1

Tissue-specific versus pleiotropic enhancers within the *bric-a-brac* tandem gene duplicates 2 display differential regulatory activity and evolutionary conservation 3 4 Henri-Marc G. Bourbon^{1#}, Mikhail H. Benetah^{1#}, Emmanuelle Guillou¹, Luis Humberto 5 Mojica-Vazquez^{1, 3}, Aissette Baanannou^{1, 4}, Sandra Bernat-Fabre¹, Vincent Loubiere^{2, 5}, 6 Frédéric Bantignies², Giacomo Cavalli², and Muriel Boube^{1*} 7 8 ¹Center for Integrative Biology, Molecular Cellular and Developmental (MCD) Biology Unit, 9 Federal University of Toulouse, 118 Route de Narbonne, F-31062 Toulouse, France; ²Institute 10 11 of Human Genetics, UMR9002 CNRS/University of Montpellier, 141 Rue de la Cardonille, F-34396 Montpellier, France; ³Genotoxicología Ambiental, Departamento de Ciencias 12 Ambientales, Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México, 13 Circuito Exterior s/n, Ciudad Universitaria Coyoacán, 04510, México, México; ⁴Laboratory of 14 Molecular and Cellular Screening Processes, Center of Biotechnology of Sfax, Sfax, Tunisia; 15 ⁵Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Vienna, Austria; 16 Medical University of Vienna, Vienna BioCenter (VBC), Vienna, Austria. 17 18 [#] Authors having contributed equally to this work. 19 *Corresponding author: muriel.boube-trey@univ-tlse3.fr 20 21 22 Running title: Differential enhancer conservation and activity within the *bric-a-brac* locus 23 24

25 Abstract

During animal evolution, de novo emergence and modifications of pre-existing transcriptional 26 27 enhancers have contributed to biological innovations, by implementing gene regulatory networks. The Drosophila melanogaster bric-a-brac (bab) complex, comprising the tandem 28 29 paralogous genes bab1-2, provides a paradigm to address how enhancers contribute and coevolve to regulate jointly or differentially duplicated genes. We previously characterized an 30 intergenic enhancer (named LAE) governing *bab2* expression in leg and antennal tissues. We 31 show here that LAE activity also regulates bab1. CRISPR/Cas9-mediated LAE excision reveals 32 its critical role for *bab2*-specific expression along the proximo-distal leg axis, likely through 33 paralog-specific interaction with the bab2 gene promoter. Furthermore, LAE appears involved 34 35 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 36 BER) including a series of *cis*-regulatory elements active in the leg, antennal, wing, haltere and 37 gonadal tissues. Phylogenomics analyses indicate that (i) bab2 originates from bab1 duplication 38 within the Muscomorpha sublineage, (ii) LAE and *bab1* promoter sequences have been 39 40 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 41 for transcription factors known or prone to regulate directly the paralogous bab genes in diverse 42 43 developmental contexts. This work provides new insights on enhancers, particularly about their emergence, maintenance and functional diversification during evolution. 44

45 Author summary

46 Gene duplications and transcriptional enhancer emergence/modifications are thought having greatly contributed to phenotypic innovations during animal evolution. However, how 47 48 enhancers regulate distinctly gene duplicates and are evolutionary-fixed remain largely unknown. The Drosophila bric-a-brac locus, comprising the tandemly-duplicated genes bab1-49 2, provides a good paradigm to address these issues. The twin bab genes are co-expressed in 50 51 many tissues. In this study, genetic analyses show a partial co-regulation of both genes in the 52 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 53 54 also required for bab2-specific expression. Our results also reveal the existence of partlyredundant regulatory functions of a large pleiotropic enhancer which contributes to co-regulate 55 the bab genes in distal leg tissues. Phylogenomics analyses indicate that the Drosophila bab 56 57 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-58 59 conserved among Diptera suborders. This work illuminates how transcriptional enhancers from tandem gene duplicates (i) differentially interact with distinct cognate promoters and (ii) 60 undergo distinct evolutionary changes to diversifying their respective tissue-specific gene 61 62 expression pattern.

63

64 Introduction

65 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 66 67 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 68 of the parental gene function between the duplicate copies (3-5). Phenotypic novelties are 69 thought having originated mainly from evolutionary emergence or modifications of genomic 70 Cis-Regulatory Elements (CREs) or modules, most often dubbed as "enhancer" regions, which 71 regulate gene transcription in a stage-, tissue- and/or cell-type-specific manner (6-10). How 72 73 CRE (enhancers) within gene complexes (i) are distinctly interacting with their cognate promoters and (ii) are differentially (co-)evolving remain largely unknown. 74

75 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 76 77 domains: (i) a Bric-a-brac/Tramtrack/Broad-complex (BTB) domain involved in proteinprotein interactions, and (ii) a specific DNA-binding domain (referred to as BabCD, for Bab 78 Conserved Domain), in their amino(N)- and carboxyl(C)-terminal moieties, respectively (11). 79 Bab1-2 proteins are co-expressed in many tissues (11, 12). In the developing abdominal 80 epidermal cells, within so-called histoblast nests, they jointly regulate directly vellow 81 expression in a sexually-dimorphic manner, thus controlling adult male versus female body 82 83 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 84 pattern in the abdominal segments A2-A5 of both sexes (termed AE, for "Anterior Element"), 85 and (ii) in a female-specific pattern in the A5-A7 segments (DE, for "Dimorphic Element") (Fig. 86 1A) (14, 17). In addition to controlling male-specific abdominal pigmentation traits, *bab1-2* are 87 required, singly, jointly or in a partially-redundant manner, for embryonic cardiac development, 88

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).

94 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 95 (hereafter simply referred to as leg discs) (26-28). Upon completion of the third-instar larval 96 stage (L3), each leg disc is already patterned along the proximo-distal (P-D) axis through 97 regionalized expression of the Distal-less (Dll), Dachshund (Dac) and Homothorax (Hth) 98 transcriptional regulators in the distal (center of the disc), medial and proximal (peripheral) 99 100 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 101 selector protein Dll and the tarsal Rotund protein as well as nuclear effectors of Notch and 102 103 Epidermal Growth Factor Receptor (EGFR) signaling, i.e., Bowl and C15, respectively (26, 27). 104

105 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-106 of-function legs are apparently wild-type, a null allele (*bab*^{AR07}) removing *bab2* (and *bab1*) 107 activities causes segmental transformation along the P-D leg axis, notably sex comb teeth in 108 109 tarsal segments ts2-3 of male forelegs, normally only found in ts1, as well as ts2-5 tarsal fusions 110 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 111 ts5 (11, 29). We previously showed that *bab2* expression in distal leg (and antennal) tissues is 112 governed by a 567-bp-long CRE/enhancer (termed LAE for "Leg and Antennal Enhancer") 113

