Evolutionary Plasticity of Polycomb/Trithorax Response Elements in *Drosophila* Species

Arne Hauenschild[1☯], Leonie Ringrose[2,3☯*], Christina Altmutter[2], Renato Paro[3,4], Marc Rehmsmeier[1,5*]

1 Universität Bielefeld, Center for Biotechnology (CeBiTec), Bielefeld, Germany, **2** Institute of Molecular Biotechnology (IMBA), Vienna, Austria, **3** Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH), Heidelberg, Germany, **4** Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland, **5** Gregor Mendel Institute of Molecular Plant Biology (GMI), Vienna, Austria

***cis*-Regulatory DNA elements contain multiple binding sites for activators and repressors of transcription. Among these elements are enhancers, which establish gene expression states, and Polycomb/Trithorax response elements (PREs), which take over from enhancers and maintain transcription states of several hundred developmentally important genes. PREs are essential to the correct identities of both stem cells and differentiated cells. Evolutionary differences in *cis*-regulatory elements are a rich source of phenotypic diversity, and functional binding sites within regulatory elements turn over rapidly in evolution. However, more radical evolutionary changes that go beyond motif turnover have been difficult to assess. We used a combination of genome-wide bioinformatic prediction and experimental validation at specific loci, to evaluate PRE evolution across four *Drosophila* species. Our results show that PRE evolution is extraordinarily dynamic. First, we show that the numbers of PREs differ dramatically between species. Second, we demonstrate that functional binding sites within PREs at conserved positions turn over rapidly in evolution, as has been observed for enhancer elements. Finally, although it is theoretically possible that new elements can arise out of nonfunctional sequence, evidence that they do so is lacking. We show here that functional PREs are found at nonorthologous sites in conserved gene loci. By demonstrating that PRE evolution is not limited to the adaptation of preexisting elements, these findings document a novel dimension of *cis*-regulatory evolution.**

Citation: Hauenschild A, Ringrose L, Altmutter C, Paro R, Rehmsmeier M (2008) Evolutionary plasticity of Polycomb/Trithorax response elements in *Drosophila* species. PLoS Biol 6(10): e261. doi:10.1371/journal.pbio.0060261

Introduction

cis-Regulatory DNA elements are essential for the correct activation, repression, and maintenance of gene expression. These elements typically contain multiple short DNA motifs, which are recognised by sequence-specific DNA binding proteins, that either themselves act as activators and repressors of transcription, or recruit other proteins that do so [1,2]. One class of *cis*-regulatory DNA elements is enhancers, which establish gene expression states. Another important class is Polycomb/Trithorax response elements (PREs), first identified in the *Drosophila* homeotic (hox) gene complexes [3,4], where they maintain the transcriptional states of hox genes that have been determined earlier on in development by embryonic enhancers [5–7]. The hox PREs preserve the transcription patterns of their associated genes stably over many cell generations, long after the proteins that bind the enhancers have disappeared. Thus, hox PREs are epigenetic memory elements [8]. Although PREs are similar to enhancers in many ways, the most important functional difference between these two types of elements is that enhancers respond to local differences in concentration of the transcription factors that bind them, whereas the Polycomb group (PcG) and Trithorax group (TrxG) proteins are ubiquitously expressed; thus, the PRE element responds to the transcriptional state of the promoter [3,4]. Since their initial discovery in the hox complexes, it has become clear that PREs regulate several hundred other genes in addition. In both flies and vertebrates, the targets of Polycomb regulation include genes involved in major cell-fate decisions, and in several differentiation and morphogenetic pathways [9–15]. Consistent with the nature of these target genes, the PcG proteins are essential to the correct identities of both stem cells and differentiated cells [16,17].

In *D. melanogaster*, many PRE elements that have similar functional properties in transgenic assays are enriched in preferred pairs of motifs, enabling the identification of a subset of *Drosophila* PREs by computational prediction [18,19]. However, these same elements show no preferred order or number of motif pairs, suggesting that the design of PREs in terms of linear arrangement of motifs is flexible [18]. Furthermore, fly PREs can act many tens of kilobases upstream, downstream, or in the introns of the genes they regulate [9,10], suggesting that their position relative to their cognate promoter is also flexible. This diversity of design among *D. melanogaster* PREs raises the question of whether these differences are important for function, and whether PRE position at each gene is conserved across different

Academic Editor: Michael B. Eisen, University of California Berkeley, United States of America

Received June 6, 2008; **Accepted** September 15, 2008; **Published** October 28, 2008

Abbreviations: *abd-A, abdominal-A; Abd-B, Abdominal-B; bxd, bithoraxoid;* ChIP, chromatin immunoprecipitation; *dpp, decapentaplegic;* H3K27me3, histone H3 trimethylated at lysine 27; hox, homeotic; PC, Polycomb protein; PcG, Polycomb group; PH, Polyhomeotic protein; PRE, Polycomb/Trithorax response element; *salm, spalt major; Scr, Sex-combs reduced; trh, trachealess;* TrxG, Trithorax group;

* To whom correspondence should be addressed. E-mail: marc.rehmsmeier@gmi.oeaw.ac.at (MR); ringrose@imp.univie.ac.at (LR)

☯ These authors contributed equally to this work.

Author Summary

The evolution of regulatory DNA plays a crucial role in making species different from one another. One way to study the evolution of regulatory DNA is by genome alignment, which assumes that elements with conserved function will be found in conserved pieces of DNA. Although conservation does imply function, it does not follow that all functional elements must be conserved, nor that nonconserved DNA has no function. However, computational approaches based on genome alignment alone cannot identify any kind of evolution beyond small changes in otherwise conserved elements. We have used a novel computational approach, in combination with experimental validation, to examine how regulatory DNA evolves in four *Drosophila* species. We focus on Polycomb/Trithorax response elements (PREs), which regulate several hundred developmental genes, and are vital for maintaining cell identities. We find that PRE evolution is extraordinarily dynamic: not only motif composition, but also the total number of PREs, and even their genomic positions, have changed dramatically in evolution. By demonstrating that the evolution of PREs goes far beyond the gradual adaptation of preexisting elements, this study documents a novel dimension of regulatory evolution. We propose that PRE evolution provides a rich source of potential diversity between species.

Drosophila species. The *bithoraxoid* (*bxd*) PRE, which regulates the hox gene *Ultrabithorax* (*Ubx*; FBgn0003944), shows large blocks of conserved sequence across several *Drosophila* species, supporting the idea that PRE position is evolutionarily constrained [7,20]. However, the conservation of the several hundred other PREs in the *D. melanogaster* genome has not been evaluated, and it is not known whether these PREs are also evolutionarily constrained.

The effects of evolutionary changes in enhancers and promoters have been well studied for several individual genes in diverse organisms [21,22]. Starting from a known *cis*-regulatory element in one species, the orthologous sequences in other species have been analysed in terms of evolutionary changes and their impact on regulatory function. These studies have demonstrated that many *cis*-regulatory elements show rapid motif turnover [23,24]. In cases in which function has been evaluated, these studies have shown that some enhancers tolerate evolutionary change without large differences in function [25–27]. On the other hand, there are also many examples of evolutionary differences in enhancer sequences that lead to major phenotypic changes [21,22,28–30]. Thus, *cis*-regulatory elements are a potential source of phenotypic diversity, and it has been proposed that positive selection acts primarily on *cis*-regulatory sequences rather than protein-coding sequences [31–33]. The genomic sequencing of several closely related species has enabled the study of *cis*-regulatory evolution on a genome-wide scale [34–36]. To overcome the inherent difficulties in identifying *cis*-regulatory elements in genomic sequence, much effort has been invested in comparative genomic approaches, based on the idea that in closely related species, functional elements will be more conserved than nonfunctional DNA [36–39]. Thus, to date, both gene-specific and genome-wide evaluation of *cis*-regulatory evolution have been limited to the examination of local changes within elements that are otherwise conserved between species. These studies have given rise to the view that *cis*-regulatory evolution operates on existing elements, in which small changes create novel functions [22].

However, there is also evidence that argues against local motif turnover as the only source of *cis*-regulatory evolution. First, although conservation certainly does imply function [1,35,36], it does not necessarily follow that all functional elements must be conserved, nor that nonconserved DNA has no function [2,37,40]. Indeed, it has been shown theoretically that new elements may arise at a certain frequency from nonfunctional sequences [41,42], generating functional elements that reside at nonorthologous positions in the genomes of related species. Consistent with this prediction, a recent genome-wide chromatin immunoprecipitation on chip (ChIP-chip) study in *D. melanogaster* embryos has demonstrated that many transcription factor binding sites are not evolutionarily conserved, suggesting that comparative genomics has limited ability to identify true functional *cis*-regulatory elements [2].

By definition, computational approaches based on genome alignment alone cannot identify *cis*-regulatory elements whose sequence and genomic position is not conserved. Thus, with this approach, it has not been possible to evaluate any aspect of *cis*-regulatory evolution beyond local motif turnover. An alternative means to ascertain whether more radical types of *cis*-regulatory evolution do indeed occur would be to begin by analysing single genomes using computational prediction tools, and subsequently, to compare results across several genomes. Since all computational predictions are prone to false-positive and false-negative results, an essential final step would be to validate predictions experimentally.

In this paper, we use a combination of alignment-independent prediction of *cis*-regulatory elements [18,19], comparative genomics, and experimental validation to examine *cis*-regulatory evolution beyond motif turnover for PREs in four *Drosophila* species. This analysis shows that PRE evolution is extraordinarily dynamic. We show both computationally and experimentally that the numbers of PRE elements, their motif composition, and their genomic position change rapidly in evolution. We identify at least two classes of PREs: those whose positions are constrained in evolution (such as the hox PREs), and those that do not have constrained positions. Remarkably, despite the general conservation of the hox PREs, we identify an extra functional PRE in the Bithorax complex of *D. pseudoobscura*. By demonstrating that PRE evolution is not limited to the adaptation of preexisting elements, these findings document a novel dimension of *cis*-regulatory evolution. The implications of these findings for evolutionary diversity are discussed.

Results

The DNA Sequence Criteria for PRE Function Are Essentially Identical in Four *Drosophila* Species

We have previously developed an algorithm that predicts PREs in the genome of *D. melanogaster* by scoring for favoured pairs of binding sites for proteins that act on them [18,19]. In [18], 43 predicted PREs were selected for experimental analysis; 29 of these were enriched for PcG proteins in ChIP experiments in S2 cells. A further 12 of those 14 sites that were not enriched in [18] were found to be strongly enriched

for PcG proteins in other cell types, or were confirmed in transgenic assays [10,11,18]. Thus, over 95% of these 43 predictions were functional in one cell type or another, confirming the predictive power of the algorithm for correctly identifying PRE elements.

