Tissue-specific versus pleiotropic enhancers within the *bric-a-brac* tandem gene duplicates display differential regulatory activity and evolutionary conservation

Henri-Marc G. Bourbon^{1#}, Mikhail H. Benetah^{1#}, Emmanuelle Guillou¹, Luis Humberto Mojica-Vazquez^{1, 3}, Aissette Baanannou^{1, 4}, Sandra Bernat-Fabre¹, Vincent Loubiere^{2, 5}, Frédéric Bantignies², Giacomo Cavalli², and Muriel Boube^{1*}

¹Center for Integrative Biology, Molecular Cellular and Developmental (MCD) Biology Unit, Federal University of Toulouse, 118 Route de Narbonne, F-31062 Toulouse, France; ²Institute of Human Genetics, UMR9002 CNRS/University of Montpellier, 141 Rue de la Cardonille, F-34396 Montpellier, France; ³Genotoxicología Ambiental, Departamento de Ciencias Ambientales, Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria Coyoacán, 04510, México, México; ⁴Laboratory of Molecular and Cellular Screening Processes, Center of Biotechnology of Sfax, Sfax, Tunisia; ⁵Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Vienna, Austria; Medical University of Vienna, Vienna BioCenter (VBC), Vienna, Austria.

[#] Authors having contributed equally to this work.

*Corresponding author: muriel.boube-trey@univ-tlse3.fr

Running title: Differential enhancer conservation and activity within the *bric-a-brac* locus

Abstract

During animal evolution, de novo emergence and modifications of pre-existing transcriptional enhancers have contributed to biological innovations, by implementing gene regulatory networks. The *Drosophila melanogaster* *bric-a-brac* (*bab*) complex, comprising the tandem paralogous genes *bab1-2*, provides a paradigm to address how enhancers contribute and co-evolve to regulate jointly or differentially duplicated genes. We previously characterized an intergenic enhancer (named LAE) governing *bab2* expression in leg and antennal tissues. We show here that LAE activity also regulates *bab1*. CRISPR/Cas9-mediated LAE excision reveals its critical role for *bab2*-specific expression along the proximo-distal leg axis, likely through paralog-specific interaction with the *bab2* gene promoter. Furthermore, LAE appears involved but not strictly required for *bab1-2* co-expression in leg tissues. Phenotypic rescue experiments, chromatin features and a gene reporter assay reveal a large “pleiotropic” *bab1* enhancer (termed BER) including a series of *cis*-regulatory elements active in the leg, antennal, wing, haltere and gonadal tissues. Phylogenomics analyses indicate that (i) *bab2* originates from *bab1* duplication within the Muscomorpha sublineage, (ii) LAE and *bab1* promoter sequences have been evolutionarily-fixed early on within the Brachycera lineage, while (iii) BER elements have been conserved more recently among muscomorphans. Lastly, we identified conserved binding sites for transcription factors known or prone to regulate directly the paralogous *bab* genes in diverse developmental contexts. This work provides new insights on enhancers, particularly about their emergence, maintenance and functional diversification during evolution.

Author summary

Gene duplications and transcriptional enhancer emergence/modifications are thought having greatly contributed to phenotypic innovations during animal evolution. However, how enhancers regulate distinctly gene duplicates and are evolutionary-fixed remain largely unknown. The *Drosophila bric-a-brac* locus, comprising the tandemly-duplicated genes *bab1-2*, provides a good paradigm to address these issues. The twin *bab* genes are co-expressed in many tissues. In this study, genetic analyses show a partial co-regulation of both genes in the developing legs depending on tissue-specific transcription factors known to bind a single enhancer. Genome editing and gene reporter assays further show that this shared enhancer is also required for *bab2*-specific expression. Our results also reveal the existence of partly-redundant regulatory functions of a large pleiotropic enhancer which contributes to co-regulate the *bab* genes in distal leg tissues. Phylogenomics analyses indicate that the *Drosophila bab* locus originates from duplication of a dipteran *bab1*-related gene, which occurred within the Brachycera (true flies) lineage. *bab* enhancer and promoter sequences have been differentially-conserved among Diptera suborders. This work illuminates how transcriptional enhancers from tandem gene duplicates (i) differentially interact with distinct cognate promoters and (ii) undergo distinct evolutionary changes to diversifying their respective tissue-specific gene expression pattern.

Introduction

Gene duplications have largely contributed to create genetic novelties during evolution (1, 2). Intra-species gene duplicates are referred to as “paralogs”, which eventually diverged functionally during evolution in a phylogenetic manner. Gene family expansion has facilitated phenotypic innovation through (i) acquisition of new molecular functions or (ii) the subdivision of the parental gene function between the duplicate copies (3-5). Phenotypic novelties are thought having originated mainly from evolutionary emergence or modifications of genomic Cis-Regulatory Elements (CREs) or modules, most often dubbed as “enhancer” regions, which regulate gene transcription in a stage-, tissue- and/or cell-type-specific manner (6-10). How CRE (enhancers) within gene complexes (i) are distinctly interacting with their cognate promoters and (ii) are differentially (co-)evolving remain largely unknown.

The *Drosophila melanogaster* *bric-a-brac* (*bab*) locus comprises two tandemly-duplicated genes (Fig 1A), *bab1-2*, which encode paralogous transcription factors sharing two conserved domains: (i) a Bric-a-brac/Tramtrack/Broad-complex (BTB) domain involved in protein-protein interactions, and (ii) a specific DNA-binding domain (referred to as BabCD, for Bab Conserved Domain), in their amino(N)- and carboxyl(C)-terminal moieties, respectively (11). Bab1-2 proteins are co-expressed in many tissues (11, 12). In the developing abdominal epidermal cells, within so-called histoblast nests, they jointly regulate directly *yellow* expression in a sexually-dimorphic manner, thus controlling adult male versus female body pigmentation traits (13-16). *bab1-2* co-expression in the developing epidermal histoblast nests is partially governed by two CREs which drive reporter gene expression (i) in a monomorphic pattern in the abdominal segments A2-A5 of both sexes (termed AE, for “Anterior Element”), and (ii) in a female-specific pattern in the A5-A7 segments (DE, for “Dimorphic Element”) (Fig 1A) (14, 17). In addition to controlling male-specific abdominal pigmentation traits, *bab1-2* are required, singly, jointly or in a partially-redundant manner, for embryonic cardiac development,

sexually-dimorphic larval somatic gonad formation, salivary glue gene repression, female oogenesis, wing development as well as distal leg (tarsal) and antennal segmentation (11, 13, 17-24). In addition to abdominal AE and DE, two other *bab* enhancers, termed CE and LAE (see Fig 1A), have been characterized, which recapitulate *bab2* expression in embryonic cardiac cells and developing tarsal as well as distal antennal cells, respectively (17, 21, 25).

Adult T1-3 legs, on the pro-, meso- and meta-thoraces, respectively, are derived from distinct mono-layered epithelial cell sheets, organized as sac-like structures, called leg imaginal discs (hereafter simply referred to as leg discs) (26-28). Upon completion of the third-instar larval stage (L3), each leg disc is already patterned along the proximo-distal (P-D) axis through regionalized expression of the Distal-less (*Dll*), Dachshund (*Dac*) and Homothorax (*Hth*) transcriptional regulators in the distal (center of the disc), medial and proximal (peripheral) regions, respectively (26). The five (ts1-5) tarsal and the single pretarsal (distalmost) segments are patterned through genetic cascades mobilizing transcription factors, notably the distal selector protein *Dll* and the tarsal Rotund protein as well as nuclear effectors of Notch and Epidermal Growth Factor Receptor (EGFR) signaling, i.e., Bowl and C15, respectively (26, 27).

While both *bab* genes are required for dimorphic abdominal pigmentation traits and somatic gonad specification (13, 22), only *bab2* is critical for tarsal segmentation (11). While *bab1* loss-of-function legs are apparently wild-type, a null allele (*bab*^{AR07}) removing *bab2* (and *bab1*) activities causes segmental transformation along the P-D leg axis, notably sex comb teeth in tarsal segments ts2-3 of male forelegs, normally only found in ts1, as well as ts2-5 tarsal fusions in both genders (11). While the two *bab* genes are co-expressed within ts1-4 cells, *bab2* is expressed more proximally than *bab1* in ts1, and in a graded manner along the P-D leg axis in ts5 (11, 29). We previously showed that *bab2* expression in distal leg (and antennal) tissues is governed by a 567-bp-long CRE/enhancer (termed LAE for “Leg and Antennal Enhancer”)

which is situated between the *bab1-2* transcription units (Fig 1A) (17, 25). However, LAE enhancer contribution to *bab1-2* co-regulation in the developing distal legs remains to be investigated in tarsal segments ts3-4 where expression levels of both paralogous BTB-BabCD proteins are the highest (see Fig 1B) (11).

Here, we show that *bab1* expression in the developing distal leg also depends on the Rotund, Bowl and C15 proteins, three transcription factors known to regulate directly *bab2* expression, by binding to dedicated LAE sequences (17, 25). LAE excision by CRISPR/Cas9-mediated genome editing indicates that this enhancer is partly involved in *bab1-2* co-regulation and, more unexpectedly, is also required for their differential expression along the P-D leg axis. Additionally, we show that LAE acts redundantly with a large enhancer signature region (termed BER), located within the *bab1* transcription unit, which is bound by dedicated transcription factors involved in diverse developmental processes and thus BER is prone to act as a “pleiotropic” enhancer region. Our phylogenomics analyses indicate that LAE and *bab1* promoter sequences have been fixed early on during dipteran evolution, well before *bab1* duplication. Conversely, BER and *bab2* promoter sequences have been fixed much later. Lastly, within *D. melanogaster* BER, we identified conserved binding sites for many transcriptional regulators known or prone to regulate *bab1* and/or *bab2* expression in the developing leg and antenna, but also in wing, haltere, mesodermal and gonadal tissues. This work illuminates how transcriptional enhancers from tandem gene duplicates (i) differentially interact with distinct cognate promoters and (ii) undergo distinct evolutionary changes to diversifying their respective tissue-specific gene expression pattern.

Results

The tandem *bab1-2* gene paralogs are co-regulated in the developing distal leg

In addition to the distal selector homeodomain (HD) protein Distal-less, we and others have previously shown that the C15 HD protein (homeoprotein) as well as Rotund and Bowl Zinc-Finger (ZF) transcription factors (TFs) bind dedicated sequences within LAE to ensure precise *bab2* expression in four concentric tarsal rings within the leg discs (Fig 1B) (17, 25). *bab1-2* are co-expressed in ts2-4 tarsal segments, while *bab2* is specifically expressed in ts5 and more proximally than *bab1* in ts1, both in a graded manner along the P-D leg axis (Fig 1C and S1A Fig) (11). Given *bab1-2* co-expression in ts1-4, we first asked whether *C15*, *rotund* and *bowl* activities are also controlling *bab1* expression in the developing distal leg. To this end, we compared Bab1 expression with that of a X-linked *LAE-GFP* (or *LAE-RFP*) reporter gene faithfully reproducing the *bab2* expression pattern there (17, 25), in homozygous mutant leg discs for a null *C15* allele or in genetically-mosaic leg discs harboring *rotund* or *bowl* loss-of-function mutant cells (Fig 1D-F).

C15 is specifically activated in the distalmost (center) part of the leg disc giving rise to the pretarsal (pt) segment (see Fig 1B) (30, 31). We have previously shown that the C15 homeoprotein down-regulates directly *bab2* to restrict its initially broad distal expression to the tarsal segments (25). Bab1 expression analysis in a homozygous *C15* mutant leg disc revealed that both *bab1* and *LAE-RFP* (*bab2*) are similarly de-repressed in the pretarsus (Fig 1, compare panels C-D).

In contrast to *C15*, *rotund* expression is restricted to the developing tarsal segments (32) and the transiently-expressed Rotund ZF protein contributes directly to *bab2* up-regulation in proximal (ts1-2) but has no functional implication in distal (ts3-5) tarsal cells (17). Immunostaining of genetically-mosaic leg discs at the L3 stage revealed that *bab1* is cell-autonomously down-regulated in large *rotund* mutant clones in ts1-2, but not in ts3-4 segments (Fig 1E), as it is the case for *LAE-GFP* reflecting *bab2* expression. Lastly, we examined whether the Bowl ZF protein, a repressive TF active in pretarsal but not in most tarsal cells, is down-

regulating *bab1* expression there (33), like *bab2* (25). Both *bab1* and *LAE-RFP (bab2)* appeared cell-autonomously de-repressed in *bowl* loss-of-function pretarsal clones (Fig 1F).

In addition to loss-of-function, we also conducted gain-of-function experiments for *bowl* and *rotund*. Bowl TF gain-of-function was achieved by down-regulating *lines* which encodes a related but antagonistic ZF protein (i) destabilizing nuclear Bowl and is specifically expressed in the tarsal territory (33). As previously shown for *LAE-GFP* (and *bab2*) expression, nuclear Bowl stabilization in the developing tarsal region appears sufficient to down-regulate cell-autonomously *bab1* (S1C Fig). Prolonged expression of the Rotund protein in the entire distal part of the developing leg disc, i.e., tarsal in addition to pretarsal primordia, induces ectopic *bab1* expression-in the presumptive pretarsal territory, albeit with some differences with *bab2* expression (S1B Fig, differentially-expressing cells are indicated with arrows), thus suggesting differential sensitivity of the two gene duplicates to Rotund TF levels (see discussion).

Taken together, these data indicate that the C15, Bowl and Rotund transcription factors, previously shown to interact physically with specific LAE sequences and thus to regulate directly *bab2* expression in the developing distal leg, are also regulating *bab1* expression there. These results suggest that the limb-specific intergenic LAE enhancer activity regulates directly both *bab* genes.

LAE activity regulates both *bab1* and *bab2* gene paralogs along the proximo-distal leg axis

To test the role of LAE in regulating *bab1-2*, we deleted precisely the LAE sequence through CRISPR/Cas9 genome editing (see Materials and Methods) (Fig 2A). Two independent deletion events (termed Δ LAE-M1 and -M2; see S2 Fig for deleted DNA sequences) were selected for phenotypic analysis. Both are homozygous viable and give rise to fertile adults with identical fully-penetrant distal leg phenotypes, namely ectopic sex-comb teeth on ts2 (normally

only found on ts1) tarsal segment in the male prothoracic (T1) legs (Fig 2B), which are typical of *bab2* hypomorphic alleles (11). The $\Delta LAE-M1$ allele was selected for detailed phenotypic analyses and is below referred to as *bab*^{LAE}.

First, we quantified *bab1* and *bab2* mRNAs prepared from dissected wild-type and homozygous *bab*^{LAE} mutant leg discs. As shown in Fig 2C, both mRNAs were detected in mutant discs, although *bab1* levels were two times lower than wild-type. Second, Bab1-2 expression patterns were analyzed in homozygous *bab*^{LAE} leg discs (Fig 3). To identify leg cells that should normally express *bab2*, we used the X-linked *LAE-GFP* reporter. In homozygous *bab*^{LAE} mutant leg discs, *bab2* specific expression (see Fig. 1B) is no longer observed (Fig 3B-C), while Bab1-2 shared expression is very low in ts3-4 to undetectable in ts1-2. Nevertheless, residual *bab1-2* co-expression in homozygous *bab*^{LAE} mutant discs indicates that additional *cis*-regulatory region(s) within the *bab* locus act(s), at least partly, redundantly with the LAE enhancer.

Taken together, our data indicate that LAE enhancer activity is (i) required for *bab1-2* co-expression in the two proximal-most tarsal segments, particularly ts1, (ii) dispensable for their co-expression in ts3-4, suggesting the presence of redundant *cis*-regulatory information and (iii) critically required for *bab2*-specific tarsal expression both proximally and distally. Thus, LAE activity governs both shared and paralog-specific expression of the *bab1-2* gene duplicates.