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, 118 Bowl and C15 proteins, three transcription factors known to regulate directly bab2 expression, 119 by binding to dedicated LAE sequences (17, 25). LAE excision by CRISPR/Cas9-mediated 120 genome editing indicates that this enhancer is partly involved in bab1-2 co-regulation and, more 121 unexpectedly, is also required for their differential expression along the P-D leg axis. 122 123 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 124 125 transcription factors involved in diverse developmental processes and thus BER is prone to act 126 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 127 128 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 129 regulators known or prone to regulate bab1 and/or bab2 expression in the developing leg and 130 antenna, but also in wing, haltere, mesodermal and gonadal tissues. This work illuminates how 131 transcriptional enhancers from tandem gene duplicates (i) differentially interact with distinct 132 cognate promoters and (ii) undergo distinct evolutionary changes to diversifying their 133 respective tissue-specific gene expression pattern. 134

135

136 **Results**

137 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 138 previously shown that the C15 HD protein (homeoprotein) as well as Rotund and Bowl Zinc-139 Finger (ZF) transcription factors (TFs) bind dedicated sequences within LAE to ensure precise 140 bab2 expression in four concentric tarsal rings within the leg discs (Fig 1B) (17, 25). bab1-2 141 142 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 143 Fig) (11). Given bab1-2 co-expression in ts1-4, we first asked whether C15, rotund and bowl 144 activities are also controlling *bab1* expression in the developing distal leg. To this end, we 145 compared Bab1 expression with that of a X-linked LAE-GFP (or LAE-RFP) reporter gene 146 147 faithfully reproducing the *bab2* expression pattern there (17, 25), in homozygous mutant leg 148 discs for a null C15 allele or in genetically-mosaic leg discs harboring rotund or bowl loss-offunction mutant cells (Fig 1D-F). 149

150 *C15* is specifically activated in the distalmost (center) part of the leg disc giving rise to the 151 pretarsal (pt) segment (see Fig 1B) (30, 31). We have previously shown that the C15 152 homeoprotein down-regulates directly *bab2* to restrict its initially broad distal expression to the 153 tarsal segments (25). Bab1 expression analysis in a homozygous *C15* mutant leg disc revealed 154 that both *bab1* and *LAE-RFP* (*bab2*) are similarly de-repressed in the pretarsus (Fig 1, compare 155 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 cellautonomously 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 165 166 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 167 in the tarsal territory (33). As previously shown for LAE-GFP (and bab2) expression, nuclear 168 Bowl stabilization in the developing tarsal region appears sufficient to down-regulate cell-169 autonomously *bab1* (S1C Fig). Prolonged expression of the Rotund protein in the entire distal 170 part of the developing leg disc, i.e., tarsal in addition to pretarsal primordia, induces ectopic 171 172 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 173 174 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.

180

181 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 $\Delta LAE-M1$ 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*^{ΔLAE}.

190 First, we quantified *bab1* and *bab2* mRNAs prepared from dissected wild-type and homozygous bab^{ΔLAE} mutant leg discs. As shown in Fig 2C, both mRNAs were detected in mutant discs, 191 although *bab1* levels were two times lower than wild-type. Second, Bab1-2 expression patterns 192 were analyzed in homozygous bab^{ΔLAE} leg discs (Fig 3). To identify leg cells that should 193 normally express bab2, we used the X-linked LAE-GFP reporter. In homozygous $bab^{\Delta LAE}$ 194 mutant leg discs, bab2 specific expression (see Fig. 1B) is no longer observed (Fig 3B-C), while 195 196 Bab1-2 shared expression is very low in ts3-4 to undetectable in ts1-2. Nevertheless, residual bab1-2 co-expression in homozygous babALAE mutant discs indicates that additional cis-197 regulatory region(s) within the *bab* locus act(s), at least partly, redundantly with the LAE 198 enhancer. 199

Taken together, our data indicate that LAE enhancer activity is (i) required for *bab1-2* coexpression 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.

205

206 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.

217

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 220 existence of other redundant *cis*-regulatory elements, presumably located also within the *bab* 221 222 locus. On one side, a X-linked Bacterial Artificial Chromosome (BAC) construct, BAC26B15^{ZH2A}, encompassing the bab2 gene and downstream intergenic sequence including 223 224 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, 225 a BAC26B15 construct (BAC26B15ALAE^{ZH2A}) inserted at the same genomic landing site and 226 specifically lacking LAE sequence did not (Fig 4B-D). These results confirmed that (i) in 227 228 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 229 region covered by BAC26B15. 230

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 *BAC26B15* Δ LAE^{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.

244

245 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 <u>High-throughput chromosome</u> conformation <u>Capture (Hi-C)</u> 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 eyeantennal 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 <u>Chromatin Immuno-Precipitation (ChIP-Seq)</u>, <u>Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-Seq) and Assay for</u> <u>Transposase-Accessible Chromatin (ATAC-Seq) experiments (38-41)</u>, 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 "<u>bab1 Enhancer Region</u>", 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-262 published ChIP-Seq data from L3 leg discs (42) for binding sites for Dll, Sp1 and Hth proteins, 263 known to regulate *bab1* and/or *bab2* expression in the developing legs (17, 42-44). Strong Dll 264 binding is detected throughout BER, including over the *bab1* promoter (Fig 5B, lane 10; see 265 also S3B Fig). In leg discs, Dll binding is detected over 8 out of 10 Open Chromatin Subregions 266 (OCS) within BER (Fig 6C and S3B Fig) and six of those eight are also bound by Sp1 ZF 267 protein. Of note, all nucleosome-depleted (i.e., OCS) BER subregions in the leg are also 268 accessible in the eye-antennal discs (40) (Fig 5B, compare lanes 2 and 9, and Fig 6C, two upper 269 lanes). Lastly, of six OCS sequences co-bound by Dll and Sp1, four are also bound by Hth 270 271 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 272 273 and/or haltere discs (Fig 5B, lanes 13, 16, 18, and Fig 6C, four upper lines; see also S3B Fig). 274 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 275 276 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.

280

281 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 282 analysis of Gal4 reporter lines (45). Out of six lines containing BER fragments (Fig 6B), only 283 two, 73C11 and 73C05, overlapping OCS1-3, are active in the leg and eye-antennal discs 284 (FlyLight database; (http://flweb.janelia.org/cgi-bin/flew.cgi; S4B Fig). Nevertheless, none 285 286 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 287 288 sequences (14) indicate that OCS1-7 (see Fig 6B) are not sufficient to properly drive bab1/2expression in the developing legs and suggest the requirement of additional BER elements, 289 particularly the *bab1* promoter region (i.e., OCS9). This hypothesis is consistent with binding 290 291 of the known bab1-2 leg regulators Dll and Sp1 throughout BER, in addition to LAE (Fig 5, 292 lane 10; and S3B Fig).

293 The 73C05 BER fragment also drives reporter gene expression in the wing, haltere and genital 294 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 295 296 OCS1-3 are nucleosome-free and bind the Ultrabithorax (Ubx) Hox-type homeoprotein known 297 to activate directly bab2 expression in haltere tissues (Fig 5B, lanes 16-17, and Fig 6C) (46, 298 47). Furthermore, ChIP-Seq data from whole L3 larvae (modENCODE), showed binding over 299 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 300 by the mesodermal transcription factors Mef2, Slp1 and Tinman in late embryos (14, 49) as 301 well as (ii) a sequence element (overlapping OCS5) which confers enhancer activity in ovarian 302 somatic cells (50) (see Fig 6D). 303

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 proposedto act as a pleiotropic enhancer region.

