PLOS ONE

RESEARCH ARTICLE

Molecular population genetics of the Polycomb genes in *Drosophila subobscura*

Juan M. Calvo-Martín, Montserrat Papaceit, Carmen Segarra*

Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain

* csegarra@ub.edu

Citation: Calvo-Martín JM, Papaceit M, Segarra C (2017) Molecular population genetics of the Polycomb genes in *Drosophila subobscura*. PLoS ONE 12(9): e0185005. https://doi.org/10.1371/journal.pone.0185005

Editor: Arnar Palsson, University of Iceland, ICELAND

Received: June 1, 2017

Accepted: September 4, 2017

Published: September 14, 2017

Data Availability Statement: Complete sequences have been deposited in EMBL/GenBank Data Libraries under accession numbers LT856235 to LT856459.

Funding: This work was supported by a predoctoral fellowship from Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya, Catalonia, Spain, to JMC-M.; and grants BFU2012-35168 and BFU2015-63732 from Ministerio de Economía y Competitividad, Spain, and 2009SGR-1287 and 2014SGR10555 from Comissió Interdepartamental

Abstract

Polycomb group (PcG) proteins are important regulatory factors that modulate the chromatin state. They form protein complexes that repress gene expression by the introduction of posttranslational histone modifications. The study of PcG proteins divergence in *Drosophila* revealed signals of coevolution among them and an acceleration of the nonsynonymous evolutionary rate in the lineage ancestral to the *obscura* group species, mainly in subunits of the Pcl-PRC2 complex. Herein, we have studied the nucleotide polymorphism of PcG genes in a natural population of *D. subobscura* to detect whether natural selection has also modulated the evolution of these important regulatory genes in a more recent time scale. Results show that most genes are under the action of purifying selection and present a level and pattern of polymorphism consistent with predictions of the neutral model, the exceptions being *Su(z)12* and *Pho*. MK tests indicate an accumulation of adaptive changes in the SU(Z)12 protein during the divergence of *D. subobscura* and *D. guanche*. In contrast, the HKA test shows a deficit of polymorphism at *Pho*. The most likely explanation for this reduced variation is the location of this gene in the dot-like chromosome and would indicate that this chromosome also has null or very low recombination in *D. subobscura*, as reported in *D. melanogaster*.

Introduction

The identification of genes under adaptive selection is a major goal of evolutionary genetics. Several methods and statistical tests have been developed to detect the footprint left by the action of positive selection at the molecular level (reviewed in [1]). These approaches were first applied to single candidate genes with a clear adaptive function. In *Drosophila*, these studies consistently corroborated that adaptive selection has shaped the evolution of genes involved in processes such as immunity [2], reproduction [3] and stem cell maintenance [4]. However, molecular evolution of proteins and thus of their encoding genes is not solely determined by selection on protein function [5]. In fact, proteins do not act in isolation, but they are members of complex metabolic, regulatory or interaction networks that control biological processes. The evolution of a protein can be affected by its location in the network, the connectivity with network partners and the evolution of the network as a whole [6]. The strength of purifying

de Recerca i Innovació Tecnològica, Generalitat de Catalunya, Catalonia, Spain, to Montserrat Aguadé.

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

selection on genes coding for enzymes depends on the protein position either in the metabolic pathway [7] or in the signal transduction pathway [8,9]. In addition, physically interacting proteins tend to evolve co-ordinately. Accordingly, pervasive adaptive evolution has been detected among genes coding for subunits of nuclear pore complexes [10] and centromeric interacting proteins [11]. This concordant evolution also results in coincident increases or decreases of the evolutionary rate over particular branches of a phylogeny, as has been detected among genes encoding either interacting proteins [12] or members of protein networks [13].

Polycomb group (PcG) genes are involved in the epigenetic silencing of chromatin by the introduction of posttranslational histone modifications [14]. They code for proteins that form diverse protein repressive complexes. These interacting proteins are expected to evolve co-ordinately to maintain the integrity and functionality of the complexes, which can cause a correlation in their evolutionary rates. Evidence of coevolution among the subunits of these complexes has been detected in a previous study of interspecific nucleotide divergence in the *Drosophila* genus [15]. Moreover, this study revealed an acceleration in the nonsynonymous substitution rate in the lineage ancestral to the *obscura* group species, mainly in genes encoding subunits of the Pcl-PRC2 complex. Therefore, signatures of adaptive selection have been detected in some PcG genes prior diversification of the *obscura* species, i.e. in a rather distant time scale. Accordingly, it can be expected that positive selection have also modulated PcG genes evolution in a more recent time scale. The action of selection in the recent past can be detected by the study of intraspecific nucleotide polymorphism. The fixation of an adaptive mutation results in a strong reduction of linked neutral variation around the selected site (selective sweep). In contrast, the maintenance of different variants by balancing selection causes an increase of linked neutral variation [1].

The main aim of this study was to ascertain whether the action of positive selection detected in some Polycomb group genes in the lineage leading to the *obscura* group species is also detected by the analysis of nucleotide polymorphism in a species of this group. With this aim, we have analyzed by high-quality Sanger sequencing the nucleotide polymorphism of 16 PcG genes that code for the subunits of four Polycomb repressive complexes (PhoRC, Pcl-PRC2, PRC1 and dRAF) in *D. subobscura* as representative of the *obscura* group.

Nucleotide polymorphism in *D. subobscura* has been previously studied in multiple gene regions. Several of these studies have focused on analyzing how the level and pattern of nucleotide variation is affected by the presence of chromosomal inversions. Indeed, natural populations of *D. subobscura* have a rich chromosomal inversion polymorphism. More than 60 inversions that form about 90 complex chromosomal arrangements of overlapping inversions have been described in the five major chromosomes of the species: A, J, U, E and O [16]. The average number of inversions in heterozygosis per individual ranges from 2.4 to 5.1 in European populations, being 3.8 the mean value [17]. This rich chromosomal polymorphism causes nucleotide variation in *D. subobscura* to be strongly structured in regions affected by the inversions. This effect is due to the reduced recombination along the inverted region in heterokaryotypes, which prevents gene flow and results in a strong genetic differentiation between inverted and non-inverted (standard) chromosomes, mainly in gene regions located near the inversion breakpoints [18] but also along the whole inversion [19]. Accordingly, nucleotide variation at a particular gene in order to detect adaptive selection has to be studied in a random sample of arrangements that are homosequential for the region where the gene is located. This sampling strategy prevents any effect of inversion polymorphism on nucleotide variation.

In addition, one of the PcG genes studied (*Pho*) is located in the dot-like chromosome in *D. melanogaster* and other species of the *melanogaster* group [20]. This chromosome has in

these species a wide footprint of reduced variation relative to the other autosomes [21–23]. This low variation is consistent with a virtual lack of recombination in this chromosome that was previously identified by genetic analysis in *D. melanogaster* [24]. No data on nucleotide variation at genes located in the dot-like chromosome of *D. subobscura* are available. As the gene content of chromosomal elements is highly conserved in the Drosophila genus [25], *Pho* is also expected to be located in the dot-like chromosome of *D. subobscura*. In this case, the level and pattern of nucleotide variation at *Pho* would enable to contrast whether the dot-like chromosome of this species also has a reduced recombination rate.

