Different Evolutionary Strategies To Conserve Chromatin Boundary Function in the Bithorax Complex

Fabienne Cleard,*[,1] Daniel Wolle,[†,1] Andrew M. Taverner,[‡] Tsutomu Aoki,[†] Girish Deshpande,[†] Peter Andolfatto,[‡,§] Francois Karch,* and Paul Schedl[†,2]

*Department of Genetics and Evolution, University of Geneva, 1211-4 Geneva, Switzerland, [†]Department of Molecular Biology, [‡]Lewis-Sigler Institute for Integrative Genomics, and [§]Department of Evolution and Ecology, Princeton University, Princeton, New Jersey 08544

ABSTRACT Chromatin boundary elements subdivide chromosomes in multicellular organisms into physically independent domains. In addition to this architectural function, these elements also play a critical role in gene regulation. Here we investigated the evolution of a *Drosophila* Bithorax complex boundary element called *Fab-7*, which is required for the proper parasegment specific expression of the homeotic *Abd-B* gene. Using a "gene" replacement strategy, we show that *Fab-7* boundaries from two closely related species, *D. erecta* and *D. yakuba*, and a more distant species, *D. pseudoobscura*, are able to substitute for the *melanogaster* boundary. Consistent with this functional conservation, the two known *Fab-7* boundary factors, Elba and LBC, have recognition sequences in the boundaries from all species. However, the strategies used for maintaining binding and function in the face of sequence divergence is different. The first is conventional, and depends upon conservation of the 8 bp Elba recognition sequence. The second is unconventional, and takes advantage of the unusually large and flexible sequence recognition properties of the LBC boundary factor, and the deployment of multiple LBC recognition elements in each boundary. In the former case, binding is lost when the recognition sequence is altered. In the latter case, sequence divergence is accompanied by changes in the number, relative affinity, and location of the LBC recognition elements.

KEYWORDS DNA binding proteins; boundary element; conservation of function; evolution; insulator

SPECIAL elements called chromatin boundaries or insulators play a central role in the architectural and functional organization of chromosomes in multi-cellular eukaryotes (Bartkuhn and Renkawitz 2008; Ghirlando *et al.* 2012; Van Bortle and Corces 2013; Chetverina *et al.* 2014; Maksimenko and Georgiev 2014; Matzat and Lei 2014; Schoborg and Labrador 2014). Genome wide chromatin immunoprecipitations (ChIPs) with known insulator proteins, together with chromatin conformation experiments have shown that insulators are a pervasive feature of eukaryotic genomes from Drosophila to humans (Holohan *et al.* 2007; Kim *et al.* 2007; Cuddapah *et al.* 2009; Jiang *et al.* 2009; Smith *et al.* 2009; Negre *et al.* 2010; Chen *et al.* 2012; Schwartz *et al.* 2012). As architectural elements, they physically interact with their neighbors to delimit topologically independent loops (or topologically associated domains: TADs). In humans, the average size of the loops defined by paired insulators is about 180 kb, while in *Drosophila*, loop size is smaller between 10 and 100 kb (Hou *et al.* 2012; Sexton *et al.* 2012; Rao *et al.* 2014). Coupled to their role in determining chromosome architecture, boundaries/insulators have genetic functions. The genetic activities that can be ascribed to most boundaries include an enhancer-blocking or insulator activity, a silencer-blocking or barrier activity, and, when paired in appropriate combinations, an ability to bring distant chromosomal DNA segments into close proximity (Ghirlando *et al.* 2012; Chetverina *et al.* 2014).

One of the most thoroughly characterized *Drosophila* insulators is the *Fab-7* boundary of the Bithorax complex (BX-C). The BX-C encodes three homeotic genes, *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*), that are responsible for specifying parasegments PS5-13 (Lewis 1978; Sanchez-Herrero 1985; Maeda and Karch 2006; Mihaly *et al.* 2006). Expression of the three genes is orchestrated by a ~300 kb DNA sequence that can be subdivided

doi: 10.1534/genetics.116.195586
Manuscript received September 2, 2016; accepted for publication December 12, 2016; published Early Online December 20, 2016.
Available freely online through the author-supported open access option.
Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.195586/-/DC1.
[1]These authors contributed equally to this work.
[2]Corresponding author: Princeton University, Lewis Thomas Laboratory, Washington Road, Princeton, NJ 08544. E-mail: pschedl@princeton.edu

into nine *cis*-regulatory domains. Each of these domains is responsible for regulating the expression of its target homeotic gene in a specific parasegment. For example, the four *Abd-B cis*-regulatory domains *iab-5*, *iab-6*, *iab-7*, and *iab-8* direct *Abd-B* expression in PS10, PS11, PS12, and PS13, respectively (Figure 1A) (Maeda and Karch 2006; Mihaly *et al.* 2006). In order for these *cis*-regulatory domains to properly specify parasegment identity, the domains must be able to function autonomously, and this is one of the functions of *Fab-7*, and of the other BX-C boundary elements. The *Fab-7* boundary is located between *iab-6* and *iab-7*, and is responsible for ensuring their autonomy (Figure 1A) (Gyurkovics *et al.* 1990; Galloni *et al.* 1993; Mihaly *et al.* 1997). Deletions of *Fab-7* exhibit a complex mixture of gain-of-function (GOF) and loss-of-function (LOF) phenotypes in PS11, which arise due to crosstalk between regulatory elements in the *iab-6* and *iab-7* regulatory domains. In addition to preventing crosstalk between adjacent regulatory domains, BX-C insulators must also be permissive (insulator bypass) for interactions between the regulatory domains and their homeotic gene targets (Hogga *et al.* 2001; Iampietro *et al.* 2008; Kyrchanova *et al.* 2015, 2016). For example, three of the *Abd-B* regulatory domains (*iab-5*, *iab-6*, and *iab-7*) must be able to bypass one or more of the *Abd-B* insulators to contact the *Abd-B* promoter (Figure 1A). Like other fly boundary elements, *Fab-7* and other BX-C insulators also function in transgene enhancer/silencer-blocking and insulator bypass assays.

Although all of the fly insulators that have been tested have both blocking and bypass activities, these functions are context dependent. This context dependence can be illustrated by experiments in which the *Fab-7* insulator in BX-C is replaced by the *gypsy*, *su*(*Hw*), and *scs* insulators (Hogga *et al.* 2001). While both of these heterologous insulators are able to block crosstalk between *iab-6* and *iab-7*, neither is permissive for interactions between the two downstream regulatory domains, *iab-5* and *iab-6*, and the *Adb-B* gene. Moreover, the ability of the *su*(*Hw*) insulator to prevent crosstalk between *iab-6* and *iab-7* is tissue specific, and is lost in the embryonic CNS. The *Fab-7* insulator has also been replaced with multimerized dCTCF sites and boundaries from BX-C (Kyrchanova *et al.* 2016). Surprisingly, only *Fab-8* was able to fully substitute for *Fab-7*. The context dependence evident in these replacement experiments is thought to arise because the architectural functions of insulators depend upon their ability to physically interact with insulators in their neighborhood (Cai and Shen 2001; Muravyova *et al.* 2001; Kyrchanova *et al.* 2007; Kyrchanova *et al.* 2008; Gohl *et al.* 2011; Fujioka *et al.* 2016). These physical contacts are thought to be mediated by protein–protein interactions between factors bound to neighboring insulators, and thus require at least some degree of compatibility between the factors associated with each insulator (Blanton *et al.* 2003; Pai *et al.* 2004; Bartkuhn and Renkawitz 2008; Cuartero *et al.* 2014; Vogelmann *et al.* 2014; Savitsky *et al.* 2016) .

As the context-dependent activity of boundary elements may impose unusual evolutionary constraints, we wondered what strategies might be used to conserve *Fab-7* boundary function. The *Fab-7* boundary in *D. melanogaster* spans a 1.2 kb DNA segment that includes three chromatin specific major nuclease hypersensitive sites, "*," HS1, and HS2 (Figure 1B) (Galloni *et al.* 1993; Hagstrom *et al.* 1996; Zhou *et al.* 1996; Rodin *et al.* 2007). While all of these sequences are required for full boundary function in transgene assays, replacement experiments have shown that HS1 alone is sufficient in an otherwise wild-type background (Wolle *et al.* 2015). Molecular and biochemical experiments have shown that the boundary activity of HS1 depends upon two subelements that function at different stages of development. The proximal half of HS1, pHS1, has boundary activity in early embryos in transgene assays, but does not have boundary activity during midembryogenesis or in adult flies (see Figure 1B). The distal half of HS1, dHS1, has weak to moderate blocking activity in early embryos, while it has strong blocking activity during midembryogenesis and in adult flies (Schweinsberg and Schedl 2004; Schweinsberg *et al.* 2004; Aoki *et al.* 2008, 2012; Wolle *et al.* 2015).

The developmentally restricted activities detected *in vivo* are recapitulated *in vitro* in electrophoretic mobility shift assay (EMSA) experiments with staged nuclear extracts. While several probes derived from pHS1 generate shifts in nuclear extracts prepared from both early and late embryos, one of the probes only gave a shift with early extracts (Aoki *et al.* 2008, 2012). This probe contains an 8 bp sequence, GGAATAAG, which is both necessary and sufficient for early boundary activity. It is recognized by the heterotrimeric Elba factor. Elba consists of two BEN domain DNA binding proteins, Elba1 and Elba2, that are bridged together by a third protein, Elba3. Elba insulating activity *in vivo* is developmentally restricted because Elba1 and Elba3 are midblastula transition genes.

Probes from dHS1 generate both early and late stage specific shifts. As illustrated in Figure 1B, dHS1 contains four GAGA factor (GAF) recognition motifs, GAGA3, GAGA4, GAGA5, and GAGA6. The early shift maps to an 8 bp palindromic sequence, CCAATTGG, located just proximal to the GAGA5 motif, and it is generated by the Elba factor. Shifts in late nuclear extracts are generated by an >800 kDa complex called the LBC, and are observed with probes spanning GAGA3, GAGA4, and GAGA5, but not GAGA6 (Wolle *et al.* 2015). The LBC has unusual sequence recognition properties. Whereas the Elba binding sites in *Fab-7* are conventional 8 bp sequences, the minimal recognition sequence for the LBC is >65 bp in length. Moreover, other than the GAGAG motifs, the probes spanning GAGA3, GAGA4, and GAGA5 have no obvious sequence similarity. The same is true for the LBC recognition sequences in the *Fab-8* boundary, and in an insulator-like element upstream of the *Abd-B* promoter (Wolle 2015). LBC binding to GAGA3 and GAGA4 in nuclear extracts requires the GAGAG motifs (GAGA5 motif was not tested), as does late boundary activity in transgene assays (Wolle *et al.* 2015). However, we were unable to identify any other short sequences in either GAGA3 or GAGA4 that are

critical for LBC binding, and we suspect that the other recognition motifs in these probes are likely to be redundant.