309

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

311 Bab proteins interact with A/T-rich DNA sequences through their BabCD DNA-binding 312 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 313 indicated that Bab2 protein expression is unaffected in *bab1* loss-of-function mutants (11). 314 315 Given that protein null *bab2* alleles are not available, we used RNA interference coupled to 316 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, 317 both LAE-RFP and bab1 were up-regulated cell-autonomously in most tarsal mitotic clones 318 319 (n=17/20) (Fig 7, panels A-A"). Moreover, *bab2* down-regulation in proximal-most RFP+ ts1 320 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 321 the bab2 paralog specifically down-regulates its own expression through partial repression of 322 323 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 324 reporter up-regulation could be observed in all examined *bab*^{AR07} clones (detected with anti-325 Bab2 antibodies; n>20) (Fig 7, panels C-C"), independently of their size and position within 326 *bab2*-expressing tarsal cells (see white arrows and arrowheads). 327

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.

331

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

333 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, 334 we identified proteins orthologous to *Dmel* Bab1 or Bab2, i.e., displaying an N-terminal BTB 335 336 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 337 strongly related to Dmel Bab1 and Bab2, respectively, were identified in the Muscomorpha 338 339 (higher flies, also known as Cyclorrhapha) superfamily, both within the Schizophora (in Calyptratae, such as *Musca domestica* and *Glossina morsitans*, and in Acalyptratae, particularly 340 among Drosophilidae) and Aschiza subsections (Fig 8A-B and Supplementary data). In 341 contrast, a single BTB-BabCD protein could be identified in evolutionarily-distant dipteran 342 species within (i) the brachyceran Asilomorpha and Stratiomyomorpha superfamilies (such as 343 344 Proctacanthus coquilletti and Hermetia illucens, respectively), collectively referred to as Orthorrhapha; (ii) the Nematocera suborder families (with rare exceptions, in Psychodomorpha 345 and Bibionomorpha, see below); (iii) other Insecta orders (e.g., Coleoptera, Hymenoptera and 346 347 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.

364

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

366 Having traced back the bab gene duplication raised the question of the evolutionary origin of 367 the LAE enhancer, which regulates both *bab1* and *bab2* expression (17) (this work). We have 368 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 369 370 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, 371 while the 41 bp CR2 comprises contiguous binding sites for Dll as well as the ZF protein Bowl 372 (S7 Fig, panels B and C, respectively) (17, 25). 373

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 tandemlyduplicated 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. coquilletti* (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.

389

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

391 Given their differential interplay with the long-lasting LAE enhancer, we next analyzed the 392 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 393 (TTCAGTC), its *bab2* paralog displays tandemly-duplicated Inr sequences (ATTCAGTTCGT) 394 395 (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 396 397 (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* 398 399 promoter (56-58) probably also occurs for bab1 promoter.

Homology searches revealed that *bab1* promoter sequences have been strongly conserved inthe 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 Acalyptratae 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.
- 411

412 Unlike LAE, other *bab* CRE sequences have not been conserved beyond the 413 Muscomorpha

414 The broad LAE sequence conservation led us to also trace back the evolutionary origins of the 415 pleiotropic BER enhancer region as well as the cardiac CE, abdominal anterior AE and 416 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 417 (52), in Calyptratae schizophorans and even in the Muscomorpha Aschiza subsection (e.g., 418 OCS3) (S10-18 Fig). Unlike the LAE enhancer, homologous BER sequence elements (except 419 the bab1 promoter) could not be detected in non-muscomorphan families. Cardiac CE and 420 421 abdominal DE are even less conserved given that related sequences could be only detected within schizophoran (excepted in Calyptratae) (Fig 9B and S9 Fig, panels B and C, 422 423 respectively), whereas abdominal AE sequences could be only identified among drosophilids (Supplementary data) but not in aschizan, asilomorphan and nematoceran bab loci. 424

In conclusion, contrary to the LAE enhancer which among the Diptera emerged early on in the 425 426 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 427 long-lasting brachyceran LAE (CR1-2) sequences, these data suggest that other enhancer 428 429 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 430 region, BER sequence conservation allowed us to predict binding sites for transcription factors 431 known, or so far unsuspected, prone to regulate directly the two bab genes in many distinct 432 developmental contexts, and which are presented hereafter. 433

434

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

We gathered our data from TF binding site evolutive conservation (described in Supplementary 436 data and S10-19 Fig) with ChIP-Seq experiments from the literature (GEO datasets; see 437 438 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 439 compilation presented in Fig. 10 allows to propose new models for the TF code involved in 440 441 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 442 (Sd) in the wing and Ubx in the haltere (ChIP-Seq data shown in S3 Fig); (ii) It suggests that 443 BER also acts as an enhancer region for bab gene regulation in other developmental contexts 444 and that common TF sets (notably Abd-B together with Dsx) are acting through distinct cis-445 446 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 447 448 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.

454

455 **Discussion**

In this work, we have addressed the issue of the emergence and functional diversification of 456 457 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 458 duplication in the Muscomorpha/Cyclorrhapha. The early-fixed brachyceran bab1 LAE has 459 been co-opted lately to regulate also bab2 expression. Furthermore, this unique enhancer is also 460 responsible for paralog-specific *bab2* expression along the P-D leg axis presumably through 461 462 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 463 a large pleiotropic enhancer driving *bab1* and/or *bab2* expression in highly-diverse tissues, by 464 465 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 466 enhancers evolutionary adapt with distinct cognate gene promoters, and (iii) which functional 467 links can be predicted between dedicated transcription factors and chromatin dynamics during 468 development. 469

470

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

Here, we have shown that a single enhancer, LAE, regulates two tandem gene paralogs at the 473 474 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 475 foci composed of high concentrations of TFs and proteins from the basal RNA polymerase II 476 477 (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), 478 479 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 480 regulation. In contrast, no strong chromosome contacts could be detected between LAE and 481 482 any of the two *bab* promoter regions, indicating that this enhancer is not stably associated to 483 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 484 485 bab1-2 genes are much more broadly expressed.

In addition to be required for *bab1-2* co-expression in proximal tarsal segments, we showed 486 487 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 488 through enhancer emergence, our present work indicates that differential expression of two 489 490 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-491 expressing ts1 cells (44) and more distally in the ts5 segment expressing nuclear Bowl protein, 492 493 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 494 495 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 496 transiently-expressed Rotund activating TF may antagonize Bowl (and eventually Dac) 497

repressive activity to precisely delimit *bab2* expression among ts1 cells (17, 61). Given that 498 499 *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 500 sequences (discussed below). Thus, we speculate that the *bab2* promoter responds to Rotund 501 502 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 503 504 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 505 pre-initiation complex stably-associated to the bab2 core promoter. We envision that the LAE-506 507 bound ZF protein Rotund, the chromatin-remodeling ZF protein GAF (for GAGA-associated 508 Eactor) and the PolII-associated TFIID subunit TAF3, the latter known to interact physically 509 with GAF and Bab2 BTB proteins (62, 63), are parts of the underlying promoter targeting 510 molecular mechanism.