LAE paralog-specific activity requires the *bab2* core promoter

Whereas enhancer emergence has been proposed to account for acquisition of novel tissue- or paralog-specific functions for gene duplicates (34-36), LAE regulatory function provides an example of a single enhancer responsible both for shared and differential expression of two tandemly-repeated gene paralogs. Previously tested LAE reporter constructs fused *bab2* core

promoter sequences to the minimal *Hsp70* promoter region (*pHsp70*) (17, 25). To examine the contribution of the *bab2* promoter to LAE activity we compared the expression of two LAE reporters containing (*LAE-RFP*) or not (*LAE-pHsp70only-GFP*) the *bab2* promoter sequence (Fig 3D). Strikingly, the *LAE-pHsp70only-GFP* reporter was no longer activated in *RFP+* (*bab2*-expressing) ts1 and ts5 cells (Fig 3E; see white brackets and arrows). These data indicate that *bab2*-specific regulation by LAE activity requires the *bab2* core promoter sequences.

In addition to the intergenic LAE, other leg-specific enhancer elements are present within the *bab1* first intron

Since LAE appeared dispensable for *bab1-2* co-expression in ts3-4 cells, our data suggested the existence of other redundant *cis*-regulatory elements, presumably located also within the *bab* locus. On one side, a X-linked Bacterial Artificial Chromosome (BAC) construct, *BAC26B15^{ZH2A}*, encompassing the *bab2* gene and downstream intergenic sequence including LAE (see Fig 4A), could rescue (i) Bab2 expression in the tarsal primordium and (ii), distal leg phenotypes detected in homozygous animals for the null allele *bab^{AR07}* (17). On the other side, a *BAC26B15* construct (*BAC26B15ΔLAE^{ZH2A}*) inserted at the same genomic landing site and specifically lacking LAE sequence did not (Fig 4B-D). These results confirmed that (i) in absence of redundant *cis*-regulatory information, LAE is essential for *bab1-2* expression in the tarsal segments and (ii) the *cis*-information redundant with LAE is located outside the genomic region covered by *BAC26B15*.

To identify limb-specific redundant *cis*-regulatory information within the *bab* complex, we first tested the capacity of another BAC, *BAC69B22*, which overlaps *bab1* and lacks LAE (see Fig 4A), to restore Bab1 expression in *bab^{AR07}* mutant leg discs. As shown in Fig 4E-F, the X-linked *BAC69B22^{ZH2A}* could restore *bab1* expression in ts2-4, indicating that it contains *cis*-regulatory

information redundant with LAE activity in these segments. To test the capacity of *BAC69B22* sequences to also regulate *bab2* expression in ts2-4, we placed *BAC69B22^{ZH2A}* across *BAC26B15ΔLAE^{ZH2A}*, to allow pairing-dependent *trans*-interactions (i.e., transvection) between the two X chromosomes in females. This configuration partially restored Bab2 expression in ts2-4 cells from *bab^{AR07}* mutant L3 leg discs, albeit in salt and pepper patterns (Fig 4G), diagnostic of transvection effects (37).

From these data, we predicted the existence of *cis*-regulatory information within the *69B22* chromosomal interval capable to drive some *bab1-2* expression in distal leg tissues and acting redundantly with the LAE enhancer.

Chromatin features predict limb-specific *cis*-regulatory elements within *bab1*

Next, we sought to identify *cis*-regulatory information acting redundantly with LAE by taking advantage of available genome-wide chromatin features and High-throughput chromosome conformation Capture (Hi-C) experiments performed from L3 eye-antennal and/or leg discs (Fig 5). *bab1-2* are indeed co-expressed in distal antennal cells within the composite eye-antennal imaginal disc (11). A topologically-associating domain covering the entire *bab* locus was detected in Hi-C data from eye-antennal discs (Fig 5A and S3A Fig) (38), revealing particularly strong interactions between *bab1-2* promoter regions.

We then used published genome-wide data from Chromatin Immuno-Precipitation (ChIP-Seq), Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-Seq) and Assay for Transposase-Accessible Chromatin (ATAC-Seq) experiments (38-41), looking for active enhancer marks (H3K4me1 and H3K27Ac) and nucleosome-depleted chromatin regions (thus accessible to transcription factors). Active enhancer signatures are mainly associated with a ~15-kb-long genomic region that we termed BER, for “*bab1* Enhancer Region”, encompassing

the *bab1* promoter, first exon and part of its first intron (Fig 5B, lanes 1-2 and 5-6, respectively; see also S3B Fig for peak calling data). Note that LAE is also accessible to transcription factors and carries H3K4me1 marks, consistently with enhancer activity in distal antennal cells (17).

To more precisely locate putative enhancer element(s) within BER, we analyzed previously-published ChIP-Seq data from L3 leg discs (42) for binding sites for Dll, Sp1 and Hth proteins, known to regulate *bab1* and/or *bab2* expression in the developing legs (17, 42-44). Strong Dll binding is detected throughout BER, including over the *bab1* promoter (Fig 5B, lane 10; see also S3B Fig). In leg discs, Dll binding is detected over 8 out of 10 Open Chromatin Subregions (OCS) within BER (Fig 6C and S3B Fig) and six of those eight are also bound by Sp1 ZF protein. Of note, all nucleosome-depleted (i.e., OCS) BER subregions in the leg are also accessible in the eye-antennal discs (40) (Fig 5B, compare lanes 2 and 9, and Fig 6C, two upper lanes). Lastly, of six OCS sequences co-bound by Dll and Sp1, four are also bound by Hth protein (Fig 5B, lane 8, and Fig 6C, bottom lane). FAIRE-Seq data indicate that, in addition to LAE, only OCS7 is nucleosome-depleted in the leg and eye-antennal discs but not the wing and/or haltere discs (Fig 5B, lanes 13, 16, 18, and Fig 6C, four upper lines; see also S3B Fig). Importantly, ventral limb-specific OCS7 is co-bound by Dll, Sp1 and Hth transcription factors (Fig 6C). Thus, within the entire *bab* locus, only LAE and BER OCS7 are specifically bound by transcriptional regulators in the developing leg and antenna.

In summary, data mining indicates that BER includes a cluster of enhancer elements bound by Dll and Sp1 in leg discs and thus are good candidates for acting redundantly with LAE in regulating *bab* genes in a ventral limb-specific manner.

BER includes multiple *cis*-regulatory elements active in diverse developmental contexts

To further ascribe regulatory roles to BER subregions, we took advantage of a systematic analysis of Gal4 reporter lines (45). Out of six lines containing BER fragments (Fig 6B), only two, *73C11* and *73C05*, overlapping OCS1-3, are active in the leg and eye-antennal discs (FlyLight database; (<http://flweb.janelia.org/cgi-bin/flew.cgi>; S4B Fig). Nevertheless, none reproduce the *bab1/2* leg or antennal expression patterns in four or two concentric distal rings, respectively. These and other published data from reporter constructs including *bab1* first intron sequences (14) indicate that OCS1-7 (see Fig 6B) are not sufficient to properly drive *bab1/2* expression in the developing legs and suggest the requirement of additional BER elements, particularly the *bab1* promoter region (i.e., OCS9). This hypothesis is consistent with binding of the known *bab1-2* leg regulators Dll and Sp1 throughout BER, in addition to LAE (Fig 5, lane 10; and S3B Fig).

The *73C05* BER fragment also drives reporter gene expression in the wing, haltere and genital discs (S4 Fig, panels E and H-J) in patterns strikingly similar to those described for *bab2* (11, 44). Consistently, FAIRE- and ChIP-Seq data from haltere discs indicate respectively that OCS1-3 are nucleosome-free and bind the Ultrabithorax (Ubx) Hox-type homeoprotein known to activate directly *bab2* expression in haltere tissues (Fig 5B, lanes 16-17, and Fig 6C) (46, 47). Furthermore, ChIP-Seq data from whole L3 larvae (modENCODE), showed binding over the entire BER region of the Hox Abd-B genital selector (Fig5B, lane 20) (48). Lastly, BER includes (i) nucleosome-depleted sequences governing expression in adult muscles and bound by the mesodermal transcription factors Mef2, Srp1 and Tinman in late embryos (14, 49) as well as (ii) a sequence element (overlapping OCS5) which confers enhancer activity in ovarian somatic cells (50) (see Fig 6D).

Taken together, our data indicate that BER sequences drive *bab1-2* expression in developing limbs but also in other tissues such as wing, haltere, genitalia and mesoderm. Moreover, owing to the presence of binding sites for transcription factors known to regulate *bab* gene expression

in these respective tissues, spread out over the whole BER sequence, the latter is thus proposed to act as a pleiotropic enhancer region.

Cross- and auto-regulations among the *bab* genes

Bab proteins interact with A/T-rich DNA sequences through their BabCD DNA-binding domain, including binding sites within their own locus (51). We therefore tested whether the Bab1-2 proteins autoregulate and/or cross-regulate their own expression. Previous data indicated that Bab2 protein expression is unaffected in *bab1* loss-of-function mutants (11). Given that protein null *bab2* alleles are not available, we used RNA interference coupled to flip-out (FO) Gal4 expression to down-regulate clonally *bab2* expression within developing legs, and examine *LAE-RFP* and *bab1* expression in mosaic L3 leg discs (Fig 7). Strikingly, both *LAE-RFP* and *bab1* were up-regulated cell-autonomously in most tarsal mitotic clones (n=17/20) (Fig 7, panels A-A"). Moreover, *bab2* down-regulation in proximal-most RFP+ ts1 cells (expressing only *bab2*) activated cell-autonomously *bab1*, in addition to up-regulating *LAE-RFP* expression (Fig 7, white arrows in panels B-B"). These results suggested to us that the *bab2* paralog specifically down-regulates its own expression through partial repression of LAE enhancer activity. To confirm these observations, we generated mutant clones for the *bab*^{AR07} null allele, lacking both *bab2* and *bab1* activities. A slight cell-autonomous *LAE-GFP* reporter up-regulation could be observed in all examined *bab*^{AR07} clones (detected with anti-Bab2 antibodies; n>20) (Fig 7, panels C-C"), independently of their size and position within *bab2*-expressing tarsal cells (see white arrows and arrowheads).

Altogether, we conclude that *bab2* down-regulates its own expression, likely via partial repression of LAE activity, thus ensuring appropriate levels of both paralogous BTB-BabCD transcription factors in distal leg tissues, and most likely in other appendages as well.

The *bab* gene complex arose from *bab1* duplication in the Muscomorpha infraorder

The different levels of *cis*-regulatory element redundancy within the *bab* locus led us to trace back the evolutionary origin of the *bab* duplication found in *D. melanogaster* (*Dmel*). To start, we identified proteins orthologous to *Dmel* Bab1 or Bab2, i.e., displaying an N-terminal BTB associated to a C-terminal BabCD domain (collectively referred to as BTB-BabCD proteins) (11) within highly diverse dipteran species (see Fig 8A). Two distinct BTB-BabCD proteins strongly related to *Dmel* Bab1 and Bab2, respectively, were identified in the Muscomorpha (higher flies, also known as Cyclorrhapha) superfamily, both within the Schizophora (in Calypterae, such as *Musca domestica* and *Glossina morsitans*, and in Acalypterae, particularly among Drosophilidae) and Aschiza subsections (Fig 8A-B and Supplementary data). In contrast, a single BTB-BabCD protein could be identified in evolutionarily-distant dipteran species within (i) the brachyceran Asilomorpha and Stratiomyomorpha superfamilies (such as *Proctacanthus coquilleti* and *Hermetia illucens*, respectively), collectively referred to as Orthorrhapha; (ii) the Nematocera suborder families (with rare exceptions, in Psychodomorpha and Bibionomorpha, see below); (iii) other Insecta orders (e.g., Coleoptera, Hymenoptera and Lepidoptera), and in crustaceans (e.g., *Daphnia pulex*) (see Supplementary data).

To analyze the phylogenetic relationships between these different Bab-related proteins, their primary sequences were aligned and their degree of structural relatedness examined through a maximum likelihood analysis. As expected from an ancient duplication, muscomorphan Bab1-2 paralogs cluster separately, while singleton asilomorphan BTB-BabCD proteins are more related to muscomorphan Bab1 than Bab2 (Fig 8B and S5 Fig), indicating that muscomorphan *bab2* originated from *bab1* duplication.

Interestingly, contrary to most nematocerans, two or even three *bab1* paralogs are present in the fungus gnat *Coboldia fuscipes* (Psychodomorpha) and the gall midge *Mayetiola destructor* (Bibionomorpha), respectively. Significantly, *M. destructor* and *C. fuscipes bab1* paralogs (i) cluster separately in our phylogenetic analysis (Fig 8B and S5 Fig) and (ii) two are arrayed in the same chromosomal contexts for both species (S6 Fig), indicating that they have likely been generated through independent gene duplication processes in the Bibionomorpha and Psychodomorpha, respectively.

Taken together, and updating a previous work (13), our phylogenomics analysis (summarized in Fig 9B-C) indicates that a single ancestral *bab* gene related to *bab1* has been duplicated to give rise to *bab2* within the Muscomorpha (Cyclorrhapha) infraorder.

LAE sequences have been fixed in the Brachycera, thus predating *bab1* duplication

Having traced back the *bab* gene duplication raised the question of the evolutionary origin of the LAE enhancer, which regulates both *bab1* and *bab2* expression (17) (this work). We have previously shown that LAE includes three subsequences highly-conserved among twelve reference Drosophilidae genomes (52), termed CR1-3 (for Conserved Regions 1 to 3; see S7A Fig and Supplementary data), of which only two, CR1 and 2, are critical for tissue-specificity (17, 25). The 68 bp CR1 includes contiguous binding sites for Dll and C15 homeoproteins, while the 41 bp CR2 comprises contiguous binding sites for Dll as well as the ZF protein Bowl (S7 Fig, panels B and C, respectively) (17, 25).

To trace back the LAE evolutionary origin, we then systematically searched for homologous CR1-3 sequences (>50% identity) in dipteran genomes. Importantly, conserved LAE sequences have not been yet reported outside drosophilids. Small genomic regions with partial or extensive homologies to the CR1 (encompassing the C15 and Dll binding sites) and CR2 (particularly the

Dll and Bowl binding sites) could be detected in all examined Brachycera families but not in any nematoceran (Fig 9B and S7B-C Fig). Contrary to closely-associated CR1-2 homologous sequences, no CR3-related sequence could be identified nearby, in any non-Drosophilidae species. Significantly, homologous LAE sequences are situated (i) in between the tandemly-duplicated paralogs in muscomorphan species for which the entire *bab* locus sequence was available to us, suggesting an evolutionarily-conserved enhancer role, or (ii) 20 kb upstream of the *bab1*-related singleton in the asilomorphan *P. coquillettii* (see Fig 9C).

Taken together, as summarized in Fig 9A-C, these data suggest that a LAE-like enhancer with CR1- and CR2-related elements emerged early on in the Brachycera suborder, 180-200 million years ago, and has been since fixed within or upstream the *bab* locus in the Muscomorpha and Asilomorpha infraorders, respectively.

Like LAE elements, *bab1* promoter sequences have been fixed early on in the Brachycera

Given their differential interplay with the long-lasting LAE enhancer, we next analyzed the evolutionary conservation of *Dmel bab1-2* promoter core sequences (Fig 9B and S8 Fig). Both *bab* promoters are TATA-less. Whereas *bab1* has a single transcriptional initiator (Inr) element (TTCAGTC), its *bab2* paralog displays tandemly-duplicated Inr sequences (ATTCA GTTCGT) (53, 54) (S8 Fig). Both promoters display 64% sequence identity over 28 base pairs, including Inr (TTCAGT) and downstream putative Pause Button (PB; consensus CGNNCG) sequences (55) (see S8A Fig). These data suggested that (i) the duplication process having yielded *bab2* included the ancestral *bab1* promoter and (ii) PolII pausing ability previously shown for *bab2* promoter (56-58) probably also occurs for *bab1* promoter.