Here, we use a combination of genetic and biochemical approaches to study the evolutionary conservation of the *Fab-7* boundary. For these studies we have selected *Fab-7* boundaries from two species, *D. erecta* and *D. yakuba*, which are closely related to *D. melanogaster*, and a *Fab-7* boundary from a more distantly related species, *D. pseudoobscura*. Sequence alignments indicate that the *erecta* and *yakuba* boundaries have undergone relatively limited divergence, while the *Fab-7* boundary from *pseudoobscura* has diverged much more extensively, so that there is <50% identity. We used an insulator replacement strategy to test the functional properties of the *Fab-7* boundaries. As might be expected, the *Fab-7* boundaries from the two more closely related species can substitute for the *melanogaster Fab-7* boundary. Surprisingly, in spite of the significant changes in the sequence of the *pseudoobscura* boundary, it is also able to substitute. Consistent with conservation of function *in vivo*, the *Fab-7* boundaries from all three species have recognition sequences for both Elba and LBC. However, the differences in the DNA binding properties of these two factors are reflected in the evolution of their recognition sequences. Whereas the key 8 bp Elba binding motif in pHS1 is precisely preserved in all four species (as well as in other more distantly related *Drosophilids*; Aoki *et al.* 2008) the sequence of the three *melanogaster* LBC recognition elements have diverged, especially in *pseudoobscura*. In spite of this sequence divergence, boundary function and LBC binding to the dHS1 region are preserved, though the number and arrangement of LBC recognition elements is all three species differs from that found in *melanogaster*.

Materials and Methods

Electrophoretic mobility shift assay

Probe (1 pmol) was 5′ end labeled with (γ-^{32}P)ATP (MP Biomedicals) using T4 polynucleotide kinase (New England Biolabs) in a 50 μl total reaction at 37° for a period of 1 hr. Columns packed with Sephadex G-50 fine gel (Amersham Biosciences) were used to separate free ATP from the labeled probes. The volume of the sample eluted from the column was adjusted to 100 μl using deionized water so that the final concentration of the probe was 10 fmol/μl. Binding reactions were performed in a 20 μl volume consisting of 25 mM Tris-Cl (pH 7.4), 100 mM KCl, 1 mM EDTA, 0.1 mM dithiothreitol, 0.1 mM PMSF, 0.03 mg/ml bovine serum albumin, 10% glycerol, 0.25 mg/ml poly(dA-dT)/poly(dA-dT), and 20 μg of protein derived from *D. melanogaster* nuclear extract, or an equal volume of 360 mM nuclear extraction buffer. In some samples, unlabeled competitor DNA was included, so that the final concentration of the competitor would be in 5- to 100-fold excess. The reaction mixture containing the ^{32}P labeled DNA probes were incubated for 30 min at room temperature, with or without 20 μg of nuclear extracts derived from 0 to 6 hr (early) and 6 to 18 hr (late) embryos, and loaded onto a precleared 4% acrylamide:bis-acrylamide gels in 0.5× TBE–2.5% glycerol gel. Binding reactions were electrophoresed at 180 V for 3–4 hr with a 0.5× TBE-2.5% glycerol running buffer at 4°, dried and imaged using a Typhoon 9410 scanner and Image Gauge software and/or X-ray film.

Integration of modified Fab-7 elements within the Fab-7^{attP50} platform

As described in Wolle *et al.* (2015), integration of the KSattB^{FL}Fab7ry plasmid within *Fab-7^{attP50}* was achieved by injecting the plasmid into progeny from a cross of *Fab-7^{attP50}* males to females carrying a P(*y*$^+$; *nos-ΦC31-nos*) transgene inserted onto the X chromosome as source of integrase (Bischof *et al.* 2007). These females also carried two third chromosomes balancers MKRS/TM2. The emerging Go individuals were then crossed to TM2/MKRS flies, and the resulting integrants were recognized on the basis of their *ry*$^+$ eyes. The *ry*$^+$ and plasmids sequences were then flipped introducing a source of flippase (Golic and Linquist 1989), and selecting *ry*$^-$ progeny.

Data and reagent availability

All data required to confirm the main findings presented in the article are included with the article and in the Supplemental Material, Figure S1, Figure S2, Figure S3, Figure S4, Figure S5, Figure S6, Figure S7, Figure S8, and Figure S9. Fly strains, DNAs and plasmids are available upon request.

Results

Sequence evolution of the Fab-7 boundaries from D. erecta, D. yakuba, and D. pseuoobscura

Sequence alignment of the HS1 region of *D. melanogaster* (mel_HS1) with that of *D. erecta* (*ere*_HS1), *D. yakuba* (*yak*_HS1), and *D. pseudoobsucra* (*pse*_HS1), revealed that the HS1 region of *D. erecta* and *D. yakuba* share extensive sequence homology with that of *D. melanogaster* (Figure 2). The overall identity for *erecta* is 88%, while that for *yakuba* is 83%. By contrast, though blocks of homology between the *pseudoobscura* and *melanogaster* sequences are present, there are also many nucleotide substitutions and gaps. Accordingly, the overall identity of *pseudoobscura* Fab-7 with *melanogaster* is only 44%.

The proximal half of the *melanogaster* HS1 (pHS1) contains the Elba recognition sequence (CCAATAAG) (Aoki *et al.* 2008) flanked by two GAGAG motifs, GAGA1 and GAGA2, that are known to contribute to early insulator activity of the boundary in *D. melanogaster* (Schweinsberg *et al.* 2004). The 8 bp Elba sequence is located at one edge of a 17 bp sequence that is conserved in all four species. By contrast, the two GAGAG motifs are not well conserved. GAGA1 is present in *erecta* and *yakuba*, but not *pseudoobscura*, while GAGA2 is absent in all three species (Figure 2); *pseudoobscura* does, however, have a GAGAG motif located 1 bp closer to the Elba binding sequence than GAGA2.

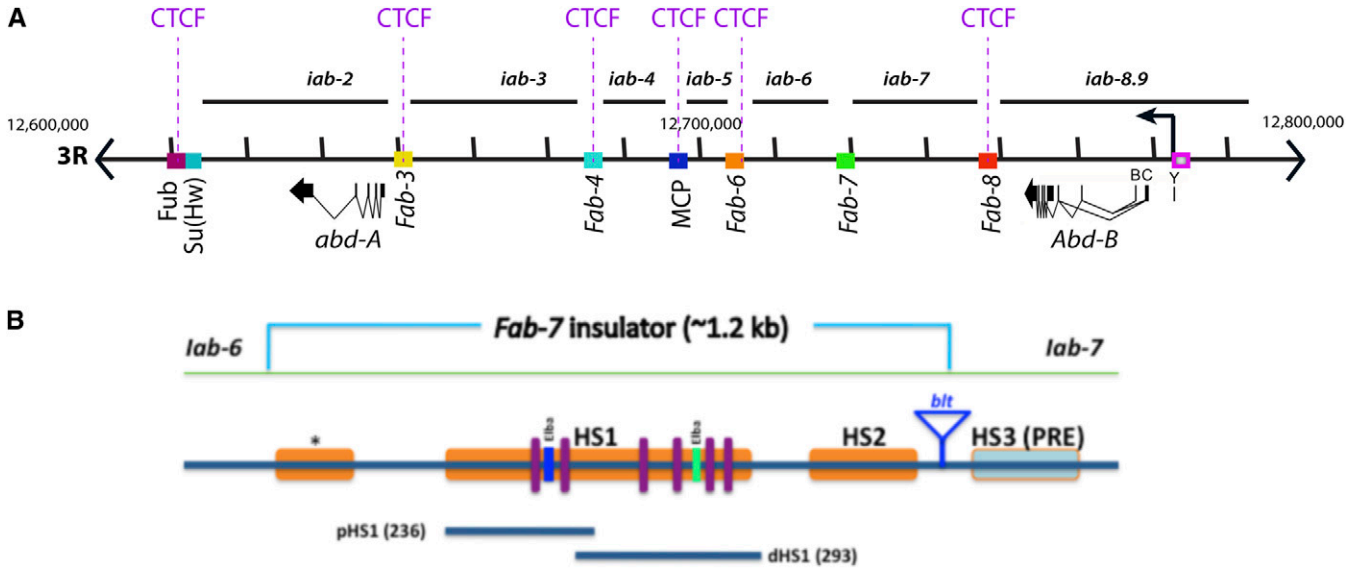

Figure 1 The distal two-thirds of the Bithorax complex, and the *Fab-7* boundary. (A) Genomic map showing the *cis*-regulatory regions within the Bithorax Complex (BX-C) located on the 3R chromosome. The two BX-C genes, *abd-A* and *Abd-B*, along with their associated *cis*-regulatory domains (*iab-2*–*iab-9*) are indicated. BX-C insulators that are interposed between the *cis*-regulatory domains are depicted as rectangular boxes of different colors. (B) A schematic drawing of the Fab-7 insulator (~1.2 kb). DNase I-hypersensitive regions, "*," HS1 and HS2 are shown as rectangular orange boxes. Within HS1, the six binding sites for GAGA factor (GAF) are shown as purple lines (GAGAG). The Elba binding site (CCAATAAG) in pHS1 is shown as a light red line, while the palindromic Elba site (CCAATTGG) in dHS1 is indicated by a green line.

The "late" region of *melanogaster*, dHS1, has four GAGAG motifs 3–6, and a palindromic Elba binding sequence, CCAATTGG, located 2 bp proximal to GAGA5 (Figure 2). In this region of *erecta* and *yakuba*, there are five GAGAG motifs, while *pseudoobscura* has nine GAGAG motifs. Of the four *melanogaster* GAGAG motifs, only GAGA4 and GAGA6 are present in all species. GAGA3 is found in *erecta*, but is absent in *yakuba* and *pseudoobscura*. Instead, these two species have a single (*yakuba*), or multiple GAGA (*pseudoobscura*), motifs located on the proximal side of dHS1. All three species differ from *melanogaster* in that they have a second GAGAG motif, just distal to GAGA4. (In *melanogaster*, this sequence is GAGAA.) While the GAGA5 motif is present in *erecta* and *yakuba*, it is absent in *pseudoobscura*. Instead, there is a nearby GAGAG motif. As for the palindromic Elba sequence, it is retained only in *yakuba*.

Fab-7 boundaries from D. erecta, D. yakuba, and D. pseudoobscura are functional in D. melanogaster

To test the functional properties of the *erecta*, *yakuba*, and *pseudoobscura* boundaries, we took advantage of a phiC31-based integration platform (*Fab-7*^{attp50}), in which the region spanning the minor (*) and three major nuclease hypersensitive sites (HS1–HS3) has been deleted, and replaced by a minimal attP target site for the phiC31 integrase (Mihaly *et al.* 1997; Bischof *et al.* 2007; Iampietro *et al.* 2010; Wolle *et al.* 2015). We introduced the tested boundary sequences into the attP target site using an attB-based integration plasmid carrying the boundary sequences, and a *rosy*$^{+}$ (*ry*$^{+}$) gene to detect integration events. Once integrated, the *ry*$^{+}$ marker and plasmid DNA are inserted within BX-C between the end of HS3, and the remainder of the *iab-7* region. These foreign sequences are then excised by using the FRT/FLP site-specific recombination system (Golic and Linquist1989).

Using this scheme, we have shown that the DNA segment spanning the first major nuclease hyper-sensitive HS1 (Figure 1) contains all the sequences required for *Fab-7* boundary activity in an otherwise wild-type background. Indeed, Figure 2B shows that a deletion of HS1 (*Fab-7*$^{\Delta HS1}$) results in a phenotype that is indistinguishable from *Fab-7*^{2}—the founding member of the *Fab-7* boundary mutation (class 2). These deletions exhibit a complex mixture of GOF and LOF phenotype in PS11 that corresponds to A6 in the adult fly. As shown in Figure 3B, most of the cuticle in A6 is absent due to the GOF transformation into A7 (which is absent in WT males). However, there are several small patches of tergite (white circles) and sternite (black circle) that have an A5 (PS10) identity (Wolle *et al.* 2015). Relevant to this study, mutations of all six HS1 GAGAG motifs (plus 2 GAGA-like motifs) result in the same phenotype (Figure 3C).