The results obtained indicate that nucleotide variation of the PcG genes has been mainly modelled by purifying selection in *D. subobscura*. In fact, evidence of positive selection was only detected at *Su(z)12*, which encodes a subunit of the Pcl-PRC2 complex. In addition, an extremely low level of variation was detected at *Pho* relative to the other PcG genes, which would support that the dot-like chromosome of *D. subobscura* also has null or very low recombination rate.

Material and methods

Genes and fly stocks

The sixteen Polycomb group (PcG) genes studied code for subunits of four Polycomb repressive complexes: PhoRC, Pcl-PRC2, PRC1 and dRAF. These genes are: *Pho*, its paralog *Phol* and *Sfmbt* (PhoRC complex); *Caf1-55*, *E(z)*, *Esc*, its paralog *Escl*, *Su(z)12* and *Pcl* (Pcl-PRC2 complex); *Psc*, *Sce*, *Pc*, *Ph-p*, its paralog *Ph-d* and *Scm* (PRC1 complex), and *Kdm2* (dRAF complex that also contains PSC and SCE). The sequences of these genes (except *Ph-d)* in *D. subobscura*, *D. madeirensis* and *D. guanche* have been previously reported [15] and are available in EMBL/GenBank Data Libraries. The location of these genes in the polytene chromosomes of the *chcu* strain of *D. subobscura* that is homokaryotypic for known chromosomal arrangements was determined by *in situ* hybridization using biotinylated probes, as described in [26]. The exact chromosomal section where the different probes hybridized was determined according to the Kunze-Mühl and Müller cytological map [27] that has the standard (or reference) arrangement for each chromosome.

Isofemale lines were established after sampling a natural population of *D. subobscura* in the Observatori Fabra (outskirts of Barcelona, Catalonia, Spain). Thereafter, individuals from these lines were used to obtain highly inbred lines by at least 12 generations of sib mating. After inbreeding, it was confirmed that lines were homokaryotypic and the gene arrangement of the different chromosomes was determined as described in [28].

The location of the genes in the chromosomes was considered to select the lines studied. This selection was made to ensure that the lines studied for each gene were homosoquential for the chromosomal region where the gene is located. This selection is important to prevent any effect of the inversion polymorphism on the level and pattern of nucleotide variation when a gene maps near the breakpoints of an inversion. For instance, *Ph-p* and *Ph-d* are located in the distal half of the A (= X) chromosome near one of the breakpoints of inversion 2, which is the single inversion present in the A_2 arrangement. According to this location, lines A_2 were excluded from the analysis of variation in these genes. Consequently, variation at *Ph-p* and *Ph-d* was studied in a random sample including only lines with the A_{st} and A_1 arrangements that are homosequential for the chromosomal region were the genes map. S1 Table indicates the chromosome and chromosomal section where each gene is located and the arrangement for this chromosome of the selected lines. On average 15 lines were sequenced for each gene.

DNA sequencing and analysis

Genomic DNA of the selected *D. subobscura* lines was purified with the Puregen®Core Kit B (Qiagen). The polymerase chain reaction was used to amplify the complete coding region of the PcG genes here studied and the obtained PCR products were purified with Multiscreen plates (Millipore). The ABI Prism BigDye Terminators 3.0 Cycle kit (Applied Biosystems) and internal primers were used to sequence completely both strands of the amplicons. The sequences of the primers used in the PCR amplification and sequencing are available on request to the authors. Sequencing reactions were run on an ABI PRISM 3700 sequencer at Serveis Científico-Tècnics de la Universitat de Barcelona. Partial sequences of each gene were assembled with the SEQMAN program of the DNASTAR v6.0 Lasergene package [29]. Complete sequences have been deposited in EMBL/GenBank Data Libraries under accession numbers LT856235 to LT856459. Finally, sequences were multiply aligned with the MEGA v6.0 software [30]. Nucleotide variation at *Caf1-55* in *D. subobscura* has been previously reported [31] and sequences are also available in EMBL/GenBank Data Libraries.

The DnaSP v5 program [32] was used to estimate the basic summary statistics that describe nucleotide variation, to measure the extent of codon bias and to perform most neutrality tests. The level and pattern of variation was estimated by the number of polymorphic sites (S), nucleotide diversity (π), haplotype diversity (Hd), Tajima's D statistic [33], the minimum number of recombination events (R_m) inferred by the four-gamete test [34] and the linkage disequilibrium between parsimony informative sites measured by the $|D'|$ parameter [35]. Codon bias was estimated by the effective number of codons (ENC) as proposed in [36]. ENC ranges from 20 when a single codon is used for each amino acid (maximum codon bias) to 64 when all synonymous codons are used equally (no codon bias). Nucleotide diversity and divergence with *D. guanche* along each gene studied were also analyzed by the sliding window method using 50-nucleotide windows at five-nucleotide intervals with DnaSP v5.

Different neutrality tests were applied to the data to detect signatures of selection. The Tajima's test [33] based only on polymorphism data and the different Fu and Li's tests [37] were used to infer whether the observed pattern of variation conforms to that expected under the neutral model in a population at mutation-drift equilibrium. Fu and Li's tests based on the D^* and F^* statistics can be performed when only polymorphism data are available, in contrast to those based on the D and F statistics that also need data on an outgroup species. *D. guanche* was used as outgroup to perform the neutrality tests that require interspecific data. Significance levels of Tajima's D and Fu and Li's D and F tests statistics were estimated according to the standard neutral model as implemented in DnaSP v5 program [32]. In addition, as *D. subobscura* has a genome wide pattern of variation consistent with an expansion process, significance levels of these tests statistics were estimated after 10000 computer simulations using the software mlcoalsim v1.42 [38] with the parameters of the expansion model inferred in [28]. In addition, the HKA test [39] and the MK test [40] were applied to detect a putative decoupling of polymorphism and divergence among genes (HKA test) and between synonymous and nonsynonymous sites within a gene (MK test). The multilocus HKA test was performed by the HKA program developed by J. Hey and available as a software resource at https://bio.cst.temple.edu/~hey. This program conducts coalescence simulations to infer the significance of the observed χ^2-like test statistic and to perform the maximum cell value test that was proposed to detect significant outliers in the multilocus HKA test [41]. The test is based on the maximum absolute cell value among genes for the standardized discrepancies between observations and expectations detected for polymorphism when applying the multilocus HKA test. The null distribution of maximum cell values is inferred by computer simulations to assess the probability of the observed value.

Results

The chromosomal location in *D. subobscura* of the sixteen studied Polycomb group (PcG) genes is shown in Fig 1. These genes map at the expected chromosome according to their cytological location in *D. melanogaster* and the well-established chromosomal homologies between this species and *D. subobscura* [42]. Only the two tandem paralogs *Ph-p* and *Ph-d* are X-linked, all other genes are located in autosomes and *Pho* maps in the dot-like chromosome (Muller's element F). Chromosomal locations were taken into account when selecting the chromosomal arrangement of the lines sequenced for each gene to prevent any effect of inversions on nucleotide variation (S1 Table).