Homology searches revealed that *bab1* promoter sequences have been strongly conserved in the three extant Muscomorpha families and even partially in some asilomorphans (e.g., *P.*

coquellitti), for which a *bab1*-related singleton gene is present (Fig 9B and S8B Fig). In striking contrast to *bab1*, sequence conservation of the *bab2* promoter could only be detected among some Acalyptatae drosophilids (Fig 9B and S8C Fig). In agreement with a fast-evolutionary drift for *bab2* promoter sequences, the duplicated Inr is even only detected in *Drosophila* group species.

Taken together, these evolutionary data (summarized in Fig 9B) indicate that, likewise for the LAE enhancer, *bab1* promoter sequences have been under strong selective pressure among the Brachycera, both in the Muscomorpha and Asilomorpha infraorders, while paralogous *bab2* promoter sequences diverged rapidly among muscomorphans.

Unlike LAE, other *bab* CRE sequences have not been conserved beyond the Muscomorpha

The broad LAE sequence conservation led us to also trace back the evolutionary origins of the pleiotropic BER enhancer region as well as the cardiac CE, abdominal anterior AE and sexually-dimorphic DE *cis*-regulatory elements (see Fig 9). Sequences homologous to half of the BER OCS subregions could be detected among the 12 reference Drosophilidae genomes (52), in Calyptatae schizophorans and even in the Muscomorpha Aschiza subsection (e.g., OCS3) (S10-18 Fig). Unlike the LAE enhancer, homologous BER sequence elements (except the *bab1* promoter) could not be detected in non-muscomorphan families. Cardiac CE and abdominal DE are even less conserved given that related sequences could be only detected within schizophoran (excepted in Calyptatae) (Fig 9B and S9 Fig, panels B and C, respectively), whereas abdominal AE sequences could be only identified among drosophilids (Supplementary data) but not in aschizan, asilomorphan and nematoceran *bab* loci.

In conclusion, contrary to the LAE enhancer which among the Diptera emerged early on in the Brachycera suborder, other so-far identified *bab* cis-regulatory sequences have not been conserved beyond the Muscomorpha infraorder. Thus, as summarized in Fig 9B, and unlike the long-lasting brachyceran LAE (CR1-2) sequences, these data suggest that other enhancer sequences have been fixed within the Muscomorpha concomitantly (BER) or even after (CE, DE and AE) the *bab2* paralog emergence. Moreover, as expected for a pleiotropic enhancer region, BER sequence conservation allowed us to predict binding sites for transcription factors known, or so far unsuspected, prone to regulate directly the two *bab* genes in many distinct developmental contexts, and which are presented hereafter.

Predictive TF combinatorial code governing *bab* gene expression in diverse tissues

We gathered our data from TF binding site evolutive conservation (described in Supplementary data and S10-19 Fig) with ChIP-Seq experiments from the literature (GEO datasets; see Materials and Methods) (Fig 5B and S3B Fig). Associated with our precise knowledge of *bab* locus enhancer sequences and with previous genetic data also gained from the literature, our compilation presented in Fig. 10 allows to propose new models for the TF code involved in *Dmel bab* locus regulation: (i) It provides new insights into limb-specific *bab1/2* regulation proposing additional direct regulators such as Sp1 in the legs, Hth in the antenna, Scalloped (Sd) in the wing and Ubx in the haltere (ChIP-Seq data shown in S3 Fig); (ii) It suggests that BER also acts as an enhancer region for *bab* gene regulation in other developmental contexts and that common TF sets (notably Abd-B together with Dsx) are acting through distinct cis-regulatory elements within BER to drive *bab* gene expression in distinct tissues (e.g., in abdominal histoblast nests versus genitalia); (iii) It proposes a direct Bab2 binding on LAE for *bab* gene auto- and cross regulation (tested above); (iv) Finally, analysis of sequences conserved

among brachyceran *bab* loci identified predicted binding sites for more broadly-expressed transcriptional regulators, i.e., GAF, Pho and CTCF (directly interacting with BER OCS; see Fig 5B, lanes 3, 14 and 19, respectively, as well as S3 Fig for GAF), as well as Eip93F, Eip74EF, Chinmo, all related to chromatin organization whose putative roles in *bab* locus regulation are discussed hereafter.

Discussion

In this work, we have addressed the issue of the emergence and functional diversification of enhancers and promoters from two tandem gene duplicates. Using the *Drosophila bab* locus as a model, we showed that the paralogous genes *bab1-2* originate from an ancient *bab1* duplication in the Muscomorpha/Cyclorrhapha. The early-fixed brachyceran *bab1* LAE has been co-opted lately to regulate also *bab2* expression. Furthermore, this unique enhancer is also responsible for paralog-specific *bab2* expression along the P-D leg axis presumably through privileged interactions with the *bab2* promoter. Finally, LAE regulates only some aspects of *bab1-2* expression in the developing limbs because redundant information has emerged within a large pleiotropic enhancer driving *bab1* and/or *bab2* expression in highly-diverse tissues, by binding common sets of developmental transcription factors. This work brings some cues about (i) how a single enhancer can drive specificity among tandem gene duplicates, (ii) how enhancers evolutionary adapt with distinct cognate gene promoters, and (iii) which functional links can be predicted between dedicated transcription factors and chromatin dynamics during development.

A shared enhancer differentially regulating two tandem gene paralogs through distinct promoter targeting specificities

Here, we have shown that a single enhancer, LAE, regulates two tandem gene paralogs at the same stage and in the same expression pattern. How can this work? It has been proposed that enhancers and their cognate promoters are physically associated within phase-separated nuclear foci composed of high concentrations of TFs and proteins from the basal RNA polymerase II (PolII) initiation machinery inducing strong transcriptional responses (59, 60). Our Hi-C data from eye-antennal discs show a strong interaction between *bab1-2* promoter regions (Fig 5), suggesting that both *bab* promoters could be in close proximity within such phase separated droplets, thus taking advantage of shared transcriptional regulators and allowing concerted gene regulation. In contrast, no strong chromosome contacts could be detected between LAE and any of the two *bab* promoter regions, indicating that this enhancer is not stably associated to the *bab2* or *bab1* promoter in the eye-antennal disc (where only the antennal distal part expresses both genes). It would be interesting to gain Hi-C data from leg discs, in which the *bab1-2* genes are much more broadly expressed.

In addition to be required for *bab1-2* co-expression in proximal tarsal segments, we showed here that the LAE enhancer is also responsible for paralog-specific *bab2* expression along the proximo-distal leg axis. While it has been proposed that expression pattern modifications occur through enhancer emergence, our present work indicates that differential expression of two tandem gene paralogs can depend on a shared pre-existing enhancer (i.e., LAE). How this may work? Relative to its *bab1* paralog, *bab2* expression extends more proximally within the Dac-expressing ts1 cells (44) and more distally in the ts5 segment expressing nuclear Bowl protein, whereas both Dac and Bowl proteins have been proposed to act as *bab2* (and presumably *bab1*) repressors (25, 33, 61). CRISPR/Cas9-mediated LAE excision allowed us to establish that this enhancer is critically required for paralog-specific *bab2* leg expression proximally and distally, in ts1 and ts5 cells, respectively. In this context, we and others have previously proposed that transiently-expressed Rotund activating TF may antagonize Bowl (and eventually Dac)

repressive activity to precisely delimit *bab2* expression among *ts1* cells (17, 61). Given that *bab1-2* are distinctly expressed despite being both regulated by Bowl and Rotund, we propose that paralog-specific LAE activity depends on privileged interactions with *bab2* promoter sequences (discussed below). Thus, we speculate that the *bab2* promoter responds to Rotund transcriptional activity differently from its *bab1* counterpart. Consistent with this view, ectopic Rotund expression reveals differential regulatory impacts on the two *bab* gene promoters (S1B Fig). Genetic together with Hi-C experiments indicate that this could occur through specific interactions between LAE-bound TFs (e.g., Rotund) and dedicated proteins within the PolII pre-initiation complex stably-associated to the *bab2* core promoter. We envision that the LAE-bound ZF protein Rotund, the chromatin-remodeling ZF protein GAF (for GAGA-associated Factor) and the PolII-associated TFIID subunit TAF3, the latter known to interact physically with GAF and Bab2 BTB proteins (62, 63), are parts of the underlying promoter targeting molecular mechanism.

In this context, despite that sequence homologies between both promoters (consistent with an ancient duplication event mobilizing the ancestral *bab1* promoter) are still detectable, it is significant that the *bab2* promoter evolves much faster than its *bab1* counterpart. While the *bab1* promoter sequence has been strongly conserved among brachycerans, predating *bab2* gene emergence in the Muscomorpha, the *bab2* promoter sequence has only been fixed recently among Drosophilidae, notably through the Initiator (*Inr*) sequence duplication, indicating very fast evolutionary drift after the gene duplication process which yielded the *bab2* paralog. We envision that this evolutionary ability has largely contributed to allow novel expression patterns for *bab2*, presumably through differential enhancer-promoter pairwise interplay.

Differential evolutionary conservation of tissue-specific versus pleiotropic enhancers

Our comprehensive phylogenomics analyses from highly diverse Diptera families indicate that the *bab* gene complex has been generated through tandem duplication from an ancestral *babl*-related gene singleton within the Muscomorpha (Cyclorrhapha), about 100-140 years ago. This result contrasts with published data reporting that the duplication process having yielded the tandem *bab* genes occurred much earlier in the Diptera lineage leading to both the Brachycera (true flies; i.e., with short antenna) and Nematocera (long horned “flies”, including mosquitos) suborders (13). In fact, tandem duplication events implicating the *bab* locus did occur in the Bibionomorpha, as reported (13)), and even in the Psychodomorpha with three *babl*-related gene copies (Figure 8 and S6 Fig), but our phylogenetic analysis supports independent events. Thus, within the Diptera, the ancestral *babl* singleton had a high propensity to duplicate locally.

In this study, we have shown a strong evolutionary conservation of LAE subsequences among brachycerans, notably its CR2 element containing Dll and Bowl binding sites (S7C Fig). This conservation suggests a long-lasting enhancer function in distal limb-specific regulation of ancestral singleton *babl* genes. In striking contrast, BER sequence conservation could only be detected among extant muscomorphan *bab* loci. We assume that during evolution large pleiotropic enhancers may better assimilate binding sites for gene-specific transcription factors, thus generating regulatory novelties in distinct imaginal discs.

Gene duplication is a major source to generate phenotypic innovations during evolution, through diverging expression and molecular functions, and eventually from single gene copy translocation to another chromosomal site. Emergence of tissue-specific enhancers not shared between the two gene duplicates, as well as of “shadow” enhancers, have been proposed to be evolutionary novelty sources (64) (6). Our work indicates that the presumably long-lasting brachyceran LAE enhancer has recently been co-opted in drosophilids to allow differential *bab* gene expression. Conversely, the large BER region has apparently accumulated regulatory

sequence elements bound by diverse tissue-specific transcription factors (e.g., Dll, Hth, Abd-B and Dsx) acting in different cellular contexts.

A pleiotropic enhancer region overlapping with a PcG-response element

ChIP-Seq analysis for histone H3 modifications (H3K4me3, H3K27Ac and H3K4me1 enhancer/promoter marks; H3K27me3 chromatin repressive mark) from eye-antennal discs has revealed the pleiotropic BER enhancer region but also an overlapping repressive PcG (Polycomb Group family) domain, indicating that BER encompasses a bivalent chromatin domain, while another one is detected over the *bab2* promoter region. A dual enhancer/silencing function for PcG-Response Element (PRE) during embryogenesis has recently been established genome-wide (65), and the authors have proposed that reuse of enhancer regulatory elements by PcG proteins may help fine-tune gene expression and ensure the timely maintenance of cell identities throughout development. More recently, we have shown widespread enhancer-PcG domain interplay during developmental gene activation through chromatin looping in eye-antennal discs (38). Altogether these data suggest that the *bab1-2* genes might be poised for activation throughout the eye-antennal disc, and possibly other imaginal discs as well.

The Pleiohomeotic (Pho) protein is a DNA-binding PcG complex recruiter, critical for gene silencing maintenance during development (66). Pho interaction with several BER subregions, as detected in ChIP-Seq experiments from L3 tissues, as well as the presence of many evolutionarily-conserved predicted Pho binding sites, support a role for Pho in PcG repression throughout BER. We thus propose that Pho-containing PcG repressive complex bound at PREs within the *bab* bivalent locus is counteracted by one or several tissue-specific transcriptional activators identified in this work, which remain(s) to be characterized.

In this context, it is significant that in the eye-antennal disc, the ZF protein CTCF, acting redundantly with other chromatin insulator proteins, strongly interacts directly with the two flanking regions of the TAD covering the *bab* locus and also with several BER OCSs (Fig 5B). Significantly, two of these predicted CTCF interacting sites overlap with putative optimal binding sites for the PcG-recruiter Pho (S18-19 Fig). These data suggest that the CTCF architectural protein and the PcG-recruiter Pho may functionally interact to regulate the *bab* locus chromosome topology. Interestingly, the human Pho homolog (YY1) is a structural enhancer-promoter looping regulator (67) and orchestrates, together with the CTCF protein, a 3D chromatin looping switch during early neural lineage commitment (68). To our knowledge, functional relationships between CTCF and Pho proteins have not been investigated genome-wide in *Drosophila*.

Dynamic *bric-a-brac* locus chromatin accessibility during development

Recent data indicate that chromatin accessibility is dynamic during *Drosophila* larval development, being triggered by the ecdysone hormone (69). Dynamic enhancer activity and chromatin accessibility have been proposed to be regulated by the ecdysone-induced Eip93F (Ecdysone-induced protein 93F, also called E93) transcriptional regulator (70). ChIP data from early pupal wings indicate that Eip93F binds many BER OCS as well as the *babl* promoter region (70). Consistently, many putative Eip93F binding sites are present in these BER subregions and several have been conserved beyond Drosophilidae (Fig 10C). Interestingly, the human Eip93F homolog interacts with CtBP through a conserved motif (71), and *Drosophila* CtBP is known to recruit diverse chromatin-modifying complexes, notably to participate in Pho-mediated PcG recruitment to PREs (72). Thus, Eip93F binding to BER *cis*-regulatory elements may impact the proposed dual PcG activity at the *bab* locus. In addition to Eip93F,

BER regulatory sequences include many evolutionarily-conserved putative binding sites for the Eip74EF protein (Fig 10C), another ecdysone-induced TF, including one which overlaps with a conserved putative Pho binding site, suggesting again functional correlation between Ecdysone regulation and PcG activity.

Lastly, the *cis*-regulatory landscape within the *bab* locus (i.e., AE, DE, CE, LAE and BER) includes one or several evolutionarily-conserved predicted binding sites for the Chinmo BTB-ZF protein participating to developmental timing, notably through interplay with ecdysone signaling (73, 74). Consistently, ChIP-Seq experiments from embryos (ModENCODE data; <http://www.modencode.org/>) indicate Chinmo binding to BER sequences (75). Intriguingly, the Chinmo ZF protein is an additional BTB-containing TF prone to regulate directly the *bab* genes, possibly through molecular partnerships with the chromatin organizer GAF (another BTB-ZF protein interacting directly with both *bab* promoter regions; Fig 5B and S3 Fig) and the twin Bab BTB-BabCD proteins themselves. Thus, Chinmo implication in chromatin organization and enhancer activity within the *bab* locus undoubtedly deserves to be investigated.

In summary, the *bab* locus offers a good paradigm to investigate molecularly in great details how chromatin structure, particularly higher-order chromosome organization, impacts on transcriptional memory during development and selective enhancer-promoter interplay in diverse tissular contexts. Indeed, our comprehensive predictive combinatorial code for tissue-specific, as well as broadly-expressed architectural transcription factors (e.g., CTCF, Pho and GAF) regulating two tandem gene paralogs, offers the opportunity to dissect underlying molecular mechanisms, which are prone to be conserved during animal evolution and thus to be of broad biological significance.