We tested the functioning of the HS1 sequences from the other species by rescuing the *melanogaster* ΔHS1 deletion. As shown in Figure 3, D and E, HS1 sequences from both *erecta* and *yakuba* block cross-talk between *iab-6* and *iab-7*, and rescue the GOF phenotypes of the *Fab-7*^{attp50} mutant (Figure 3, D and E). There is also no evidence of the LOF phenotypes (A6–A5) that are seen when *Fab-7* is replaced by *scs* or *su* (*Hw*). In adult wild-type males, the A6 sternite and tergite are morphologically distinct from that of A5. The A6 sternite has a banana shape, and lacks bristles, while the A5 sternite has a more circular shape, and is covered in bristles. For the tergite, trichome hairs are only found along the ventral and anterior

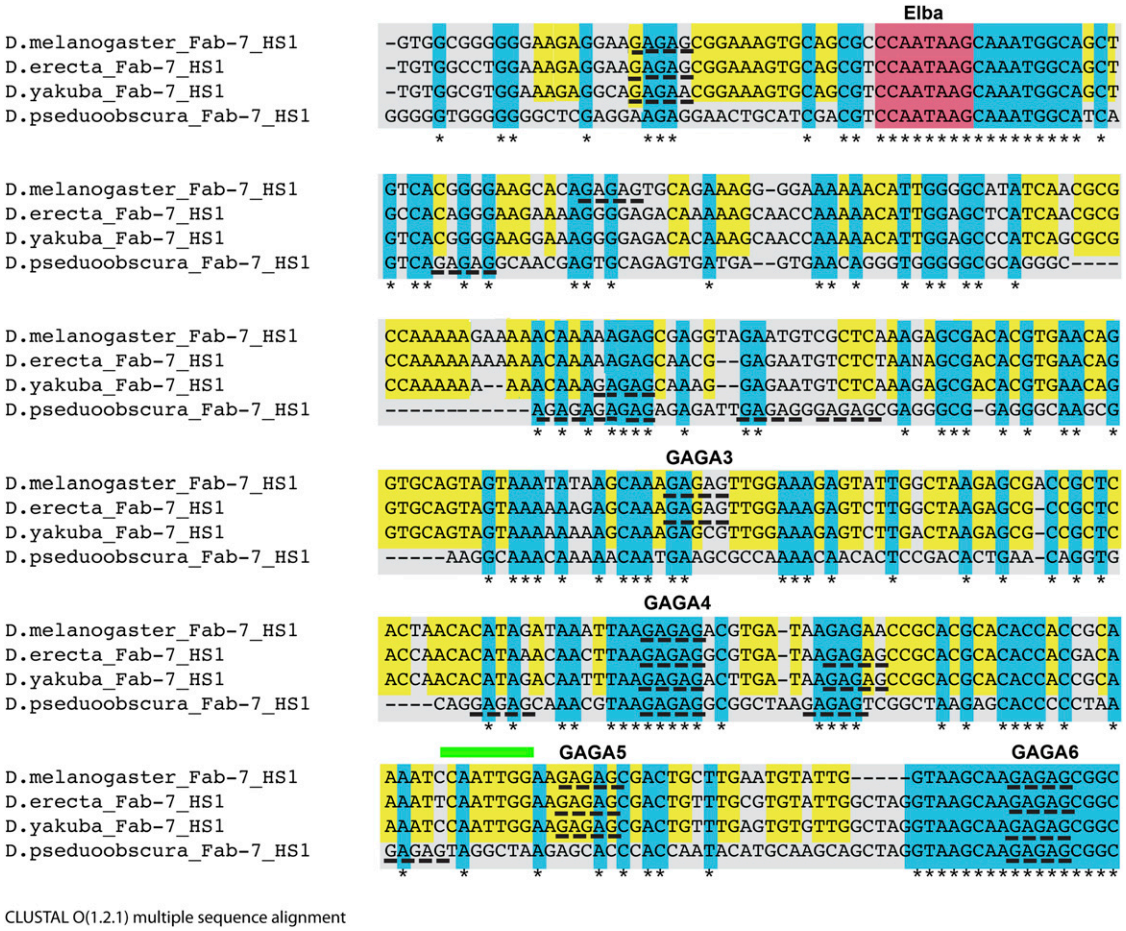

Figure 2 Sequence comparison of the HS1 sequence of *Fab-7* boundaries from *melanogaster*, *erecta*, *yakuba*, and *pseudoobscura*. The *Fab-7* HS-1 sequences of *D. melanogaster* and three other *Drosophila* species were assembled, and stacked up with the ClustalO program to align the high-homology region conserved among the four species. Sequences that are homologous in all four *Drosophila* species are highlighted in blue and marked with an asterisk. Sequences that are homologous in *melanogaster*, *erecta*, and *yakuba*, but not *pseudoobscura* are highlighted in yellow. The 8 bp Elba recognition sequence in dHS1 is aligned and highlighted in light red. The 5 bp GAGAG motifs (GAGAG) are underlined with a dashed line and labeled in dHS1. The Elba palindrome next to GAGA5 is indicated by a green bar.

edges of the A6 tergite, while the A5 tergite is completely covered with trichomes. As can be seen in Figure 3, D and E, the A6 segment in both *Fab-7*$^{+(erecta)}$ and *Fab-7*$^{+(yakuba)}$ resembles that seen in wild type *melanogaster*; the sternites have a banana shape and lack bristles, while the tergites only have trichomes along the ventral and anterior edges. In spite of the much more extensive sequence divergence, the *pseudoobscura* insulator also rescues *Fab-7*^{attp50} mutant. However, unlike either *Fab-7*$^{+(erecta)}$ and *Fab-7*$^{+(yakuba)}$, not all *Fab-7*$^{+(pseudoobcura)}$ males are fully wild type. As illustrated in the example shown in Figure 3F, the A6 sternite in the *Fab-7*$^{+(pseudoobcura)}$ substitution has a few bristles. This weak *iab-6* LOF is not fully penetrant, and is observed in only a small fraction (~2–5%) of the *Fab-7*$^{+(pseudoobcura)}$ adult males.

Elba binds to pHS1 sequences from D. erecta, D. yakuba and D. pseudoobscura

The fact that *Fab-7* HS1 sequences from *erecta*, *yakuba*, and *pseudoobscura* can fully or largely substitute for the *melanogaster* HS1 region poses the question of whether their insulator activity in *melanogaster* is generated by the known *Fab-7* stage specific boundary factors. In the case of the early boundary activity of the pHS1 region, this seems likely, as all four species have an 8 bp Elba motif that is part of a conserved 17 bp sequence block. This sequence block is embedded in a larger 35 bp region that has 77% identity between the four species. To confirm that this pHS1 sequence is recognized by the Elba factor, we performed EMSA experiments with staged nuclear extracts, and a 100 bp probe spanning this conserved sequence from each species. As observed for *melanogaster*, the *erecta*, *yakuba*, and *pseudoobscura* pHS1 probes give multiple shifts. As shown in Figure 4, one of these shifts (which is a doublet: black arrowhead) is enriched in nuclear extracts from 0 to 6 hr. Supershift experiments with rabbit antibodies directed against the Elba1 subunit indicate that this shift is generated by the Elba factor. Other shifts besides Elba are also observed. One of these (a faster migrating shift: blue arrowhead) appears to be present in all four species, though the yield is higher with the *erecta* and *pseudoobscura* probes. The other (more slowly

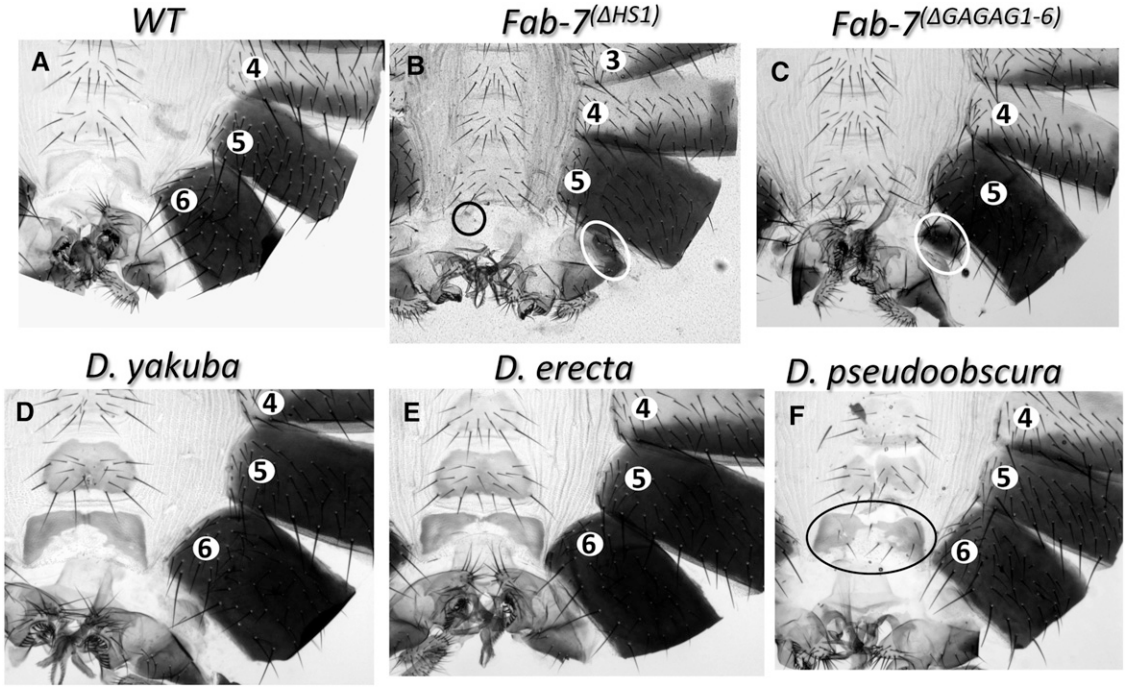

Figure 3 HS1 is functionally conserved in other fly species. The ΦC31 *Fab-7*^{attP50} platform was used to introduce HS1 sequences derived from *D. erecta*, *D. yakuba*, and *D. pseudoobscura*. The first two panels show the cuticles of the posterior abdominal segments of WT *Fab-7* (A) and of the ΔHS1 deletion (B). As indicated by the circles, there is residual tergite (white circle) and sternite (black circle) tissue in the ΔHS1 males. This residual tissue has an A5 (PS10) not an A6 (PS11) identity. See, for example, the bristles on the "A6" sternite. In wild type (A), the sternite has no bristles. A similar phenotype is observed when all of the GAGA motifs in HS1 are mutated (C). (D–F) HS1 replacements with sequences from *erecta*, *yakuba*, and *pseudoobscura*, as indicated (see text for full description). As can be seen in these panels, the HS1 sequence from these species rescues the ΔHS1 mutant. Note that *pseudoobscura* male fly shown in (F) has a few bristles on the sternite, indicative of an LOF transformation toward an A5 identity. This phenotype is only observed in a small fraction of the *Fab-7*$^{+(pseudoobcura)}$ males. Cuticles were prepared as described in Duncan (1982).

migrating: green arrowhead) shift can be detected with the *melanogaster*, *pseudoobscura*, and probably also *erecta* probes, but not with *yakuba*.

dHS1 sequences have LBC recognition elements

While the Elba recognition sequence is completely conserved in all four species, the sequences for the recognition elements for the late boundary factor, LBC, have changed to a lesser or greater extent in *erecta, yakuba, and pseudoobscura*, respectively. To test the DNA binding properties of the dHS1 sequences from these species, we designed a set of probes that span the "late" region of the *Fab-7* insulators. In previous studies, we subdivided the dHS1 sequence of *melanogaster* into five overlapping probes of about 80 bp long, pdHS1A, GAGA3, GAGA4, GAGA5, and GAGA6 (see Figure 5). While the GAGA3–6 probes each span one of the four dHS1 GAGAG motifs, pdHS1A is derived from the proximal end of dHS1, and, unlike the other probes, does not contain a GAGAG motif. LBC shifts were detected in late nuclear extracts for GAGA3, GAGA4, and GAGA5 (Figure S1; and see Wolle *et al.* 2015). By contrast, GAGA6 did not give an LBC shift, while only a very weakly labeled LBC shift was observed for pHS1A. As expected, probe five was shifted by the Elba factor in early nuclear extracts (Figure 1; and see Wolle *et al.* 2015).