In this context, despite that sequence homologies between both promoters (consistent with an 511 512 ancient duplication event mobilizing the ancestral bab1 promoter) are still detectable, it is 513 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 514 515 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 516 fast evolutionary drift after the gene duplication process which yielded the *bab2* paralog. We 517 envision that this evolutionary ability has largely contributed to allow novel expression patterns 518 519 for *bab2*, presumably through differential enhancer-promoter pairwise interplay.

520

521 Differential evolutionary conservation of tissue-specific versus pleiotropic enhancers

Our comprehensive phylogenomics analyses from highly diverse Diptera families indicate that 522 523 the *bab* gene complex has been generated through tandem duplication from an ancestral *bab1*related gene singleton within the Muscomorpha (Cyclorrhapha), about 100-140 years ago. This 524 result contrasts with published data reporting that the duplication process having yielded the 525 526 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) 527 528 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 bab1-related 529 gene copies (Figure 8 and S6 Fig), but our phylogenetic analysis supports independent events. 530 531 Thus, within the Diptera, the ancestral *bab1* singleton had a high propensity to duplicate locally. In this study, we have shown a strong evolutionary conservation of LAE subsequences among 532 533 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 534 ancestral singleton *bab1* genes. In striking contrast, BER sequence conservation could only be 535 536 detected among extant muscomorphan bab loci. We assume that during evolution large

thus generating regulatory novelties in distinct imaginal discs.

537

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

pleiotropic enhancers may better assimilate binding sites for gene-specific transcription factors,

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

548

549 A pleiotropic enhancer region overlapping with a PcG-response element

ChIP-Seq analysis for histone H3 modifications (H3K4me3, H3K27Ac and H3K4me1 550 551 enhancer/promoter marks; H3K27me3 chromatin repressive mark) from eye-antennal discs has 552 revealed the pleiotropic BER enhancer region but also an overlapping repressive PcG (Polycomb Group family) domain, indicating that BER encompasses a bivalent chromatin 553 554 domain, while another one is detected over the *bab2* promoter region. A dual enhancer/silencing 555 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 556 by PcG proteins may help fine-tune gene expression and ensure the timely maintenance of cell 557 identities throughout development. More recently, we have shown widespread enhancer-PcG 558 559 domain interplay during developmental gene activation through chromatin looping in eyeantennal discs (38). Altogether these data suggest that the bab1-2 genes might be poised for 560 activation throughout the eye-antennal disc, and possibly other imaginal discs as well. 561

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 569 570 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). 571 Significantly, two of these predicted CTCF interacting sites overlap with putative optimal 572 573 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 574 575 locus chromosome topology. Interestingly, the human Pho homolog (YY1) is a structural 576 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, 577 578 functional relationships between CTCF and Pho proteins have not been investigated genome-579 wide in Drosophila.

580

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

582 Recent data indicate that chromatin accessibility is dynamic during Drosophila larval development, being triggered by the ecdysone hormone (69). Dynamic enhancer activity and 583 chromatin accessibility have been proposed to be regulated by the ecdysone-induced Eip93F 584 585 (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 bab1 promoter 586 region (70). Consistently, many putative Eip93F binding sites are present in these BER 587 subregions and several have been conserved beyond Drosophilidae (Fig 10C). Interestingly, the 588 human Eip93F homolog interacts with CtBP through a conserved motif (71), and Drosophila 589 590 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 591 elements may impact the proposed dual PcG activity at the bab locus. In addition to Eip93F, 592

593 BER regulatory sequences include many evolutionarily-conserved putative binding sites for the 594 Eip74EF protein (Fig 10C), another ecdysone-induced TF, including one which overlaps with 595 a conserved putative Pho binding site, suggesting again functional correlation between 596 Ecdysone regulation and PcG activity.

597 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-598 599 ZF protein participating to developmental timing, notably through interplay with ecdysone 600 signaling (73, 74). Consistently, ChIP-Seq experiments from embryos (ModENCODE data; http://www.modencode.org/) indicate Chinmo binding to BER sequences (75). Intriguingly, the 601 602 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 603 protein interacting directly with both *bab* promoter regions; Fig 5B and S3 Fig) and the twin 604 605 Bab BTB-BabCD proteins themselves. Thus, Chinmo implication in chromatin organization and enhancer activity within the *bab* locus undoubtedly deserves to be investigated. 606

607 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 608 609 transcriptional memory during development and selective enhancer-promoter interplay in diverse tissular contexts. Indeed, our comprehensive predictive combinatorial code for tissue-610 specific, as well as broadly-expressed architectural transcription factors (e.g., CTCF, Pho and 611 GAF) regulating two tandem gene paralogs, offers the opportunity to dissect underlying 612 613 molecular mechanisms, which are prone to be conserved during animal evolution and thus to be of broad biological significance. 614

615

616 Material sand Methods

617 Fly stocks, culture and genetic manipulations

618 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 LAEpHsp70-619 620 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 621 (X chromosome) or ZH86Fb (third chromosome) attP landing platforms, and displaying 622 identical expression patterns, have been previously described (17, 25). $C15^2$ /TM6B, Tb^1 stock 623 624 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-625 626 *Flp/+; FRT82B, Ub-RFP/FRT82B rn*¹² (i.e., *rn* mutant clones are RFP negative; Fig 1E) and v w LAE-RFP; DllGal4^{EM2012}, UAS-Flp/+; Ub-GFP, FRT40A/bowl¹ FRT40A (i.e., bowl mutant 627 clones are GFP negative; Fig 1F), respectively. Rotund protein gain-of-function within the Dll-628 expressing domain was obtained with the following genotype: y w LAE-GFP; DllGal4^{EM2012}; 629 UAS- Rn^{1} /+. The Dll^{EM212}-Gal4 line was provided by M. Suzanne, while the UAS- Rn^{1} line was 630 631 obtained from the Bloomington stock center. "Flip-out" (FO) mitotic clones over-expressing 632 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-dsRNAlines/pAct>y+>Gal4, UAS-GFP (i.e., 633 FO clones express GFP in S1C Fig). Mutant mitotic clones for the null *bab*^{AR07} allele were 634 generated by 30 mn heat shocks at 38°C, in early first to late second-instar larvae of genotypes: 635 y w LAE-GFP, hsFlp; FRT80B/bab^{AR07}, FRT80B. FO mitotic clones over-expressing dsRNA 636 against bab1 or bab2 were generated by 40 mn heat shocks at 38°C, in early to early-mid L3 637 larvae of genotypes: *y w LAE-RFP hsFlp*; *UAS-dsbab2 /Pact>y+>Gal4*, *UAS-GFP* and *y w* 638 639 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 640 (#35707) and *bab2* (#37260) were obtained from the Bloomington stock Center. 641