Material sand Methods

Fly stocks, culture and genetic manipulations

D. melanogaster stocks were grown on standard yeast extract-sucrose medium. The *vasa*-PhiC31 ZH2A *attP* stock (kindly provided by F. Karch) was used to generate the *LAE-Hsp70-GFP* reporter lines and the *BAC69B22* construct as previously described (17). *LAE-GFP* and *LAE-RFP* constructs (including both *Hsp70* and *bab2* core promoters) inserted on the ZH2A (X chromosome) or ZH86Fb (third chromosome) *attP* landing platforms, and displaying identical expression patterns, have been previously described (17, 25). *C15²/TM6B*, *Tb¹* stock was kindly obtained from G. Campbell. Mutant mitotic clones for null alleles of *bowl* and *rotund* were generated with the following genotypes: *y w LAE-GFP; DllGal4^{EM2012}, UAS-Flp/+; FRT82B, Ub-RFP/FRT82B rn¹²* (i.e., *rn* mutant clones are RFP negative; Fig 1E) and *y w LAE-RFP; DllGal4^{EM2012}, UAS-Flp/+; Ub-GFP, FRT40A/bowl¹ FRT40A* (i.e., *bowl* mutant clones are GFP negative; Fig 1F), respectively. Rotund protein gain-of-function within the *Dll*-expressing domain was obtained with the following genotype: *y w LAE-GFP; DllGal4^{EM2012}; UAS-Rn¹/+*. The *Dll^{EM212}-Gal4* line was provided by M. Suzanne, while the *UAS-Rn¹* line was obtained from the Bloomington stock center. “Flip-out” (FO) mitotic clones over-expressing dsRNA against *lines* were generated by 40 mn heat shocks at 38°C, in mid-late L2 to early-mid L3 larvae of genotypes: *y w LAE-RFP hsFlp; UAS-dsRNA^{lines}/pAct>y+>Gal4, UAS-GFP* (i.e., FO clones express GFP in S1C Fig). Mutant mitotic clones for the null *bab^{AR07}* allele were generated by 30 mn heat shocks at 38°C, in early first to late second-instar larvae of genotypes: *y w LAE-GFP, hsFlp; FRT80B/bab^{AR07}, FRT80B*. FO mitotic clones over-expressing dsRNA against *bab1* or *bab2* were generated by 40 mn heat shocks at 38°C, in early to early-mid L3 larvae of genotypes: *y w LAE-RFP hsFlp; UAS-dsbab2 /Pact>y+>Gal4, UAS-GFP* and *y w LAE-RFP, hsFlp; Pact>y+>Gal4, UAS-GFP/+; UAS-dsbab1/+* (i.e., FO clones express GFP in Fig 7). *UAS-dsRNA* stocks used to obtain interfering RNA against *lines* (#40939), *bab1* (#35707) and *bab2* (#37260) were obtained from the Bloomington stock Center.

642

643 **Immuno-histochemistry and microscopy**

644 Leg discs were dissected from wandering (late third instar stage) larvae (L3). Indirect immuno-
645 fluorescence was carried out as previously described (17) using a LEICA TCS SP5 or SPE
646 confocal microscope. Rat anti-Bab2 (11), rabbit anti-Bab1 (14), rabbit anti-Dll (76), rabbit anti-
647 Bowl (61), and rabbit anti-C15 (31) antibodies were used at 1/2000, 1/500, 1/200, 1/1000 and
648 1/200, respectively.

649

650 **CRISPR/Cas9-mediated chromosomal deletion**

651 Guide RNAs (gRNAs) were designed with CHOPCHOP at the Harvard University website
652 (<https://chopchop.cbu.uib.no/>). Four gRNA couples were selected that cover two distinct
653 upstream and downstream LAE positions: TCGTGAGCCTTCTTCGCCAGG or
654 TGGAGCCTTCTTCGCCAGGCGG; and TATACTGTTGAGATCCCATGCGG or
655 TTAGGCGCACATAAGGAGGCAGG (the PAM protospacer adjacent motif sequences are
656 underlined), respectively. Targeting tandem chimeric RNAs were produced from annealed
657 oligonucleotides inserted into the pCFD4 plasmid, as described in
658 (<http://www.crisprflydesign.org/>). Each pCFD4-LAE-KO construct was injected into 50 *Vasa*-
659 *Cas9* embryos (of note the *vasa* promoter sequence is weakly expressed in somatic cells). F0
660 fertile adults and their F1 progeny, with possible somatic LAE-deletion events and candidate
661 mutant chromosomes (balanced with *TM6B*, *Tb*), respectively, were tested by polymerase chain
662 reactions (PCR) with the following oligonucleotides: AGTTTTTCATCCCCCTTCCA and
663 GTATTCTTTGCCTTGCCATCG (predicted wild-type amplified DNA: 2167 base pairs).

664

Quantitative RT-PCR analysis

T1-3 leg imaginal discs were dissected from homozygous *white*¹¹¹⁸ and *bab*^{ALAE-M1} late L3 larvae in PBS 0.1% Tween. 50 discs of each genotype were collected and frozen in nitrogen. Total messenger RNAs were purified using RNeasy kit (Qiagen) and reverse transcribed by SuperScript II (ThermoFisher). *bab1*, *bab2* or *rp49* cDNA levels were monitored by quantitative PCR using the following oligonucleotides: Bab1Fw: CGCCCAAGAGTAACAGAAGC; Bab1Rev: TCTCCTTGTCCTCGTCCTTG; Bab2Fw: CTGCAGGATCCAAGTGAGGT; Bab2Rev: GACTTCACCAGCTCCGTTTC; RP49Fw: GACGCTTCAAGGGACAGTATCTG; RP49Rev: AAACGCGGTTCTGCATGAG. A Wilcoxon test was performed to evaluate the difference between samples.

Homology searches, sequence alignments and phylogenetic analyses

Homology searches were done at the NCBI Blast site (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Protein or nucleotide sequence alignments were done using MAFFT (Multiple Alignment using Fast Fourier Transform) (<https://mafft.cbrc.jp/alignment/server/>). Phylogenetic relationships were inferred through a maximum likelihood analysis with W-IQ-Tree (<http://iqtree.cibiv.univie.ac.at/>) and visualized with the ETE toolkit (<http://etetoolkit.org/treeview/>).

Transcription factor binding prediction

DNA binding predictions were done using the motif-based sequence analysis tool TomTom from the MEME suite (<http://meme-suite.org/tools/tomtom>) and the Fly Factor Survey database (<http://mccb.umassmed.edu/ffs/>).

688

689 **Gene expression omnibus datasets**

690 The following gene expression omnibus (GEO) datasets were extracted from the NCBI website
 691 (<https://www.ncbi.nlm.nih.gov/gds/>): GSE59078; GSM1261348; GSM1426265; GSE126985;
 692 GSM3139658; GSM948715; GSE113574; GSM948718; GSM948717; GSE38594;
 693 GSM948720; GSM948716; GSM659162; GSM948719; GSE102339; GSE50363.

694

695 **Hi-C and histone tail mark analyses from L3 eye-antennal imaginal discs**

696 Hi-C and histone mark ChIP-Seq analyses from L3 eye-antennal discs have been recently
 697 published in (38).

698

699 **Acknowledgments**

700 We thank F. Karch, M. Suzanne, T.M. Williams, G. Boekhoff-Falk, S. Bray, G. Campbell, T.
 701 Kojima, F. Laski, J.L. Couderc and the Bloomington Stock Center for fly stocks and reagents.
 702 We are grateful to Alain Vincent for his proofreading of the manuscript. We thank Julien Favier
 703 for technical assistance, particularly in managing the transgenic facility. Lastly, we
 704 acknowledge Brice Ronsin and the Toulouse RIO Imaging platform.

705

706

References

1. Ohno S. **Evolution by Gene Duplication**. Springer-Verlag; 1970.
2. Zhang J. Evolution by gene duplication - an update. *Trends Ecol. Evol.*; 2003. p. 292-8.
3. Lundin LG. Gene duplications in early metazoan evolution. *Semin Cell Dev Biol.* 1999;10(5):523-30.
4. He X, Zhang J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics.* 2005;169(2):1157-64.
5. Sundström G, Larsson TA, Larhammar D. Phylogenetic and chromosomal analyses of multiple gene families syntenic with vertebrate Hox clusters. *BMC Evol Biol.* 2008;8:254.
6. Long HK, Prescott SL, Wysocka J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell.* 2016;167(5):1170-87.
7. Rubinstein M, de Souza FS. Evolution of transcriptional enhancers and animal diversity. *Philos Trans R Soc Lond B Biol Sci.* 2013;368(1632):20130017.
8. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell.* 2008;134(1):25-36.
9. Levine M. Transcriptional enhancers in animal development and evolution. *Curr Biol.* 2010;20(17):R754-63.
10. Vlad D, Kierzkowski D, Rast MI, Vuolo F, Dello Ioio R, Galinha C, et al. Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. *Science.* 2014;343(6172):780-3.
11. Couderc JL, Godt D, Zollman S, Chen J, Li M, Tiong S, et al. The bric à brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in *Drosophila*. *Development.* 2002;129(10):2419-33.
12. Salomone JR, Rogers WA, Rebeiz M, Williams TM. The evolution of Bab paralog expression and abdominal pigmentation among *Sophophora* fruit fly species. *Evol Dev.* 2013;15(6):442-57.

13. Roeske MJ, Camino EM, Grover S, Rebeiz M, Williams TM. Cis-regulatory evolution integrated the Bric-à-brac transcription factors into a novel fruit fly gene regulatory network. *Elife*. 2018;7.
14. Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, Carroll SB. The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell*. 2008;134(4):610-23.
15. Gompel N, Carroll SB. Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. *Nature*. 2003;424(6951):931-5.
16. Kopp A, Duncan I, Carroll SB. Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature*. 2000;408(6812):553-9.
17. Baanannou A, Mojica-Vazquez LH, Darras G, Couderc JL, Cribbs DL, Boube M, et al. *Drosophila* distal-less and Rotund bind a single enhancer ensuring reliable and robust bric-a-brac2 expression in distinct limb morphogenetic fields. *PLoS Genet*. 2013;9(6):e1003581.
18. Godt D, Couderc JL, Cramton SE, Laski FA. Pattern formation in the limbs of *Drosophila*: bric a brac is expressed in both a gradient and a wave-like pattern and is required for specification and proper segmentation of the tarsus. *Development*. 1993;119(3):799-812.
19. Godt D, Laski FA. Mechanisms of cell rearrangement and cell recruitment in *Drosophila* ovary morphogenesis and the requirement of bric à brac. *Development*. 1995;121(1):173-87.
20. Sahut-Barnola I, Godt D, Laski FA, Couderc JL. *Drosophila* ovary morphogenesis: analysis of terminal filament formation and identification of a gene required for this process. *Dev Biol*. 1995;170(1):127-35.
21. Junion G, Bataillé L, Jagla T, Da Ponte JP, Tapin R, Jagla K. Genome-wide view of cell fate specification: ladybird acts at multiple levels during diversification of muscle and heart precursors. *Genes Dev*. 2007;21(23):3163-80.
22. Camara N, Whitworth C, Dove A, Van Doren M. Doublesex controls specification and maintenance of the gonad stem cell niches in. *Development*. 2019;146(11).
23. Duan J, Zhao Y, Li H, Habernig L, Gordon MD, Miao X, et al. Bab2 Functions as an Ecdysone-Responsive Transcriptional Repressor during *Drosophila* Development. *Cell Rep*. 2020;32(4):107972.

758 24. Zhao Y, Duan J, Dziedziech A, Büttner S, Engström Y. Bab2 activates JNK signaling to
759 reprogram *Drosophila* wing disc development. 2021.

760 25. Mojica-Vázquez LH, Benetah MH, Baanannou A, Bernat-Fabre S, Deplancke B, Cribbs DL, et al.
761 Tissue-specific enhancer repression through molecular integration of cell signaling inputs. PLoS Genet.
762 2017;13(4):e1006718.

763 26. Kojima T. The mechanism of *Drosophila* leg development along the proximodistal axis. Dev
764 Growth Differ. 2004;46(2):115-29.

765 27. Kojima T. Developmental mechanism of the tarsus in insect legs. Curr Opin Insect Sci.
766 2017;19:36-42.

767 28. Estella C, Voutev R, Mann RS. A dynamic network of morphogens and transcription factors
768 patterns the fly leg. Curr Top Dev Biol. 2012;98:173-98.

769 29. Li Q, Barish S, Okuwa S, Maciejewski A, Brandt AT, Reinhold D, et al. A Functionally Conserved
770 Gene Regulatory Network Module Governing Olfactory Neuron Diversity. PLoS Genet.
771 2016;12(1):e1005780.

772 30. Campbell G. Regulation of gene expression in the distal region of the *Drosophila* leg by the
773 Hox11 homolog, C15. Dev Biol. 2005;278(2):607-18.

774 31. Kojima T, Tsuji T, Saigo K. A concerted action of a paired-type homeobox gene, *aristaleless*, and
775 a homolog of Hox11/*tlx* homeobox gene, *clawless*, is essential for the distal tip development of the
776 *Drosophila* leg. Dev Biol. 2005;279(2):434-45.

777 32. Natori K, Tajiri R, Furukawa S, Kojima T. Progressive tarsal patterning in the *Drosophila* by
778 temporally dynamic regulation of transcription factor genes. Dev Biol. 2012;361(2):450-62.

779 33. Greenberg L, Hatini V. Essential roles for lines in mediating leg and antennal proximodistal
780 patterning and generating a stable Notch signaling interface at segment borders. Dev Biol.
781 2009;330(1):93-104.

782 34. Prud'homme B, Gompel N, Carroll SB. Emerging principles of regulatory evolution. Proc Natl
783 Acad Sci U S A. 2007;104 Suppl 1:8605-12.

784 35. Hittinger CT, Carroll SB. Gene duplication and the adaptive evolution of a classic genetic switch.
785 Nature. 2007;449(7163):677-81.

786 36. Carroll SB. From pattern to gene, from gene to pattern. Int J Dev Biol. 1998;42(3 Spec No):305-
787 9.

788 37. Lim B, Heist T, Levine M, Fukaya T. Visualization of Transvection in Living Drosophila Embryos.
789 Mol Cell. 2018;70(2):287-96.e6.

790 38. Loubiere V, Papadopoulos GL, Szabo Q, Martinez AM, Cavalli G. Widespread activation of
791 developmental gene expression characterized by PRC1-dependent chromatin looping. Sci Adv.
792 2020;6(2):eaax4001.

793 39. Yeung K, Wang F, Li Y, Wang K, Mardon G, Chen R. Integrative genomic analysis reveals novel
794 regulatory mechanisms of eyeless during Drosophila eye development. Nucleic Acids Res.
795 2018;46(22):11743-58.

796 40. McKay DJ, Lieb JD. A common set of DNA regulatory elements shapes Drosophila appendages.
797 Dev Cell. 2013;27(3):306-18.

798 41. Davie K, Jacobs J, Atkins M, Potier D, Christiaens V, Halder G, et al. Discovery of transcription
799 factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open
800 chromatin profiling. PLoS Genet. 2015;11(2):e1004994.

801 42. Newcomb S, Voutev R, Jory A, Delker RK, Slattery M, Mann RS. cis-regulatory architecture of a
802 short-range EGFR organizing center in the Drosophila melanogaster leg. PLoS Genet.
803 2018;14(8):e1007568.

804 43. Córdoba S, Requena D, Jory A, Saiz A, Estella C. The evolutionarily conserved transcription
805 factor Sp1 controls appendage growth through Notch signaling. Development. 2016;143(19):3623-31.