Comparison of the full set of 167 predictions [18] with genome-wide binding profiles of PcG proteins performed in different cell types or in embryos [9–11] revealed a partial overlap. Using the most statistically stringent score cutoff (a score of 157, corresponding to an *E*-value, or expected number of false positives, of 1.0), PREs were correctly predicted at 20% (37 of 186) of experimentally defined binding sites in Sg4 cells [10]. Lower score cutoffs gave higher coverage of ChIP sites [8]; however, it is not clear how many of the detected PcG binding sites in [10] contain functional PREs. Indeed, a recent ChIP-chip analysis of transcriptional regulators in *Drosophila* embryos demonstrated that many detected binding sites appear not to be functional [2]. In addition, we predict many PREs at sites at which no ChIP enrichment was observed [10]. These include, for example, the well-characterised *Fab-7* PRE [43]. For a selection of these predicted sites, ChIP in other cell types (9/12 positive) and transgene analysis (3/3 positive) have confirmed that they are indeed bona fide PRE elements and not false-positive predictions [9,18].

The fact that these predicted and verified PREs were not all enriched in any one cell type is consistent with the partial overlap observed between three recent genome-wide Polycomb binding profiles (28% to 34%) generated by ChIP or DNA adenine methyltransferase mapping (DamID) on different *D. melanogaster* cell types [8–11,15]. Other studies have also observed discrepancies between genome-wide ChIP data and conserved *cis*-regulatory elements identified by comparative genomics [2,36]. Together, these comparisons show that neither ChIP nor computational analysis provides a comprehensive list of all *cis*-regulatory elements in the genome: computational analysis can identify sites of potential function, whereas ChIP gives a measure of cell-type– or developmental-stage–specific deployment of these elements. For this reason, in the present study, we combine computational prediction of PRE elements with ChIP and transgenic analysis of specific loci.

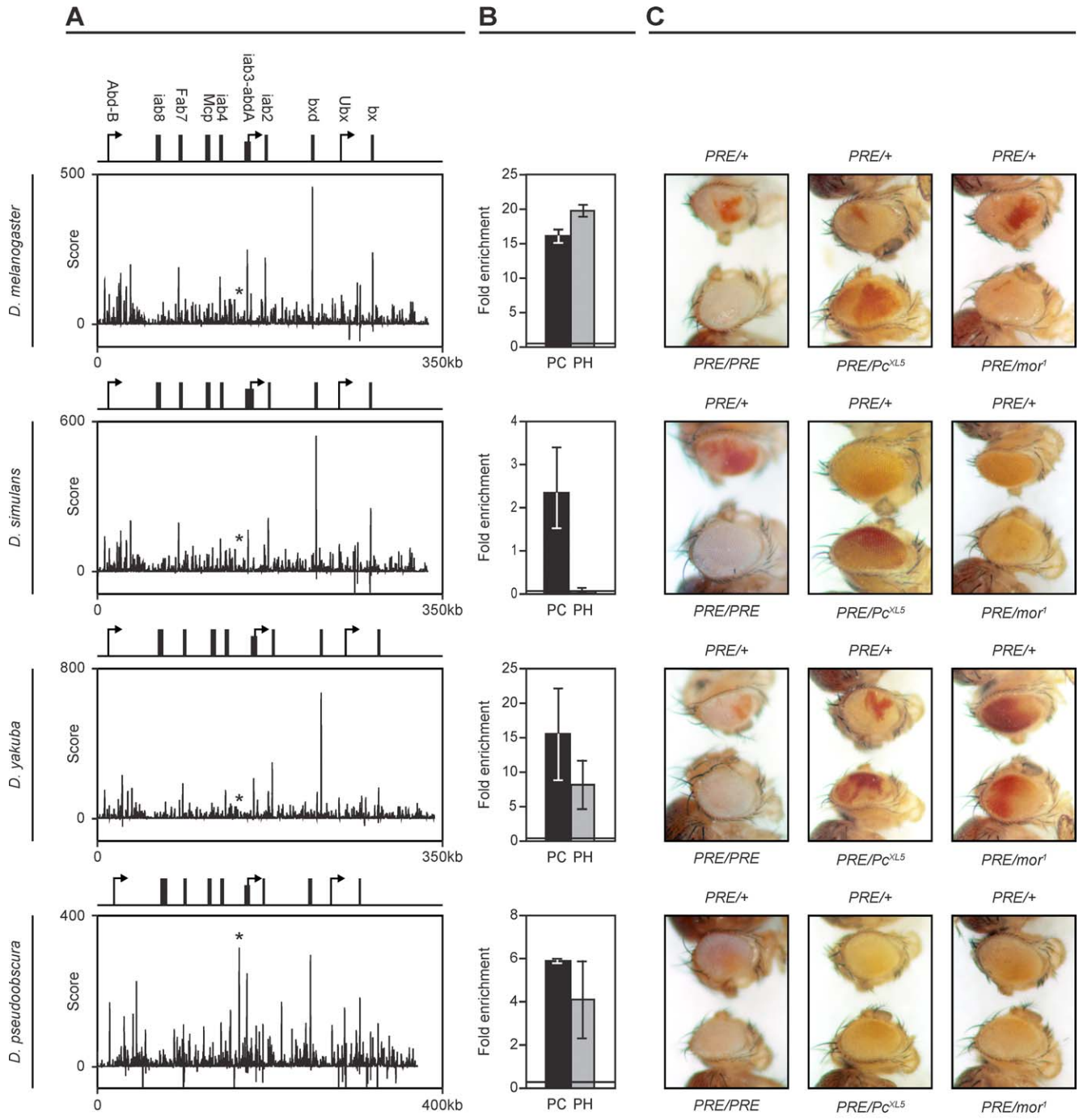
To assess the evolutionary behaviour of PREs independent of genome alignment, we applied the algorithm to four *Drosophila* genomes: *D. melanogaster*, *D. simulans*, *D. yakuba*, and *D. pseudoobscura*. The algorithm was trained on *D. melanogaster* PRE sequences. Its performance on other *Drosophila* genomes was confirmed by comparison of PRE predictions in the homeotic Bithorax complexes of all four species, showing that well-characterised PREs in *D. melanogaster* are also predicted with high significance at orthologous sites in the three other genomes (Figure 1A). In addition, antibodies raised against *D. melanogaster* PcG proteins were confirmed in the other three species by western blot (Figure 2F) and were used for ChIP. This analysis showed that PcG proteins were enriched on the predicted PREs of the Bithorax complex in embryos of all four species (Figures 1B and S1, and unpublished data). Interestingly, Polycomb protein (PC) and Polyhomeotic protein (PH) were detected at similar levels on the *bxd* PRE in *D. melanogaster*, but at different levels on the *bxd* PRE in the other species (Figure 1B). Similar behaviour was also detected in other ChIP experiments (Figures 3C, 4B, and 4D). It is unlikely that these differences arise from different antibody affinities in the different species, because both the PC and PH antibodies gave essentially identical results in western blots on embryonic extracts of the four species (Figure 2F). Furthermore, the differences in ChIP enrichments are not consistently higher for a given antibody or species (see, for example, Figure 4B and 4D). We reason that these differences may arise from the fact that we used embryos for the ChIP experiments. The ChIP results represent an average of binding levels for a mixture of cell types, and a range of embryonic stages from 0–16 h. We observed that embryonic development in the four species proceeds at slightly different rates, which would affect the distribution of embryonic stages in a 0–16-h collection, and may therefore affect the observed binding levels of PC and PH. Alternatively, the different binding of PC and PH may reflect different species-specific compositions of PcG complexes at different PREs.

In order to measure PRE function by independent means, we used a transgenic reporter assay in which a PRE sequence is linked to the *miniwhite* gene. The predicted *bxd* PRE (Figure 1A) from all four species showed typical PRE behaviour in this assay in *D. melanogaster*, giving pairing-sensitive repression and variegation of *miniwhite*, and response to *PcG* and *trxG* mutations (Figure 1C and 1D). Taken together, these results indicate that the DNA sequence criteria for PRE function are essentially identical in all four species, and that the *D. melanogaster* PRE prediction algorithm is applicable to the other three genomes examined here.

A Dynamic Scoring System Increases the Sensitivity of PRE Prediction by Using Comparative Genomic Information

For PRE prediction in a single genome, we previously used a stringent score cutoff of 157, corresponding to an *E*-value (expected number of false positives) of 1.0 [18]. This emphasis on specificity had costs for sensitivity: with a score cutoff at 157, only 20% of sites identified by a later ChIP study were predicted [10]. Aiming to improve sensitivity without costs for specificity, we took steps to adapt the algorithm. We first tested binding sites for other proteins such as DSP1 (FBgn0011764) [44], GRH (FBgn0259211) [45], and SP1/KLF (FBgn0020378; FBgn0040765) [46]. However, the inclusion of these sites did not improve the predictive power of the algorithm, but merely lowered the stringency (M. Rehmsmeier, T. Fiedler, and A. Hauenschild, unpublished data). The original motif set [18] was thus used for further experiments.

We reasoned that the inclusion of comparative genomic data could increase the predictive power of the algorithm. The presence of a high-scoring hit at an orthologous or close position in a second genome would increase statistical confidence. Thus, for the present study, we employed this principle to calculate a sliding scale of score thresholds (Figure 2A; Materials and Methods) which in effect gives a bonus to low-scoring predictions in one genome that are close to high-scoring predictions in another genome. This indeed improved the predictive power of the algorithm, without costs for specificity. At an *E*-value of 1.0, the overlap between *D. melanogaster* predictions and published ChIP data [10] was increased from 20% to 34%. In summary, this "dynamic" scoring system increases the sensitivity of the algorithm by taking account of comparative genomic information, but does not exclude elements that occur at nonconserved positions.



D

transgene	PSS	var
Dm bxd	2/3	2/4
Ds bxd	2/3	3/4
Dy bxd	3/7	4/8
Dp bxd	3/3	2/3

Figure 1. Validation of PRE Prediction in Four Species

(A) Score plots of PRE predictions in the Bithorax complex of four *Drosophila* species (genome versions used in this work are: *D. melanogaster*: 4.0, *D. simulans*: 1.0, *D. yakuba*: 1.0, and *D. pseudoobscura*: 2.0). Coordinates of sequences used plotted from left to right of the figure are as follows: *D. melanogaster*: 12807958–12472096, *D. simulans*: 8665200–8999760, *D. yakuba*: 12260626–12603034, and *D. pseudoobscura*: 731941–359993. Positions of experimentally verified PREs and homeotic gene promoters are shown above each plot. Asterisks show the region of *iab3* analysed in Figure 4E to 4G. (*D. melanogaster* position 12663000).

(B) ChIP analysis of Polycomb (PC; FBgn0003042) and Polyhomeotic (PH; FBgn0004861) enrichments on the *bxd* PRE in embryos of four species. Region analysed is shown in Figure 3A. Error bars indicate standard deviation. Horizontal lines represent mean enrichments of negative control fragments that were present at detectable levels in all samples.