642

643 Immuno-histochemistry and microscopy

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

649

650 CRISPR/Cas9-mediated chromosomal deletion

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

664

665 **Quantitative RT-PCR analysis**

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

675

676 Homology searches, sequence alignments and phylogenetic analyses

- Homology searches were done at the NCBI Blast site (https://blast.ncbi.nlm.nih.gov/Blast.cgi). 677 Protein or nucleotide sequence alignments were done using MAFFT (Multiple Alignment using 678 Fast Fourier Transform) (https://mafft.cbrc.jp/alignment/server/). Phylogenetic relationships 679 inferred through maximum likelihood analysis with W-IQ-Tree 680 were a (http://iqtree.cibiv.univie.ac.at/) visualized with ETE toolkit 681 and the 682 (http://etetoolkit.org/treeview/).
- 683

684 Transcription factor binding prediction

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

688

689 Gene expression omnibus datasets

- 690 The following gene expression omnibus (GEO) datasets were extracted from the NCBI website
- 691 (<u>https://www.ncbi.nlm.nih.gov/gds/</u>): 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

Hi-C and histone mark ChIP-Seq analyses from L3 eye-antennal discs have been recentlypublished 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
- for technical assistance, particularly in managing the transgenic facility. Lastly, weacknowledge Brice Ronsin and the Toulouse RIO Imaging platform.

705

707 **References**

- 1. Ohno S. Evolution by Gene Duplication. Springer-Verlag; 1970.
- 709 2. Zhang J. Evolution by gene duplication an update. Trends Ecol. Evol.; 2003. p. 292-8.
- 710 3. Lundin LG. Gene duplications in early metazoan evolution. Semin Cell Dev Biol. 1999;10(5):523-
- 711 30.

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

715 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

717 Development and Evolution. Cell. 2016;167(5):1170-87.

718 7. Rubinstein M, de Souza FS. Evolution of transcriptional enhancers and animal diversity. Philos
719 Trans R Soc Lond B Biol Sci. 2013;368(1632):20130017.

720 8. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of
721 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.

73012.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.

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

739 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

742 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.

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.

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.

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.

754 22. Camara N, Whitworth C, Dove A, Van Doren M. Doublesex controls specification and 755 maintenance of the gonad stem cell niches in. Development. 2019;146(11).

Duan J, Zhao Y, Li H, Habernig L, Gordon MD, Miao X, et al. Bab2 Functions as an EcdysoneResponsive 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.

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

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

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

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

33. Greenberg L, Hatini V. Essential roles for lines in mediating leg and antennal proximodistal
patterning and generating a stable Notch signaling interface at segment borders. Dev Biol.
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):305787 9.

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

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

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

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

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

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

43. Córdoba S, Requena D, Jory A, Saiz A, Estella C. The evolutionarily conserved transcription

factor Sp1 controls appendage growth through Notch signaling. Development. 2016;143(19):3623-31.

Kongen Marken Marken

45. Jenett A, Rubin GM, Ngo TT, Shepherd D, Murphy C, Dionne H, et al. A GAL4-driver line resource
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.

49. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE. Combinatorial binding predicts spatio-

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

activity maps identified by STARR-seq. Science. 2013;339(6123):1074-7.

51. Lours C, Bardot O, Godt D, Laski FA, Couderc JL. The Drosophila melanogaster BTB proteins bric

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.

S2. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al. Evolution of genes and
genomes on the Drosophila phylogeny. Nature. 2007;450(7167):203-18.

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

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

S5. Gilchrist DA, Dos Santos G, Fargo DC, Xie B, Gao Y, Li L, et al. Pausing of RNA polymerase II
disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell.
2010;143(4):540-51.

- 835 56. Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, et al. RNA polymerase is
- 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
- 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.
- Kim S, Shendure J. Mechanisms of Interplay between Transcription Factors and the 3D
 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.
- 61. de Celis Ibeas JM, Bray SJ. Bowl is required downstream of Notch for elaboration of distal limb
 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
- 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.
Beagan JA, Duong MT, Titus KR, Zhou L, Cao Z, Ma J, et al. YY1 and CTCF orchestrate a 3D
chromatin looping switch during early neural lineage commitment. Genome Res. 2017;27(7):1139-52.
G9. Uyehara CM, Nystrom SL, Niederhuber MJ, Leatham-Jensen M, Ma Y, Buttitta LA, et al.
Hormone-dependent control of developmental timing through regulation of chromatin accessibility.
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).

Quinlan KG, Verger A, Kwok A, Lee SH, Perdomo J, Nardini M, et al. Role of the C-terminal
binding protein PXDLS motif binding cleft in protein interactions and transcriptional repression. Mol
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.

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

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

37

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.436949; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

- 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

A



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 C15² 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 nonexpressing 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).

Bourbon et al



pt

ts3

ts2

ts4

ts5 pt

ts2

ts3

ts4

ts5



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 ($bab^{\Delta LAE}$) is shown in beneath (deleted LAE is depicted as a broken line). (**B**) Photographs of wild-type and homozygous $bab^{\Delta LAE}$ T1 distal legs from adult males. The regular sexcomb (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 $bab^{\Delta LAE}$ mutant legs. (**C**) Overall bab1-2 expression from wild-type and homozygous $bab^{\Delta LAE}$ L3 leg discs, as determined from reverse transcription quantitative PCR analyses. The bab1-2 expression levels were quantified relative to rp49 mRNA abundance.

Bourbon et al



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) openreading 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).

Bourbon et al

Figure 3



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).

Bourbon et al



Fig 5. A topologically-associating domain encompasses the *bab* locus in the eyeantennal 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).

Bourbon et al

Figure 5



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].

Bourbon et al





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 eyeantennal (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).

bab2 RNAi clones (GFP positive)



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 9. Conservation of enhancer/promoter sequences and evolutionary history of the *bab* locus among the **Brachycera**. (A) Organization of the *Dmel bab* gene paralogs and enhancers. The locus is depicted as in Fig 1A, except that *bab2* is represented in pink instead of red. (B) Evolutionary conservation of the *bab* gene paralogs, enhancers and promoters among diverse dipterans. Infraorders, sections, subsections and superfamilies are indicated on the left, arranged in a phylogenetic series from the "lower" Nematocera to the "higher" Brachycera suborders. Presence of *bab1* and/or *bab2* paralogs and conservation of enhancer and promoter sequences are indicated by filled or hatched boxes colored as depicted in (A). (C) Evolutionary scenario for the *bab* locus within the Brachycera suborders. A scheme depicting chromosomal fate of an ancestral *bab1*-like gene which gives rise to derived extant orthorrhaphan singletons (Asilomorpha) and Muscomorpha-specific paralogous (Calyptratae and Acalyptratae) genes. Locations of conserved enhancer sequences are shown, as depicted in (A).

Bourbon et al

Figure 9



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) Evolutive 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).