Polymorphic sites and indels present in the multiple alignment of each gene are shown in S1 Fig. Summary statistics of nucleotide variation are shown in Table 1. Nucleotide diversity (π) ranges from 0.0009 at *Pho* to 0.0117 at *Escl*. Therefore, variation at *Pho* is about one order of magnitude lower than at *Escl*. The low level of nucleotide variation at *Pho* relative to the other PcG gene regions is also evident when nucleotide diversity is normalized by nucleotide divergence (K) with *D. guanche* to correct for differences in the mutation rate across gene regions. The π/K ratio at *Pho* is 0.0263 that is also the lowest estimate and about 3.2 times lower than the next value present at *Pcl* (0.0833). In addition to nucleotide diversity, *Pho* also has the lowest number of haplotypes (10 in 17 lines) and haplotype diversity (*Hd* = 0.9). No recombination events were inferred in the *Pho* region ($R_m = 0$) by the four-gamete test. *Pho* also present a rather low codon bias (ENC = 54.98). This estimate is slightly higher (indicating a lower codon bias) than the median ENC value (53.92) for the *D. melanogaster* genes located in the dot-like element [43]. Tajima's *D* statistic was negative in all regions (Table 1), indicating an excess of low frequency variants at polymorphic sites in all genes. Negative Tajima's *D*

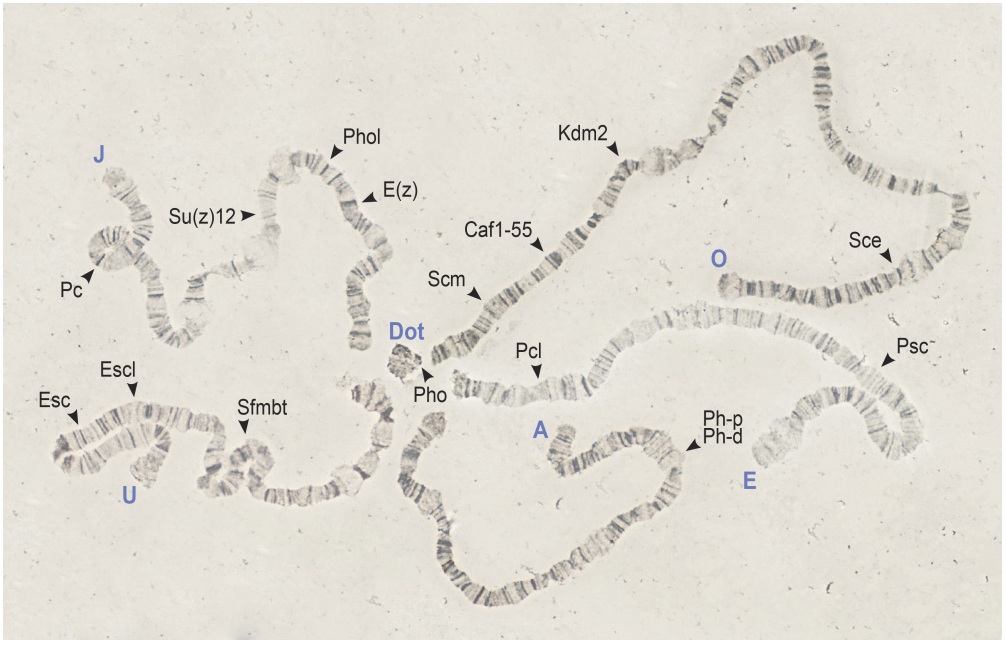

Fig 1. Location of the PcG genes on the *D. subobscura* chromosomes. The location of the sixteen studied PcG genes is indicated on the polytene chromosomes of *D. subobscura* (*ch cu* strain) by arrowheads. All chromosomes have the standard arrangement shown in the cytological map of the species [27] except the O chromosome that has the O_{3+4} arrangement that includes the overlapping inversions O_3 and O_4. Chromosomes are identified by capital letters (A, J, U, E and O) at the distal end. The dot-like chromosome is also indicated (Dot).

https://doi.org/10.1371/journal.pone.0185005.g001

Table 1. Summary estimates of the level and pattern of polymorphism.

Complex *Gene*	n° lines	n° sites	*S*	*h*	*Hd*	π	π/K	Tajima's *D*	R_m
PhoRC									
Pho	17	2842	11	10	0.90	0.0009	0.0263	-0.8763	0
Sfmbt	15	5617	76	14	0.99	0.0033	0.1055	-0.9055	9
Phol	16	2293	34	15	0.99	0.0030	0.1130	-1.3772	5
Pcl-PRC2									
Caf1-55	14	2621	85	13	0.99	0.0083	0.1857	-0.9447	9
E(z)	16	3652	143	15	0.99	0.0113	0.2162	-0.3766	16
Esc	15	2105	78	15	1.00	0.0098	0.3617	-0.7133	10
Su(z)12	16	3763	58	15	0.99	0.0039	0.1788	-0.7093	15
Pcl	16	3491	39	16	1.00	0.0027	0.0833	-0.8365	5
Escl	15	1756	66	14	0.99	0.0117	0.3400	-0.1027	10
PRC1									
Psc	15	11839	325	15	1.00	0.0060	0.1799	-1.3784	47
Sce	9	2714	50	9	1.00	0.0064	0.1455	-0.4042	6
Pc	17	3749	161	16	0.99	0.0085	0.2312	-1.4846	20
Ph-p	17	9203	176	16	0.99	0.0038	0.1348	-1.4433	20
Scm	16	4161	123	15	0.99	0.0060	0.1716	-1.4928	18
Ph-d	16	5752	63	15	0.99	0.0022	0.1007	-1.4728	6
dRAF*									
Kdm2	16	8507	186	15	0.99	0.0046	0.1403	-1.3780	20

S: number of segregating sites, *h*: number of haplotypes, *Hd*: haplotype diversity, π: nucleotide diversity per site, π/K: nucleotide diversity per site normalized by nucleotide divergence with *D. guanche*, R_m: minimum number of recombination events.

*dRAF complex also contains PSC and SCE.

https://doi.org/10.1371/journal.pone.0185005.t001

estimates are expected after the fixation of an advantageous mutation by positive selection (selective sweep). However, demographic factors as bottlenecks or population expansions can also cause the same pattern of variation. *D. subobscura* has a genome wide pattern of variation consistent with an expansion process [28] and, therefore, present results on the Polycomb genes are more likely explained by demographic than by selective factors. None of the performed Tajima's *D* test was significant ($P > 0.05$) assuming either a constant population size or an expansion process. Fu and Li's *D* and *F* tests using *D. guanche* as outgroup were significant at *Psc* ($P = 0.0121$ and $P = 0.0212$, respectively) and *Ph-d* ($P = 0.0184$ and $P = 0.0187$, respectively) under the unrealistic assumption of a constant population size but not significant under the expansion model (S2 Table), which indicates the importance of using an accurate demographic model to calculate the significance of the neutrality tests.

Nucleotide variation in the coding region was also analyzed independently for synonymous (π_s) and nonsynonymous (π_a) sites (Table 2). *Caf1-55* without nonsynonymous polymorphism and the lowest nonsynonymous divergence (K_a) is the most conserved PcG gene. On the other hand, a lack of synonymous polymorphism is present at *Pho*. In fact, all polymorphic sites in the *Pho* coding region present nonsynonymous variants (S1 Fig). However, the overall nonsynonymous diversity at *Pho* is not high ($\pi_a = 0.0006$) given the low number of polymorphic sites (only 4). The highest nonsynonymous diversity is detected at *Psc* ($\pi_a = 0.0019$), which also has a high nonsynonymous divergence ($K_a = 0.0097$) only overcome by *Pho* ($K_a = 0.0181$) and *Pcl* ($K_a = 0.0129$).The ratios π_a/π_s and K_a/K_s in the different PcG genes are represented in Fig 2. These ratios are low when purifying selection against nonsynonymous mutations is strong. The genes encoding the core subunits of the PRC2 complex (*Caf1-55*, *E(z)*, *Esc* and *Su*

Table 2. Summary of synonymous and nonsynonymous polymorphism and divergence with *D. guanche* in the PcG genes.