(C) Transgenic reporter assay for PRE function. All transgenic PREs comprise 1.6 kb centred on the region shown in Figure 3A. Top row: the *D. melanogaster bxd* PRE [4] was cloned upstream of the *miniwhite* reporter gene. As previously reported, the eyes of transgenic flies show variegation, pairing-sensitive silencing (left panel), loss of silencing in a *PcG* mutant background (middle panel), and loss of activation in a *trxG* mutant background (right panel). Bottom three rows: *miniwhite* reporter constructs containing 1.6 kb of *D. simulans*, *D. yakuba*, and *D. pseudoobscura* sequences orthologous to the *D. melanogaster bxd* PRE were injected into *D. melanogaster* embryos. All show behaviour similar to the *D. melanogaster bxd* PRE. Several independent lines were analysed for each construct. Each photograph shows the line that displayed the strongest effect.

(D) The number of total lines tested that showed a given behaviour are listed. PSS, pairing-sensitive silencing; var, variegation. For *Dm*, *Ds*, and *Dp*, few lines were recovered, due to strong silencing of the *miniwhite* transformation marker by the *bxd* PRE.

doi:10.1371/journal.pbio.0060261.g001

D. pseudoobscura Has More PREs Than Three Other Species

Using this approach, we performed PRE predictions on the four genomes in all possible pairwise combinations. In each search, the starting point was a set of predictions that scored highly (above 157) in a single genome. Remarkably, in these single-genome analyses, the number of predicted PREs in *D. pseudoobscura* (560) was over twice that predicted in any of the other species (*D. melanogaster*: 201, *D. simulans*: 143, and *D. yakuba*: 203), despite almost identical genome size [35]. To evaluate interspecies differences in PRE number by independent experimental means, we examined the distribution of PC by immunofluorescence on polytene chromosomes prepared from third instar larvae of the four species. This analysis detected over twice as many PC bands in *D. pseudoobscura* as in the other three species, consistent with the prediction of over twice as many PREs (Figure 2C and 2E). Essentially identical results were obtained for PH (unpublished data). The anti-PC and -PH antibodies were raised against the *D. melanogaster* proteins, but detected the PC and PH proteins equally efficiently in a western blot of all three other species (Figure 2F). Nevertheless, to confirm that differences in band number do not reflect differences in antibody behaviour, polytene stainings were also performed with antibodies against histone H3 trimethylated at lysine 27 (H3K27me3; Figure 2D and 2E). This epitope is identical in all four species and is a hallmark of PcG action that is conserved from flies to vertebrates [8]. The band numbers, calculated from analysis of multiple chromosome spreads, were similar for PC and H3K27me3 in all four species, and were consistently approximately twice as high in *D. pseudoobscura* as in the other three species (Figure 2E). This analysis confirms that in salivary glands, *D. pseudoobscura* has at least twice as many binding sites for PC protein as any other species, and is consistent with the results of the prediction. To ascertain how many genes with predicted PREs *D. melanogaster* and *D. pseudoobscura* have in common, we compared genes that were in the neighbourhood of the PRE, not farther than 10 kb from its closest end. Including PREs predicted with a fixed genome-wide cutoff and PREs predicted with our dynamic scoring scheme, we thus determined 166 genes unique to *D. melanogaster*, 349 genes unique to *D. pseudoobscura*, and 112 genes common to both species. This indicates that not only the numbers of PREs differ between *D. pseudoobscura* and the *melanogaster* subgroup, but also the identities of the genes they regulate.

The Genomic Position of PREs Is Predicted to Change Rapidly in Evolution

Despite these differences in PRE number, we expected that a large proportion of PREs would have conserved genomic position. To ascertain whether this is indeed the case, we compared each predicted PRE in a given genome to its nearest counterpart, identified by dynamic scoring in a second genome. For each PRE hit in the first genome, a BLAST search was performed on the second genome, and the distance between the BLAST hit and the nearest statistically significant PRE was calculated (Figure 2A). Figure 2G shows the distribution of these distances for *D. melanogaster* versus *D. yakuba* (triangles) and for *D. melanogaster* versus *D. pseudoobscura* (squares). Surprisingly, despite the statistical bonus given to PRE pairs with conserved position, this analysis predicts that many PREs do not have conserved position. For example, in the *D. melanogaster–D. yakuba* comparison, although approximately 140 PRE pairs are within 1 kb of each other, we predict 30 pairs that are separated by over 10 kb. In the *D. melanogaster–D. pseudoobscura* comparison, PRE positions are less conserved still, with approximately 80 pairs within 1 kb, and approximately 80 that are over 10 kb apart in the two genomes. The PREs that have the highest conservation of position in all four genomes are listed in Table S1. In summary, these data predict that there are at least two classes of PRE elements: those whose positions are evolutionarily constrained, and those whose positions change rapidly in evolution.

PREs with Constrained Position Show Motif Turnover

To test these predictions experimentally, we performed ChIP on embryos from all four species to evaluate binding of PcG proteins to predicted PRE sites in vivo. We focused on specific examples of two classes of predicted PRE: those that have conserved position, and those that do not. For PREs with conserved position, we selected *bxd* and *spalt major* (*salm*; FBgn0004579) as examples of PREs that have been confirmed in *D. melanogaster* [4,10,47]. ChIP analysis in embryos from all four species demonstrated robust PcG binding to these predicted PREs (*bxd*, Figure 1B; *salm*, Figure 3C), indicating that these sites do indeed have PRE function in all four species. In the case of the *salm* PRE, the *D. pseudoobscura* prediction has a score that is significant only in the context of the double-genome search, and would not have been retrieved in a search of the *D. pseudoobscura* genome alone, demonstrating the value of the dynamic scoring system.

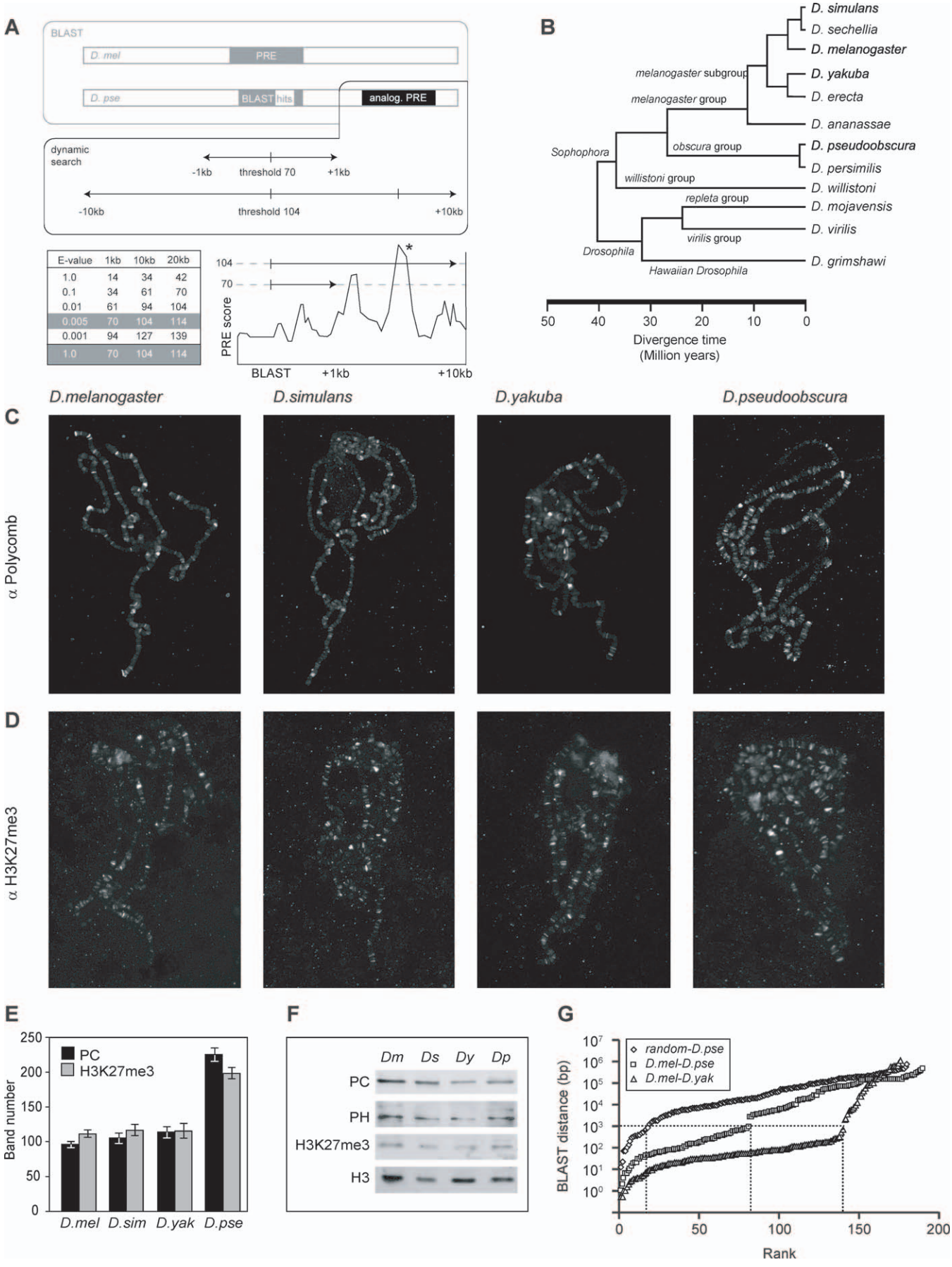

A
BLAST
D. mel
PRE
D. pse
BLAST hits
analog. PRE
dynamic search
-1kb
threshold 70
+1kb
-10kb
threshold 104
+10kb
E-value 1kb 10kb 20kb
1.0 14 34 42
0.1 34 61 70
0.01 61 94 104
0.005 70 104 114
0.001 94 127 139
1.0 70 104 114
104
70
PRE score
BLAST
+1kb
+10kb
B
D. simulans
D. sechellia
D. melanogaster
D. yakuba
D. erecta
D. ananassae
D. pseudoobscura
D. persimilis
D. willistoni
D. mojavensis
D. virilis
D. grimshawi
melanogaster subgroup
melanogaster group
obscura group
willistoni group
repleta group
virilis group
Sophophora
Drosophila
Hawaiian Drosophila
50 40 30 20 10 0
Divergence time (Million years)
C
D.melanogaster
D.simulans
D.yakuba
D.pseudoobscura
α Polycomb
D
α H3K27me3
E
PC
H3K27me3
Band number
250
200
150
100
50
0
D.mel D.sim D.yak D.pse
F
Dm Ds Dy Dp
PC
PH
H3K27me3
H3
G
random-D.pse
D.mel-D.pse
D.mel-D.yak
BLAST distance (bp)
Rank
0 50 100 150 200

Figure 2. Differences in Number and Genomic Position of PREs between *Drosophila* Species

(A) Dynamic scoring system for PRE predictions in pairs of genomes. Top panel: for each predicted PRE from one species, its orthologous region in another species, if present, is determined by BLAST search (see Materials and Methods). Around this orthologous region, PREs are predicted in sequence areas of increasing sizes (1 kb, 10 kb, 20 kb, and chromosome-wide) and with increasing score cutoffs (70, 104, 114, and 157). The diagram shows 1 kb and 10 kb as examples. Score plot, bottom right: PRE scores are calculated around the orthologous region (marked "BLAST"), and the closest region that scores above the relevant threshold is taken as the putative functional analog of the original PRE. In the example shown, no PRE scoring over 70 is found within 1 kb of the BLAST site. The search is extended to 10 kb, and the closest PRE scoring over 104 (asterisk) is taken. Table, bottom left: score cutoffs increase with increasing search radii (columns) and decreasing *E*-values (rows). The cutoffs of 70, 104, and 114 correspond to an *E*-value of 0.005 per search, such that 200 searches (about the number of PREs predicted in a genome-wide scan in *D. melanogaster*) correspond to an overall *E*-value of 1.0 (bottom row of table).