Bourbon et al

Figure 10

Wild-type leg disc



rotund gain-of-function (Dll domain)



bowl gain-of-function clones (FO lines RNAi; GFP+)



	Up targeted site	LAE_	Down targeted site
Wt	TAGGGTATTTGCGTGGAGCCTTCTTCGCCAGGCCGGCCTTTTGT	ITTTATTTCCTTGACTC//CCGGITGTTCTAATT	GTTGGCTAGTTAGGCGCACATAAGGAGGCAGGTCTCTGAACCC
M1	TAGGGTATTTGCGT	////////	caaaagagtcaGTCTCTGAACCC
M2	TAGGGTATTTGCGTGGAGCaaat	////	agagaCTGAACCC
M3	TAGGGTATTTGCGTGGAGCCTTC	//	ItcgcAGGTCTCTGAACCC
M4	TAGGGTATTTGCGTGGAGCCTTC	//	tagacctgcGCAGGTCTCTGAACCC
M5	TAGGGTATTTGCGTGGAGCCTTCTTgcgcaaaat	 //	
M6	TAGGGTATTTGCGTGGAGCCTTCTTCGC	////	gcGGCAGGTCTCTGAACCC





73C11-Gal4 + UAS-GFP



Leg disc

Eye-antennal disc

73C05-Gal4 + UAS-GFP





Female genital discs

Male genital disc







С





Bourbon et al

S8 Figure





ATCAAA Pan

ig i c

D. melanogaster (Dmel) bric-a-brac enhancers

Mdom

Α

S9 Figure



CS element(s) conserved beyond Drosophilidae



Bourbon et al

S11 Figure



Bourbon et al

S12 Figure



S13 Figure



Leg/Eye-antenna/Wing CS1 within OCS5





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





Eye-antenna-specific CS1 within OCS7



Eye-antenna-specific CS2 within OCS7



Eye-antenna-specific CS6 within OCS7



S16 Figure



Leg/Eye-antenna/Wing/Haltere CS3 within OCS9

Leg/Eye-antenna/Wing/Haltere CS2 within OCS9



Leg/Eye-antenna/Wing/Haltere CS4 within OCS9



S17 Figure





bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.436949; this version posted March 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 fourletter 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.

Dmel: Drosophila melanogaste Dsim: Drosophila simulans Dsec: Drosophila sechellia Dyak: Drosophila yakuba Dere: Drosophila erecta Dsuz: Drosophila suzukii Drho: Drosophila rhopaloa Dele: Drosophila rhopaloa Dele: Drosophila elegans Dbip: Drosophila bipectinata Dana: Drosophila bipectinata Dana: Drosophila persimilis Dpse: Drosophila persimilis Dpse: Drosophila pseudoobscu Dwil: Drosophila virilis Droi: Drosophila virilis Dmoj: Drosophila mojavensis Dgri: Drosophila grimshawi Tmin: Themira minor	r Drosophilidae	Acalyptratae	Schizophora	orpha/Cyclorrhapha	
Tdal: Teleopsis dalmanni Blat: Bactrocera latifrons				uscom	
Ccap : Ceratitis capitata				Σ	
Mdom: Musca domestica					Brachycera
Scal: Stomoxys calcitrans					
Lcup: Lucilia cuprina		ae			
Preg: Phormia regina		rat			
Chom: Cochliomyia hominivoro	X	ypt			
Pmac: Paykullia maculata		Cal			
Gbre: Glossina brevipalpis					
Gmor: Glossina morsitans					
Edim: Eristalis dimidiata			Aschiza		
Mabd: Megaselia abdita					
Cpat: Condylostylus patibulatu	S			ua l	
Cmol: Chrysotimus molliculus			Acilometraha	Jap	
Pcoq: Proctacanthus coquilletti			Asilomorpha	orr	
Dala: Dasypogon alaaema				th	
Hill: Hormotia illusons				0	
nili. Hermetia materis			Stratiomyomorpha	I	
Mdes: Mayetiala destructor			 Dinionomorpho		1
Cfus: Coholdia fuscines			ыппопопногрна І		
Pnat: Phlehotomus nanatasi			Psychodomorpha		
I lon: I utzomvia longinalnis			r sychodomorpha		Nematocera
Aaam: Anopheles aambiae			L Culicomorpha		
Aaea: Aedes aeavoti					
Dpul: Daphnia pulex	Crustacea		-		-

Bab1 sequence conservation among muscomorphans (Part1)



Bab1 sequence conservation among muscomorphans (Part2)

Dmell Dysel Dvirl Blat1 Ccapl Tminl Scall Lcupl Pregl Choml Pmac1 Gbrel Gbrel Gmorl Ediml Mabdl	464 502 499 646 681 594 623 577 568 567 568 567 546 546 522 260	P H H H G G G V G G G V G G G A G V G S G G S S L A D D L E I F P G I A EM I R E E B A R M M D N S H A M G A T A G S L B A . D S Y O Y O L O S M W O K C H N T Y O N L M H H M F R R G D L K S H R D E T M A F H H H G G V G G G A . G V G G G G V G S G V G Y G Y O L O S M W O K C H N T Y O N L M H H M F R R G D L K S H R D E T M A F H H H S O F Y G G A . G A Y G A G A . D A Y O Y O L O S M W O K C H N T Y O N L M H H M F R R G D L K S H R D E T M A F H H H S O F Y G G A . G A Y G A Y G A G A . D A Y O Y O L O S M W O K C H N T Y O N L M H H M F R R R G D L K S H R D E T M A F H H H S O F Y G G A . G A Y
Dmell Dpsel Dvirl Blatl Ccapl Tminl Tdall Mdoml Scall Scall Choml Pregl Choml Pmacl Gbrel Gbrel Gmorl Ediml Mabdl	576 612 598 736 730 770 683 713 727 658 651 656 634 634 634 605 342	
Dmell Dysel Dvirl Blatl Ccapl Tminl Tdall Mdoml Scall Lcupl Pregl Choml Pmacl Gbrel Gmorl Ediml Mabdl	688 710 847 847 878 878 878 793 825 834 777 768 761 766 744 714 452	A A S G A L A G A P S S M A C P N C S G P O T G V G V A C C H S O E T A A V A V A H N H I R O M O M A N V PPP C C A S G G T P P G A S L S A L A P N M G S G G G G V T G P G G V G V P G V D (H S O E T A A V A V A H Y A H I R I R O M O M A A V V H O H G I E A G P P VV P P P G A C M S G C T P G A S L S A L A P N M G S G G G G V T G P G G V G V P G V D (H S O E T A A V A V A H Y A H I R I R O M O M A A V V H O H G I E A G P P VV P P P G A C M S G P P N M G S G G G G A V G A S (G G G G A V A A V A H I R I R O M O M A A V O H O H G I E A G P P PPP G A C M S G P P N M P T P T G D O M S (E T A A V A A V A H I R I R O M O A A N A O H O H F E E G P F PP N M P S P G M G G A S G M P N A M A A H M H I R O M O A A N A O H O H F E E G P F G P F G A G G G G G G G G G G G G G G G G G G
Dmell Dpsel Dvirl Blat1 Ccap1 Tmin1 Tdal1 Mdom1 Scal1 Lcup1 Preg1 Chom1 Pmac1 Gbre1 Gbre1 Mabd1	748 806 794 904 898 959 890 908 8356 819 824 801 801 804 507	LUNL PHP
Dmell Dpsel Dvirl Blat1 Ccap1 Tmin1 Scal1 Lcup1 Pregl Chom1 Pregl Gbre1 Gbre1 Edim1 Edim1	811 870 996 990 10537 925 916 908 911 895 895 895 895	
Dmell Dpsel Dvirl Blatl Ccapl Tminl Scall Lcupl Pregl Chonl Pregl Gbrel Gbrel Gbrel Mabdl	870 940 940 1072 1066 1115 1012 1009 1075 979 972 963 968 951 953 953 933 535	SSBMGOHHAPKAKS.SPL RSETP. RUHSPGDUGTD MA. SYKR.FFS. HSKPHSKBHSGEVOKKG.SPH RSETP. RUHSPGDUGTE MS. SYKR.FFS. COQLOCKSKSKG.SPL RSETP. RUHSPGDUGTE MS. SYKR.FFS. COQLOCKSKSKG.SPL RSETP. RUHSPGDUGTE MS. SYKR.FFS. COQLOCKSKSKG.SPL RSSPSSLEH. RSSPSGN.SHHSPGTE LEGTD MS.KTRSSSSYS. COMPOCOLOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO
Dmell Dpsel Dvirl Blatl Ccapl Tdall Mdoml Scall Lcupl Pregl Cbrel Gbrel Gbrel Gmorl Ediml Mabdl	913 988 982 1144 1159 1185 1085 1058 1129 1035 1025 1025 1019 1025 1009 1011 992 584	SRUE AB DIACLVGASVS
Dmell Dpsel Dvirl Blatl Ccapl Tminl Scall Lcupl Pregl Choml Pregl Gbrel Gbrel Gbrel Ediml	961 1029 1201 1217 1257 1142 1164 1240 1120 1104 1108 1071 1073 1021 608	S S S G G I V S F I . T S S S S S G G I V S F I . T S S S S S G G I V S F I . T S S S S S G G I V S S F I . T S S S S S G G I V S S F I . T S S S S S S S S S S S S S S S S S S S