806 44. Chu J, Dong PD, Panganiban G. Limb type-specific regulation of bric a brac contributes to
807 morphological diversity. Development. 2002;129(3):695-704.

808 45. Jenett A, Rubin GM, Ngo TT, Shepherd D, Murphy C, Dionne H, et al. A GAL4-driver line resource
809 for Drosophila neurobiology. Cell Rep. 2012;2(4):991-1001.

810 46. Slattery M, Ma L, Négre N, White KP, Mann RS. Genome-wide tissue-specific occupancy of the
811 Hox protein Ultrabithorax and Hox cofactor Homothorax in *Drosophila*. *PLoS One*. 2011;6(4):e14686.

812 47. Boube M, Hudry B, Immarigeon C, Carrier Y, Bernat-Fabre S, Merabet S, et al. *Drosophila*
813 *melanogaster* Hox Transcription Factors Access the RNA Polymerase II Machinery through Direct
814 Homeodomain Binding to a Conserved Motif of Mediator Subunit Med19. *PLoS Genet*.
815 2014;10(5):e1004303.

816 48. Estrada B, Sanchez-Herrero E. The Hox gene Abdominal-B antagonizes appendage
817 development in the genital disc of *Drosophila*. *Development*. 2001;128(3):331-9.

818 49. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE. Combinatorial binding predicts spatio-
819 temporal cis-regulatory activity. *Nature*. 2009;462(7269):65-70.

820 50. Arnold CD, Gerlach D, Stelzer C, Boryń Ł, Rath M, Stark A. Genome-wide quantitative enhancer
821 activity maps identified by STARR-seq. *Science*. 2013;339(6123):1074-7.

822 51. Lours C, Bardot O, Godt D, Laski FA, Couderc JL. The *Drosophila melanogaster* BTB proteins bric
823 a brac bind DNA through a composite DNA binding domain containing a pipsqueak and an AT-Hook
824 motif. *Nucleic Acids Res*. 2003;31(18):5389-98.

825 52. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al. Evolution of genes and
826 genomes on the *Drosophila* phylogeny. *Nature*. 2007;450(7167):203-18.

827 53. Ohler U, Liao GC, Niemann H, Rubin GM. Computational analysis of core promoters in the
828 *Drosophila* genome. *Genome Biol*. 2002;3(12):RESEARCH0087.

829 54. Meers MP, Adelman K, Duronio RJ, Strahl BD, McKay DJ, Matera AG. Transcription start site
830 profiling uncovers divergent transcription and enhancer-associated RNAs in *Drosophila melanogaster*.
831 *BMC Genomics*. 2018;19(1):157.

832 55. Gilchrist DA, Dos Santos G, Fargo DC, Xie B, Gao Y, Li L, et al. Pausing of RNA polymerase II
833 disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell*.
834 2010;143(4):540-51.

835 56. Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, et al. RNA polymerase is
836 poised for activation across the genome. *Nat Genet.* 2007;39(12):1507-11. Epub 2007 Nov 11.

837 57. Lee C, Li X, Hechmer A, Eisen M, Biggin MD, Venters BJ, et al. NELF and GAGA factor are linked
838 to promoter-proximal pausing at many genes in *Drosophila*. *Mol Cell Biol.* 2008;28(10):3290-300.

839 58. Kwak H, Fuda NJ, Core LJ, Lis JT. Precise maps of RNA polymerase reveal how promoters direct
840 initiation and pausing. *Science.* 2013;339(6122):950-3.

841 59. Kim S, Shendure J. Mechanisms of Interplay between Transcription Factors and the 3D
842 Genome. *Mol Cell.* 2019;76(2):306-19.

843 60. Sabari BR, Dall'Agnese A, Young RA. Biomolecular Condensates in the Nucleus. *Trends Biochem*
844 *Sci.* 2020.

845 61. de Celis Ibeas JM, Bray SJ. Bowl is required downstream of Notch for elaboration of distal limb
846 patterning. *Development.* 2003;130(24):5943-52.

847 62. Pointud JC, Larsson J, Dastugue B, Couderc JL. The BTB/POZ domain of the regulatory proteins
848 Bric à brac 1 (BAB1) and Bric à brac 2 (BAB2) interacts with the novel *Drosophila* TAF(II) factor
849 BIP2/dTAF(II)155. *Dev Biol.* 2001;237(2):368-80.

850 63. Chopra VS, Srinivasan A, Kumar RP, Mishra K, Basquin D, Docquier M, et al. Transcriptional
851 activation by GAGA factor is through its direct interaction with dmTAF3. *Dev Biol.* 2008;317(2):660-70.

852 64. Hong JW, Hendrix DA, Levine MS. Shadow enhancers as a source of evolutionary novelty.
853 *Science.* 2008;321(5894):1314.

854 65. Erceg J, Pakozdi T, Marco-Ferreres R, Ghavi-Helm Y, Girardot C, Bracken AP, et al. Dual
855 functionality of. *Genes Dev.* 2017;31(6):590-602.

856 66. Schuettengruber B, Bourbon HM, Di Croce L, Cavalli G. Genome Regulation by Polycomb and
857 Trithorax: 70 Years and Counting. *Cell.* 2017;171(1):34-57.

858 67. Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, et al. YY1 Is a Structural
859 Regulator of Enhancer-Promoter Loops. *Cell.* 2017;171(7):1573-88.e28.

860 68. Beagan JA, Duong MT, Titus KR, Zhou L, Cao Z, Ma J, et al. YY1 and CTCF orchestrate a 3D
861 chromatin looping switch during early neural lineage commitment. *Genome Res.* 2017;27(7):1139-52.

862 69. Uyehara CM, Nystrom SL, Niederhuber MJ, Leatham-Jensen M, Ma Y, Buttitta LA, et al.
863 Hormone-dependent control of developmental timing through regulation of chromatin accessibility.
864 *Genes Dev.* 2017;31(9):862-75.

865 70. Nystrom SL, Niederhuber MJ, McKay DJ. Expression of E93 provides an instructive cue to
866 control dynamic enhancer activity and chromatin accessibility during development. *Development.*
867 2020;147(6).

868 71. Quinlan KG, Verger A, Kwok A, Lee SH, Perdomo J, Nardini M, et al. Role of the C-terminal
869 binding protein PXDLS motif binding cleft in protein interactions and transcriptional repression. *Mol*
870 *Cell Biol.* 2006;26(21):8202-13.

871 72. Basu A, Atchison ML. CtBP levels control intergenic transcripts, PHO/YY1 DNA binding, and PcG
872 recruitment to DNA. *J Cell Biochem.* 2010;110(1):62-9.

873 73. Marchetti G, Tavosanis G. Steroid Hormone Ecdysone Signaling Specifies Mushroom Body
874 Neuron Sequential Fate via Chinmo. *Curr Biol.* 2017;27(19):3017-24.e4.

875 74. Narbonne-Reveau K, Maurange C. Developmental regulation of regenerative potential in
876 *Drosophila* by ecdysone through a bistable loop of ZBTB transcription factors. *PLoS Biol.*
877 2019;17(2):e3000149.

878 75. Nègre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, et al. A cis-regulatory map of the
879 *Drosophila* genome. *Nature.* 2011;471(7339):527-31.

880 76. Panganiban G, Sebring A, Nagy L, Carroll S. The development of crustacean limbs and the
881 evolution of arthropods. *Science.* 1995;270(5240):1363-6.

882 77. Jory A, Estella C, Giorgianni MW, Slattery M, Lavery TR, Rubin GM, et al. A survey of 6,300
883 genomic fragments for cis-regulatory activity in the imaginal discs of *Drosophila melanogaster*. *Cell*
884 *Rep.* 2012;2(4):1014-24.

885 78. Slattery M, Voutev R, Ma L, Nègre N, White KP, Mann RS. Divergent transcriptional regulatory
 886 logic at the intersection of tissue growth and developmental patterning. PLoS Genet.
 887 2013;9(9):e1003753.

888

Fig 1. *C15*, *rotund* and *bowl* all regulate both *bab1* and *bab2*. (A) Schematic view of the *Dmel* *bab* locus on the 3L chromosomal arm (Chr3L). The tandem *bab1* (blue) and *bab2* (red) transcription units (filled boxes and broken lines represent exons and introns, respectively), the previously known CRE/enhancers are depicted by filled dots (abdominal DE and AE in dark and light orange, respectively; leg/antennal LAE in dark green and cardiac CE in purple), and the telomere and centromere directions are indicated by arrows. (B) A scheme depicted *C15*, *Bowl* and *Rn* TF activities in regulating *bab2* expression as a four-ring pattern within the developing distal leg, is shown. (C) Medial confocal view of a wild-type L3 leg disc. Merged *Bab1* (blue) and *Bab2* (red) immunostainings, as well as each marker in isolation in (C') and (C''), respectively, are shown. Positions of *bab2*-expressing ts1-5 cells and the pretarsal (pt) field are indicated in (C''). Brackets indicate paralog-specific expression in proximalmost and distalmost *bab2*-expressing cells. (D) Distal confocal view of a homozygous *C15*² mutant L3 leg disc expressing *LAE-RFP*. Merged *Bab1* immunostaining (in blue) and RFP fluorescence (red), and each marker in isolation in (D') and (D''), are shown. *Bab2*-expressing mutant pt cells are circled with a dashed line in (D') and (D''). (E) Medial confocal view of a mosaic L3 leg disc expressing *LAE-GFP* and harboring *rotund* mutant clones. Merged *Bab1* (blue) immunostaining, GFP (green) and RFP (red) fluorescence, as well as each marker in isolation in (E'), (E'') and (E'''), respectively, are shown. Mutant clones are detected as black areas, owing to the loss of RFP. The respective ts1-5 fields are indicated in (E). White and yellow arrows indicate *bab1* (*bab2*) still- and non-expressing *rotund*^{-/-} clones, respectively. (F) Distal confocal view of a mosaic L3 leg disc expressing *LAE-RFP* and harboring *bowl* mutant clones. Merged *Bab1* (blue) immunostaining, RFP (red) and GFP (green) fluorescence, as well as a higher magnification of the boxed area for each marker in isolation in (F'), (F'') and (F'''), respectively, are shown. Mutant clones are detected as black areas, owing to the loss of GFP. White arrows indicate pretarsal *bowl*^{-/-} clones ectopically expressing both *bab1* and *LAE-RFP* (*bab2*).

Fig 2. LAE is not critically required for tarsal segmentation and for overall *bab2* expression in the leg disc. (A) Schematic view of the *Dmel* *bab* locus on the 3L chromosomal arm (Chr3L). The tandem *bab1* and *bab2* transcription units (filled boxes and broken lines represent exons and introns, respectively), the intergenic LAE enhancer (in green), as well as the telomere and centromere directions, are depicted as in Fig 1A, except that both genes are depicted in grey. The small CRISPR/Cas9-mediated chromosomal deficiency (*bab^{ΔLAE}*) is shown in beneath (deleted LAE is depicted as a broken line). (B) Photographs of wild-type and homozygous *bab^{ΔLAE}* T1 distal legs from adult males. The regular sex-comb (an array of about 10 specialized bristles on the male forelegs) on distal ts1 is indicated with asterisks, while ectopic sex-comb bristles on distal ts2 from the mutant leg (right) is indicated by an arrow. Note that the five tarsal segments remain individualized in homozygous *bab^{ΔLAE}* mutant legs. (C) Overall *bab1-2* expression from wild-type and homozygous *bab^{ΔLAE}* L3 leg discs, as determined from reverse transcription quantitative PCR analyses. The *bab1-2* expression levels were quantified relative to *rp49* mRNA abundance.

Fig 3. LAE is mostly critically required for paralog-specific *bab2* expression in the developing distal legs. (A-C) Medial (A-B) and distal (C) confocal views of wild-type (A) and homozygous *bab^{DLAE}* mutant (B and C) L3 leg discs expressing *LAE-GFP*. Merged GFP fluorescence (green), Bab1 (blue) and Bab2 (red) immunostainings, as well as the two latter in isolation in (A'-C') and (A''-C''), respectively, are shown. Brackets indicate positions of paralog-specific expression in proximalmost (ts1) and distalmost (ts5) *bab2*-expressing (GFP+) cells. Green asterisks in (B'-B'') and (C'-C'') indicate weaker expression of both *bab* paralogs in GFP+ ts2 cells. (D) Modular structures of the *LAE-pHsp70onlyGFP* and *LAE-RFP* reporter constructs. GFP (green box) and RFP (red box) open-reading frames (ORFs) have been fused with ORFs for the Transformer (Tra) nuclear localization signal (NLS) and the histone H2B, respectively (see white boxes). The SV40 polyadenylation signal region is boxed in grey. The *Dmel LAE* sequence is boxed in blue. The classical non-heat-inducible basal *Hsp70* promoter sequence is boxed in black, while the *bab2* core promoter sequence is depicted in orange. Note that both promoters are juxtaposed in the *LAE-RFP* construct. (E) LAE activity requires functionally the *bab2* promoter to ensure paralog-specific expression in the developing legs. A lateral confocal view of merged GFP (green) and RFP (red) fluorescence, as well as each marker in isolation in (E') and (E''), respectively, of the distal part of an early pupal leg expressing both the *LAE-pHsp70onlyGFP^{ZH2A}* and *LAE-RFP^{ZH86Fb}* reporter constructs (depicted in (D)), are shown. Brackets indicate tarsal RFP+ cells expressing *bab2* in a paralog-specific manner, which never express the *LAE-pHsp70onlyGFP^{ZH2A}* reporter which lacks *bab2* core promoter sequences (see white arrows).

Fig 4. *bab1* includes partially-redundant limb-specific *cis*-regulatory information. (A) Chromosomal deficiency and BAC constructs covering the *bab* locus. The tandem gene paralogs and intergenic LAE are depicted as shown in Fig. 1A, except that *bab2* is depicted in pink instead of red. The *bab^{AR07}* 3L chromosomal deficiency is shown in beneath, with known deleted portion indicated by a dashed line. Note that the breakpoints have not been precisely mapped. The two overlapping BAC constructs 69B22 and 26B15, as well as a mutant derivative of the latter specifically-deleted for LAE, are shown further in beneath. (B-G) Medial confocal views of wild-type (B-E) and homozygous *bab^{AR07}* mutant (C-D and F-G) L3 leg discs, harboring singly or combined X-linked BAC construct(s) shown in (A), as indicated above each panel. Bab2 (pink) and Bab1 (blue) immunostainings are shown. Positions of *bab1*- and *bab2*-expressing ts1-4 cells are indicated. Note stochastic *bab2* expression in (G).

Fig 5. A topologically-associating domain encompasses the *bab* locus in the eye-antennal disc and genome-wide chromatin features identify an enhancer signature region (BER) within *bab1*. (A) Hi-C screenshot of a 160 kb region covering the *Dmel bab* gene complex. Score scale is indicated on the right (yellow to dark blue from positive to negative). (B) ATAC-, FAIRE- and/or ChIP-Seq profiles from L3 eye-antennal (ED), leg, wing and haltere discs as well as from adult pharate appendages (leg, wing and haltere) and from whole larval tissues, as indicated on the left side. As referred in the main text, lanes are numbered on the right side. ChIP-Seq peak calling data are shown in lanes 8, 17-18. Otherwise, normalized open chromatin, histone H3 post-translational modifications and TF binding profiles are shown. Positions of the tandem *bab1-2* genes are indicated on the bottom. The respective locations of the BER and LAE sequences are highlighted with vertical dashed lines. Of note, according to normalized FAIRE-Seq signals, LAE is not fully accessible in the pharate T3 leg (see grey arrow in lane 12). Strongest CTCF ChIP-Seq signals are indicated by horizontal black arrows (lane 3).