Complex *Gene*	π_s	π_a	π_a/π_s	K_s	K_a	K_a/K_s
PhoRC						
Pho	0.0000	0.0006	-	0.0663	0.0181	0.2732
Sfmbt	0.0115	0.0009	0.0742	0.0890	0.0070	0.0786
Phol	0.0094	0.0012	0.1261	0.1051	0.0045	0.0428
Pcl-PRC2						
Caf1-55	0.0211	0.0000	0.0000	0.1417	0.0010	0.0070
E(z)	0.0242	0.0004	0.0145	0.1743	0.0036	0.0206
Esc	0.0234	0.0004	0.0176	0.0876	0.0022	0.0256
Su(z)12	0.0149	0.0006	0.0382	0.0621	0.0070	0.1133
Pcl	0.0071	0.0012	0.1704	0.0881	0.0129	0.1464
Escl	0.0334	0.0014	0.0407	0.1292	0.0034	0.0266
PRC1						
Psc	0.0120	0.0019	0.1586	0.0890	0.0097	0.1093
Sce	0.0149	0.0007	0.0443	0.0470	0.0033	0.0704
Pc	0.0128	0.0006	0.0494	0.0757	0.0061	0.0811
Ph-p	0.0070	0.0010	0.1355	0.0425	0.0070	0.1642
Scm	0.0137	0.0010	0.0736	0.0774	0.0063	0.0820
Ph-d	0.0056	0.0006	0.1097	0.0478	0.0092	0.1923
dRAF*						
Kdm2	0.0120	0.0006	0.0507	0.0901	0.0030	0.0334

π_s: synonymous nucleotide diversity, π_a: nonsynonymous nucleotide diversity, π_a/π_s: nonsynonymous/synonymous diversity ratio, K_s: synonymous divergence, K_a: nonsynonymous divergence, K_a/K_s: nonsynonymous/synonymous divergence ratio.
*dRAF complex also contains PSC and SCE.

https://doi.org/10.1371/journal.pone.0185005.t002

(z)12) have the lowest π_a/π_s ratios. In addition, *Caf1-55*, *E(z)* and *Esc* also have low K_a/K_s ratios, which indicates a strong purifying selection acting on these genes according to both polymorphism and divergence. By contrast, *Pcl* and *Psc* have the two highest π_a/π_s ratios and also a high K_a/K_s ratio. PSC is a structurally disordered protein mainly in its C-terminal region [44],

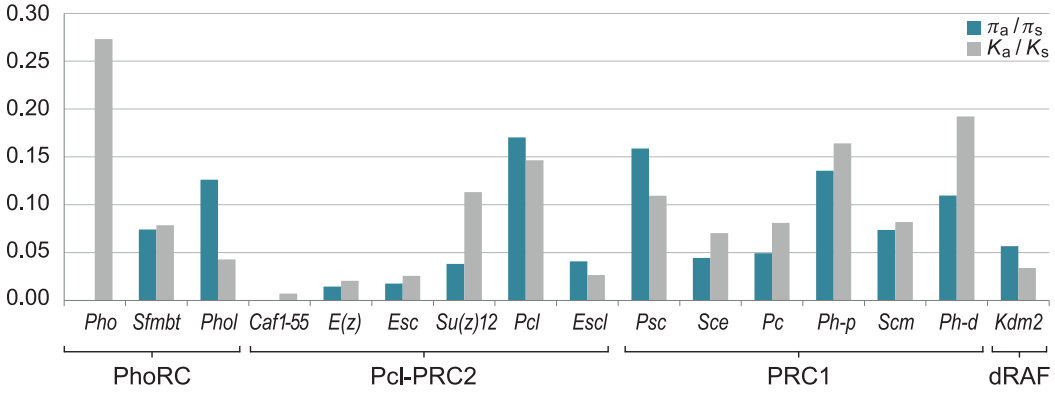

Fig 2. Polymorphism and divergence in the coding region of the PcG genes. Ratio of nonsynonymous to synonymous polymorphism in *D. subobscura* (π_a/π_s) and of nonsynonymous to synonymous divergence between *D. subobscura* and *D. guanche* (K_a/K_s) for each of the Polycomb genes studied. The *x*-axis indicates gene identity and the *y*-axis indicates π_a/π_s and K_a/K_s values. *Pho* lacks synonymous polymorphism and *Caf1-55* nonsynonymous polymorphism. Genes on the *x*-axis are grouped according to Polycomb complex. dRAF also contains PSC and SCE.

https://doi.org/10.1371/journal.pone.0185005.g002

in which the distribution of the charged amino acids and not the sequence itself is important for protein function.

Under strict neutrality, the ratio between intraspecific polymorphism and interspecific divergence is expected to be homogeneous across loci in a constant size population at mutation-drift equilibrium. Based on this expectation, the HKA neutrality test [39] was developed to detect a putative decoupling in the polymorphism to divergence ratio between two gene regions. Assuming that a region evolves under neutrality, a significant excess of polymorphism at the other region would indicate balancing selection. Conversely, a recent selective sweep would result in a deficit of polymorphism. The multilocus HKA test performed considering silent variation in the sixteen PcG genes was marginally significant using *D. guanche* as outgroup ($\chi^2 = 22.09$, df = 15, $P = 0.0812$). The HKA multilocus test is conservative when a single gene from a large sample present a ratio influenced by selection. A close inspection to the data indicates that this is the case for the PcG genes studied since *Pho* deviates considerably from the other genes (S2 Fig). In fact, *Pho* has the largest contribution (54%) to the overall χ^2-like test statistic and the contribution of the other PcG genes is quite uniform and small. In addition, the absolute maximum discrepancies between expected and observed polymorphism in the χ^2-like test statistic was detected at *Pho* (3.667) and is significant ($P = 0.0129$) using the maximum cell value test [41]. According to this result, the HKA test was applied in pairwise comparisons between *Pho* and each one of the other PcG genes. The 15 performed HKA tests were statistically significant ($P < 0.05$). This result clearly supports that *Pho* has a significant deficit of nucleotide polymorphism. Finally, the multilocus HKA test was not significant when excluding *Pho* from the analysis ($\chi^2 = 9.96$, df = 14, $P = 0.7402$).

The McDonald and Kreitman test [40] of neutrality (MK test) was also performed for each gene independently. This test is based on the comparison of the ratio of nonsynonymous to synonymous polymorphisms (P_a/P_s) and the ratio of nonsynonymous to synonymous fixed differences between species (F_a/F_s). Both ratios are expected to be equal under strict neutrality and a *G*-test of independence can be used to detect putative deviations from this expectation. A significant difference in both ratios was inferred only at *Su(z)12* ($G = 4.909$; 1 df, $P = 0.0267$). The direction of departure from neutrality can be inferred by the neutrality index (*NI*) estimated as the ratio between P_a/F_a and P_s/F_s [45]. Assuming that synonymous mutations are neutral, $NI > 1$ indicates an excess of nonsynonymous polymorphism and $NI < 1$ an excess of nonsynonymous divergence. The neutrality index at *Su(z)12* is 0.29, which indicates adaptive selection favouring nonsynonymous changes during the divergence of *D. subobscura* and *D. guanche* (see also Fig 2). The proportion of nonsynonymous substitutions fixed by natural selection at *Su(z)12* in these lineages is 0.71 according to the α-parameter [46]. The statistical power of the MK test when applied to *Pho* is affected by the low number of polymorphic sites, which might explain why the test is not significant despite the strong decoupling between polymorphism and divergence (Fig 2).