(B) Phylogenetic tree showing divergence times between *Drosophila* species. Adapted from http://flybase.bio.indiana.edu/static_pages/species/muller_synteny.html (C), Immunostaining with anti-PC antibody and (D) anti-H3K27me3 antibody, on polytene chromosomes from four species.

(E) Average band numbers from polytene chromosomes stained as in (C and D). Seven to ten genomes were counted per species and per antibody. Error bars show standard deviation.

(F) Western blot with PC, PH, and H3K27me3 antibodies on equivalent quantities of protein extract from embryos of four species: *D. melanogaster* (*Dm*), *D. pseudoobscura* (*Dp*), *D. simulans* (*Ds*), and *D. yakuba* (*Dy*). Anti-histone H3 (FBgn0001199) is shown as loading control.

(G) Distances between orthologous regions and predicted analogs (see Figure 2A). Triangles: genome-wide–predicted *D. melanogaster* PREs versus *D. yakuba* analogs. Boxes: genome-wide–predicted *D. melanogaster* PREs versus *D. pseudoobscura* analogs. Diamonds: 1-kb sequences randomly chosen from the *D. melanogaster* genome versus *D. pseudoobscura* PREs. The numbers of random sequences on each chromosome equal the numbers of PREs on that chromosome. Horizontal dotted line indicates a BLAST distance of 1 kb. Vertical dotted lines indicate the number of PRE pairs in each category that have a BLAST distance of 1 kb.

doi:10.1371/journal.pbio.0060261.g002

The *bxd* and *salm* PREs reside in orthologous regions in all four genomes, enabling us to ask whether the motifs that contribute to PRE function are located in the regions of highest conservation [20]. Unexpectedly, this was not the case (Figure 3A and 3D). The highest conserved regions (dark-grey boxes on *D. melanogaster* and *D. pseudoobscura* diagrams in Figure 3A and 3D) are typically devoid of PRE motifs. We examined other PREs that have conserved positions, and all showed a similar clustering of motifs in the less conserved regions (Figure S1, Table S1, and unpublished data). Where minimal functional PRE fragments have been defined [48–51], these do not map to the sites of highest conservation (Figures 3A and S1, and unpublished data). This raises the question of whether these highly conserved regions are important for other functions. Although specific roles have not been reported for these sequences, they may contain promoter targeting sequences, boundary elements, or specific enhancers. Alternatively, they may contain unidentified motifs that are important for endogenous PRE function, but that are not required for minimal PRE function in reporter assays [7].

Furthermore, although each PRE has one or more clusters of motifs, the position and order of motifs within the cluster is not conserved. This is most striking in the *D. melanogaster–D. pseudoobscura* comparison (red motifs, Figure 3A and 3D), but is also true to a lesser extent for pairs of PREs in more closely related species (*D. melanogaster*, *D. simulans*, and *D. yakuba*; Figure 3A and 3D). Other PREs that have conserved positions showed similar motif turnover (Figure S1, Table S1, and unpublished data). This rapid evolutionary turnover of motifs in PREs has been noted for the *bxd* PRE [7] and is similar to the turnover that has been observed in enhancer and promoter sequences [2,26,29,37,38], which suggests that motif turnover is a general feature of many classes of regulatory elements.

PREs at the *trh* and *dpp* Loci Have Changed Position during Evolution

We next selected examples of PREs that are predicted not to have conserved position, and used ChIP and transgenic assays to evaluate PRE function of the orthologous and nonorthologous sequences within selected loci. For this analysis, the *trachealess* (*trh*; FBgn0003749), *decapentaplegic* (*dpp*, FBgn0000490), and *abdominal-A* (*abd-A*; FBgn0000014) loci were selected (Figure 4). At the *trh* locus, a PRE is predicted close to the promoter in the three most closely related species, *D. melanogaster*, *D. simulans*, and *D. yakuba* (Figure 4A, top three panels, site 2). This predicted PRE was also robustly bound by PcG proteins in embryos of these three species (Figure 4B, site 2). However, in *D. pseudoobscura*, although the *trh* coding region is well conserved, no PRE was predicted at the promoter (Figure 4A, bottom panel, site 2). Consistent with this prediction, ChIP analysis showed only moderate enrichment for PcG proteins at this site (Figure 4B, site 2). Instead, the strongest PRE prediction in the *D. pseudoobscura trh* locus is within the second intron (Figure 4A, site 1). Higher PcG enrichment at this intronic PRE than at the promoter site was detected in *D. pseudoobscura*, whereas this site was less enriched than the promoter site in the other three species (Figure 4B, site 1). This analysis suggests that whereas in *D. melanogaster*, *D. simulans*, and *D. yakuba*, the main site of PRE function is at the promoter, in *D. pseudoobscura*, PRE function is situated at the intron site some 5 kb away.

For *dpp*, the situation is more complex: there are three predicted PRE sites, which have different scores in different species. Site 1 is approximately 12 kb upstream of the *dpp* promoter, site 2 is 5 kb upstream, and site 3 is at the promoter (Figure 4C). In *D. melanogaster* and *D. simulans*, the predicted PRE score and the enrichment for PcG proteins at site 1 are higher than at sites 2 and 3 (Figure 4C and 4D, top two panels). Of the three sites in *D. yakuba*, site 2 has the highest PRE score and showed the highest PcG enrichment. In *D. pseudoobscura*, the highest PRE prediction is at site 3 (the promoter site, Figure 4C). This site is bound by PcG proteins in *D. pseudoobscura*, but no binding above background was detected in the other three species (Figure 4D). Taken together, these results indicate that, like those of the *trh* locus, the *dpp* PREs are at different sites in different species, suggesting that gain or loss of PRE function at orthologous sites has occurred during evolution.

The *D. pseudoobscura* Bithorax Complex Contains an Additional PRE

In several cases, a PRE was predicted in one species, but had no detectable counterpart in other species. Two such examples are shown in Figure S2 (in the *unpaired 2* locus) and in Figure 4E–4H (in the Bithorax complex). The Bithorax

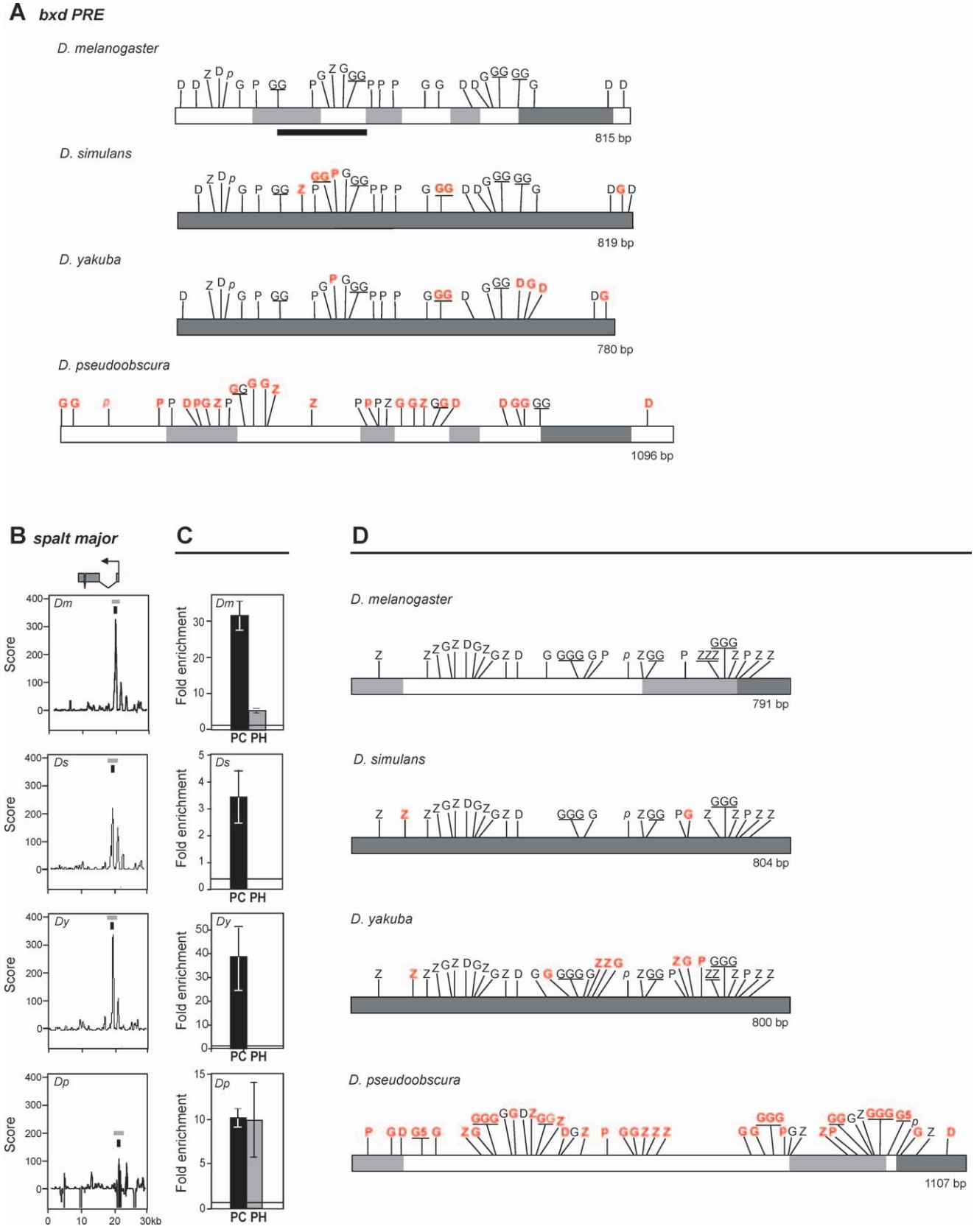


Figure 3. PREs with Conserved Position Show Motif Turnover

(A) *bxd* PRE. Motif occurrence is independent of sequence conservation. The core *D. melanogaster bxd* PRE and the orthologous regions from the other three species are shown. Coordinates of sequences shown from left to right of the figure are as follows: *D. melanogaster*:12590182–12589368, *D. simulans*: 8886059–8886877, *D. yakuba*: 12487238–12488017, and *D. pseudoobscura*: 485021–483926. Black bar below *D. melanogaster* diagram indicates minimal PRE fragment [48]. Conservation between *D. melanogaster* and *D. pseudoobscura* is marked on the diagrams for these two species:

Dark grey: regions of over 70% identity. Light grey: 50%–70% identity. *D. simulans* and *D. yakuba* conservation to *D. melanogaster* is indicated (the *D. melanogaster, D. simulans*, and *D. yakuba* sequences are over 90% identical). Motif positions are indicated above the figure. Motifs shown in red on *D. simulans*, *D. yakuba*, and *D. pseudoobscura* are not present in the *D. melanogaster* PRE. D, Dsp1 [44]; G, GAF (FBgn0013263); P, PHO extended site (PF or PM as in [18]; *p*, PHO core site (GCCAT) (FBgn0002521); Z, Zeste (FBgn0004050). Underlined motifs indicate overlapping runs of motif separated by two bases. G5 indicates five consecutive GAs.