Fig 6. BER behaves as a composite pleiotropic enhancer. (A) The BER enhancer signature region includes the abdominal AE enhancer. Organization of the *Dmel bab* locus, with the tandem gene paralogs as depicted as in Fig. 2A. The characterized enhancers are depicted by filled dots (abdominal DE and AE in dark and light orange, respectively; leg/antennal LAE in dark green and cardiac CE in purple). BER is boxed in light green. The genomic portions of the overlapping *69B22* and *26B15* BAC constructs are shown in beneath. (B) Transgenic lines covering BER identify *cis*-regulatory elements driving reporter gene expression in diverse larval imaginal tissues. Genomic fragments covered by relevant Janelia Farm FlyLight reporter lines [77] and the DE- and/or AE-containing PB-c17 and PB-c29 genomic constructs, described in [14], are shown above a scheme of the BER region. *bab1* protein coding and 5'-untranslated sequences within the first exon are filled or hatched in dark grey, respectively, while the intronic region is in light gray. The AE sequence is in orange, as depicted in (A). FlyLight reporter lines driving reporter expression in diverse imaginal discs (see S4 Fig) are filled in light green. (C) BER includes open chromatin sequences (OCS) and is bound by Dll, Sp1, Hth TFs in diverse developing appendages (leg, eye-antenna, wing and haltere). OCS and TF-bound sequences are depicted by filled grey/black boxes. Numbers refer to OCSs detected in the leg discs (see main text). The black boxes represent OCSs detected in the eye-antennal (EA) and leg but not in wing and haltere discs. OCS and Dll or Sp1-bound regions, as determined from peak calling (FAIRE-Seq GSE38727 and ChIP-Seq GSE113574 GEO dataset series, respectively), are from [40] and [42], respectively. ChIP-Seq data for Hth are from [78]. (D) BER includes pleiotropic *cis*-regulatory elements (CREs). Locations of predicted CREs (see text) are indicated by light green boxes. The hatched part of the predicted leg/antennal CRE is inferred from data obtained with the PB-c17 construct reported in [14].

Fig 7. Auto- and cross-regulation among the two *bab* paralogs in the developing legs. (A) Distal confocal view of a L3 leg disc expressing *LAE-RFP* and harboring flip-out (FO) clones expressing interfering RNA against *bab2*. Merged RFP (red) and GFP (green) fluorescence, as well as the former in isolation and under two distinct signal magnification (A' and A''), are shown. A FO clone (GFP+) within *LAE-RFP* (*bab2*)-expressing ts4 cells is circled with a dashed line (see arrow). (B) Proximal confocal view of a L3 leg disc expressing *LAE-RFP* and harboring FO clones expressing interfering RNA against *bab2*. Merged Bab1 (blue) immunostaining, RFP (red) and GFP (green) fluorescence, as well as the two formers in isolation, in (B') and (B''), respectively, are shown. Note that a single-cell FO clone (GFP+), within *LAE-RFP* (*bab2*)-expressing ts1 cells, is sufficient to upregulate *bab1* (see arrows). (C-D) L3 leg (C) and eye-antennal (D) discs expressing *LAE-GFP* and harboring mutant clones for the protein null allele *bab^{AR07}*. Merged Bab2 (red) immunostaining and GFP (green) fluorescence as well as the latter in isolation under two distinct signal magnifications (C'-C'' and D'-D''), are shown. Tiny and larger clones (Bab2 negative) are circled with dashed lines (arrowheads and arrows, respectively).

Fig 8. Phylogenetic relationships among dipteran *bab* paralogs and orthologs. (A) Dipteran families studied in this work. Species abbreviations are described in Supplementary data. (B) Phylogenetic relationships of the *bab* paralogs and orthologs inferred from a maximum likelihood consensus tree constructed from 1000 bootstrap replicates. Support values (percentage of replicate trees) are shown in red. Scale bar represents substitution per site. Clustered positions of *bab2* paralogs and *bab1* paralogs/orthologs are shown in pink and light blue, respectively.

Fig 10. A comprehensive predictive tissue-specific TF code governing *bab* paralog expression. (A) Organization of the *Dmel bab* gene paralogs and enhancers, as depicted in Fig 9A. Expression tissue specificities conferred by each enhancer are shown in beneath. (B) BER structural and chromatin state organizations. The *bab1* first exon is hatched. OCS regions (see Fig 6C), as defined in leg tissues; are represented by light green boxes. The abdominal AE CRE (not detected in FAIRE-Seq data from leg and eye-antennal discs) is depicted as a light orange box. (C) Evolutionary conservation of predicted TF binding sites within *Dmel bab* cis-regulatory sequences. Site conservation among and beyond drosophilids of transcriptional regulators involved in tissue-specific morphogenetic processes, cell signaling, developmental timing and chromatin organization. Predicted/validated TF site numbers are indicated within colored or hatched boxes reflecting their relative conservation are indicated in beneath. Experimentally-validated direct *bab* regulators are colored according to their well characterized bound-enhancer sequences (see A).

A

B

73C11-Gal4 + UAS-GFP

73C05-Gal4 + UAS-GFP

A

B

C

A

B

B

Muscomorphan families

OCS : 1 2 3 4 5 6 7 8 9 10

Schizophora	Acalypttratae	Drosophilidae	4	2	8	1	4	3	6	2	4	3
		Tephritidae (Blat, Ccap)	2	1	5		1	1	5		1	
		Diopsidae (Tdal)	2	1	5		1	1	5		1	
		Sepsidae (Tmin)	1		2			1	1		1	
	Calyptratae	Glossinidae (Gbre, Gmor)	1		3			1	3		1	
		Muscidae (Mdom, Scal)		1	1			1	1		1	
		Calliphoridae (Lcup, Preg)		1	2			1	3		1	
		Rhinophoridae (Pmac)		1	2			1	2		1	
	Aschiza	Syrphidae (Edim)			1						1	
		Phoridae (Mabd)			1						1	

Sequence element(s) conserved among 12 Drosophilids (dubbed CS)

CS element(s) conserved beyond Drosophilidae

Leg/Eye-antenna/Wing/Haltere CS1 within OCS3

Leg/Eye-antenna/Wing/Haltere CS3 within OCS3

Leg/Eye-antenna/Wing/Haltere CS2 within OCS3

B Leg/Eye-antenna/Wing/Haltere CS1 within OCS4

Leg/Eye-antenna/Wing CS1 within OCS5

Leg/Eye-antenna/Wing/Haltere 3' next-to-CS1 within OCS6

B Leg and eye-antennal-specific CS4 within OCS7

Leg and eye-antennal-specific CS5 within OCS7

Eye-antenna-specific CS1 within OCS7

Eye-antenna-specific CS2 within OCS7

Eye-antenna-specific CS6 within OCS7

B

Leg/Eye-antenna/Wing/Haltere CS1 within OCS10

Leg/Eye-antenna/Wing/Haltere CS3 within OCS10

Supplementary data

Four letter abbreviations for investigated species	page 2
Predicted sequences for BTB-BabCD proteins	pages 3-20
Bab1 sequence conservation among muscomorphans	pages 21-22
Bab2 sequence conservation among muscomorphans	pages 23-24
Sequence conservation between Bab1/2 paralogs	pages 25-29
Sequence conservation between paralogous Bab1/2 proteins among muscomorphans. The four-letter species abbreviations are as listed below (page 2). Strictly conserved amino-acid residues are indicated by white characters on a red background while partially conserved ones are in black characters on a yellow background. Locations of the strongly-conserved BTB and BabCD domains are indicated along the right side (see black lines).	
BPE ^{OCS1} sequence conservation among Drosophilidae	pages 30-31
BPE ^{OCS2} sequence conservation among Drosophilidae	page 32
BPE ^{OCS3} sequence conservation among Drosophilidae	pages 33-34
BPE ^{OCS4} sequence conservation among Drosophilidae	page 35
BPE ^{OCS5} sequence conservation among Drosophilidae	page 36
BPE ^{OCS6} sequence conservation among Drosophilidae	pages 37-38
BPE ^{OCS7} sequence conservation among Drosophilidae	pages 39-40
BPE ^{OCS8} sequence conservation among Drosophilidae	pages 41-42
BPE ^{OCS9} sequence conservation among Drosophilidae	pages 43-45
BPE ^{OCS10} sequence conservation among Drosophilidae	pages 46-47
Enhancer sequence conservation among Drosophilidae	pages 48-57

Conservation among twelve reference drosophilids of *D. melanogaster* BER OCS1-10 as well as LAE, CE, AE and DE sequences. The four-letter Drosophilidae species abbreviations are as listed below (page 2). Locations of conserved sequence elements (CS) are indicated by underneath black bars. Sequence LOGOs of predicted binding sites for the Dll, Bowl, C15, Sp1, Rn, Sal/Salr, BabCD, GAF, Pho, Eip74EF, Dsx, Abd-B, Sd, Chinmo, Pan, Mad, GATA factors, Twist-Da and Lbe transcription factors are depicted above or below the alignments.

Four letter abbreviations for investigated species

<i>Dmel: Drosophila melanogaster</i>	Drosophilidae	Acalypttratae	Schizophora	Muscomorpha/Cyclorrhapha	Brachycera
<i>Dsim: Drosophila simulans</i>					
<i>Dsec: Drosophila sechellia</i>					
<i>Dyak: Drosophila yakuba</i>					
<i>Dere: Drosophila erecta</i>					
<i>Dsuz: Drosophila suzukii</i>					
<i>Drho: Drosophila rhopaloa</i>					
<i>Dele: Drosophila elegans</i>					
<i>Dbip: Drosophila bipectinata</i>					
<i>Dana: Drosophila ananassae</i>					
<i>Dper: Drosophila persimilis</i>					
<i>Dpse: Drosophila pseudoobscura</i>					
<i>Dwil: Drosophila willistoni</i>					
<i>Dvir: Drosophila virilis</i>					
<i>Dmoj: Drosophila mojavensis</i>					
<i>Dgri: Drosophila grimshawi</i>					
<i>Tmin: Themira minor</i>	Calypttratae	Aschiza	Orthorrhapha	Nematocera	
<i>Tdal: Teleopsis dalmanni</i>					
<i>Blat: Bactrocera latifrons</i>					
<i>Ccap : Ceratitis capitata</i>					
<i>Mdom: Musca domestica</i>					
<i>Scal: Stomoxys calcitrans</i>					
<i>Lcup: Lucilia cuprina</i>					
<i>Preg: Phormia regina</i>					
<i>Chom: Cochliomyia hominivorax</i>					
<i>Pmac: Paykullia maculata</i>					
<i>Gbre: Glossina brevipalpis</i>					
<i>Gmor: Glossina morsitans</i>					
<i>Edim: Eristalis dimidiata</i>					
<i>Mabd: Megaselia abdita</i>					
<i>Cpat: Condylostylus patibulatus</i>					
<i>Cmol: Chrysotimus molliculus</i>					
<i>Pcoq: Proctacanthus coquilletti</i>					
<i>Ddia: Dasypogon diadema</i>					
<i>Hfus: Holcocephala fusca</i>					
<i>Hill: Hermetia illucens</i>					
<i>Bval: Beris vallata</i>					
<i>Mdes: Mayetiola destructor</i>					
<i>Cfus: Coboldia fuscipes</i>					
<i>Ppat: Phlebotomus papatasi</i>					
<i>Llon: Lutzomyia longipalpis</i>					
<i>Agam: Anopheles gambiae</i>					
<i>Aaeg: Aedes aegypti</i>					
<i>Dpul: Daphnia pulex</i>	Crustacea				

B/B

21

BabCI

22

BTB

2

Bab2 sequence conservation among muscomorphans (Part2)

Dmel2	488	AMH.HH.....AQQLAAQ.HQL.....AMSH.A	AMASALGAA	AAGGAGAAGAGGAGSGSGSGASAPTGGTVAGSGAGAAVGS	H	DM	TKPTIAEMTRE
Dspe2	540	AMH.HH.....AQQLAAQ.HQL.....AMSH.A	AMASALGAA	AAGGAGAAGAGGAGSGSGSGASAPTGGTVAGSGAGAAVGS	H	DM	TKPTIAEMTRE
Dvir2	537	AAHHAASAGAVPQADPPF.....AMH.HH.....AQQLAAQ.HQL.....AMSH.A	AMASALGAA	AAGGAGAAGAGGAGSGSGSGASAPTGGTVAGSGAGAAVGS	H	DM	TKPTIAEMTRE
Blaac2	621	AAALG.....VNHHPHHPHHPH.HHAAHAAQAQLAAQ.HQLHAAHSHAA	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Ccap2	639	AAALG.....VNHHPHHPHHPH.HHAAHAAQAQLAAQ.HQLHAAHSHAA	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Tm12	673	LG.....GHSHH.....HGHSHH.....HGHSHH.....HGHSHH.....	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Td12	590	LG.....GHSHH.....HGHSHH.....HGHSHH.....HGHSHH.....	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Mdm2	601	AAALA.....NPHHPH.....HHAAHAAQAQLAAQ.HQLHAAHSHAA	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Scal2	584	AAALA.....NPHHPH.....HHAAHAAQAQLAAQ.HQLHAAHSHAA	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Lcup2	523	AAALA.....NPHHPH.....HHAAHAAQAQLAAQ.HQLHAAHSHAA	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Pre2	529	AAALA.....NPHHPH.....HHAAHAAQAQLAAQ.HQLHAAHSHAA	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Chom2	531	AAALA.....NPHHPH.....HHAAHAAQAQLAAQ.HQLHAAHSHAA	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Pmac2	530	AAALA.....NPHHPH.....HHAAHAAQAQLAAQ.HQLHAAHSHAA	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Gbre2	512	ANALT.....DPHHPH.....HQAFAAAQAQLAAQ.HQFHTAAQAQSHAV	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Gmor2	511	GNALA.....DPHHPH.....HQAFAAAQAQLAAQ.HQFHTAAQAQSHAV	AMASALGAA	LAAGAGAGAA.....GTGAAGAAIIS.....AVG	H	DM	TKPTIAEMTRE
Meu2	414	LSMSALGT	LSMSALGT	LSMSALGT	H	DM	TKPTIAEMTRE
Edim2	393	LSMSALGT	LSMSALGT	LSMSALGT	H	DM	TKPTIAEMTRE
Mab2	345	STPST.....MAQQPS.....TTAGVTITTFNLIA.....VTFPEVITNLGMGN	LSMSALGT	LSMSALGT	H	DM	TKPTIAEMTRE
Dmel2	578	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Dspe2	637	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Dvir2	631	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Blaac2	714	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Ccap2	730	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Pre2	751	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Td12	637	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Mdm2	683	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Scal2	665	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Lcup2	606	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Pre2	612	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Chom2	614	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Pmac2	613	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Gbre2	599	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Gmor2	599	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Meu2	440	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Edim2	419	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Mab2	426	EEAKMIEGCGFHCWCAAAATG.....ASVAD	DEYQYQLOS	MMQKCNWNTNQNLVQQLRFRERGPKLSWRPEAMAAEAFSV	KEGLSLGQAARKD	DI	PTPTFVLYANRVHNM
Dmel2	689	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Dspe2	746	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Dvir2	738	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Blaac2	818	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Ccap2	835	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Tm12	855	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Td12	842	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Mdm2	788	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Scal2	770	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Lcup2	711	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Pre2	716	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Chom2	718	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Pmac2	717	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Gbre2	703	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Gmor2	703	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Meu2	543	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Edim2	523	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Mab2	538	GPGLDGGTDPPEPKKARGPDRILLGMNPEELIRSVIKAVVFDYREK	KDMGLGF	PANGQAH.....AALFS	TS	PGSSG	VNFGFN
Dmel2	778	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Dspe2	835	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Dvir2	822	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Blaac2	912	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Ccap2	929	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Tm12	959	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Td12	880	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Mdm2	835	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Scal2	864	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Lcup2	801	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Pre2	807	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Chom2	808	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Pmac2	808	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Gbre2	793	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Gmor2	793	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Meu2	737	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Edim2	620	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Mab2	633	APPDASPPSPISMTTLRLROI	MSQ	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH	QOQGHQOQHQQAHGQOQFSDH
Dmel2	876	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Dspe2	923	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Dvir2	909	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Blaac2	981	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Ccap2	998	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Tm12	1052	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Td12	900	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Mdm2	967	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Scal2	941	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Lcup2	868	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Pre2	874	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Chom2	875	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Pmac2	875	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Gbre2	862	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Gmor2	862	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Meu2	705	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Edim2	687	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Mab2	696	FDLSALGLMLGIPLVNMPFS	GS	GGG	GGG	GGG	GGG
Dmel2	939	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Dspe2	989	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Dvir2	993	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Blaac2	1044	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Ccap2	1081	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Tm12	1106	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Td12	983	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Mdm2	1062	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Scal2	1040	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Lcup2	943	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Pre2	946	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Chom2	953	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Pmac2	947	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Gbre2	926	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Gmor2	926	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Meu2	763	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Edim2	744	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Mab2	740	QSRGSSGG.....SGAKSLTASQ.....RFG.AA	SPYSA	AH.YAKHO.....ASAY	NKRFLSLP	AGID	FAFANGLOSV
Dmel2	1005	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Dspe2	1072	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Dvir2	1066	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Blaac2	1115	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Ccap2	1137	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Tm12	1168	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Td12	1051	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Mdm2	1138	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Scal2	1104	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Lcup2	1002	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Pre2	1005	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Chom2	1012	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Pmac2	1006	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Gbre2	995	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Gmor2	995	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Meu2	826	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Edim2	838	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED
Mab2	779	NKSPPFEDFPGG.....QMSLFLFNDASAA.....AAAAAAYA.....PPGAIN	ESPLMKIKL	QOQAH	ATE	LP	PHED