The distribution of nucleotide diversity and divergence along the sixteen Polycomb group genes was also analyzed by the sliding window approach. Fig 3 shows the results of this analysis in the two genes with a nonneutral variation: *Pho* and *Su(z)12*. The results for the other genes are shown in S3 Fig. Peaks of polymorphism at *Pho* are very low and mainly located in noncoding regions. The highest divergence peak at *Pho* is detected at the end of the first intron. In *Su(z)12* there is a rather good correspondence between peaks of polymorphism and divergence. The highest divergence peak is present in the short fourth intron. Divergence is also rather high at the end of exon 5 and in exon 6. The gene regions coding for protein domains are highly conserved in the two genes.

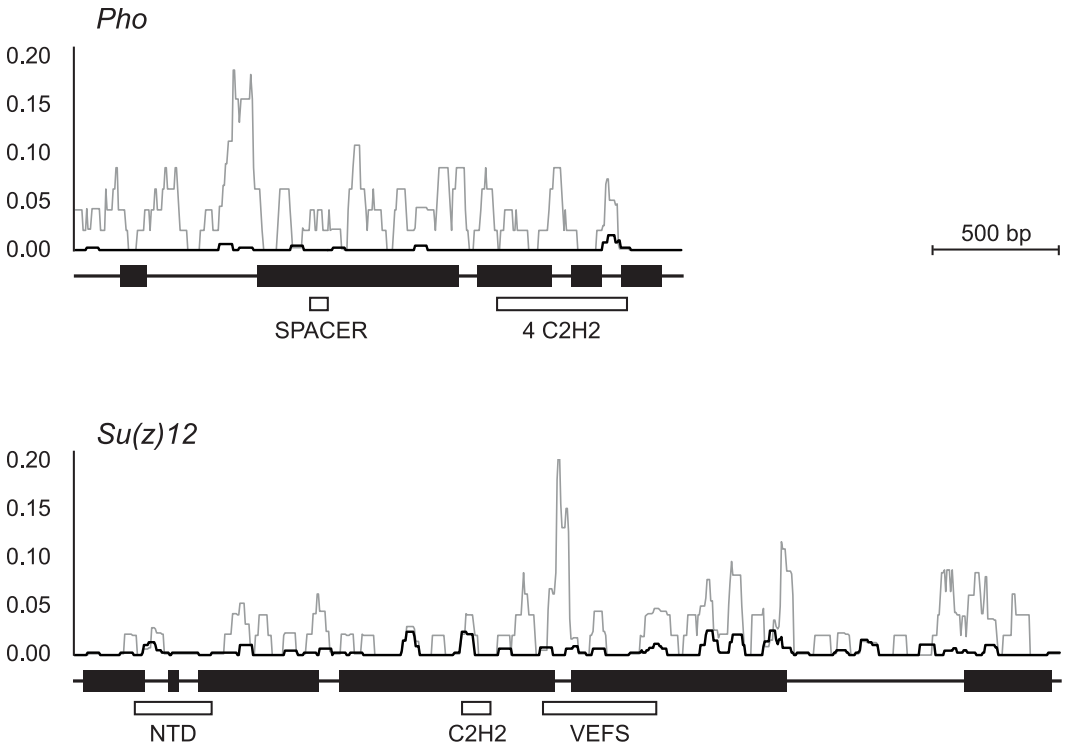

Fig 3. Nucleotide diversity and divergence along *Pho* and *Su(z)12*. Sliding window plots of the distribution of nucleotide diversity in *D. subobscura* (π, black line) and of nucleotide divergence between *D. subobscura* and *D. guanche* (*K*, gray line). Windows include 50 sites with successive displacements of 5 sites. The *x*-axis indicates nucleotide sites across the gene region and the *y*-axis indicates nucleotide diversity or divergence. Solid boxes in the lower part of the figure indicate the coding exons and thin lines show flanking regions and introns. The protein domains of the encoded proteins are indicated below the gene structure with empty boxes.

https://doi.org/10.1371/journal.pone.0185005.g003

Discussion

Selection at the Polycomb group genes

Polycomb group (PcG) proteins form different complexes that maintain the repressive state of chromatin. The most relevant Polycomb complexes are Pho-RC, PRC1, Pcl-PRC2 and dRAF. The genes encoding the subunits of these complexes are mainly under the action of purifying selection as reflected by the π_a/π_s and K_a/K_s ratios (Fig 2). However, the strength of purifying selection differs both among the genes encoding the subunits of the same complex and among the genes encoding subunits of the different complexes. *Caf1-55*, *E(z)*, *Esc* and *Su(z)12* that code for the core subunits of the PRC2 complex are under a strong purifying selection and therefore highly conserved, except *Su(z)12* with a rather high K_a/K_s ratio caused by positive selection (see below). PRC2 is a catalytic complex responsible for the histone H3 methylation at lysine 27, which is a chromatin repressive mark. The methyltransferase activity of PRC2 requires the E(Z) subunit with the active site for histone methylation and the noncatalytic subunits ESC and SU(Z)12 [47]. The critical role of the three subunits for the PRC2 function would explain their high conservation. PRC1 and dRAF are also catalytic complexes. The catalytic subunit of PRC1 is SCE that monoubiquitylates lysine 118 of histone H2A introducing a second chromatin silencing mark. *Sce* is also under strong purifying selection and it is the most conserved gene encoding subunits of the PRC1 complex. SCE is also a subunit of the dRAF complex that has an additional catalytic subunit that is KDM2. The gene encoding

KDM2 (*Kdm2*) is also under strong purifying selection further confirming that the genes encoding the catalytic subunits of the Polycomb complexes are highly conserved.

Although most genes coding for the subunits of the Polycomb complexes are under puryfying selection, the footprint of adaptive selection was detected by the MK test in *Su(z)12* during the divergence of *D. subobscura* and *D. guanche*. SU(Z)12 is a noncatalytic subunit of the PRC2 complex that potentiates the enzymatic activity of its partner E(Z) and contributes to nucleosome binding [48,49]. The protein has three domains: an N-terminal domain, a zinc finger C2H2 and a VEFS box. The N-terminal domain is responsible for the interaction with CAF1-55 [50]. The VEFS domain enhances the methyltransferase activity of E(Z) and is critical for the SU(Z)12/E(Z) interaction [51]. Most of the fourteen amino acid replacements detected in the SU(Z)12 protein of *D. subobscura* and *D. guanche* are located at the poorly conserved C-terminal region of the protein (Fig 3) and none of them affect the described protein domains. Therefore, it is not expected that these replacements affect the interaction of SU(Z) 12 with the other subunits of PRC2.

Polycomb proteins showed evidence of coevolution in a divergence study in the Drosophila genus [15]. In addition, a concordant acceleration of the evolutionary rate was detected in different PcG proteins mainly of the Pcl-PRC2 complex in the branch ancestral to the species of the *obscura* group. In contrast to this previous study, the present analysis of nucleotide variation in *D. subobscura* (as representative of the obscura group) in sixteen Polycomb groups genes failed to detect signs of coevolution and positive selection could only be inferred in *Su (z)12*. Coevolution between proteins is expected when amino replacements in the interacting interface of one protein are compensated by additional replacements in the interacting protein to maintain the integrity and function of the protein complex. The adaptive selection detected at *Su(z)12* during the divergence of *D. subobscura* and *D. guanche* is not evident in *Caf1-55* and *Esc*, which are two genes that code for subunits known to interact with SU(Z)12. The location of the amino acid replacement fixed between the two species outside the interacting interface of SU(Z)12 would explain the lack of coevolution with the partners of this protein.