BabCD

Bab2 sequence conservation among muscomorphans (Part1)

Dme12 Dpse2 Dvir2 Blac2 Ccap2 Tmin2 Scal2 Scal2 Preg2 Chom2 Gbre2 Gbre2 Gbre2 Gbre2 Edim2 Mequ2 Edim2	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	DMITKQINDFERK.SEIIGS
Dme12 Dpse2 Dvir2 Ccap2 Tmin2 Scal2 Lcup2 Scal2 Lcup2 Gbre2 Gbre2 Gbre2 Gbre2 Edim2 Madu2	77 66 64 79 108 69 72 551 56 58 54 47 47 8 7	KDIESVGEVKSPEREDVETELVKEKSAS.IIIIIII PRODUCTS EVVGETAD. DE EMENC. ILEAKES ELIVE OF OF ORI. AT OF AT KKAAADLDE ADDE FON. AT OF AT KKAAADLDE ADDE KON. AT OF AT KKAAADLDE ADDE KON. AT OF AT KKAAADLDE ADDE KON. AT OF AD EXAMPLE OF PAA. ET ANSM. SVOKTSEAT AT AAAKTVG. AT MSGDIT PE DE DETSEVVGETAD. DE EMENC. ILEAKES ELIVES OF DAVA. NI KHAAODLE DANSME FOF AA. ET ANSM. SVOKTSEAT AT AAAKTVG. AT MSGDIT PE DE OFTSEVVGETSES EL. DE EMENC. VIE DE DANSKES ELIVES OF DAVA. SAE OG OC OLEO AKSTOANME FOF PAA. ET ANSM. SVOKTSEAT AT AAAKTVG. AT MSGDIT PE DE OFTSEVVGENSSALDD PE DUKVO. VOLF KS. SAE OC OC OLEO AKSTOANME FOF AL. ET ANSM. SVOKTSEAT AT AAAKTVG. AT MSGDIT PE DE OFTSEVVGENSSALDD PE DUKVO. VOLF KS. SAE OC OC OLEO AKSTOADAAVOOP RE. EL NDTVAAVE DE LKATASOP PKTVGAAVVSGET TAT DE PE OFTSEVVGENSSALDD PE DUKVO. VOLF KS. SAE OC OC OLEO AKSTOADAAVOOP RE. EL NDTVAAVE DE LKATASOP PKTVGAAVVSGET TAT DE PE OFTSEVVGENSSALDD PE DUKVO. TOV FK SE CLASON HERSELDSAS SAE OC OC OLEO ASSTOADAAVOOP RE. EL NDTVAAVE DE LKATASOP PKTVGAAVVSGET TAT DE PE OFTSE VVGET SATURES TG SAAATTSAVLET NTAT VKTPF VLENT TO LLENTADE SOV. NOOKKNENSTEDE LIKME LEVVQVPT TT OLLENTADE SSAKA LOSGOIS OF PE OF LISSEVVGET TAT. DE PE MOLDAS LLASE CLASON HEAR ALSSNEN HAR KF FARMENTAD OLOK PTT OLLENTANV. STLK. PAMGROSIS PE OF LISSEVVGET TAT. DE PE MOLDAS LLASE CLASON HEAR ALSSNEN HAR KF FARMENTAD OLOK PTT OLLENTANV. STLK. PT VG SIG IS OF PE OF LISSEVVGET TAT. DE PE MOLDAS LLASE CLASON HEAR ALSSNEN HAR KF FARMENTAD OLOK PTT OLLENTANV. STLK. PAMGROSIS PE OF LISSEVVGET TAT. DE PE MOLDAS LLASE LASEN HEAR TISSNEGTIS PE OLOK SE LIASEN VK. FV PAGVELSEKTAYNES VERVENTE AL OLOK PT POLLEN KANV. STLK. PT VG SIG IS OF PE OLITSSEVVGET TAT. DE PE MOLDAS LLASEN HEAR TISSNEGTIS PE OLITSSEVGET TAT. DE PE MOLDAS LASEN HEAR TISSNEGTIS PE OLITSSEVGET TAT. DE PE MOLDAS LASEN HEAR STATAS ALSEN HEAR ANV. STLK. PAMGROSIS PE OLITSSEVGET TAT. DE PE MOLDAS LASEN HEAR TISSNEGTIS PE OLITSSEVGET TAT. DE PE MOLDAS LASEN HEAR STATAS ALSEN HAR ANV. STLK. PAMGROSIS PE OLITSSEVGET TAT.
Dme12 Dpse2 Dvir2 Blac2 Ccap2 Tmin2 Scal2 Scal2 Preg2 Chom2 Pmac2 Gbre2 Gbre2 Gbre2 Gbre2 Edim2 Mabd2	154 143 141 185 200 201 155 167 153 145 150 152 148 140 140 64 64 44	LKEAA II. GSALEFFGGRSS
Dme12 Dpse2 Dvir2 Blac2 Ccap2 Tmin2 Tdal2 Mcdom2 Scal2 Lcup2 Chom2 Prag2 Ghor2 Ghor2 Ghor2 Mequ2 Edim2 Mabd2	205 204 200 263 290 285 214 230 227 199 204 206 202 193 193 193 128 126 82	O S LL "N PE DE LOS S S FVO VT C 0 C C S L A HENV S A CS P F O A HEVE N P C 0 HE TI M RD VS C DEL KALV PEM KGE I N C 000 T N C L KUKETL K RG LA EVS A CS C L , GC V S N LT N PE DE LOS S FVO VT C 1 C S C S KAREM V S A CS P F O A HEVE N P C 0 HE TI M RD VN C D L KALV PEM KGE I N C 000 T N L KUKETL K RG LA EVS A CS S L , GC V S N LT N PE DE LOS S FVO VT C 1 C S C S KAREM V S A CS P F O A HEVE N P C 0 HE TI M RD VN C D L KALV PEM KGE I N C 000 T N L KUKETL K RG LA EVS A CS S L , A R N C A CS P F O A HEVE N P C 0 HE TI M RD VN C D L KALV PEM KGE I N C 000 T N L KUKETL K RG LA EVS A CS S L , A R N