(B) PRE prediction score plots for *spalt major* (*salm*) PRE at orthologous regions of the four genomes. The *salm* transcription unit is indicated. Grey bars at the top of each score plot indicate the regions shown in detail in (D). Black boxes indicate PCR fragments used for real time PCR detection in ChIP analysis in (C).

(C) ChIP enrichments at *salm* PRE in embryos of four species (see also legend to Figure 1).

(D) Motif occurrences in *salm* PREs, annotation as in (A). Coordinates of sequences shown from left to right of the figure are as follows: *D. melanogaster*: 11446402–11445612, *D. simulans*: 11255943–11256747, *D. yakuba*: 7893940–7893140, and *D. pseudoobscura*: 6845260–6844154 .

doi:10.1371/journal.pbio.0060261.g003

complex of *D. melanogaster* contains the best-characterised PREs, which act to maintain expression domains of the three hox genes *Abdominal-B* (*Abd-B*; FBgn0000015), *abd-A*, and *Ubx*. In all four species examined, the PREs of the Bithorax complex were predicted at well-conserved positions (Figure 1A), with one notable exception: an extra PRE 10 kb upstream of *abd-A* is predicted in *D. pseudoobscura* (Figure 1A, bottom panel, asterisks; Figure 4E). Strikingly, the orthologous sequences in *D. melanogaster*, *D. simulans*, and *D. yakuba* have PRE scores of less than 20 (Figure 1A, top three panels, asterisks) and have very few PRE motifs (Figure 4E).

The predicted extra *D. pseudoobscura* PRE was bound by PcG proteins in *D. pseudoobscura* embryos (Figure 4F, top panel), indicating that it may indeed be a functional element. The orthologous sequences showed no detectable PcG binding in any of the other species, suggesting that this element does not function as a PRE in *D. melanogaster*, *D. simulans*, or *D. yakuba* (Figure 4F, bottom three panels). To test these observations by independent means, we generated transgenic reporter flies carrying either the predicted *D. pseudoobscura* PRE or the orthologous *D. melanogaster* sequence (Figure 4G and 4H). Whereas the *D. melanogaster* sequence did not show any typical PRE behaviour in this assay, the *D. pseudoobscura* element showed pairing-sensitive silencing, variegation, and response to *PcG* and *trxG* mutations (Figure 4G and 4H), all typical features of PRE elements [4,52,53]. Thus, we conclude that this extra *D. pseudoobscura* element is indeed a functional PRE.

Genome-Wide Comparisons Predict That PREs Can Arise from Nonfunctional Sequence

The presence of an additional functional PRE in the *D. pseudoobscura* Bithorax complex is intriguing, particularly since the positions of other PREs at this locus are so well conserved. This PRE may be a remnant of an ancestral Bithorax complex, which has lost the PRE at that position in some lineages. Alternatively, the *D. pseudoobscura* PRE may have arisen from nonfunctional sequence and been fixed by positive selection. To evaluate these two possibilities, PRE scores were calculated for the orthologous sequences at this position in eight *Drosophila* genomes [35]. This analysis showed a statistically significant PRE score for this site in *D. ananassae*, *D. pseudoobscura*, and *D. persimilis*, but not in the *melanogaster* subgroup. A maximum likelihood analysis suggests that the PRE was present in the common ancestor of the species under consideration and was lost in the *melanogaster* subgroup (Figure S3). To gain further insight into global gain and loss of PREs during the evolution of the *D. melanogaster* lineage, we carried out genome-wide comparisons with eight genomes as described in Materials and Methods. From this analysis, it can be inferred that 33 PREs have been gained in *D. melanogaster* (Figure S4 and Table S2). For only one of these 33 PREs, the nearest gene, *scribbled* (*scrib*; FBgn0026178), has another PRE, and gene *CG12852* (FBgn0085383) has gained two PREs, without having a further one. Thus, 30 of these PREs are associated with genes that previously had no PRE. Taken together, these data indicate that PREs can arise from nonfunctional sequence, and furthermore suggest that genes can newly acquire PcG regulation.

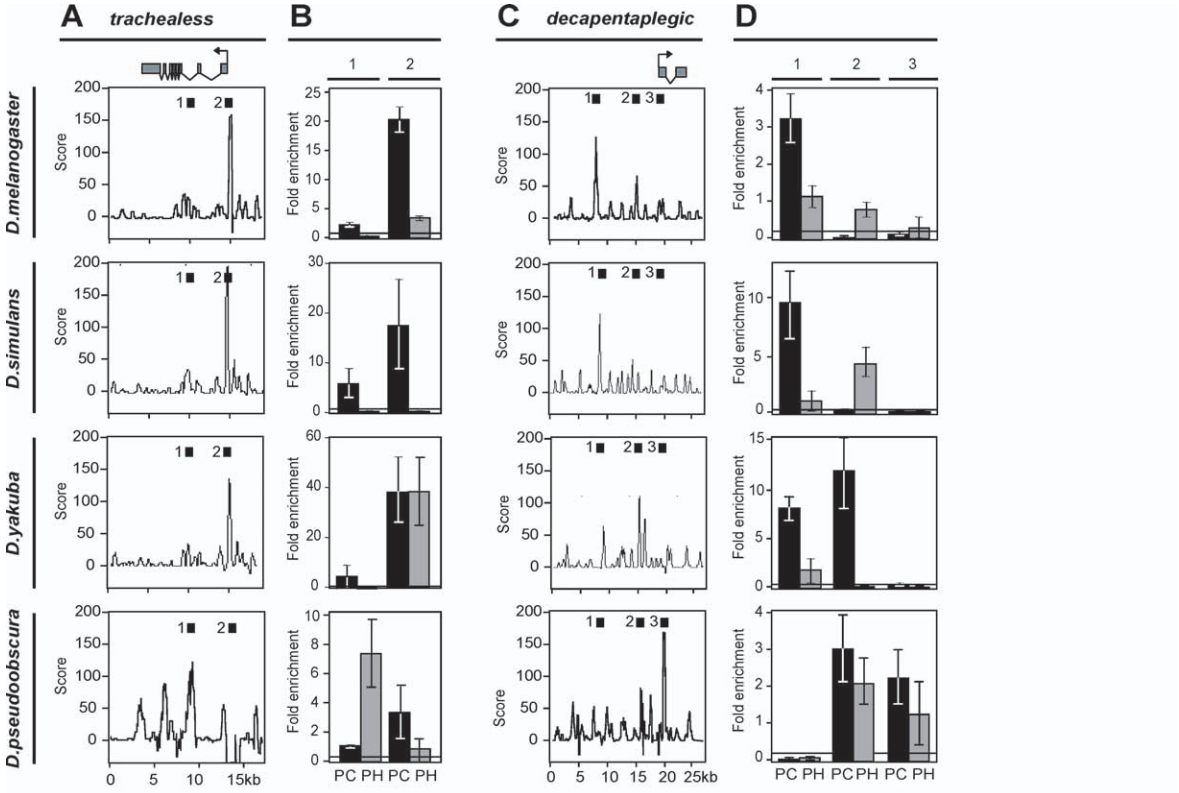
Discussion

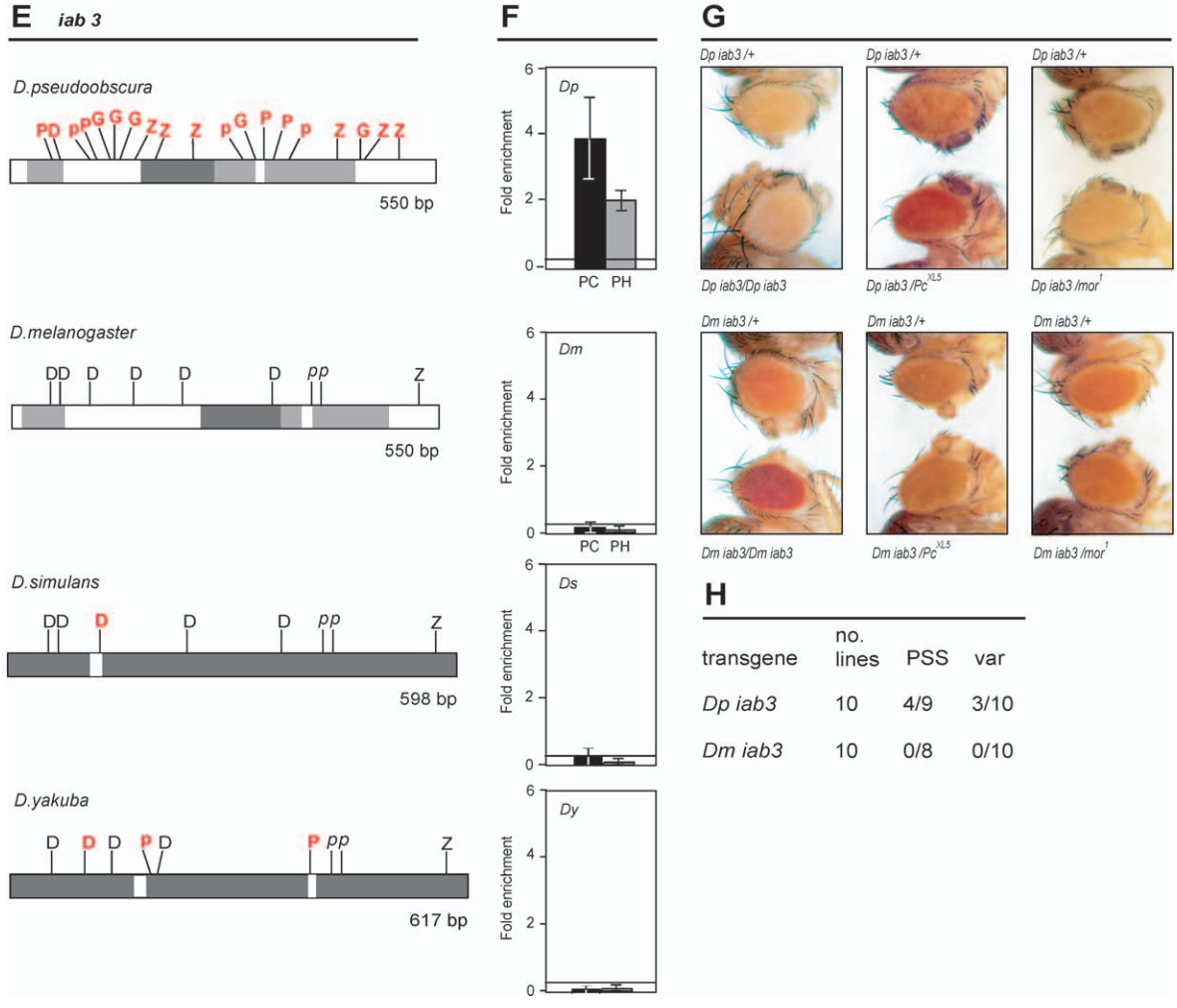
By using predictive methods that identify *Drosophila* PREs independent of their genomic position, in combination with experimental validation at selected loci, we document three kinds of evolutionary plasticity: the numbers of PRE elements, their motif composition, and their genomic position all change rapidly in evolution. By demonstrating that PRE evolution is not limited to the adaptation of preexisting elements [22], these findings document a novel dimension of *cis*-regulatory evolution.