BabCD

Sequence conservation between Bab1/2 paralogs (Part1)

Dmel1	1	ASA	QAEIN
Dpsel	1	SA	QAEIN
Dvir1	1	SA	QAEIN
Blat1	1	SN	QAEIN
Ccap1	1	SN	QAEIN
Tmin1	1	SN	QAEIN
Tdall1	1	SN	QAEIN
Mdom1	1	SN	QAEIN
Scall1	1	SN	QAEIN
Lcup1	1	SN	QAEIN
Preg1	1	SN	QAEIN
Chom1	1	SN	QAEIN
Pmac1	1	SN	QAEIN
Gbre1	1	SN	QAEIN
Gmor1	1	SN	QAEIN
Edin1	1	SN	QAEIN
Mabd1	1	SN	QAEIN
Dmel2	1	SA	QAEIN
Dpsel2	1	SA	QAEIN
Dvir2	1	SA	QAEIN
Blac2	1	SA	QAEIN
Ccap2	1	SA	QAEIN
Tmin2	1	SA	QAEIN
Tdall2	1	SA	QAEIN
Mdom2	1	SA	QAEIN
Scal2	1	SA	QAEIN
Lcup2	1	SA	QAEIN
Preg2	1	SA	QAEIN
Chom2	1	SA	QAEIN
Pmac2	1	SA	QAEIN
Gbre2	1	SA	QAEIN
Gmor2	1	SA	QAEIN
Edin2	1	SA	QAEIN
Megu2	1	SA	QAEIN
Edin2	1	SA	QAEIN
Mabd2	1	SA	QAEIN
Dmel1	10	VLAS	EQGF
Dpsel	10	VLAS	EQGF
Dvir1	22	AQNA	EQGF
Blat1	21	AQNA	EQGF
Ccap1	21	AQNA	EQGF
Tmin1	15	TEAATQFRVLSCAAASTVITTSADSDQGHGIRKNTSSG	STTPTTQ
Tdall1	18	AADSSVFFKRLKE	STTPTTQ
Mdom1	30	AADSVFFKRLKE	STTPTTQ
Scal1	34	QADSVFFKRLKE	STTPTTQ
Lcup1	18	ATGVRVDSPPFR	STTPTTQ
Preg1	18	ATGVRVDSPPFR	STTPTTQ
Chom1	18	ATGVRVDSPPFR	STTPTTQ
Pmac1	18	ATGVRVDSPPFR	STTPTTQ
Gbre1	17	ATGVRVDSPPFR	STTPTTQ
Gmor1	17	ATGVRVDSPPFR	STTPTTQ
Edin1	23	HMHVSV	STTPTTQ
Mabd1	10	VLAS	EQGF
Dmel2	81	SVGEVSSPQKQV	EQGF
Dpsel2	70	ATEKAAADLDEADEEAEPKQ	EQGF
Dvir2	68	QQAQARLEVLKLEEEQLEKPTIS	EQGF
Blac2	80	AAQQLQHKSTQANAMEQFRA	EQGF
Ccap2	93	QSQQLRELVKQAGAAVQQR	EQGF
Tmin2	106	NELTICKADNNLSAALQTSKPKQ	EQGF
Tdall2	73	AATTSVAVLETNAP	EQGF
Mdom2	76	TEITATTTMKPTPEVMAKPTTQ	EQGF
Scal2	58	NEHNAEKFEAKMFTAQQLKPTTQ	EQGF
Lcup2	55	NEHNAEKFEAKMFTAQQLKPTTQ	EQGF
Preg2	60	LEONTERIEKMKPTTQ	EQGF
Chom2	62	LEONTERIEKMKPTTQ	EQGF
Pmac2	58	LEONTERIEKMKPTTQ	EQGF
Gbre2	51	AGVEFSKRTTAYEREKVRNPSTPELLEN	EQGF
Gmor2	51	AGVEFSKRTTAYEREKVRNPSTPELLEN	EQGF
Megu2	8	LEPFLKQLS	EQGF
Edin2	8	LEPFLKQLS	EQGF
Mabd2	5	LEPFLKQLS	EQGF
Dmel1	43	QAEKRIK	EQGF
Dpsel	55	ADDRRGS	EQGF
Dvir1	66	TCGCGGDSLATLEHESKAGRC	EQGF
Blat1	92	TCGCGGDSLATLEHESKAGRC	EQGF
Ccap1	92	TCGCGGDSLATLEHESKAGRC	EQGF
Tmin1	115	TCGCGGDSLATLEHESKAGRC	EQGF
Tdall1	84	TCGCGGDSLATLEHESKAGRC	EQGF
Mdom1	99	TCGCGGDSLATLEHESKAGRC	EQGF
Scall1	86	TCGCGGDSLATLEHESKAGRC	EQGF
Lcup1	70	TCGCGGDSLATLEHESKAGRC	EQGF
Preg1	71	TCGCGGDSLATLEHESKAGRC	EQGF
Chom1	68	TCGCGGDSLATLEHESKAGRC	EQGF
Pmac1	64	TCGCGGDSLATLEHESKAGRC	EQGF
Gbre1	66	TCGCGGDSLATLEHESKAGRC	EQGF
Gmor1	66	TCGCGGDSLATLEHESKAGRC	EQGF
Edin1	77	TCGCGGDSLATLEHESKAGRC	EQGF
Mabd1	23	TCGCGGDSLATLEHESKAGRC	EQGF
Dmel2	143	TCGCGGDSLATLEHESKAGRC	EQGF
Dpsel2	132	TCGCGGDSLATLEHESKAGRC	EQGF
Dvir2	130	TCGCGGDSLATLEHESKAGRC	EQGF
Blac2	168	TCGCGGDSLATLEHESKAGRC	EQGF
Ccap2	183	TCGCGGDSLATLEHESKAGRC	EQGF
Tmin2	185	TCGCGGDSLATLEHESKAGRC	EQGF
Tdall2	144	TCGCGGDSLATLEHESKAGRC	EQGF
Mdom2	156	TCGCGGDSLATLEHESKAGRC	EQGF
Scal2	142	TCGCGGDSLATLEHESKAGRC	EQGF
Lcup2	134	TCGCGGDSLATLEHESKAGRC	EQGF
Preg2	139	TCGCGGDSLATLEHESKAGRC	EQGF
Chom2	141	TCGCGGDSLATLEHESKAGRC	EQGF
Pmac2	137	TCGCGGDSLATLEHESKAGRC	EQGF
Gbre2	129	TCGCGGDSLATLEHESKAGRC	EQGF
Gmor2	129	TCGCGGDSLATLEHESKAGRC	EQGF
Megu2	48	TCGCGGDSLATLEHESKAGRC	EQGF
Edin2	48	TCGCGGDSLATLEHESKAGRC	EQGF
Mabd2	30	TCGCGGDSLATLEHESKAGRC	EQGF
Dmel1	76	GGGSSPVS	EQGF
Dpsel	87	GGGSSPVS	EQGF
Dvir1	81	GGGSSPVS	EQGF
Blat1	148	GGGSSPVS	EQGF
Ccap1	147	GGGSSPVS	EQGF
Tmin1	163	GGGSSPVS	EQGF
Tdall1	117	GGGSSPVS	EQGF
Mdom1	128	GGGSSPVS	EQGF
Scal1	113	GGGSSPVS	EQGF
Lcup1	104	GGGSSPVS	EQGF
Preg1	101	GGGSSPVS	EQGF
Chom1	94	GGGSSPVS	EQGF
Tdall1	98	GGGSSPVS	EQGF
Gbre1	97	GGGSSPVS	EQGF
Gmor1	97	GGGSSPVS	EQGF
Edin1	101	GGGSSPVS	EQGF
Mabd1	30	GGGSSPVS	EQGF
Dmel2	170	GGGSSPVS	EQGF
Dpsel2	167	GGGSSPVS	EQGF
Dvir2	163	GGGSSPVS	EQGF
Blac2	223	GGGSSPVS	EQGF
Ccap2	250	GGGSSPVS	EQGF
Tmin2	238	GGGSSPVS	EQGF
Tdall2	177	GGGSSPVS	EQGF
Mdom2	187	GGGSSPVS	EQGF
Scal2	172	GGGSSPVS	EQGF
Lcup2	164	GGGSSPVS	EQGF
Preg2	169	GGGSSPVS	EQGF
Chom2	171	GGGSSPVS	EQGF
Pmac2	167	GGGSSPVS	EQGF
Gbre2	163	GGGSSPVS	EQGF
Gmor2	163	GGGSSPVS	EQGF
Megu2	89	GGGSSPVS	EQGF
Edin2	90	GGGSSPVS	EQGF
Mabd2	50	GGGSSPVS	EQGF

BTB

Sequence conservation between Bab1/2 paralogs (Part2)

Dmel1	155	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Dpsel	165	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Dvir1	163	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Blat1	229	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Ccap1	228	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Tmin1	244	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Tdal1	195	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Mdom1	212	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Scall1	193	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Lcup1	184	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Preg1	181	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Chom1	174	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Pmac1	178	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Gbre1	177	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Gmor1	177	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Edim1	179	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Mabd1	106	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Dps2	250	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Dvir2	246	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Blac2	309	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Ccap2	336	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Tmin2	331	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Tdal2	260	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Mdom2	276	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Scal2	273	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Lcup2	245	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Preg2	250	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Chom2	252	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Pmac2	248	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Gbre2	239	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Gmor2	239	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Mequ2	174	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Edim2	172	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Mabd2	128	HLAEPCPCPPVIMRDRVMDLRAIEMFYRGEINVSQDQGLPIRIEMIKVRGLADVTHMEAA.TAAAA.AASS.....ERMFPSSPEESTST..SRTE.....	
Dmel1	246	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Dpsel	256	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Dvir1	258	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Blat1	320	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Ccap1	319	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Tmin1	344	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Tdal1	291	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Mdom1	311	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Scall1	294	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Lcup1	280	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Preg1	278	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Chom1	271	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Pmac1	275	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Gbre1	268	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Gmor1	268	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Edim1	275	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Mabd1	177	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Dps2	343	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Dvir2	347	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Blac2	398	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Ccap2	426	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Tmin2	420	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Tdal2	356	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Mdom2	359	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Scal2	372	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Lcup2	337	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Preg2	346	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Chom2	348	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Pmac2	344	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Gbre2	333	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Gmor2	333	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Mequ2	257	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Edim2	255	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Mabd2	215	HDLERE.....AEELDAFMQF.....KLR.....TSQND.....EAEELRSLF.....RQGGKRVKRRW.....ADT.....	
Dmel1	300	FNPFP..SPLSSLIAAERMELE.....QKERE.....RQRDC.....SMT.....PK.....	
Dpsel	326	FNPFP..SPLSSLIAAERLELE.....QKERE.....RQRDC.....SMT.....PK.....	
Dvir1	314	FNPFP..SPLSSLIAAERLEVE.....QKERE.....RQRDC.....SMT.....PK.....	
Blat1	422	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Ccap1	419	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Tmin1	431	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Tdal1	380	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Mdom1	392	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Scall1	378	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Lcup1	367	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Preg1	365	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Chom1	358	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Pmac1	363	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Gbre1	357	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Gmor1	357	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Edim1	348	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Mabd1	199	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Dps2	381	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Dvir2	391	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Blac2	450	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Ccap2	476	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Tmin2	478	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Tdal2	408	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Mdom2	415	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Scal2	419	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Lcup2	383	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Preg2	390	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Chom2	392	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Pmac2	391	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Gbre2	376	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Gmor2	376	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Mequ2	292	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Edim2	290	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Mabd2	263	FNPFP..SPLSSLIAAERAEERER.....ERERE.....RQRDC.....SMT.....PK.....	
Dmel1	340	PP.....MSGS.....TVGA.....TRLETAI.....ALDMPSPAATGCL.....RSS..RPS.....SQPQQ.....DAQQCGQSPPLP	
Dpsel	364	PQ.....LQCG.....GVGCGGGGCTPR.....RLTP.....ALESAI.....GLDMPSPAATGSP.....RSS..RALA.....PSPQQH.....HQQQQQQQQQLHMP	
Dvir1	356	PL.....GSGS.....GVSPTR.....RLTPFAAAALTAH.....GLDMPSPAATGPAVAAAAA.....GRQC..RLAQLSGLSQPQQQ.....HQQQQQQQQQLHMP	
Blat1	464	TT.....AAGT.....AFSSTTSSVST.....RLTP.....SLLETAH.....LDLSPSTFAPSLM.....RTF..PFGV.....THLSQSQ.....QAHHQ.....	
Ccap1	460	AA.....AAGT.....AFSSTTSSVST.....RLTP.....SLLETAH.....LDLSPSTFAPSLM.....RTF..PFGV.....THLSQSQ.....QAHHQ.....	
Tmin1	473	PP.....SSEH.....PHEHQAHHL.....GLDMPSPAATFAPGSSASSITVSGRTF..TSL.....ATSPLL.....	
Tdal1	425	PT.....TSCTA.....SATAQLS.....SST.....RLTP.....SVLEAVSR.....LDLSPSTPTPNLLAP.....RTF..SSTL.....TPSPQLQ.....QGHSHQ.....	
Mdom1	434	PT.....TSCTA.....SATAQLS.....SST.....RLTP.....SVLEAVSR.....LDLSPSTPTPNLLAP.....RTF..SSTL.....TPSPQLQ.....QGHSHQ.....	
Scall1	428	PV.....TSAGS.....SATAQLS.....SST.....RLTP.....SVLEAVSR.....LDLSPSTPTPNLLAP.....RTF..SSTL.....TPSPQLQ.....QGHSHQ.....	
Lcup1	409	PV.....TSACS.....AATSSLSGSGMS.....TNVLDTSV.....MDLSPSTPTPNLLP.....RTF..SGML.....TPSPQLQ.....QGHSHQ.....	
Preg1	409	PV.....TSACS.....GHTSGL.....TSALTTVH.....MDLSPSTPTPNLLP.....RTF..SGML.....TPSPQLQ.....QGHSHQ.....	
Chom1	402	PV.....TSACS.....GHTSGL.....TSALTTVH.....MDLSPSTPTPNLLP.....RTF..SGML.....TPSPQLQ.....QGHSHQ.....	
Pmac1	407	PV.....TSACS.....GHTSGL.....TSALTTVH.....MDLSPSTPTPNLLP.....RTF..SGML.....TPSPQLQ.....QGHSHQ.....	
Gbre1	393	AI.....SSAT.....NSATHLS.....SLSATVQ.....LIPSSSLA.....LPPA.....RTF..SGML.....TPSPQLQ.....QGHSHQ.....	
Gmor1	393	AI.....TSBTI.....NSATHLS.....SLSATVQ.....LIPSSSLA.....LPPA.....RTF..SGML.....TPSPQLQ.....QGHSHQ.....	
Edim1	394	PT.....QALCNM.....SSSLINSMS.....RTF.....TSALTAH.....GA.....LPPA.....RTF..SGML.....TPSPQLQ.....QGHSHQ.....	
Mabd1	222	PP.....TLGR.....SATAQLS.....SST.....RLTP.....SVLEAVSR.....LDLSPSTPTPNLLP.....RTF..SSTL.....TPSPQLQ.....QGHSHQ.....	
Dps2	428	PQ.....LQAHLLTGITSSASSSSGGT.....SSTVGTAAVGT.....SSSBAAS.....SAGAT.....VTA..ARAC.....SPPPHQ.....QHSHSGSGSGS	
Dvir2	437	PQ.....LQAHLLTGITSSASSSSGGT.....SSTVGTAAVGT.....SSSBAAS.....SAGAT.....VTA..ARAC.....SPPPHQ.....QHSHSGSGSGS	
Blac2	516	PN.....TINNA.....AAMNONSQSAI.....ATASPTAASSSTPTF..HCSATPTADPTSLATSHS.....YRP.....SMESA.....TAAPTTVA	
Ccap2	538	PN.....PNHNA.....GAATNQNQIQAAT.....AIPAFSAVSSSTPTF..HCSATPTADPTSLATSHS.....YRP.....SMESA.....TAAPTTVA	
Tmin2	550	PN.....SAAAP.....SATSSSSSSSSSS.....SITSSSSSSSBAIVASVSGSSSTASNTPASG.....YRA..QVAC.....TPPL.....TAAPTTVA	
Tdal2	468	PL.....TGTAT.....SATSSSSSSSSSS.....SITSSSSSSSBAIVASVSGSSSTASNTPASG.....YRA..QVAC.....TPPL.....TAAPTTVA	
Mdom2	478	PL.....TGTAT.....SATSSSSSSSSSS.....SITSSSSSSSBAIVASVSGSSSTASNTPASG.....YRA..QVAC.....TPPL.....TAAPTTVA	
Lcup2	474	PT.....QGTONS.....SATSANAPTAQSSS.....VNSTATSSISA.....AAAAAASNVVVALP.....YRSL..ERSC.....SPPMTPPQ.....ARACSSVNT	
Scall2	434	PSNOFAATNA.....SATSANAPTAQSSS.....VNSTATSSISA.....AAAAAASNVVVALP.....YRSL..ERSC.....SPPMTPPQ.....ARACSSVNT	
Preg2	441	PSNOFAATNA.....SATSANAPTAQSSS.....VNSTATSSISA.....AAAAAASNVVVALP.....YRSL..ERSC.....SPPMTPPQ.....ARACSSVNT	
Chom2	443	PSNOFAATNA.....SATSANAPTAQSSS.....VNSTATSSISA.....AAAAAASNVVVALP.....YRSL..ERSC.....SPPMTPPQ.....ARACSSVNT	
Pmac2	442	PSNOFAATNA.....SATSANAPTAQSSS.....VNSTATSSISA.....AAAAAASNVVVALP.....YRSL..ERSC.....SPPMTPPQ.....ARACSSVNT	
Gbre2	431	QINOSDDVQV.....KAPLITSSSSSSS.....AASSDASSS.....SVLP.....FRSM..TRSC.....SPL.....AAAAITSR.....	
Gmor2	430	QINOSDDVQV.....KAPLITSSSSSSS.....AASSDASSS.....SVLP.....FRSM..TRSC.....SPL.....AAAAITSR.....	
Mequ2	362	SH.....	
Edim2	344	PH.....	
Mabd2	324	PL.....SPTK.....SASKEPPFAVSTP.....STPTMAQ.....	