For *erecta* and *yakuba*, we generated five overlapping probes, *erecta2-6* and *yakuba2-6*, that correspond closely in their length and relative position to the *melanogaster* probes pdHS1A, GAGA3–6 respectively (Figure 5A and Figure S2). In the case of *pseudoobsucra*, the dHS1 homology region is shorter, and, unlike the *melanogaster* group species, has diverged substantially in sequence. The five *pseudoobsucra* probes were designed so that they matched the corresponding *melanogaster*-group probes as closely as possible, and, with the exception of *pseudo3*, positioned the *pseudoobscura* GAGAG motifs toward the center of the probe (Figure 5A). Since the LBC recognition sequence is unusually large, and its binding to *melanogaster* sequences can be affected by the endpoints and/or length of the probes, we also generated variants of several of the *erecta*, *yakuba*, and *pseudoobsucra* probes in which the end-points and/or length were altered (see Figure S3).

Consistent with the replacement experiments, the shifts generated by the *erecta* and *yakuba* probes from the dHS1 region display obvious similarities to those observed for *melanogaster* probes. Like the corresponding *melanogaster* probes (Figure S1) *erecta3/yakuba3*, *erecta4/yakuba4*, and *erecta5/yakuba5*, give LBC-like shifts, while *erecta6/yakuba6* do not (Figure 5, B and C). The shift has the same stage specificity as the *melanogaster* LBC, and, as is further documented

below, is generated by the LBC. Supporting the identification of these *erecta* and *yakuba* LBC recognition sequences, an LBC shift is also detected with the *erecta* and *yakuba* variants, *erecta3-2*, *erecta4-2*, *yakuba3-2*, and *yakuba4-2*, which differ in their endpoints and lengths (Figure S4, A and B). Interestingly, unlike their *melanogaster* and *erecta* counterparts, neither *yakuba3* nor its variant *yakuba3-2* (Figure S2) has a GAGAG motif, yet both of these *yakuba* probes are shifted by the LBC (Figure 5C and Figure S4B). Another difference is that both *erecta* and *yakuba* have a fourth LBC recognition element that it is not found in *melanogaster*. Thus, while a stable LBC complex is not formed with the *melanogaster* probe from the proximal side of dHS1, pdHS1A (see Wolle *et al.* 2015), an LBC shift is observed for the corresponding *erecta2* and *yakuba2* probes (Figure 5, B and C: see also Figure S4, A and B). While the yield of the *erecta2* LBC shift is less than that of *yakuba2*, *yakuba2* has a GAGAG motif that is absent in *erecta2*. Finally, the early shift detected with GAGA5, and *yakuba5* is not observed for *erecta5*. This fits with the sequence differences as GAGA5 and *yakuba5* both have the CCAATTGG palindrome, while, in *erecta*, the sequence is TCAATTGG (Figure 2).

As might be expected from the more extensive sequence divergence of the *pseudoobscura* dHS1 region and the changes in the arrangement of GAGAG motifs, the DNA binding activities detected with the *pseudoobscura* probes differ most from the corresponding *melanogaster* probes (Figure 5D and Figure S4C). First, like *erecta5*, we did not detect a shift in early nuclear extracts with the *pseudo5* probe. Second, only two of the *pseudoobscura* probes, *pseudo2* and *pseudo4*, give prominent LBC shifts. One of these, *pseudo4*, corresponds by position, and the presence of several short conserved sequence blocks, to the *melanogaster* GAGA4 probe that is shifted by the LBC. By contrast, the *melanogaster* probe corresponding by position to *pseudo2*, pdHS1A, does not give a stable LBC shift. On the other hand, though they differ substantially in sequence, the corresponding *erecta* and *yakuba* probes, *erecta2* and *yakuba2*, also give an LBC shift. Moreover, like *yakuba2*, *pseudo2* has GAGAG motifs. Somewhat surprisingly, the *pseudoobscura* probe corresponding to GAGA3, *pseudo3*, does not give an LBC shift. Though *pseudo3* lacks the *melanogaster* GAGA3 motif, it does have a GAGAG motif at its proximal end. Moreover, the corresponding *yakuba* probe, *yakuba3*, does not have any GAGAG motifs, yet it can generate the LBC shift. Similar results were obtained with the larger (88 bp) variant, *pseudo3-2*, which extends distally and includes a second GAGAG motif (see Figure S4C). Finally, though *pseudo5* had two GAGAG motifs (including the motif corresponding to GAGA5), it gives only a very weak LBC shift.

LBC recognition elements from different species cross- compete

It was important to confirm that the LBC shifts seen with the *erecta*, *yakuba*, and *pseudoobscura* probes in late nuclear extracts are, in fact, generated by the LBC. For this purpose, we

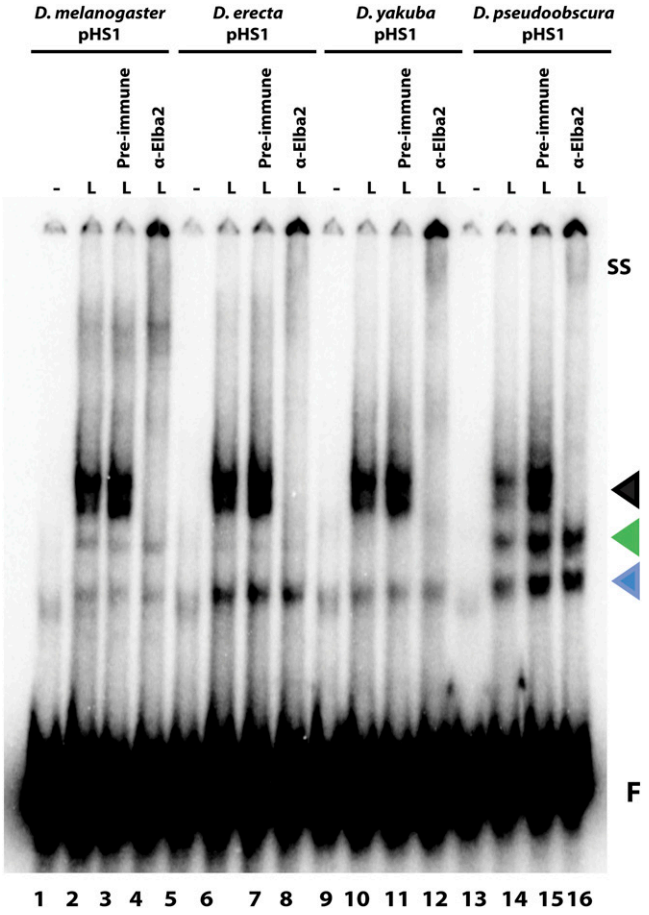

Figure 4 Elba recognition sequence in pHS1 is conserved. We used staged nuclear extracts (0–6 and 6–18 hr), and 100 bp probes spanning the Elba1 recognition sequence in the pHS1 region of *Fab-7* boundaries of *melanogaster*, *erecta*, *yakuba*, and *pseudoobscura*. In the experiment shown here, the four probes were incubated with 0–6 hr extracts, either without serum, with preimmune rabbit serum, or with rabbit antibody against Elba2, as indicated. The Elba shift is marked by a black arrowhead.

used cross-competition experiments to show that probes from these three species, which generate an LBC-like shift, are able to compete with each other, and are competed by probes from *melanogaster* that are known to shifted by the LBC. Figure 6A shows *erecta4* competed with the following unlabeled DNAs: a GAGA containing sequence from the *hsp70* promoter (*hsp70 GAGAG*), itself, *erecta2*, and *melanogaster* GAGA4. While addition of excess cold *hsp70 GAGAG* has no effect on the LBC shift, the *erecta4* shift is competed by itself, by *erecta2*, and by GAGA4. As shown in Figure S5A, cross-competition experiments with three other *erecta* probes, *erecta3-2*, *erecta4-2*, and *erecta5* indicate that they compete not only with each other, but are also competed by the *melanogaster* probe GAGA3-65. Similar results were obtained in cross-competition experiments with *yakuba2* (Figure 6B) and *yakuba4* (Figure S5B). The LBC shift generated by *yakuba2* is competed by itself, *yakuba4*, and GAGA4. Similarly, *yakuba4* is competed by excess *yakuba3-2*, *yakuba2*,