How Do PREs Change Position?

For the PREs that have changed position, there are several possible mechanisms by which a PRE may be lost from one site and gained at another, all of which may be at play in shifting the PRE landscape between species. For example, PREs may move by a simple microinversion event [54]. However, the evolutionary plasticity that we document here mainly involves the loss or gain of PRE function from orthologous sequences that do not contain inversions, thus other mechanisms must be considered. First, PREs may move by "creeping" from one site to the other. In this model, a sequence adjacent to a PRE may acquire new functional motifs, thus shifting the centre of PRE function to a slightly different location. By accumulation of such small shifts, the PRE could effectively move to a new position. Sequence insertions could accelerate this process. We observe such an insertion in the *salm* PRE (Figure 3C and 3F), in which a single motif cluster spanning approximately 600 bp in the *D. melanogaster* PRE has split into two clusters in *D. pseudoobscura*, which are separated by an insertion of a few hundred base pairs.

Second, ancestral PREs may lose their function at different sites in different lineages, resulting in an apparent change of position. Third, a PRE could change its position by de novo evolution from nonfunctional sequence. We infer from comparative genomics that this is the case for at least 35 PREs in *D. melanogaster*. It has been shown theoretically that enhancers could evolve rapidly from nonfunctional sequence, provided that the DNA motifs are simple, and that there is sufficient raw material in the form of "presites" that differ from functional sites by a single nucleotide [41]. This suggests





H

transgene	no. lines	PSS	var
Dp iab3	10	4/9	3/10
Dm iab3	10	0/8	0/10

Figure 4. PRE Function at Nonorthologous Regions

(A and B) *trachealess* (*trh*). (A) PRE prediction score plots for orthologous regions of the *trh* locus in four species. Coordinates of sequences shown from left to right of the figure are as follows: *D. melanogaster*: 362619–382619, *D. simulans*: 8587236–8607236, *D. yakuba*: 368751–388751, and *D. pseudoobscura*: 1070270–1066807. The *trh* transcription unit is shown. Black boxes 1 and 2 at the top of each plot show sites analysed by ChIP in (B). (B) ChIP analysis of PcG enrichments on sites 1 and 2, performed as in Figure 1. PCR primers were designed to detect orthologous regions in all four species for each site.

(C and D) *decapentaplegic* (*dpp*). (C) Score plots, as for (A), with boxes 1, 2, and 3 indicated. Coordinates of sequences shown from left to right of the figure are as follows: *D. melanogaster*: 2454316–2459382, *D. simulans*: 2391870–2418870, *D. yakuba*: 2451500–2478500, and *D. pseudoobscura*: 1720818–1747818. (D) ChIP analysis of sites 1, 2, and 3 as for (B).

(E, F, and G) *D. pseudoobscura* has a functional PRE in the *iab3* region of the Bithorax complex that is absent in the other three species. Asterisks in Figure 1A indicate the region in question. (E) The predicted *D. pseudoobscura* PRE is shown (top). The orthologous region was identified in the three other species by sequence alignment. Coordinates of sequences shown from left to right of the figure are as follows: *D. melanogaster*: 12663856–12663307, *D. simulans*: 8809340–8809937. *D. yakuba*: 12410998–12411614, and *D. pseudoobscura*: 567475–566926. Motifs as in Figure 3A. Motifs that are not present in the *D. melanogaster* sequence are shown in red for the other three species. Conservation between *D. melanogaster* and *D. pseudoobscura* is marked as in Figure 3A. The *D. melanogaster*, *D. simulans*, and *D. yakuba* sequences are over 90% identical. Short insertions in the *D. simulans* and *D. yakuba* sequences with respect to *D. melanogaster* are shown as white boxes.

(F, G, and H) the *D. pseudoobscura* PRE is functional, the orthologous sequences from other species are not. (F) ChIP analysis of PcG enrichments in embryos of four species on the regions shown in (E). (G and H) transgenic reporter assay; 1.6 kb of either the predicted *D. pseudoobscura* PRE or the orthologous region from *D. melanogaster*, centred on the region shown in (E), were cloned upstream of the *miniwhite* reporter gene and injected into *D. melanogaster* embryos. (G) Top row: the *D. pseudoobscura* PRE shows variegation (middle panel, top), pairing-sensitive silencing (left panel), loss of silencing in a *PcG* mutant background (middle panel) and loss of activation in a *trxG* mutant background (right panel). Several independent lines were analysed for each construct. Each photograph shows the line that displayed the strongest effect. (H) Top row: this behaviour was observed in several independent transgenic lines. PSS, pairing-sensitive silencing; var, variegation. (G and H) Bottom rows: the orthologous region from *D. melanogaster* has none of these properties.

doi:10.1371/journal.pbio.0060261.g004

that, as proposed [55], nonfunctional sequences may be "elected" to take up a role as PREs by relatively few nucleotide changes. We have examined this possibility for selected *Drosophila* PREs that occur at nonorthologous positions in different species by allowing single base changes in any motif and plotting sites of "pre-PRE" potential. We find that sites of PRE function in one species correspond to sites of high potential in a second species, so that a new PRE could theoretically emerge with very few nucleotide changes (Figure S5).

Why Do PREs Evolve So Rapidly?

What is the evolutionary significance of PRE plasticity? Many studies of enhancers have shown that small differences in sequence can lead to large phenotypic differences [21,22,28–30], thus one may expect the same to be true for PREs. However, it is important to bear in mind one important functional difference between enhancers and PREs, namely that enhancers respond to differences in cellular concentrations of the transcription factors that bind them, whereas PREs respond to the activity state of their cognate promoter, and not to local differences in the concentrations of the PcG and TrxG proteins [8]. Thus, PREs may be more tolerant than enhancers to changes in number of binding sites, and indeed to changes in the number of PREs at a given locus. On the other hand, the only feature of enhancers that has been studied is motif turnover. It remains to be seen whether enhancers display evolutionary plasticity similar to that of PREs.

Given the flexible nature of PRE design, we envision several possible effects of evolutionary plasticity, which may operate differently at different PREs. First, many differences in PRE number and sequence between species may be tolerated by the organism without causing large phenotypic differences. Indeed, the body plans of the different species are very similar. Thus, some PREs may work to maintain phenotype in the face of environmental differences. For example, one of the most important environmental constraints on different *Drosophila* species from different latitudes is temperature. In *D. melanogaster*, the PcG proteins are profoundly sensitive to the temperature at which the flies are raised [52], giving more potent silencing at higher temperatures. Thus, for some PREs, the plasticity in design that we observe may play a role in "buffering" the system against different temperatures, such that the transcriptional output of the locus is conserved. In addition, PREs may mediate phenotypic plasticity for thermosensitive traits such as pigmentation. Several of the loci involved in the plasticity of pigmentation (e.g., *Abd-B*) are regulated by PREs [56].

On the other hand, for some PREs, differences in design may have a direct effect on phenotype. Several studies have documented large effects on PRE function caused by changes in one or a few binding sites [44,57,58]. Thus, we propose that some of the changes we observe would affect the silencing or activation response of the PRE, thus in turn affecting the level of target gene transcription that is maintained, and giving selectable effects on phenotype. For example, one of the major phenotypic differences between *Drosophila* species is the male sex combs. The sex comb is one of the most rapidly diversifying organs in *Drosophila* species, and is important for male reproductive success [59]. Evolutionary diversity in sex comb number is associated with diversity in regulation of the hox gene *Sex-combs reduced* (*Scr*), which is a well-characterised target of PcG regulation [60,61]. In *D. melanogaster*, *D. simulans*, and *D. yakuba*, a single row of sex comb teeth is present, whereas *D. pseudoobscura* has two such rows. Interestingly, a microinversion event on the 3′ side of the *D. pseudoobscura Scr* locus [54] has removed 3′ regulatory sequences, including one of a cluster of three *Scr* PREs, to a new position. The *D. pseudoobscura* PREs also show many sequence changes compared to the other three species (unpublished data). Thus, differences in PRE sequence, number, and position at the *Scr* locus correlate well with phenotypic differences, and will provide an excellent model for further study of the effects of PRE plasticity on phenotype.

In summary, PREs act on several hundred genes in *Drosophila*, many of which are master developmental regulators. We propose that the extraordinary plasticity in PRE design that we observe may provide a rich capacity for transcriptional buffering, phenotypic plasticity, and phenotypic diversity between species.

Materials and Methods

Bioinformatics methods. *BLAST search.* The BLAST search takes a PRE predicted in one species and determines the orthologous position in another species. Because the PRE will usually not be conserved as a continuous sequence, multiple adjacent high-scoring pairs (HSPs) have to be grouped together. The grouping is done according to the following criteria: only HSPs with a BLAST *E*-value not larger than 0.01 are considered. HSPs of one group are on the same strand. The distance between adjacent HSPs of one group is below 1 kb. Groups are maximal in the sense that no HSPs can be added that fulfil these three criteria. From all groups that correspond to one initial PRE, we choose the one with the largest sum of HSP lengths. From several groups with the same length sum, the one is taken that happens to be the first processed (a case which has not occurred in our analysis so far). Starting with 201 PREs in *D. melanogaster* (version 4.0), this procedure resulted in 190 orthologous regions in *D. pseudoobscura* (version 2.0), 194 in *D. simulans* (version 1.0), and 176 in *D. yakuba* (version 1.0). In *D. yakuba*, an additional 20 fall into "chr2L_random," which contains clones that are not yet finished or cannot be placed with certainty at a specific place on the chromosome. These 20 hits were not included in our analysis.