In addition to *Su(z)12*, other genes involved in chromatin metabolism seem to have evolved under positive selection in *Drosophila*, such as *rhino* that belongs to the heterochromatin protein 1 (HP1) family [52]. In addition, genome wide analysis performed to detect gene regions with a strong genetic differentiation between tropical and temperate populations of *D. melanogaster* also identified chromatin organization genes such as *crm* [53], *Ph-p* [54], *chd1*, *ssrp*, *chm* and *glu* [55].

Nucleotide variation at the dot-like chromosome of *D. subobscura*

The Polycomb gene *Pho* is located at the dot-like chromosome in *D. subobscura* near the distal or telomeric end of the chromosome. This position is shared by *D. melanogaster*, *D. yakuba*, *D. teissieri* and *D. erecta*, which supports that this is the ancestral location in the *Sophophora* subgenus. Therefore, the proximal location of *Pho* detected in other species of the *melanogaster* group such as *D. orena*, *D. mauritiana*, *D. simulans* and *D. sechellia* would be the result of chromosomal inversions as suggested in [20].

Nucleotide variation at *Pho* in *D. subobscura* when compared to other Polycomb genes is peculiar in several aspects: low nucleotide diversity, low number of haplotypes, rather low haplotype diversity and only nonsynonymous polymorphism in the coding region. In addition, no recombination events are detected by the four gamete test and thus linkage disequilibrium estimated by $|D'|$ is equal to 1 in all pairwise comparisons between parsimony informative sites. Therefore, variation at *Pho* would support that the dot-like chromosome of *D. subobscura* also

has an exceedingly low recombination rate as previously reported in other *Drosophila* species such as *D. melanogaster* [24]. This lack of recombination has a drastic effect on nucleotide variation. In fact, several multilocus studies have confirmed that the level and pattern of nucleotide variation differ between the dot-like chromosome and the other autosomes in *D. melanogaster* [21] and other *Drosophila* species not only of the *melanogaster* group such as *D. simulans* [22] and *D. yakuba* [23] but also of the *virilis* group as *D. americana* [56]. These studies show a low level of nucleotide polymorphism and relatively inefficient selection in the genes located in the dot-like chromosome. In addition, evidence of recombination at the molecular level has been detected in these multilocus studies of the dot-like chromosome although with a recombination rate much lower than in the other autosomes.

Mean nucleotide diversity in genes located in the dot-like chromosome is $\pi = 0.0006$ in *D. melanogaster* and $\pi = 0.0009$ in *D. simulans* [23]. These estimates are consistent with the level of variation at *Pho* in *D. subobscura* ($\pi = 0.0009$). The low level of variation in the dot-like chromosome is explained by both the hitchhiking effect [57] and the background selection model [58]. In fact, in regions with low recombination the fixation of positively selected mutations and the elimination of deleterious mutations cause a reduction of neutral linked variants. This reduction of variation can also be explained by the Hill-Robertson effect [59] that proposes that a locus linked to a second locus under directional selection experiences a reduction in effective population size (N_e). The decrease in N_e, apart from reducing neutral variation, causes a relaxation of selection due to an increase in the effects of genetic drift. The presence of only nonsynonymous polymorphism at the *Pho* gene of *D. subobscura* is consistent with a relaxation of selection against slightly deleterious nonsynonymous mutations. In addition, the low codon bias at *Pho* would indicate inefficient selection acting at synonymous sites to maintain the prevalent use of particular synonymous codons [60]. The relaxed selection might represent a challenge for the genes located in the dot-like chromosome, mainly for housekeeping genes under strong purifying selection. *Pho* and its paralog *Phol* are the only Polycomb group genes that bind directly to DNA. The PHO protein (or alternatively PHOL) binds to SFMBT and forms the PhoRC complex that is crucial for anchoring the other Polycomb complexes at the regulated target sites [47]. The high divergence of PHO in the Drosophila genus [15] and when comparing *D. subobscura* and *D. guanche* (Fig 3) suggests that this protein can accumulate amino acid replacements maintaining function as long as they do not affect the protein domains.

In summary, the level and pattern of nucleotide variation at the PcG genes in *D. subobscura* support the action of positive selection only at *Su(z)12*. The comparison of the ratio between polymorphic and fixed sites differs significantly for synonymous and nonsynonymous variants, suggesting adaptive evolution in this gene during the divergence of *D. subobscura* and *D. guanche*. In addition, the characteristics of the nucleotide polymorphism at *Pho* are consistent with the location of this gene in the dot-like chromosome and would indicate that this chromosome exhibits little or no recombination in *D. subobscura*.

Supporting information

S1 Table. Chromosomal location of the PcG genes. Location on the chromosomes of *D. subobscura* (A, J, U, E, O and Dot-like), chromosomal sections and chromosomal arrangements of the sequenced lines. The dot-like chromosome has not chromosomal polymorphism. (PDF)

S2 Table. Significance levels of the Tajima's *D* and Fu and Li's *D* and *F* tests statistics according to the expansion model. Each *P*-value was calculated after 10000 computer

simulations using the software mlcoalsim v1.42 and the parameters obtained in Pratdesaba et al. (2015) indicated in the footnote.
(PDF)

S1 Fig. Nucleotide polymorphic sites of the 16 PcG genes in the *D. subobscura* sequenced lines. The blue bar above each alignment illustrates gene structure: 5' flanking region (5'), coding exons numbered by order (E), introns also numbered by order (I) and 3' flanking region (3'). The identification of each line is given on the left by the letters OF (Observatori Fabra) and a number. Sites are numbered according to the multiple alignment. Asterisks under site numbers indicate nonsynonymous polymorphic sites. Dots indicate nucleotides identical to the first sequence that is used as reference. Gaps of a single nucleotide are shown by dashes. Insertions and deletions are indicated by the letters *i* and *d*, respectively, with a number according to its length. Polymorphic microsatellites are shown in brackets followed by a number that indicates the number of repeats. The last row of the multiple alignment shows the information for the polymorphic sites in the sequences of *D. guanche* (Dgua) used as outgroup. This information is also shown for the sequences of *D. madeirensis* (Dmad), a closely relative to *D. subobscura*. A) *Pho*, B) *Sfmbt*, C) *Phol*, D) *Caf1-55*, E) *E(z)*, F) *Esc*, G) *Su(z)12*, H) *Pcl*, I) *Escl*, J) *Psc*, K) *Sce*, L) *Pc*, M) *Ph-p*, N) *Scm*, O) *Ph-d*, P) *Kdm2*.
(PDF)

S2 Fig. Contribution of each PcG gene to the multilocus HKA test. Gray bars indicate the contribution to the overall χ^2-like test statistic due to silent divergence between *D. subobscura* and *D. guanche*. Blue bars indicate the corresponding contribution due to silent polymorphism. The highest contribution of *Pho* both to divergence and polymorphism is significant by the maximum cell value test (see text).
(PDF)

S3 Fig. Nucleotide diversity and divergence along PcG genes. Sliding window plots of the distribution of nucleotide diversity in *D. subobscura* (π, black line) and of nucleotide divergence between *D. subobscura* and *D. guanche* (*K*, gray line). Windows include 50 sites with successive displacements of 5 sites. The *x*-axis indicates nucleotide sites across the gene region and the *y*-axis indicates nucleotide diversity or divergence. Solid boxes in the lower part of the figure indicate the coding exons and thin lines show flanking regions and introns. *Pho* and *Su(z)12* plots are shown in Fig 3.
(PDF)

Acknowledgments

We thank Montserrat Aguadé for critical comments on the manuscript. We also thank Serveis Científico-Tècnics, Universitat de Barcelona, for automated DNA sequencing facilities.