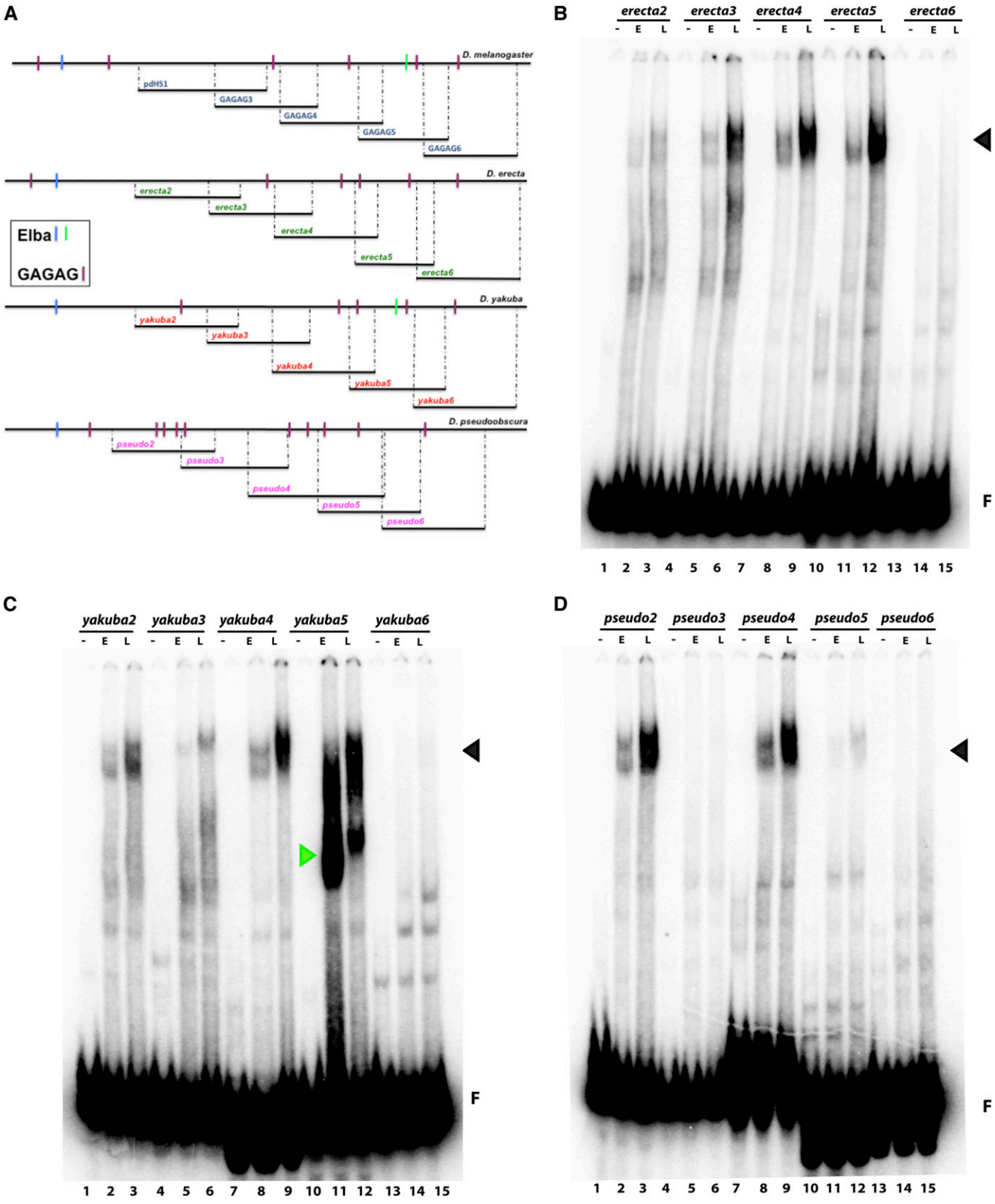

Figure 5 LBC recognition elements are present in the dHS1 region of the *Fab-7* insulators from *D. erecta*, *D. yakuba*, and *D. pseudoobscura*. (A) A schematic of the HS1 region from the *melanogaster*, *erecta*, *yakuba*, and *pseudoobscura Fab-7* boundaries, showing the location of the probes used for EMSAs. Also indicated are the recognition sites for Elba (blue and green), and GAF (light red). To assay for LBC binding, five probes were designed from the dHS1 region of each species. As indicated in the schematic, the relative position and length of the *erecta* and *yakuba* probes closely matches that of the corresponding *melanogaster* probes. Because of the more extensive sequence divergence of *pseudoobscura*, and the shorter length of the

yakuba4-2, and by GAGA4. Finally, *pseudo2* is competed by itself, by *pseudo4*, and GAGA4, but not by *hsp70 GAGAG* (Figure 6C).

The LBC complex detected with the erecta, yakuba and pseudoobscura probes contains GAF, Mod(mdg4) and E(y) 2

As a second approach for confirming the identity of the LBC shifts observed with the different *erecta*, *yakuba*, and *pseudoobscura* probes, we used supershift experiments to assay for the presence of the three factors known to be associated with the >700 kDa LBC complex. As might be expected from the effects of GAGAG motif mutations on DNA binding and boundary activity, one of these is GAF (GAGA factor or Trl). There are two major GAF isoforms, 518 aa and 581 aa in *melanogaster* (Benyajati *et al.* 1997; Greenberg and Schedl 2001; Adkins *et al.* 2006). They share an N-terminal BTB protein interaction domain and an internal zinc finger DNA binding domain, but have different C-terminal domains. Another component of the LBC is a factor previously implicated in the functioning of the *gypsy* insulator, Mod(mdg4). There are 31 predicted Mod(mdg4) isoforms, which are generated by alternative splicing (Dorn *et al.* 2001; Dorn and Krauss 2003). All share the same N-terminal BTB domain, but have different C-terminal domains. Of these, 27 isoforms have unique N-terminal FLYWCH type Zn-finger DNA binding domains. While DNA binding by Mod(mdg4) isoforms containing one of these FLYWCH Zn-fingers has not yet been demonstrated, the *Caenorhabditis elegans* PEB-1 protein has been shown to bind DNA via its FLYWCH domain (Beaster-Jones and Okkema 2004). The BTB domains of both GAF and Mod(mdg4) interact with themselves (Bonchuk *et al.* 2011). The GAF BTB domain tends to generate GAF dimers, while an octomer appears to be the preferred Mod(mdg4) oligomer. In addition to mediating self-assembly, the GAF and Mod(mdg4) BTB domains also interact with each other. Finally, the third component of the LBC is E(y)2. As is the case for Mod(mdg4), this conserved transcription factor is important for the proper functioning of the *gypsy* insulator (Kurshakova *et al.* 2007).

In the experiment shown in Figure 7, *erecta3*, *erecta4*, *yakuba2*, *yakuba3*, *yakuba3-2*, *pseudo2*, and *pseudo4* were incubated with late nuclear extracts alone, in the presence of either control rabbit or rat serum, or with antibodies directed against GAF. As would be predicted for LBC association, the shifts generated by these probes are all supershifted by GAF antibodies. Significantly, two of these probes, *yakuba3* and the *yakuba3-2* variant, do not contain a GAGAG motif, yet can bind the LBC, and are supershifted with GAF antibodies. To further confirm the identity of the LBC, we tested for the presence of the FLYWCH Mod(Mdg4) isoforms known to be in the LBC, PT (67.2) (see below), and for E(y)2. Figure 8 shows that the shifts generated by the *erecta4* and *pseudo4* probes are supershifted by the Mod(mdg4) and E(y)2 antibodies.

Discussion

In the studies reported here, we have tested the insulator function of *Fab-7* boundaries from three species, *erecta*, *yakuba*, and *pseudoobscura*. The first two are close relatives of *melanogaster*, and this is reflected in a high degree of sequence conservation. The similarity of the HS1 region of these two species is 88 and 83%, respectively. By contrast, the HS1 region of the more distantly related *pseudoobsdcura* has diverged extensively, and the similarity is <50%. Given their limited sequence divergence from *melanogaster*, it was not surprising to find that the *Fab-7* boundaries of both *erecta* and *yakuba* fully substitute for *melanogaster*. Interestingly, with the exception of a minor bristle phenotype that has very low penetrance, the much more highly diverged *pseudoobscura Fab-7* insulator is also able to substitute for *melanogaster*.

One reason that the *Fab-7* HS1 sequence from these other fly species are able to substitute for *melanogaster* is that they are recognized by the two known developmentally restricted *Fab-7* boundary factors, Elba and LBC. However, the mechanisms that ensure that the boundary function of these two factors is preserved in the face of evolutionary divergence are markedly different.

The first mechanism is conventional. It depends upon a strict conservation of the DNA binding properties of the boundary factor, in this case the Elba factor, and of its DNA recognition sequence (CCAATAAG and CCAATTGG). The BEN DNA binding domains in Elba1 and Elba2 are highly conserved in all three species (see Figure S6). In the case of *erecta* and *yakuba*, the domains are identical to those of *melanogaster*, or have only a single amino acid substitution. While there are 5 and 8 aa substitutions in the *pseudoobscura* Elba1 and Elba2 BEN domains, respectively, they are conservative changes, and also are not in regions of the DNA binding domain that are expected to interact directly with the DNA recognition sequence. Though the linker protein Elba3, and the N-terminal domains of Elba1 and Elba2 are less well conserved, this is not unusual for protein sequences that mediate interactions with

homology region, a precise matching of the position and length is not possible. However, the probes span roughly the same region as the corresponding *melanogaster* group probes, and were designed to position GAGAG motifs toward the center of the probe. (B) EMSAs were performed with five 80 bp overlapping probes, *erecta2-5* as indicated. Each probe was incubated with early and late nuclear extracts depicted as "E" and "L," respectively. Black arrowhead, LBC; F, free probe. (C) EMSAs were performed with five 80 bp overlapping probes, *yakuba2-5* as indicated. Each probe was incubated with early and late nuclear extracts depicted as "E" and "L," respectively. Black arrowhead, LBC; green arrowhead, Elba factor; F, free probe. (D) EMSAs were performed with five 80 bp overlapping probes, *pseudoobscura2-5* as indicated. Each probe was incubated with early and late nuclear extracts depicted as "E" and "L," respectively. Black arrowhead, LBC; F, free probe.

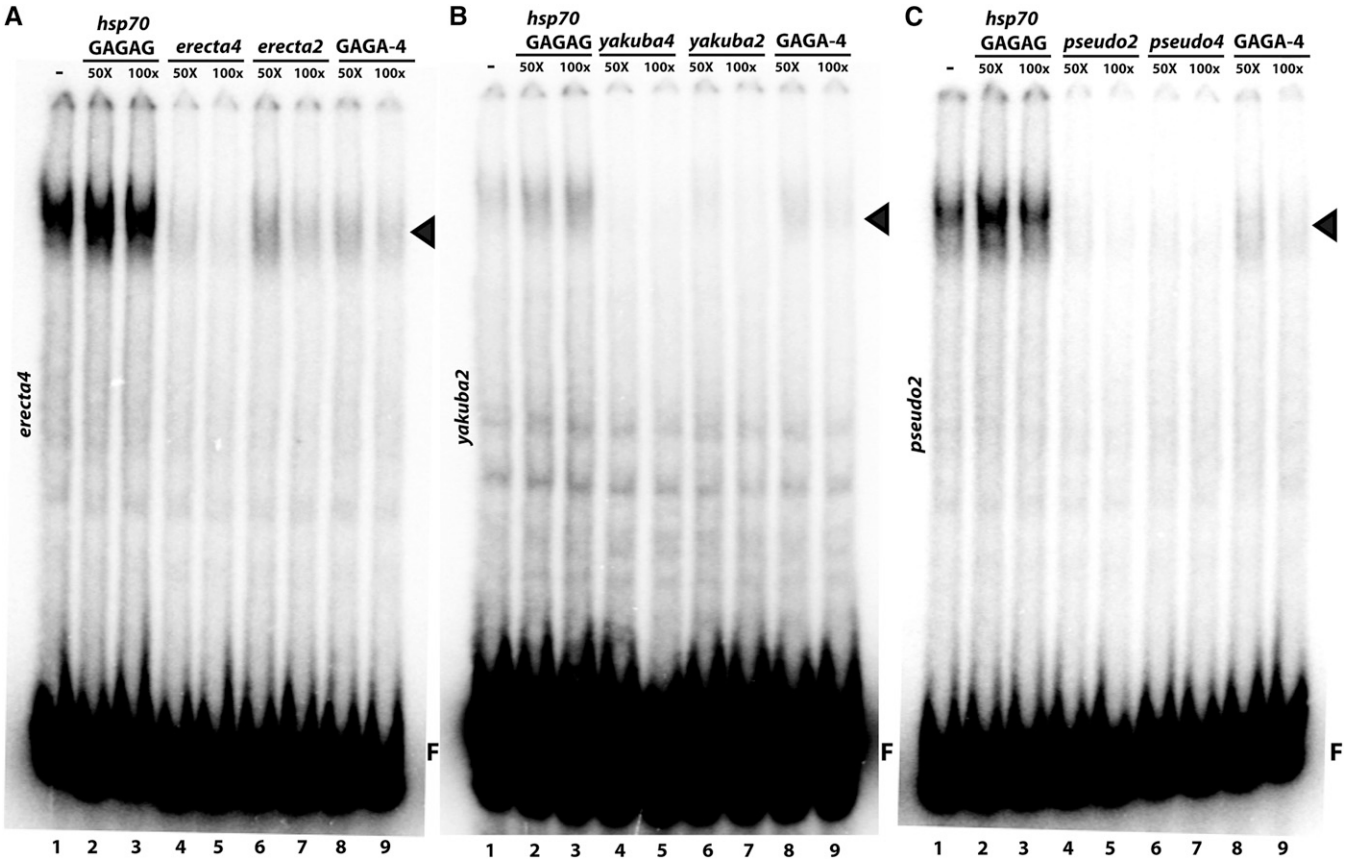

Figure 6 Competition experiments indicate that LBC binds to probes from the dHS1 region of *erecta*, *yakuba*, and *pseudoobscura Fab-7* insulator. (A) EMSAs performed with probe *erecta4* and nuclear extracts from late (6–18 hr) embryos, "L," in the absence (lane 1) and presence (lanes 2–9) of unlabeled cold competitors (*hsp70* GAGAG, *erecta4*, *erecta2*, and GAGA4), as indicated above the lanes. Black arrowhead, LBC; F, free probe. (B) EMSAs performed with probe *yakuba2* and nuclear extracts from late (6–18 hr) embryos, "L," in the absence (lane 1) and presence (lanes 2–9) of unlabeled cold competitors (*hsp70* GAGAG, *yakuba4*, *yakuba2*, and GAGA4), as indicated above the lanes. Black arrowhead, LBC; F, free probe. (C) Gel shift analysis performed with probe *pseudo2* and nuclear extracts from late (6–18 hr) embryos, "L," in the absence (lane 1) and presence (lanes 2–9) of unlabeled cold competitor (*hsp70* GAGAG, *pseudo2*, *pseudo4*, and GAGA4), as indicated above the lanes. Black arrowhead, LBC; F, free probe.