Bob

27

Sequence conservation between Bab1/2 paralogs (Part4)

[illegible]

Sequence conservation between Bab1/2 paralogs (Part5)

[illegible]

BER^{OCs1} sequence conservation among Drosophilidae (Part1)

CS1

Sp1

CS2

Mad

BER^{OCS1} sequence conservation among Drosophilidae (Part2)

BabCD

Leg-specific CS3

Leg-specific CS4

BER^{OCS2} sequence conservation among Drosophilidae

	1	10	20	30	40	50
Dmel		CGGGAGCAATTA		GCGAATAA	TACAAAAATT	ATTCCATTTCATCGTC
Dsim		GGGAGCAATTA		GCGAATAA	TACAAAAATT	ATTCCATTTCATCGTC
Dsec		GGGAGCAATTA		GCGAATAA	TACAAAAATT	ATTCCATTTCATCGTC
Dyak		GGGAGCAATTA		GCGAATAA	TACAAAAATT	ATTCCATTTCATCGTC
Dere		GGGAGCAATTA		GCGAATAA	TACAAAAATT	ATTCCATTTCATCGTC
Dana		GTGGGAACAGATACATAA		GAAATAA	TACAAAAATT	ATTCCATTTCATCGTC
Dper	CACACAAACGAGC	CGCAATACAAAA		AAACAAACAAACAAAT	TGTATTCCATTTCATCGTC	
Dspe	CACACAAACGAGC	CGCAATACAAAA		AAATGT	ATTCCATTTCATCGTC	
Dwil	TACATACATATTATAAGTATGG	GGCAGATATAAAAAA	AAAAAGAAAAAGAAAAA	TAATAA	TATATAAAATT	ATTCCATTTCATCGTC
Dvir	CACACTCGCACT	CGCACACACACAC			ACAGAAATT	ATTCCATTTCATCGTC
Dmoj	AACAT	CACACACACACAC		ACAAATGGAC	TTTAAGTATTC	ATTCCATTTCATCGTC
Dgri	CACACACACACT	CGCACACACACAA		ACGAGTGA	AAAAAGATT	ATTCCATTTCATCGTC

	60	70	80	90	100
Dmel	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA			
Dsim	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA			
Dsec	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA			
Dyak	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA			
Dere	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA			
Dana	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA			
Dper	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA		AGAAAAAATACT	ACAAACA
Dspe	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA		AAAAAAGAAAAAATAAT	ACTACA
Dwil	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA			ACACA
Dvir	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA			ACACA
Dmoj	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA			ACACA
Dgri	ATAAACTTCAATTTTCGAAAAATCATTTCCTTAG	CAGCTGCGCATTAAAAA			ACACA

CS1

	110	120	130	140
Dmel	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA
Dsim	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA
Dsec	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA
Dyak	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA
Dere	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA
Dana	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA
Dper	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA
Dspe	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA
Dwil	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA
Dvir	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA
Dmoj	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA
Dgri	AAACACAGAAAG	AGAAAAATAAGA	ACGCCAACATAAAAA	TGAA

DII

	150	170	180	190	200	210	220
Dmel	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA
Dsim	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA
Dsec	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA
Dyak	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA
Dere	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA
Dana	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA
Dper	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA
Dspe	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA
Dwil	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA
Dvir	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA
Dmoj	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA
Dgri	TAAATGATTAAAGGCT	CGACAGTGGGCA		TGATTGGCAACT	GCCAGCAAT		CAGATATCGCAATCAATA

CS2

	230	240	250	260
Dmel	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG
Dsim	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG
Dsec	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG
Dyak	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG
Dere	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG
Dana	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG
Dper	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG
Dspe	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG
Dwil	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG
Dvir	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG
Dmoj	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG
Dgri	TACCT		CACTGCCAAAGTCATTTGT	CTGTGAGTAG

	270	280	290	300	310	320
Dmel	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG
Dsim	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG
Dsec	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG
Dyak	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG
Dere	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG
Dana	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG
Dper	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG
Dspe	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG
Dwil	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG
Dvir	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG
Dmoj	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG
Dgri	TGA	GCTGCTGATTGATGT	CCGATGTC	AAGCAAAGCG	AAAGTTGGCAA	ATATGTCGGAGCTGCTAG

	330	340	350	360
Dmel	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG
Dsim	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG
Dsec	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG
Dyak	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG
Dere	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG
Dana	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG
Dper	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG
Dspe	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG
Dwil	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG
Dvir	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG
Dmoj	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG
Dgri	TCG	GAATGGGGGCAAA		CAAATGTTGTCAGTAG

BER^{OCS3} sequence conservation among Drosophilidae (Part1)

CS1

CS2

CS3

BER^{OCs3} sequence conservation among Drosophilidae (Part2)

BER^{OC}S⁴ sequence conservation among Drosophilidae

BER^{OC55} sequence conservation among Drosophilidae

BER^{OCs6} sequence conservation among Drosophilidae (Part1)

BER^{OC56} sequence conservation among Drosophilidae (Part2)

CS3

BER^{OC57} sequence conservation among Drosophilidae (Part1)

Eye/Antennal CS1

Eye/Antennal CS2

BabCD

CS3

CS4

BER^{OC57} sequence conservation among Drosophilidae (Part2)

330 340 350 360 370

Dmel ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

Dsim ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

Dsec ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

Dyak ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

Dere ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

Dana ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

Dper ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

Dpse ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

Dwll ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

Dvir ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

Dmoj ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

Dgri ATCAACAGAAAGCGAGGGGC... CAGACAT... AGAAC TGAGACCCCTGGA... CGAC

380 390 400 410 420 430

Dmel GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

Dsim GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

Dsec GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

Dyak GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

Dere GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

Dana GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

Dper GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

Dpse GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

Dwll GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

Dvir GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

Dmoj GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

Dgri GAT... CTGACACTTATCGGCTTTATCT... CTATGCGGCGGACCGAG... AGGTGAGTAAGT... AGGTGAGTAAGT...

CS5

440 450 460

Dmel AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

Dsim AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

Dsec AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

Dyak AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

Dere AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

Dana AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

Dper AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

Dpse AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

Dwll AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

Dvir AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

Dmoj AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

Dgri AGTGCATTCACACAAAC... ATATAAA... ATATAAA...

470

Dmel TATAAAT... TATAAAT... TATAAAT...

Dsim TATAAAT... TATAAAT... TATAAAT...

Dsec TATAAAT... TATAAAT... TATAAAT...

Dyak TATAAAT... TATAAAT... TATAAAT...

Dere TATAAAT... TATAAAT... TATAAAT...

Dana TATAAAT... TATAAAT... TATAAAT...

Dper TATAAAT... TATAAAT... TATAAAT...

Dpse TATAAAT... TATAAAT... TATAAAT...

Dwll TATAAAT... TATAAAT... TATAAAT...

Dvir TATAAAT... TATAAAT... TATAAAT...

Dmoj TATAAAT... TATAAAT... TATAAAT...

Dgri TATAAAT... TATAAAT... TATAAAT...

480 490

Dmel GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

Dsim GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

Dsec GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

Dyak GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

Dere GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

Dana GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

Dper GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

Dpse GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

Dwll GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

Dvir GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

Dmoj GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

Dgri GGAAAGGATTTATAAA... TATTAAGAAAGG... TATTAAGAAAGG...

500

Dmel TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

Dsim TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

Dsec TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

Dyak TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

Dere TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

Dana TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

Dper TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

Dpse TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

Dwll TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

Dvir TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

Dmoj TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

Dgri TGGTGAATCGG... TGGTGAATCGG... TGGTGAATCGG...

510 520 530 540 550 560

Dmel GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Dsim GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Dsec GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Dyak GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Dere GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Dana GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Dper GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Dpse GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Dwll GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Dvir GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Dmoj GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Dgri GCGGAGGAGGAGG... GCGGAGGAGGAGG... GCGGAGGAGGAGG...

Eye/Antennal-specific CS6

570 580 590 600 610 620 630

Dmel ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

Dsim ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

Dsec ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

Dyak ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

Dere ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

Dana ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

Dper ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

Dpse ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

Dwll ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

Dvir ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

Dmoj ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

Dgri ATCGAATGACGTTGAT... ATATTTTCATGTGCTCGAATA... TATTAATTCAGAGCTCAATCT...

BER^{OCS8} sequence conservation among Drosophilidae (Part1)

BER^{OCs8} sequence conservation among Drosophilidae (Part2)

Exon1

BER^{OCs9} sequence conservation among Drosophilidae (Part1)

BER^{OCs9} sequence conservation among Drosophilidae (Part2)

BER^{OCs9} sequence conservation among Drosophilidae (Part3)

BER^{OCs10} sequence conservation among Drosophilidae (Part1)

BER^{OCS10} sequence conservation among Drosophilidae (Part2)

CS2

CS3

LAE sequence conservation among Drosophilidae (Part1)

LAE sequence conservation among Drosophilidae (Part2)

Cardiac CE sequence conservation among Drosophilidae (Part1)

Cardiac CE sequence conservation among Drosophilidae (Part2)

Abdominal AE sequence conservation among Drosophilidae (Part1)

Abdominal AE sequence conservation among Drosophilidae (Part2)

Abdominal AE sequence conservation among Drosophilidae (Part3)

DmelAGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...
 DsainAGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...
 DsecAGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...
 DereAGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...
 DyakAGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...
 DanaAGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...
 Dper CTCTCGAGGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...
 DpseAGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...
 DwilAGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...
 DvirAGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...
 DmojAGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...
 DgriAGTT...GGGAAAAATATTTAAAAAGCATAAAAACAAAATATTTATTGAAAG...

DmelAGAGGCTAACGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DsainAGAGGCTAACGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DsecAGAGGCTAACGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DereAGAGGCTAACGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DyakAGAGGCTAACGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DanaAGAGGCTAACGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DperAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DpseAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DwilAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DvirAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DmojAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DgriAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...

DmelAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DsainAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DsecAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DereAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DyakAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DanaAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DperAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DpseAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DwilAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DvirAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DmojAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...
 DgriAGAGGCTAACGAGCCGAGCCGAGCCGCGGCGCAT...CAGAGGCCAAACAGAG...

DmelCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...
 DsainCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...
 DsecCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...
 DereCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...
 DyakCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...
 DanaCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...
 DperCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...
 DpseCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...
 DwilCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...
 DvirCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...
 DmojCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...
 DgriCAGACGACACACAGCCTTCANA...ATTGAG...ACATGAG...

DmelCATAGCT...CATAGCT...CATAGCT...
 DsainCATAGCT...CATAGCT...CATAGCT...
 DsecCATAGCT...CATAGCT...CATAGCT...
 DereCATAGCT...CATAGCT...CATAGCT...
 DyakCATAGCT...CATAGCT...CATAGCT...
 DanaCATAGCT...CATAGCT...CATAGCT...
 DperCATAGCT...CATAGCT...CATAGCT...
 DpseCATAGCT...CATAGCT...CATAGCT...
 DwilCATAGCT...CATAGCT...CATAGCT...
 DvirCATAGCT...CATAGCT...CATAGCT...
 DmojCATAGCT...CATAGCT...CATAGCT...
 DgriCATAGCT...CATAGCT...CATAGCT...

DmelATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...
 DsainATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...
 DsecATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...
 DereATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...
 DyakATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...
 DanaATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...
 DperATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...
 DpseATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...
 DwilATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...
 DvirATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...
 DmojATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...
 DgriATAAACACGGACAGGAAC...TGGACATGGCAGAGAC...

Abdominal AE sequence conservation among Drosophilidae (Part4)

Abdominal DE sequence conservation among Drosophilidae (Part1)

Abdominal DE sequence conservation among Drosophilidae (Part2)