Finding the right locus. To evaluate the validity of the BLAST search procedure, we checked whether orthologous regions were in correct loci. For each PRE from *D. melanogaster* and its orthologous region in *D. pseudoobscura*, we compared the distance between the PRE and the two genes closest to it with the distance of the orthologous region and the two genes closest to that. If a PRE was inside a gene, only that gene was included into the comparison. In the majority of cases (163 out of 190), this "locus shift" is below 10 kb, although it can become larger than 200 kb. In some cases (24), the ortholog of the *D. melanogaster* PRE and the ortholog of one of the possibly two *D. melanogaster* genes are found on different chromosomes. In general, there are legitimate doubts about the reliability of the *D. pseudoobscura* gene annotation. Frequently, one or more exons are missing, which leads to too large a distance between PRE ortholog and closest gene in *D. pseudoobscura*. Additionally, we can show that the observed rare events of chromosome changes are consistent with the gene rearrangement in the annotation. For example, the gene *CG1924* is located on chromosome X in *D. melanogaster* and on chromosome 2 in *D. pseudoobscura*, whereas the adjacent genes are on chromosome X in both species.

Calculating BLAST distances (Figure 2D). A BLAST distance is calculated as the difference between, first, the distance between the centre of the query sequence (predicted PRE or random) and the centre of the BLAST hit in the query genome (*D. melanogaster*), and second, the distance between the centre of the putative functional analog and the centre of the BLAST hit in the target genome (*D. yakuba* or *D. pseudoobscura*). We cannot directly calculate the distance between BLAST hit and analog in the target genome only, since BLAST hits are not necessarily centred around the query sequence.

PRE prediction and calculation of dynamic scoring thresholds. PRE prediction was performed using the jPREdictor software [19], which follows the PREdictor algorithm as described in [18], except that a step size of 10 bp instead of 100 bp was used. Score cutoffs and *E*-values were calculated with a nonparametric statistics on random sequence data 100 times the size of the *D. melanogaster* genome, with the *D. melanogaster* nucleotide distribution (29% A, 21% C, 21% G, and 29% T). A score *s* such that scores of *s* or better occur *r* times in the random data, corresponds to an *E*-value of *r*/100 in the single *D. melanogaster* genome. For an *E*-value of 1, this score cutoff is 157. For the dynamic scoring system, cutoffs were calculated similarly, taking into account the smaller search spaces of 1 kb, 10 kb, and 20 kb radius and the fact that about 200 such searches are performed (see Figure 2A). All PREs predicted in *D. melanogaster* and *D. pseudoobscura* will be available at http://bibiserv.techfak.uni-bielefeld.de/fly_pres upon publication.

Evolutionary gain and loss of PREs. We performed a maximum likelihood analysis of 73 *D. melanogaster* PREs in eight *Drosophila* genomes. Each of these 73 PREs had been genome-wide predicted, its orthologous regions could be determined in all the other seven species, and at least one of the other species had no functionally analogous PRE. A functionally analogous PRE was defined as a hit predicted dynamically within a 10-kb BLAST distance. The eight species comprise those for which the efficacy of our predictive method has been well established (up to *D. pseudoobscura*). We employed a probabilistic model whose separate gain and loss parameters were estimated with the Mesquite software (http://mesquiteproject.org) on the given contemporary character states: 1 for a (functionally analogous) PRE being present in the respective species, 0 for no such PRE being present. Subsequently, maximum likelihood ancestral character states were reconstructed based on the estimated parameters. Defining a *D. melanogaster* PRE whose most ancestral node (the root of the tree) has a PRE likelihood of smaller than 0.5 as being gained during evolution resulted in 33 such PREs, listed in Table S2. Figure S4 shows the trees for the 73 PREs.

Fly methods. *Strains and handling.* For polytene chromosomes and ChIP, *D. melanogaster* wild-type flies (Oregon R) were used. For the other species, the strains used for whole-genome sequencing were obtained from http://stockcenter.arl.arizona.edu/. Stock numbers: *D. yakuba* 14021-0261.01; *D. simulans* 14021-0251.195; and *D. pseudoobscura* 14011-0121.94. With the exception of *D. pseudoobscura*, all species were raised on cornmeal food. For *D. pseudoobscura*, standard banana-*Opuntia* food was prepared as specified at http://stockcenter.arl.arizona.edu/.

Transgenics. Genomic fragments of 1.5 to 1.6 kb were amplified by PCR from genomic DNA of each species and cloned using SpeI/NotI sites into the pUZ P-element vector upstream of the *miniwhite* reporter gene [18]. Embryo injections were carried out by Vanedis *Drosophila* injection service (http://www.vanedis.no). Chromosomal mapping and crosses to *PcG* and *trxG* mutants were performed as described [18]. Primer sequences, constructs, and transgenic fly lines are available on request.

Polytene chromosome staining. Polytene chromosomes were prepared from third instar larvae of all four species and stained with rabbit polyclonal anti-Polycomb antibody or anti-H3K27me3 (provided by Thomas Jenuwein) as described in [62].

Western blotting. Protein extracts were made from 0–12-h-old embryos for all four species, as described in [63]. Western blots were probed with antibodies against PC, PH, H3K27me3, or H3 (Upstate).

Chromatin immunoprecipitation (ChIP). ChIP on whole embryos of *D. melanogaster*, *D. simulans*, *D. yakuba*, and *D. pseudoobscura* was performed using anti-PC and -PH antibodies, as described [64]. Two independent chromatin preparations on 0–16-h-old embryos, and two to four independent ChIP assays were performed for each species. Enrichments of immunoprecipitated DNA over input DNA were quantified by real-time PCR using SYBR green (Sigma). Three technical replicates were performed for each primer pair on each chromatin preparation. Primers were designed to amplify a fragment of 100 to 300 bp within the highest scoring region of each predicted PRE (or the minimal PRE, if known), or of the orthologous region in the species in which no PRE was predicted. Primer sequences are available on request.

Supporting Information

Figure S1. The *Fab-7* PRE Has Conserved Position and Shows Motif Turnover

(A) PRE prediction score plots for *Fab-7* PRE at orthologous regions of *D. melanogaster* and *D. pseudoobscura* genomes. Grey bars below each score plot indicate the regions shown in detail in (C). Black boxes below plots indicate the position of PCR fragments used for real time PCR detection in ChIP analysis.
(B) ChIP enrichments of PC and PH on *Fab7* PRE in *D. melanogaster* and *D. pseudoobscura* embryos.
(C) Motif occurrence is independent of sequence conservation. The high-scoring region of each PRE is shown. Coordinates of sequences shown from left to right of the figure are as follows: *D. melanogaster*: 12725760–12724576, and *D. pseudoobscura*: 631891–630632. Annotation as for Figure 3. Black bar below *D. melanogaster* diagram indicates minimal PRE fragment [49,50].

Found at doi:10.1371/journal.pbio.0060261.sg001 (376 KB PDF).

Figure S2. A PRE Is Present in *D. melanogaster* but Absent in *D. yakuba* and *D. pseudoobscura* at the *unpaired2* (*upd2*; FBgn0030904) Locus

(A) PRE prediction score plots for *upd2* at orthologous regions of *D. melanogaster*, *D. yakuba*, and *D. pseudoobscura* genomes (the *D. simulans* sequence for this locus is incomplete). Coordinates of sequences shown from left to right of the figure are as follows: *D. melanogaster*: 18081000–18071000, *D. yakuba*: 16868497–16858497, and *D. pseudoobscura*: 6996333–7006333. The *upd2* transcription unit is shown. Black boxes at the top of plots indicate the position of PCR fragments used for real-time PCR detection in ChIP analysis.
(B) PC shows strong ChIP enrichment on predicted *upd2* PRE in *D. melanogaster*, but no detectable enrichment on the orthologous sequences in *D. yakuba* and *D. pseudoobscura* embryos, for which no PRE is predicted.

Found at doi:10.1371/journal.pbio.0060261.sg002 (336 KB PDF).

Figure S3. Evolution of the Extra *D. pseudoobscura* PRE in the *iab3* Region of the BX-C

The *D. pseudoobscura* PRE shown in Figure 4E–4H (*D. pseudoobscura*: positions 567475–566926) was traced in eight *Drosophila* genomes. The highest scores found within a 5-kb radius of the orthologous positions in each species are shown on the right of the diagram. The PRE is absent in the *melanogaster* subgroup (*D. simulans*–*D. erecta*, open circles at the leaves of the tree) and is present in *D. ananassae*, *D. pseudoobscura*, and *D. persimilis* (solid circles at the leaves of the tree). Internal nodes (those that are not leaves) indicate likelihoods of reconstructed character states, with more solid circles representing larger likelihoods of ancestral PREs. The analysis suggests that the PRE was present in the common ancestor and was lost in the *melanogaster* subgroup.

Found at doi:10.1371/journal.pbio.0060261.sg003 (217 KB PDF).

Figure S4. Phylogenetic Trees for 73 PREs in Eight Species

The analysis was performed as described in Materials and Methods. The coordinates of the PRE and name of the closest gene are given. The circular nodes at the leaves (those marked with species names) indicate absence (open circles) or presence (solid circles) of PREs in the respective species. Internal nodes (those that are not leaves) indicate likelihoods of reconstructed character states, with more solid circles representing larger likelihoods of ancestral PREs. Defining a *D. melanogaster* PRE whose most ancestral node (the root of the tree) has a PRE likelihood of smaller than 0.5 as being gained during evolution resulted in 33 such PREs, listed in Table S2.

Found at doi:10.1371/journal.pbio.0060261.sg004 (730 KB PDF).

Figure S5. The *D. pseudoobscura trh* Locus

(A) PREdictor score plot. Site 1 is predicted to be a PRE in *D. pseudoobscura*. Site 2 is not predicted to be a PRE in *D. pseudoobscura*, but in *D. melanogaster* (see Figure 4).
(B) Score plot of PRE potential ("pre-PRE score"; see Discussion). Site 2 shows strong PRE potential, coinciding with the position of the PRE in *D. melanogaster*.

Found at doi:10.1371/journal.pbio.0060261.sg005 (408 KB PDF).

Table S1. PREs Whose Positions Are Conserved in All Four Species

Maxd (column 1) is the maximum distance in base pairs found between the centres of two predicted PREs in any two genomes (see Materials and Methods). The closest annotated *D. melanogaster* gene to each PRE (column 2) and the distance in base pairs from PRE to gene (column 3) are shown. When the PRE is within the coding gene, this distance is given as zero. PREs of the homeotic complexes in *D. melanogaster* are indicated (column 4) using the nomenclature of [18]. In these cases, the gene that has been shown to be regulated by the PRE is shown.

Found at doi:10.1371/journal.pbio.0060261.st001 (49 KB DOC).

Table S2. Gained PREs in *D. melanogaster*.