other proteins, as these interactions often rely on relatively short motifs in unstructured segments. Thus, one would expect that the heterotrimeric Elba complex is present in these three other species, and that its sequence recognition properties are likely to be identical to the *melanogaster* complex. Consistent with this conclusion, the 8 bp Elba recognition sequence is present on the proximal side (at roughly the same position as *melanogaster*) in the *Fab-7* boundaries of these species. In fact, the Elba recognition sequences is also conserved in the *Fab-7* boundaries of much more distant *Drosophilids*, such as *D. virilis* and *D. grimshawi* (Aoki *et al.* 2008).

That sequence conservation is important for retaining Elba factor binding is illustrated by the second Elba binding site in the *melanogaster* HS1 sequence, which is located close to GAGA5 in dHS1. In *melanogaster*, this Elba sequence differs from the Elba site in pHS1 (CCAATAAG) in that it is a palindrome (CCAATTGG). This palindrome is conserved in *yakuba*, and Elba binding is observed in nuclear extracts from early embryos. In *erecta*, there is a single base change from CCAATTGG to TCAATTGG. This sequence difference is sufficient to abrogate Elba binding in *melanogaster* nuclear extracts. While we have not tested stage specific nuclear extracts from *erecta*, the high degree of sequence conservation evident in the *erecta* BEN domains of Elba1 and Elba2 provides a strong argument that the *erecta* Elba factor also does not recognize this sequence in the *erecta Fab-7* boundary. Much more extensive sequence alterations are evident in *pseudoobscura*, and this region of the *pseudoobscura* dHS1 (GTAGGCTA) does not resemble any of the known Elba recognition sequences, and it is not recognized by the *melanogaster* Elba factor. Again since the BEN DNA binding domains of the *pseudoobscura* Elba1 and Elba2 proteins are well conserved, the *pseudoobscura* Elba factor will probably not bind to this sequence either. From this perspective, the Elba factor is like other DNA binding proteins. The DNA binding domain changes very slowly, and this, together with compensatory mutations, tends to ensure that the sequence recognition specificity is the same even in distantly related species. Thus, in order for the functioning of the regulatory element to be preserved, the DNA recognition sequence must be conserved, even if the overall sequence of the element diverges extensively as is the case in *pseudoobscura*.

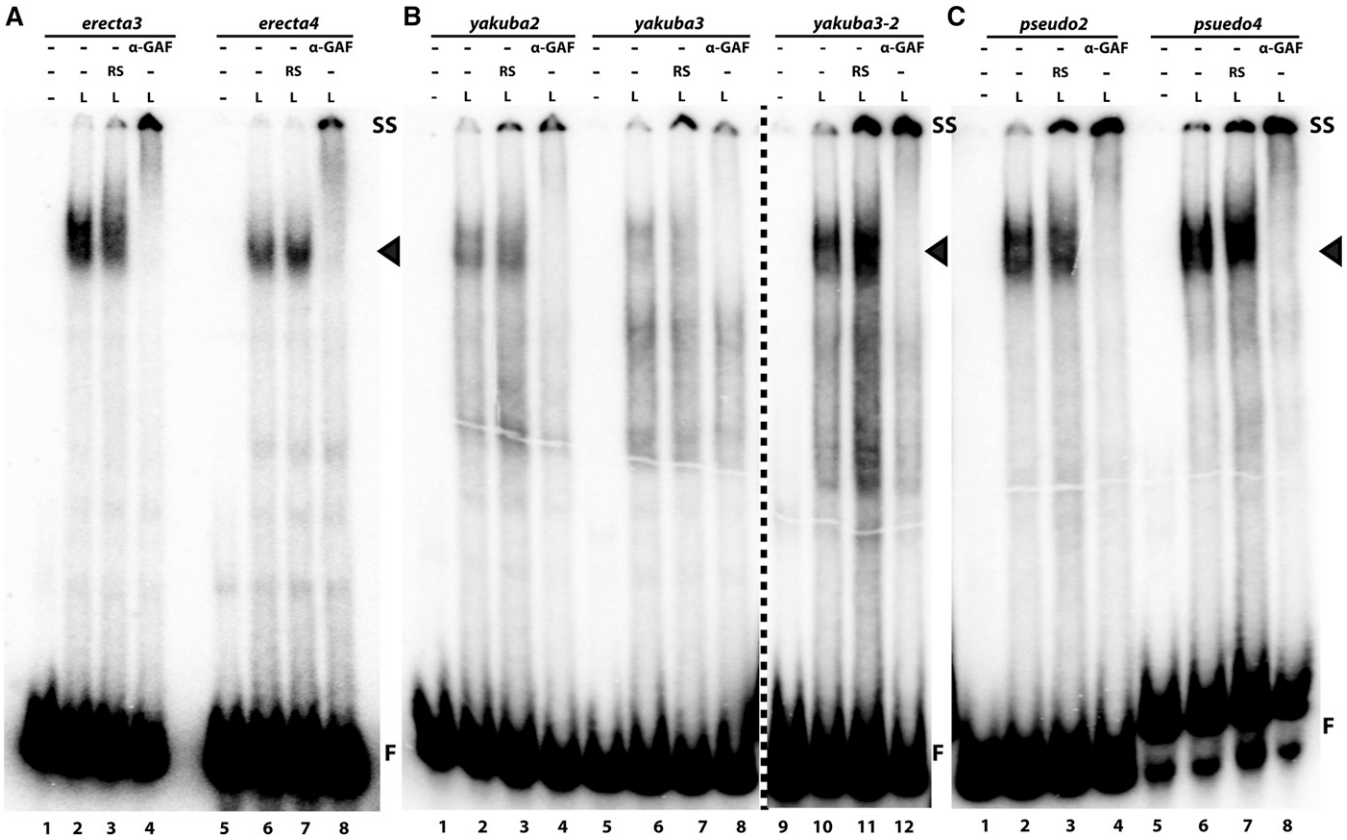

Figure 7 The LBC shifts detected with probes from *erecta*, *yakuba*, and *pseudoobscura* contains GAF. (A) *Fab-7*^{erecta} probes *erecta3* and *erecta4* were incubated with late nuclear extracts together with rat serum "RS" (lanes 3 and 7), or a rat polyclonal GAF antibody "α-GAF" (lanes 4 and 8). (B) *Fab-7*^{yakuba} fragments *yakuba2*, *yakuba3*, and *yakuba3-2* were incubated with late nuclear extracts together with rat serum "RS" (lanes 3, 7, and 11) or a rat polyclonal GAF antibody "α-GAF" (lanes 4, 8, and 12). Antibody supershifts are indicated with the letters "SS." Black arrowhead, LBC; F, free probe. (C) *Fab-7*^{psedoobscura} fragments *pseudo2* and *pseudo4* were incubated with late nuclear extracts together with rat serum "RS" (lanes 3 and 7) or a rat polyclonal GAF antibody "α-GAF" (lanes 4 and 8). L, late nuclear extracts. Antibody supershifts are indicated with the letters "SS." Black arrowhead, LBC; F, free probe.

In the second case, conservation of function is unconventional, and depends upon a novel mechanism to compensate for the potentially deleterious effects of sequence drift. This mechanism takes advantage of the unusually flexible sequence recognition properties of the LBC, and combines this flexibility with the deployment of multiple and malleable LBC recognition elements in each insulator. As was case for the Elba factor, the sequences of one of the two (known) DNA binding proteins in the LBC components, GAF is well conserved. The two major GAF isoforms, 519 aa and 582 aa, share the BTB protein interaction domain and the zinc finger DNA binding domain. Both of these domains have identical sequences in all of the species we examined (Figure S7). Thus, the DNA binding activity of GAF in *erecta*, *yakuba*, and *pseudoobscura* is expected to be identical to that of *melanogaster*. For several reasons, the situation is more complicated for the Mod(mdg4) protein isoforms (Figure S8 and Figure S9). For one, we do not know for certain which of the 31 isoforms are present in the LBC. Based on mass spectrometry of proteins associated with GAF in 0–12 hr nuclear extracts, at least 14 different Mod(mdg4) isoforms are good candidates for LBC components (see Figure S9; and D. Lomaev, personal communication). Of course, it is possible that several other Mod(mdg4) isoforms are present in the LBC at this stage of development, but were not detected by mass spectrometry. It is also possible that one or more of the isoforms found in the GAF IPs could be components of, as yet unknown, GAF-Mod(mdg4) complexes, but not the LBC. On the other hand, since the Mod(mdg4) BTB domain assembles into octomers, and this domain is also responsible for Mod(mdg4)-GAF interactions, the most plausible idea at present is that many, if not all, of the Mod(mdg4) isoforms associated with GAF in nuclear extracts will be included in LBC complexes. This means that LBCs could have different combinations of Mod(mdg4) isoforms. The number of possible combinations would depend upon the relative abundance and number of different isoforms that are present in a specific cell, and thus could differ from one cell type to the next. As 12 of the 14 isoforms have FLYWCH DNA binding domains, individual LBCs would be expected to have somewhat different sequence preferences. We suspect that this variability, together with the fact that each complex is expected

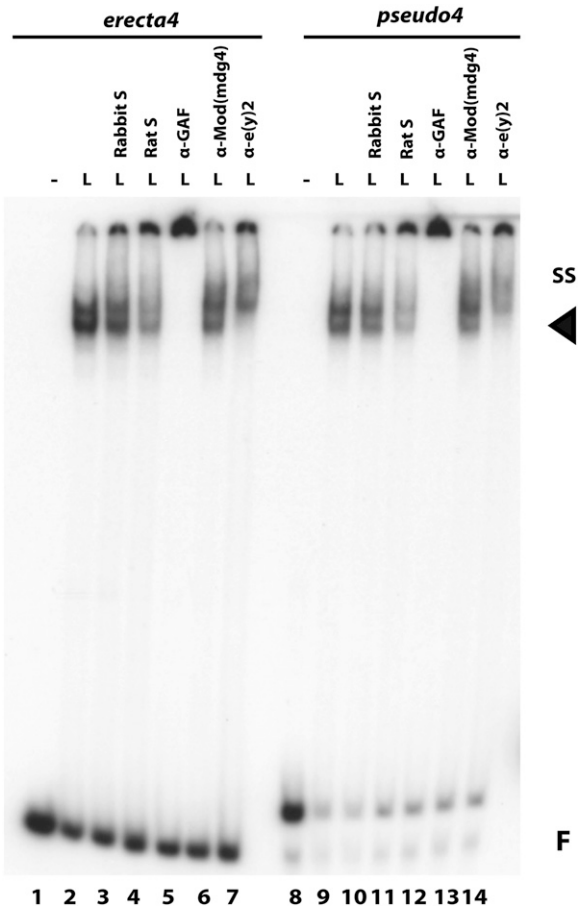

Figure 8 The LBC shift contains GAF, Mod(mdg4), and e(y)2. Supershift experiments were used to determine whether the shifts observed with the *erecta4* and *pseudo4* probes are generated by the LBC. In addition to GAF, the LBC contains Mod(mdg4) and *e(y)2* (Wolle *et al.* 2015). One of the Mod(mdg4) isoforms is PT (D. Lomaev, personal communication).

to have multiple DNA binding proteins with different specificities, contributes to the unusual DNA binding properties of the LBC.