Author Contributions

Conceptualization: Juan M. Calvo-Martín, Montserrat Papaceit, Carmen Segarra.

Formal analysis: Juan M. Calvo-Martín, Montserrat Papaceit, Carmen Segarra.

Funding acquisition: Juan M. Calvo-Martín, Montserrat Papaceit, Carmen Segarra.

Investigation: Juan M. Calvo-Martín, Montserrat Papaceit, Carmen Segarra.

Project administration: Montserrat Papaceit, Carmen Segarra.

Resources: Juan M. Calvo-Martín, Montserrat Papaceit, Carmen Segarra.

Supervision: Montserrat Papaceit, Carmen Segarra.

Visualization: Juan M. Calvo-Martín.

Writing – original draft: Juan M. Calvo-Martín, Montserrat Papaceit, Carmen Segarra.

Writing – review & editing: Juan M. Calvo-Martín, Montserrat Papaceit, Carmen Segarra.

References

1. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annu Rev Genet. 2013; 47: 97–120. https://doi.org/10.1146/annurev-genet-111212-133526 PMID: 24274750
2. Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. Dynamic evolution of the innate immune system in *Drosophila*. Nat Genet. 2007; 39: 1461–1468. https://doi.org/10.1038/ng.2007.60 PMID: 17987029
3. Wong A, Turchin MC, Wolfner MF, Aquadro CF. Evidence for positive selection on *Drosophila melanogaster* seminal fluid protease homologs. Mol Biol Evol. 2008; 25: 497–506. https://doi.org/10.1093/molbev/msm270 PMID: 18056920
4. Flores HA, DuMont VLB, Fatoo A, Hubbard D, Hijji M, Barbash DA, et al. Adaptive evolution of genes involved in the regulation of germline stem cells in *Drosophila melanogaster* and *D. simulans*. G3. 2015; 5: 583–592. https://doi.org/10.1534/g3.114.015875 PMID: 25670770
5. Pál C, Papp B, Lercher MJ. An integrated view of protein evolution. Nat Rev Genet. 2006; 7: 337–348. https://doi.org/10.1038/nrg1838 PMID: 16619049
6. Cork JM, Purugganan MD. The evolution of molecular genetic pathways and networks. BioEssays. 2004; 26: 479–484. https://doi.org/10.1002/bies.20026 PMID: 15112228
7. Eanes WF. Molecular population genetics and selection in the glycolytic pathway. J Exp Biol. 2011; 214: 165–171. https://doi.org/10.1242/jeb.046458 PMID: 21177937
8. Alvarez-Ponce D, Guirao-Rico S, Orengo DJ, Segarra C, Rozas J, Aguadé M. Molecular population genetics of the insulin/TOR signal transduction pathway: a network-level analysis in *Drosophila melanogaster*. Mol Biol Evol. 2012; 29: 123–132. https://doi.org/10.1093/molbev/msr160 PMID: 21680868
9. Riley RM, Jin W, Gibson G. Contrasting selection pressures on components of the Ras-mediated signal transduction pathway in *Drosophila*. Mol Ecol. 2003; 12: 1315–1323. PMID: 12694293
10. Presgraves DC, Stephan W. Pervasive adaptive evolution among interactors of the *Drosophila* hybrid inviability gene, *Nup96*. Mol Biol Evol. 2007; 24: 306–314. https://doi.org/10.1093/molbev/msl157 PMID: 17056646
11. Beck EA, Llopart A. Widespread positive selection drives differentiation of centromeric proteins in the *Drosophila melanogaster* subgroup. Sci Rep. 2015; 5: 17197. https://doi.org/10.1038/srep17197 PMID: 26603658
12. Clark NL, Aquadro CF. A novel method to detect proteins evolving at correlated rates: identifying new functional relationships between coevolving proteins. Mol Biol Evol. 2010; 27: 1152–1161. https://doi.org/10.1093/molbev/msp324 PMID: 20044587
13. Findlay GD, Sitnik JL, Wang W, Aquadro CF, Clark NL, Wolfner MF. Evolutionary rate covariation identifies new members of a protein network required for *Drosophila melanogaster* female post-mating responses. PLoS Genet. 2014; 10: e1004108. https://doi.org/10.1371/journal.pgen.1004108 PMID: 24453993
14. Lanzuolo C, Orlando V. Memories from the Polycomb group proteins. Annu Rev Genet. 2012; 46: 561–589. https://doi.org/10.1146/annurev-genet-110711-155603 PMID: 22994356
15. Calvo-Martín JM, Librado P, Aguadé M, Papaceit M, Segarra C. Adaptive selection and coevolution at the proteins of the Polycomb repressive complexes in *Drosophila*. Heredity. 2016; 116: 213–223. https://doi.org/10.1038/hdy.2015.91 PMID: 26486609
16. Krimbas CB. The inversion polymorphism of *Drosophila subobscura*. In: Krimbas CB, Powell JR, editors. *Drosophila* Inversion Polymorphism. Boca Raton: CRC Press; 1992. pp. 127–220.
17. Prevosti A. Chromosomal polymorphism in *Drosophila subobscura* populations from Barcelona (Spain). Genet Res. 1964; 5: 27.
18. Nóbrega C, Khadem M, Aguadé M, Segarra C. Genetic exchange versus genetic differentiation in a medium-sized inversion of *Drosophila*: The A2/Ast arrangements of *Drosophila subobscura*. Mol Biol Evol. 2008; 25: 1534–1543. https://doi.org/10.1093/molbev/msn100 PMID: 18436552

19. Munté A, Rozas J, Aguadé M, Segarra C. Chromosomal inversion polymorphism leads to extensive genetic structure: A multilocus survey in *Drosophila subobscura*. Genetics. 2005; 169: 1573–1581. https://doi.org/10.1534/genetics.104.032748 PMID: 15687280

20. Podemski L, Ferrer C, Locke J. Whole arm inversions of chromosome 4 in *Drosophila* species. Chromosoma. 2001; 110: 305–312. PMID: 11534822

21. Wang W, Thornton K, Berry A, Long M. Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. Science. 2002; 295: 134–137. https://doi.org/10.1126/science.1064521 PMID: 11778050

22. Wang W, Thornton K, Emerson JJ, Long M. Nucleotide variation and recombination along the fourth chromosome in *Drosophila simulans*. Genetics. 2004; 166: 1783–1794. PMID: 15126398

23. Arguello JR, Zhang Y, Kado T, Fan C, Zhao R, Innan H, et al. Recombination yet inefficient selection along the *Drosophila melanogaster* subgroup's fourth chromosome. Mol Biol Evol. 2010; 27: 848–861. https://doi.org/10.1093/molbev/msp291 PMID: 20008457

24. Hochman B. The fourth chromosome of *Drosophila melanogaster*. In: Ashburner M, Novitski E, editors. The genetics and biology of *Drosophila* Vol 1b. Academic Press; 1976. pp. 903–928.

