, B., Bell, I., Meadows, L., Russell, S., Manak, J.R., and White, R. (2008). Stability and dynamics of polycomb target sites in Drosophila development. PLoS Genet. *4*, e1000178.

Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat. Cell Biol. *9*, 1167–1174.

Lewis, E.B. (1985). Regulation of the genes of the bithorax complex in Drosophila. Cold Spring Harb. Symp. Quant. Biol. *50*, 155–164.

Lo Sardo, F., Lanzuolo, C., Comoglio, F., De Bardi, M., Paro, R., and Orlando, V. (2013). PcG-mediated higher-order chromatin structures modulate replication programs at the Drosophila BX-C. PLoS Genet. *9*, e1003283.

Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. *6*, e1001244.

Mohd-Sarip, A., Cléard, F., Mishra, R.K., Karch, F., and Verrijzer, C.P. (2005). Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. Genes Dev. *19*, 1755–1760.

Müller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in Drosophila. Curr. Opin. Genet. Dev. *16*, 476–484.

Nègre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A., et al. (2010). A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet. *6*, e1000814.

Nègre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R., et al. (2011). A cis-regulatory map of the Drosophila genome. Nature *471*, 527–531.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Ringrose, L., Rehmsmeier, M., Dura, J.M., and Paro, R. (2003). Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster. Dev. Cell *5*, 759–771.

Rothenberg, E.V. (2014). The chromatin landscape and transcription factors in T cell programming. Trends Immunol. *35*, 195–204.

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

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol. *7*, e13.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Stefflova, K., Thybert, D., Wilson, M.D., Streeter, I., Aleksic, J., Karagianni, P., Brazma, A., Adams, D.J., Talianidis, I., Marioni, J.C., et al. (2013). Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. Cell *154*, 530–540.

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004). Hierarchical recruitment of polycomb group silencing complexes. Mol. Cell *14*, 637–646.

Zeng, J., Kirk, B.D., Gou, Y., Wang, Q., and Ma, J. (2012). Genome-wide polycomb target gene prediction in Drosophila melanogaster. Nucleic Acids Res. *40*, 5848–5863.