As indicated in Figure S9, the sequence divergence of nine of the 12 FLYWCH domains resembles that of the Elba1 and Elba2 BEN domains—the FLYWCH domains in the *erecta* and *yakuba* isoforms are either identical to those of *melanogaster*, or have one or two amino acid substitutions, while the corresponding FLYWCH domains in *pseudoobscura* have between zero and seven amino acid substitutions. Thus, like Elba and GAF, the DNA recognition properties of these nine GAF-associated Mod(mdg4) FLYWCH isoforms are expected to be similar in all four species. On the other hand, three of the isoforms have diverged more extensively, especially in *pseudoobscura*, and thus could preferentially bind to different sequences in different species. As for the Mod(mdg4) isoforms that were not detected in GAF immunoprecipitates, all but two fall into the same category as the Elba and GAF DNA binding domains and their FLYWCH domains would be expected to recognize the same sequences in all four species.

Besides its unusually long minimal binding sequence, the other novel feature of the *melanogaster* LBC is its flexible sequence recognition properties. This flexibility is evident from a comparison of the three *melanogaster* recognition elements, GAGA3, GAGA4, and GAGA5. Other than the presence of a GAGAG motif, there are no obvious sequence similarities between them. This is also true for the different *erecta*, *yakuba*, and *pseudoobsucra* probes that are shifted by the LBC-like the corresponding *melanogaster* probes; their sequences are dissimilar to each other. Although LBC binding to the dHS1 region of *Fab-7* is conserved, the divergence in sequence between the four species alters both the relative affinity and distribution of the LBC recognition elements (see Figure 9). These changes in affinity and distribution of recognition elements arises from a series of deleterious (for binding) and compensatory mutations that are spread over a region spanning >200 bp or most of dHS1. In *melanogaster*, dHS1 has three recognition elements, GAGA3, GAGA4, and GAGA5, that form stable LBC complexes. Of the three, GAGA4 has a marginally higher affinity for the LBC than GAGA3, while both have a higher affinity than GAGA5 (Figure 9). All three recognition elements are found in *erecta* and *yakuba*. However, binding and competition experiments indicate that the relative affinities of each of these elements differs from that observed for the corresponding *melanogaster* element (Figure 9). For example, LBC binding to *yakuba3* is reduced compared to the equivalent *melanogaster* probe, GAGA3. One obvious explanation for the reduced affinity of *yakuba3* is that it no longer has a GAGAG motif (see Figure 5A and Figure S2). Compensating for the changes in *yakuba3*, LBC binding to *yakuba5* is greatly enhanced compared to the *melanogaster* probe GAGA5. This is also true for the corresponding *erecta* probe, *erecta5*. As indicated in Figure 5 (see also Figure 9 and Figure S2), *yakuba5* and *erecta5* have a second GAGAG motif that is not found in *melanogaster*.

In addition to changes in relative affinity, both *erecta* and *yakuba* have acquired a fourth recognition element that maps to the *erecta*2 and *yakuba2* probes, respectively (Figure 5A). The corresponding *melanogaster* probe, pdHS1A, does not have a GAGAG motif, and, for this reason, it was not surprising that it does not give stable LBC shift (Wolle *et al.* 2015). Indeed, *yakuba2* differs from pdHS1A in that it has acquired a GAGAG motif. However, this is most likely not the only difference between pdHS1A and *yakuba2* that is important. For one, GAGA6, *erecta6*, and *yakuba6*, all have a GAGAG motif, but are not shifted by the LBC. Additionally, the GAGAG motif is clearly not essential since the LBC also binds to both *erecta2* and *yakuba3*, which lack the GAF recognition sequence. At this point, it is not clear what other substitutions in *erecta2* enable it to form a stable LBC complex in the absence of the GAGAG motif. Though the *erecta2* sequence is slightly more related to *yakuba2* (91% identity) than it is to *melanogaster* (88% identity), a combination of both common and unique substitutions are probably important for generating this new LBC recognition element in each species.

Further evidence that the flexible sequence recognition properties of the LBC permits a conservation of function in

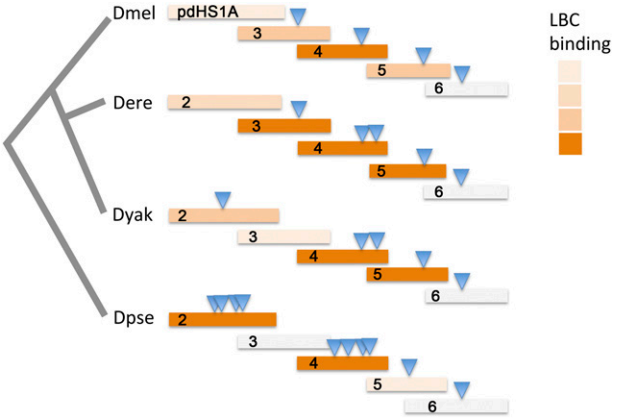

Figure 9 Schematic showing relative LBC affinity for each of the dHS1 probes from the four species. Position and name of the probes are indicated in the schematic. Probe identity as in Figure 5A. GAGAG motifs are indicated by blue arrowheads. Relative affinity indicated by the coloring of the probe. White, no binding; Orange, highest affinity.

spite of extensive sequence divergence comes from *pseudoobscura*. While the different *erecta* and *yakuba* dHS1 probes share substantial homology with the corresponding *melanogaster* probes (86–94% identity), this is not the case for *pseudoobscura*. The most closely related probe, *pseudo4*, has only 56% identity with the corresponding *melanogaster* probe, GAGA4, while *pseudo2* (30%), *pseudo3* (48%), and *pseudo5* (45%) are all <50%. Of these, only the most, *pseudo4*, and the least conserved, *pseudo2*, probes form stable complexes with the LBC. With respect to *pseudo2*, it is worth noting that, though the LBC also binds to the equivalent *erecta* and *yakuba* probes (*erecta2* and *yakuba2*), their sequences are just about as dissimilar to *pseudo2* as the *melanogaster* probe pdHS1A. Finally, the LBC binds poorly to *pseudo5* and not at all to *pseudo3*, even though both of these probes have GAGAG motifs. In fact, like *erecta5* and *yakuba5*, which are high affinity recognition elements, *pseudo5* has two GAGAG motifs (Figure 5 and Figure S2).

It is of interest to compare the evolution of the LBC recognition elements in *Fab-7* with the evolution of transcription factor binding sites in fly enhancer elements (Ludwig *et al.* 1998, 2000; Wittkopp 2006; Swanson *et al.* 2011). In the examples of rapidly evolving enhancers that have been studied in detail, function is conserved, in spite of changes in the number, physical arrangement, and relative affinity of the transcription factor binding sites. Two factors seem to be important in conserving function. One is that sequence changes that compromise a transcription factor binding site in one region of the enhancer are compensated for by mutations elsewhere in the enhancer, which generate sequences that are sufficiently close to the consensus recognition sequence either for the same factor, or for another complementary factor, to confer enhancer function. Second, there are typically multiple binding sites for key factors, and this redundancy provides a buffer in the event that one of the sites is mutated so that it is no longer functional. These deleterious and compensatory sequence alterations are piecemeal, subtracting and adding binding sites for individual transcription factors in a manner that maintains function. Additionally, many enhancers utilize clustered, low affinity binding sites, rather than sites that match the optimal sequence for factor binding. For example, Crocker *et al.* (2015) found that, for enhancers regulated by the Hox protein Ultrabithorax (Ubx), clustered low affinity Ubx binding sites increased the robustness of the enhancer response and its selectivity for the Ubx protein.

While similar principles apply for the evolution of the various LBC recognition elements in the *Fab-7* boundary, there are also important differences. Instead of reshuffling many short (4–10 bp), typically well defined, binding motifs for individual transcription factors, the recognition sequences that are lost or regenerated are ≥65 bp in length. Moreover, the sequences that could potentially constitute an LBC binding site seem to be a good deal less restrictive or specific than would be the case for typical transcription factors, which require binding sites that at least partially match a relatively short consensus motif. A good example of this permissiveness would be the *erecta2*, *yakuba2*, and *pseudo2* probes. Even though the *erecta* and *yakuba* sequences bear little resemblance to *pseudo2*, all three are bound by the LBC. An additional difference (which applies not only to the LBC, but also to the Elba factor) is, unlike enhancers, which are designed to switch from an "off" to an "on" state with high selectivity, most (but not all) boundaries are expected to function ("on") irrespective of the cell type or developmental stage. For this reason, one would expect that optimized binding sites, rather than low affinity sites would typically be deployed. In general, this requirement would be expected to increase the evolutionary demands for retaining consensus, or near consensus, binding sites if the same factors are utilized. Indeed, for the Elba site in pHS1, the sequences is conserved not only in the four species studied here, but also in much more distantly related species like *D. virilis*. Likewise, for the *melanogaster* Elba site in dHS1, a single base change eliminates Elba binding in *erecta*. As for the LBC, all four boundaries have at least two optimized sites, though the sequences of these sites and their locations differ. The fact that optimized, or near optimized, binding sites are likely to be selected for in the case of boundary elements like those in BX-C makes the LBC quite remarkable in that it can bind to sequences that seemingly bear little resemblance to each other.

Acknowledgments

We would like to thank Robert Maeda, Henrik Gyurkovics, and Ella Preger-Ben Noon for insightful discussions. We also thank Eva Favre, Benjamin Barandun, Jorge Faustino, and Gordon Grey for excellent technical assistance. This work was supported by grants from the Donation Claraz, the State of Geneva, and the Swiss National Fund for Research to F.K., and by grants from the National Institutes of Health (NIH)

to P.S. (GM043432) and P.A. (GM083228). P.S. would also like to acknowledge support from a grant to the Gene Biology Institute by the Russian Federation Ministry of Education and Science (14.B25.31.0022).

Literature Cited

Adkins, N. L., T. A. Hagerman, and P. Georgel, 2006 GAGA protein: a multi-faceted transcription factor. Biochem. Cell Biol. 84: 559–567.

Aoki, T., S. Schweinsberg, J. Manasson, and P. Schedl, 2008 A stage-specific factor confers Fab-7 boundary activity during early embryogenesis in Drosophila. Mol. Cell. Biol. 28: 1047–1060.

Bartkuhn, M., and R. Renkawitz, 2008 Long range chromatin interactions involved in gene regulation. Biochim. Biophys. Acta 1783(11): 2161–2166

Beaster-Jones, L., and P. G. Okkema, 2004 DNA binding and in vivo function of C. elegans PEB-1 require a conserved FLYWCH motif. J. Mol. Biol. 339: 695–706.