25. Papaceit M, Aguadé M, Segarra C. Chromosomal evolution of elements B and C in the *Sophophora* subgenus of *Drosophila*: evolutionary rate and polymorphism. Evolution. 2006; 60: 768–781. PMID: 16739458

26. Papaceit M, Segarra C, Aguadé M. Structure and population genetics of the breakpoints of a polymorphic inversion in *Drosophila subobscura*. Evolution. 2013; 67: 66–79. https://doi.org/10.1111/j.1558-5646.2012.01731.x PMID: 23289562

27. Kunze-Mühl E, Müller E. Weitere Untersuchungen über die chromosomale Struktur und die natürlichen Strukturtypen von *Drosophila subobscura* Coll. Chromosoma. 1958; 9: 559–570.

28. Pratdesaba R, Segarra C, Aguadé M. Inferring the demographic history of *Drosophila subobscura* from nucleotide variation at regions not affected by chromosomal inversions. Mol Ecol. 2015; 24: 1729–1741. https://doi.org/10.1111/mec.13155 PMID: 25776124

29. Burland TG. DNASTAR's Lasergene sequence analysis software. Methods Mol Biol. 2000; 132: 71–91. PMID: 10547832

30. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013; 30: 2725–2729. https://doi.org/10.1093/molbev/mst197 PMID: 24132122

31. Calvo-Martín JM, Papaceit M, Segarra C. Evidence of neofunctionalization after the duplication of the highly conserved Polycomb group gene *Caf1-55* in the obscura group of *Drosophila*. Sci Rep. 2017; 7: 40536. https://doi.org/10.1038/srep40536 PMID: 28094282

32. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25: 1451–1452. https://doi.org/10.1093/bioinformatics/btp187 PMID: 19346325

33. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123: 585–595. PMID: 2513255

34. Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics. 1985; 111: 147–164. PMID: 4029609

35. Lewontin RC. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics. 1964; 49: 49–67. PMID: 17248194

36. Wright F. The 'effective number of codons' used in a gene. Gene. 1990; 87: 23–29. PMID: 2110097

37. Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics. 1993; 133: 693–709. PMID: 8454210

38. Ramos-Onsins SE, Mitchell-Olds T. Mlcoalsim: multilocus coalescent simulations. Evol Bioinform Online. 2007; 3: 41–44. PMID: 19430603

39. Hudson RR, Kreitman M, Aguadé M. A test of neutral molecular evolution based on nucleotide data. Genetics. 1987; 116: 153–159. PMID: 3110004

40. McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature. 1991; 351: 652–654. https://doi.org/10.1038/351652a0 PMID: 1904993

41. Wang RL, Hey J. The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the period locus. Genetics. 1996; 144: 1113–1126. PMID: 8913754

42. Loukas M, Krimbas CB, Mavragani-Tsipidou P, Kastritsis CD. Genetics of *Drosophila subobscura* populations. VIII. Allozyme loci and their chromosome maps. J Hered. 70: 17–26. PMID: 112146

43. Leung W, Shaffer CD, Cordonnier T, Wong J, Itano MS, Slawson Tempel EE, et al. Evolution of a distinct genomic domain in *Drosophila*: comparative analysis of the dot chromosome in *Drosophila*

melanogaster and *Drosophila virilis*. Genetics. 2010; 185: 1519–1534. https://doi.org/10.1534/genetics.110.116129 PMID: 20479145

44. Beh LY, Colwell LJ, Francis NJ. A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence. Proc Natl Acad Sci U S A. 2012; 109: E1063–E1071. https://doi.org/10.1073/pnas.1118678109 PMID: 22517748

45. Rand DM, Kann LM. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. Mol Biol Evol. 1996; 13: 735–748. PMID: 8754210

46. Smith NGC, Eyre-Walker A. Adaptive protein evolution in *Drosophila*. Nature. 2002; 415: 1022–1024. https://doi.org/10.1038/4151022a PMID: 11875568

47. Müller J, Verrijzer P. Biochemical mechanisms of gene regulation by Polycomb group protein complexes. Curr Opin Genet Dev. 2009; 19: 150–158. https://doi.org/10.1016/j.gde.2009.03.001 PMID: 19345089

48. Ketel CS, Andersen EF, Vargas ML, Strome S, Simon JA. Subunit contributions to histone methyltransferase activities of fly and worm Polycomb group complexes. Mol Cell Biol. 2005; 25: 6857–6868. https://doi.org/10.1128/MCB.25.16.6857-6868.2005 PMID: 16055700

49. Nekrasov M, Wild B, Müller J. Nucleosome binding and histone methyltransferase activity of *Drosophila* PRC2. EMBO Rep. 2005; 6: 348–353. https://doi.org/10.1038/sj.embor.7400376 PMID: 15776017

50. Nowak AJ, Alfieri C, Stirnimann CU, Rybin V, Baudin F, Ly-Hartig N, et al. Chromatin-modifying complex component *Nurf55*/*p55* associates with histones H3 and H4 and polycomb repressive complex 2 subunit *Su(z)12* through partially overlapping binding sites. J Biol Chem. 2011; 286: 23388–23396. https://doi.org/10.1074/jbc.M110.207407 PMID: 21550984

51. Rai AN, Vargas ML, Wang L, Andersen EF, Miller EL, Simon JA. Elements of the polycomb repressor SU(Z)12 needed for histone H3-K27 methylation, the interface with E(Z), and in vivo function. Mol Cell Biol. 2013; 33: 4844–4856. https://doi.org/10.1128/MCB.00307-13 PMID: 24100017

52. Vermaak D, Henikoff S, Malik HS. Positive selection drives the evolution of *rhino*, a member of the heterochromatin protein 1 family in *Drosophila*. PLoS Genet. 2005; 1: 96–108. https://doi.org/10.1371/journal.pgen.0010009 PMID: 16103923

53. Gibert JM, Karch F, Schlötterer C. Segregating variation in the Polycomb group gene cramped alters the effect of temperature on multiple traits. PLoS Genet. 2011; 7: e1001280. https://doi.org/10.1371/journal.pgen.1001280 PMID: 21283785

54. Voigt S, Laurent S, Litovchenko M, Stephan W. Positive selection at the *polyhomeotic* locus led to decreased thermosensitivity of gene expression in temperate *Drosophila melanogaster*. Genetics. 2015; 200: 591–599. https://doi.org/10.1534/genetics.115.177030 PMID: 25855066

55. Levine MT, Begun DJ. Evidence of spatially varying selection acting on four chromatin-remodeling loci in *Drosophila melanogaster*. Genetics. 2008; 179: 475–485. https://doi.org/10.1534/genetics.107.085423 PMID: 18245821

56. Betancourt AJ, Welch JJ, Charlesworth B. Reduced effectiveness of selection caused by a lack of recombination. Curr Biol. 2009; 19: 655–660. https://doi.org/10.1016/j.cub.2009.02.039 PMID: 19285399

57. Maynard Smith J, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. 1974; 23: 23–35. PMID: 4407212

58. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993; 134: 1289–303. PMID: 8375663

59. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966; 8: 269–294. PMID: 5980116

60. Akashi H. Inferring weak selection from patterns of polymorphism and divergence at silent sites in *Drosophila* DNA. Genetics. 1995; 139: 1067–1076. PMID: 7713409