Gained PREs were identified as described in Materials and Methods. The table lists the 33 PREs that are inferred to have been gained in *D. melanogaster* using a BLAST distance of 10 kb. The coordinates of the PRE are given, and the closest gene and its distance to the PRE are listed. Distances of zero indicate that PRE and gene overlap.

Found at doi:10.1371/journal.pbio.0060261.st002 (89 KB DOC).

Accession Numbers

The FlyBase IDs for the genes and gene products mentioned in this paper are as follows: *abd-A* (FBgn0000014); *Abd-B* (FBgn0000015); CG12852 (FBgn0085383); *dpp* (FBgn0000490); DSP1 (FBgn0011764); GAF (FBgn0013263); GRH (FBgn0259211); H3 (FBgn0001199); KLF (FBgn0040765); PC (FBgn0003042); PH (FBgn0004861); PHO (FBgn0002521); *salm* (FBgn0004579); *scrib (*FBgn0026178); SP1 (FBgn0020378); *trh* (FBgn0003749); *Ubx* (FBgn0003944); *upd 2* (FBgn0030904); and ZESTE (FBgn0004050).

Acknowledgments

The authors thank Betül Hekimoglu, Heidi Ehret, and Ann Mari Voie for experimental assistance, Thomas Fiedler for bioinformatic input, and Thomas Jenuwein for providing anti H3K27me3 antibody.

Author contributions. AH, LR, and MR conceived and designed the experiments. AH performed bioinformatic analysis. CA cloned transgenic constructs. LR performed all other wet-lab experiments. AH, LR, and MR analyzed the data. MR supervised AH. LR supervised CA. RP supervised LR until December 2005. LR and MR wrote the manuscript.

Funding. AH was supported by the International NRW (North Rhine-Westphalia) Graduate School in Bioinformatics and Genome Research, LR and RP by the Deutsche Forschungsgemeinschaft and the European Union Sixth Framework Programme Network of Excellence (EU FP6 NoE) "The Epigenome," LR and CA by the Austrian Academy of Sciences and the EU FP6 NoE "The Epigenome" NET programme, and MR by the Deutsche Forschungsgemeinschaft, Bioinformatics Initiative.

Competing interests. The authors have declared that no competing interests exist.

References

1. Papatsenko D, Levine M (2007) A rationale for the enhanceosome and other evolutionarily constrained enhancers. Curr Biol 17: R955–957.
2. Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, et al. (2008) Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS Biol 6: e27. doi:10.1371/journal.pbio.0060027
3. Simon J, Chiang A, Bender W, Shimell MJ, O'Connor M (1993) Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. Dev Biol 158: 131–144.
4. Chan CS, Rastelli L, Pirrotta V (1994) A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. EMBO J 13: 2553–2564.
5. Ringrose L, Paro R (2004) Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu Rev Genet 38: 413–443.
6. Breiling A, Sessa L, Orlando V (2007) Biology of polycomb and trithorax group proteins. Int Rev Cytol 258: 83–136.
7. Schwartz YB, Pirrotta V (2007) Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet 8: 9–22.
8. Ringrose L, Paro R (2007) Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development 134: 223–232.
9. Negre N, Hennetin J, Sun LV, Lavrov S, Bellis M, et al. (2006) Chromosomal distribution of PcG proteins during Drosophila development. PLoS Biol 4: e170. doi:10.1371/journal.pbio.0040170
10. Schwartz YB, Kahn TG, Nix DA, Li XY, Bourgon R, et al. (2006) Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat Genet 38: 700–705.
11. Tolhuis B, de Wit E, Muijrers I, Teunissen H, Talhout W, et al. (2006) Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. Nat Genet 38: 694–699.
12. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, et al. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125: 301–313.
13. Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K (2006) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev 20: 1123–1136.
14. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, et al. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441: 349–353.
15. Ringrose L (2007) Polycomb comes of age: genome-wide profiling of target sites. Curr Opin Cell Biol 19: 290–297.
16. Lee N, Maurange C, Ringrose L, Paro R. (2005) Suppression of Polycomb group proteins by JNK signalling induces transdetermination in Drosophila imaginal discs. Nature 438: 234–237.
17. Sparmann A, van Lohuizen M (2006) Polycomb silencers control cell fate, development and cancer. Nat Rev Cancer 6: 846–856.
18. Ringrose L, Rehmsmeier M, Dura JM, Paro R (2003) Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster. Dev Cell 5: 759–771.
19. Fiedler T, Rehmsmeier M (2006) jPREdictor: a versatile tool for the prediction of cis-regulatory elements. Nucleic Acids Res 34: W546–550.
20. Dellino GI, Tatout C, Pirrotta V (2002) Extensive conservation of sequences and chromatin structures in the bxd polycomb response element among Drosophilid species. Int J Dev Biol 46: 133–141.
21. Wray GA (2007) The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8: 206–216.
22. Prud'homme B, Gompel N, Carroll SB (2007) Emerging principles of regulatory evolution. Proc Natl Acad Sci U S A 104: 8605–8612.
23. Dermitzakis ET, Clark AG (2002) Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. Mol Biol Evol 19: 1114–1121.
24. Costas J, Casares F, Vieira J (2003) Turnover of binding sites for

transcription factors involved in early Drosophila development. Gene 310: 215–220.
25. Romano LA, Wray GA (2003) Conservation of Endo16 expression in sea urchins despite evolutionary divergence in both cis and trans-acting components of transcriptional regulation. Development 130: 4187–4199.
26. Ludwig MZ, Bergman C, Patel NH, Kreitman M (2000) Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 403: 564–567.
27. Ludwig MZ, Palsson A, Alekseeva E, Bergman CM, Nathan J, et al. (2005) Functional evolution of a cis-regulatory module. PLoS Biol 3: e93. doi:10.1371/journal.pbio.0030093
28. Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB (2005) Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature 433: 481–487.
29. Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, et al. (2006) Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature 440: 1001–1002.
30. McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, et al. (2007) Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature 448: 587–590.
31. King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188: 107–116.
32. Carroll SB (1995) Homeotic genes and the evolution of arthropods and chordates. Nature 376: 479–485.
33. Andolfatto P (2005) Adaptive evolution of non-coding DNA in Drosophila. Nature 437: 1149–1152.
34. Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, et al. (2005) Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. Genome Res 15: 1–18.
35. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, et al. (2007) Evolution of genes and genomes on the Drosophila phylogeny. Nature 450: 203–218.
36. Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, et al. (2007) Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature 450: 219–232.
37. Emberly E, Rajewsky N, Siggia ED (2003) Conservation of regulatory elements between two species of Drosophila. BMC Bioinformatics 4: 57.
38. Berman BP, Pfeiffer BD, Laverty TR, Salzberg SL, Rubin GM, et al. (2004) Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in Drosophila melanogaster and Drosophila pseudoobscura. Genome Biol 5: R61.
39. Sinha S, Schroeder MD, Unnerstall U, Gaul U, Siggia ED. (2004) Cross-species comparison significantly improves genome-wide prediction of cis-regulatory modules in Drosophila. BMC Bioinformatics 5: 1–12.
40. King DC, Taylor J, Zhang Y, Cheng Y, Lawson HA, et al. (2007) Finding cis-regulatory elements using comparative genomics: some lessons from ENCODE data. Genome Res 17: 775–786.
41. MacArthur S, Brookfield JFY (2004) Expected rates and modes of evolution of enhancer sequences. Mol Biol Evol 21: 1064–1073.
42. Stone J, Wray GA (2001) Rapid evolution of cis-regulatory sequences via local point mutations. Mol Biol Evol 18: 1764–1770.
43. Hagstrom K, Muller M, Schedl P (1997) A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the Drosophila bithorax complex. Genetics 146: 1365–1380.
44. Dejardin J, Rappailles A, Cuvier O, Grimaud C, Decoville M, et al. (2005) Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. Nature 434: 533–538.
45. Blastyak A, Mishra RK, Karch F, Gyurkovics H (2006) Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. Mol Cell Biol 26: 1434–1444.
46. Brown JL, Grau DJ, DeVido SK, Kassis JA (2005) An Sp1/KLF binding site is important for the activity of a Polycomb group response element from the Drosophila engrailed gene. Nucleic Acids Res 33: 5181–5189.
47. Schmitt S, Prestel M, Paro R (2005) Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev 19: 697–708.
48. Horard B, Tatout C, Poux S, Pirrotta V (2000) Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. Mol Cell Biol 20: 3187–3197.
49. Mishra RK, Mihaly J, Barges S, Spierer A, Karch F, et al. (2001) The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. Mol Cell Biol 21: 1311–1318.
50. Dejardin J, Cavalli G (2004) Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. EMBO J 23: 857–868.
51. Americo J, Whiteley M, Brown JL, Fujioka M, Jaynes JB, et al. (2002) A complex array of DNA-binding proteins required for pairing-sensitive silencing by a polycomb group response element from the Drosophila engrailed gene. Genetics 160: 1561–1571.
52. Fauvarque MO, Dura JM (1993) polyhomeotic regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in Drosophila. Genes Dev 7: 1508–1520.
53. Kassis JA (1994) Unusual properties of regulatory DNA from the Drosophila engrailed gene: three "pairing-sensitive" sites within a 1.6-kb region. Genetics 136: 1025–1038.
54. Macdonald SJ, Long AD (2006) Fine scale structural variants distinguish the genomes of Drosophila melanogaster and D. pseudoobscura. Genome Biol 7: R67.
55. Zuckerkandl E (2002) Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine. Genetica 115: 105–129.
56. Gibert JM, Peronnet F, Schlotterer C (2007) Phenotypic plasticity in Drosophila pigmentation caused by temperature sensitivity of a chromatin regulator network. PLoS Genet 3: e30. doi:10.1371/journal.pgen.0030030
57. Fritsch C, Brown JL, Kassis JA, Muller J (1999) The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. Development 126: 3905–3913.
58. Busturia A, Lloyd A, Bejarano F, Zavortink M, Xin H, et al. (2001) The MCP silencer of the Drosophila Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. Development 128: 2163–2173.
59. Ng CS, Kopp A (2008) Sex combs are important for male mating success in Drosophila melanogaster. Behav Genet 38: 195–201.
60. Barmina O, Kopp A (2007) Sex-specific expression of a HOX gene associated with rapid morphological evolution. Dev Biol 311: 277–286.
61. Randsholt NB, Santamaria P (2008) How Drosophila change their combs: the Hox gene Sex combs reduced and sex comb variation among Sophophora species. Evol Dev 10: 121–133.
62. Ringrose L, Ehret H, Paro R (2004) Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. Mol Cell 16: 641–653.
63. Corona DF, Armstrong JA, Tamkun JW (2004) Genetic and cytological analysis of Drosophila chromatin-remodeling factors. Methods Enzymol 377: 70–85.
64. Orlando V, Jane EP, Chinwalla V, Harte PJ, Paro R (1998) Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis. EMBO J 17: 5141–5150.