Benyajati, C., L. Mueller, N. Xu, M. Pappano, J. Gao *et al.*, 1997 Multiple isoforms of GAGA factor, a critical component of chromatin structure. Nucleic Acids Res. 25(16): 3345–3353.

Bischof, J., R. K. Maeda, M. Hediger, F. Karch, and K. Basler, 2007 An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc. Natl. Acad. Sci. USA 104: 3312–3317.

Blanton, J., M. Gaszner, and P. Schedl, 2003 Protein:protein interactions and the pairing of boundary elements in vivo. Genes Dev. 17: 664–675.

Bonchuk, A., S. Denisov, P. Georgiev, and O. Maksimenko, 2011 Drosophila BTB/POZ domains of "ttk group" can form multimers and selectively interact with each other. J. Mol. Biol. 412: 423–436.

Cai, H. N., and P. Shen, 2001 Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. Science 291: 493–495.

Chen, H., Y. Tian, W. Shu, X. Bo, and S. Wang, 2012 Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. PLoS One 7: e41374.

Chetverina, D., T. Aoki, M. Erokhin, P. Georgiev, and P. Schedl, 2014 Making connections: insulators organize eukaryotic chromosomes into independent cis-regulatory networks. BioEssays 36(2): 163–172.

Crocker, J., N. Abe, L. Rinaldi, A. P. McGregor, N. Frankel *et al.*, 2015 Low affinity binding site clusters confer Hox specificity and regulatory robustness. Cell 160: 191–203.

Cuartero, S., U. Fresán, O. Reina, E. Planet, and M. L. Espinàs, 2014 Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function. EMBO J. 33(6): 637–647.

Cuddapah, S., R. Jothi, D. E. Schones, T. Y. Roh, K. Cui *et al.*, 2009 Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res. 19: 24–32.

Dorn, R., and V. Krauss, 2003 The modifier of mdg4 locus in Drosophila: functional complexity is resolved by trans splicing. Genetica 117: 165–177.

Dorn, R., G. Reuter, and A. Loewendorf, 2001 Transgene analysis proves mRNA trans-splicing at the complex mod(mdg4) locus in Drosophila. Proc. Natl. Acad. Sci. USA 98: 9724–9729.

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

Fujioka, M., H. Mistry, P. Schedl, and J. B. Jaynes, 2016 Determinants of chromosome architecture: insulator pairing in cis and in trans. PLoS Genet. 12(2): e1005889.

Galloni, M., H. Gyurkovics, P. Schedl, and F. Karch, 1993 The bluetail transposon: evidence for independent cis-regulatory domains and domain boundaries in the bithorax complex. EMBO J. 12: 1087–1097.

Ghirlando, R., K. Giles, H. Gowher, T. Xiao, Z. Xu *et al.*, 2012 Chromatin domains, insulators and the regulation of gene expression. Biochim. Biophys. Acta 1819(7): 644–651.

Gohl, D., T. Aoki, J. Blanton, G. Shanower, G. Kappes *et al.*, 2011 Mechanism of chromosomal boundary action: roadblock, sink, or loop? Genetics 187: 731–748.

Golic, K. G., and S. Lindquist, 1989 The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome. Cell 59: 499–509.

Greenberg, A. J., and P. Schedl, 2001 GAGA factor isoforms have distinct but overlapping functions in vivo. Mol. Cell. Biol. 21 (24): 8565–8574.

Gyurkovics, H., J. Gausz, J. Kummer, and F. Karch, 1990 A new homeotic mutation in the Drosophila bithorax complex removes a boundary separating two domains of regulation. EMBO J. 9: 2579–2585.

Hagstrom, K., M. Muller, and P. Schedl, 1996 Fab-7 functions as a chromatin domain boundary to ensure proper segment specification by the Drosophila bithorax complex. Genes Dev. 10: 3202–3215.

Hogga, I., J. Mihaly, S. Barges, and F. Karch, 2001 Replacement of Fab-7 by the gypsy or scs insulator disrupts long-distance regulatory interactions in the Abd-B gene of the bithorax complex. Mol. Cell 8: 1145–1151.

Holohan, E. E., C. Kwong, B. Adryan, M. Bartkuhn, M. Herold *et al.*, 2007 CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex. PLoS Genet. 3: e112.

Hou, C., L. Li, Z. S. Qin, and V. G. Corces, 2012 Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol. Cell 48: 471–484.

Iampietro, C., F. Cleard, H. Gyurkovics, R. K. Maeda, and F. Karch, 2008 Boundary swapping in the Drosophila bithorax complex. Development 135: 3983–3987.

Iampietro, C., M. Gummalla, A. Mutero, F. Karch, and R. K. Maeda, 2010 Initiator elements function to determine the activity state of BX-C enhancers. PLoS Genet. 6: e1001260.

Jiang, N., E. Emberly, O. Cuvier, and C. M. Hart, 2009 Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in Drosophila melanogaster links BEAF to transcription. Mol. Cell. Biol. 29: 3556–3568.

Kim, T. H., Z. K. Abdullaev, A. D. Smith, K. A. Ching, D. I. Loukinov *et al.*, 2007 Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 128: 1231–1245.

Kurshakova, M., O. Maksimenko, A. Golovnin, M. Pulina, S. Georgieva *et al.*, 2007 Evolutionarily conserved E(y)2/Sus1 protein is essential for the barrier activity of Su(Hw)-dependent insulators in Drosophila. Mol. Cell 27: 332–338.

Kyrchanova, O., D. Chetverina, O. Maksimenko, A. Kullyev, and P. Georgiev, 2008 Orientation-dependent interaction between Drosophila insulators is a property of this class of regulatory elements. Nucleic Acids Res. 36: 7019–7028.

Kyrchanova, O., V. Mogila, D. Wolle, J. P. Magbanua, R. White *et al.*, 2015 The boundary paradox in the Bithorax complex. Mech. Dev. 138(Pt. 2): 122–132.

Kyrchanova O., V. Moglia, G. Deshpande, A. Parshikov, F. Cléard *et al.*, 2016 Functional dissection of the blocking and bypass activities of the Fab-8 boundary in the Drosophila bithorax complex. PLoS Genet. 12(7): e1006188.

Lewis, E. B., 1978 A gene complex controlling segmentation in Drosophila. Nature 276: 565–570.
Ludwig, M. Z., N. H. Patel, and M. Kreitman 1998 Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. Development 125: 949–958.
Ludwig, M. Z., C. Bergman, N. H. Patel, and M. Kreitman, 2000 Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 403: 564–567.
Maeda, R. K., and F. Karch, 2006 The ABC of the BX-C: the bithorax complex explained. Development 133: 1413–1422.
Maksimenko, O., and P. Georgiev, 2014 Mechanisms and proteins involved in long distance interactions. Front Genet. 5: 28.
Matzat, L. H., and E. P. Lei, 2014 Surviving an identity crisis: a revised view of chromatin insulators in the genomics era. Biochim. Biophys. Acta 1839(3): 203–214.
Mihaly, J., I. Hogga, J. Gausz, H. Gyurkovics, and F. Karch, 1997 In situ dissection of the Fab-7 region of the bithorax complex into a chromatin domain boundary and a Polycomb-response element. Development 124: 1809–1820.
Mihaly, J., S. Barges, L. Sipos, R. Maeda, F. Cleard *et al.*, 2006 Dissecting the regulatory landscape of the Abd-B gene of the bithorax complex. Development 133: 2983–2993.
Muravyova, E., A. Golovnin, E. Gracheva, A. Parshikov, T. Belenkaya *et al.*, 2001 Loss of insulator activity by paired Su(Hw) chromatin insulators. Science 291: 495–498.
Negre, B., S. Casillas, M. Suzanne, E. Sánchez-Herrero, M. Akam *et al.*, 2005 Conservation of regulatory sequences and gene expression patterns in the disintegrating Drosophila Hox gene complex. Genome Res. 15(5): 692–700.
Negre, N., C. D. Brown, P. K. Shah, P. Kheradpour, C. A. Morrison *et al.*, 2010 A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet. 6: e1000814.
Pai, C. Y., E. P. Lei, D. Ghosh, and V. G. Corces, 2004 The centresomal protein CP190 is a component of the gypsy chromatin insulator. Mol. Cell 16(5): 737–748.
Rao, S. S., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov *et al.*, 2014 A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159: 1665–1680.
Rodin, S., O. Kyrchanova, E. Pomerantseva, A. Parshikov, and P. Georgiev, 2007 New properties of Drosophila fab-7 insulator. Genetics 177: 113–121.
Sanchez-Herrero, E., I. Vernos, R. Marco, and G. Morata, 1985 Genetic organization of Drosophila bithorax complex. Nature 313: 108–113.
Savitsky, M., M. Kim, O. Kravchuk, and Y. B. Schwartz, 2016 Distinct roles of chromatin insulator proteins in control of the Drosophila bithorax complex. Genetics 202(2): 601–617.
Schoborg, T., and M. Labrador, 2014 Expanding roles of chromatin insulators in nuclear architecture, chromatin organization and genome function. Cell. Mol. Life Sci. 71: 4089–4113.
Schwartz, Y. B., D. Linder-Basso, P. V. Kharchenko, M. Y. Tolstorukov, M. Kim *et al.*, 2012 Nature and function of insulator protein binding sites in the Drosophila genome. Genome Res. 22: 2188–2198.
Schweinsberg, S. E., and P. Schedl, 2004 Developmental modulation of Fab-7 boundary function. Development 131: 4743–4749.
Schweinsberg, S. E., K. Hagstrom, D. Gohl, P. Schedl, R. P. Kumar *et al.*, 2004 The enhancer-blocking activity of the Fab-7 boundary from the Drosophila bithorax complex requires GAGA-factor-binding sites. Genetics 168: 1371–1384.
Sexton, T., E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc *et al.*, 2012 Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148: 458–472.
Smith, S. T., P. Wickramasinghe, A. Olson, D. Loukinov, L. Lin *et al.*, 2009 Genome wide ChIP-chip analyses reveal important roles for CTCF in Drosophila genome organization. Dev. Biol. 328: 518–528.
Swanson, C. I., D. B. Schwimmer, and S. Barolo, 2011 Rapid evolutionary rewiring of a structurally constrained eye enhancer. Curr. Biol. 21: 1186–1196.
Van Bortle, K., and V. G. Corces, 2013 The role of chromatin insulators in nuclear architecture and genome function. Curr. Opin. Genet. Dev. 23(2): 212–218.
Vogelmann, J., A. Le Gall, S. Dejardin, F. Allemand, A. Gamot *et al.*, 2014 Chromatin insulator factors involved in long-range DNA interactions and their role in the folding of the Drosophila genome. PLoS Genet. 10(8):e1004544.
Wittkopp, P. J., 2006 Evolution of cis-regulatory sequence and function in Diptera. Heredity 97: 139–147.
Wolle, D., 2015 Developmental modulation of insulators in Drosophila melanogaster. Ph.D. Thesis, Princetoin University, Princeton.
Wolle, D., F. Cleard, T. Aoki, G. Deshpande, P. Schedl *et al.*, 2015 Functional requirements for Fab-7 boundary activity in the bithorax complex. Mol. Cell. Biol. 35: 3739–3752.
Zhou, J., S. Barolo, P. Szymanski, and M. Levine, 1996 The Fab-7 element of the bithorax complex attenuates enhancer-promoter interactions in the Drosophila embryo. Genes Dev. 10: 3195–3201.

Communicating editor: J. A. Birchler