C A CS P F O A HEVE N P C 0 HE TI M RD VN C D L KALV PEM KGE I N M C 000 T N L KUKETL K RG LA EVS A GS G L , A R N C A CS P F O A HEVE N P C 0 HE TI M RD VN C D L KALV PEM KGE I N M C 000 T N L KUKETL K RG LA EVS A GS G L , A R N C A CS P F O A HEVE N V S A CS P F O A HEVE N P C 0 HE TI N RD VN C D L KALV PEM KGE I N M C 000 T N L KUKETL K RG LA EVS A GS G L , A R N C A CS P F O A HEVE N P C 0 HE TI N RD VN C D L KALV PEM KGE I N M C 000 T N L KUKETL K RG LA EVS A FG G L , G N T N T N C A CS P F O A HEVE N V S A CS P F O A HEVE N RO V R D E LA LW PEM KGE I N M C 000 T N L KUKETL K RG LA EVS A FG G L , G N T N T N N S A CS P F O A HEVE N RO V R D E LA LW PEM KGE I N M C 000 T N L KUKETL K RG LA EVS A FG G L , G N T N T N V S A CS P F O A HEVE N RO V R D E LA LW PEM KGE I N M C 000 T N L KUKETL K RG LA EVS S C G G L , G N T N T N S A CS P F O A HEVE N RO V R D E LA LW PEM KGE I N M C 000 T N L KUKETL K RG LA EVS S C G G N KA R M L S A CS P F O A HEVE O R I N R N R D E LA LW PEM KGE I N M C 000 T N L KUKETL K RG LA EVS S C G G N KA R M L S A CS P F O A HEVE O R D R N R D E LA LW PEM KGE I N M C 000 T N L KUKETL K RG LA EVS S C G G N KA R M L S A CS P F O A HEVE O R D R N R D E LA LW PEM KGE I N M C 000 T N L KUKETL K RG LA EVS S C G G R A GG
Dmel2 Dpse2 Dvir2 Blac2 Ccap2 Tmin2 Tdal2 Scal2 Lcup2 Preg2 Chom2 Preg2 Ghoe2 Ghoe2 Ghoe2 Ghoe2 Edim2 Mabd2	318 317 313 376 404 398 325 341 342 310 315 317 313 304 304 241 239 195	ASALPMSAFDDEDEEEEELASATATIQOD.GDAD.PDEEMKA RPRILPDEV.LCIN.ORCRKERGOSVATPSPSL. ASAHPMLPOQRMSVYGEDEEEEELASATIQOD.GDAD.PDEEMKA RPRILPDIA.COLN.ORCRKERGOSVATPSPSL. AS.GGLGGASLPSDRWYVEDDEDEEESELASATIGODHEDPEPRAKAMMADIA.LDLN.OCRRKERGOSVATPSPSLR AS.GGLGGASLPSDRWYVEDEDEEEESALAATLEG.SED.ODENVAR RAKIMADIT.CLN.OCRRKERGOSVATPSPSLR AAPHYMPSGRMITEDEEDVSDDESA.BDGCVERAKAMDSSN.SVPFHAAA.LDLN.OCRRKERGOSVATPSPSLR AAPHYMPSGRMITEDEEDVSDDESA.BDGCVERAKAMDSSN.SVPFHAAA.LDLN.NACKORKERGOSVATPSPSLF AAAHPMLPSGRMITEDEEDVSDDESA.BDGCVERAKAMDSSN.SVPFHAAA.LDLN.NACKORKERGOSVATPSPLECR AAAHPMLPSGRMITEDEEDVSDD
Dmel2 Dpse2 Dvir2 Blac2 Ccap2 Tmin2 Tdal2 Scal2 Lcup2 Preg2 Chom2 Preg2 Gmor2 Gmor2 Gmor2 Mequ2 Edim2 Mabd2	390 394 459 485 490 417 433 436 395 402 404 403 393 393 312 310 245	QGCES. FISERGSSGT. PGSSSGD LAMTRSTIVEN PASPE PO. TLEGE. PDUVES. FAGPERTAV. TESGSGD LAMTRSTIVEN PASPE PO. SLQAHLLIGTGTSSSASSSGG. TSSIVGTAAVGT PDUVES. FAGPERTAV. TESGSGD LAMTRSTIVEN PASPE PO. SLQAHLLIGTGTSSSASSSGG. TSSIVGTAAVGT CALASS. PORTSGASSGT. TESGVGD LAMTRSTIVEN PASPE PO. SLQAHLLIGTGTSSSASSSSGG. TSSIVGTAAVGT CALASS. PORTSGASSGT. TESGVGD LAMTRSTIVEN PASPE PO. SLQAHLLIGTGTSSSASSSSGGS. AASTAGVGANCO GANASMOL.FTISSAEPNKOTTMATTDE ISGAVVSGAPOOPPUANTSTIVEN PASPE PP N. TITENA. AANANGNOGAI. ATRSAPTAASS GSNAM.SSASSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
Dmel2 Dpse2 Dvir2 Blac2 Ccap2 Tmin2 Tdal2 Mdom2 Preg2 Chom2 Preg2 Gbre2 Gbre2 Gbre2 Edim2 Mabd2	437 466 470 548 571 587 509 470 477 479 478 464 463 386 365 311	NSAMNAVANQRKSPAPTATGHENGNSGAAMHSO PGGVAVQSALPPHM.AAIV. S.SSAGSGSSAAS.SSAGTIRT.AARGCSPPHQHOHQHHHPGGSKGSSGGAQGALPPHM.AAAVAA TET.LCSSTPTAPDTPTSLATSHSYRPPSM.PS.ASSGAQQALPPHM.AAAVAA TET.LCSSTPTAPDTPTSLATSHSYRPPSM.PS.ASSLKGAKSSQAAAAAAAAAAGAYVSGWLLWHSNP.AAAAAA SAV.SMGPSSTGASNTPSABSGYRA.QVACTPFL.TAAAFTYVAAPPPPLFAQSGTQPMIPSAVGLLGSSS DSNNS.AGHLHVHSNP.AAAAAA SAV.SMGPSSTGASNTPSABSGYRA.QVACTPFL.TAAAFTYVAAPPPPLFAQSGTQPMIPSAVGLLGSSS O.TASPSIAASSNTPFSABSGYRA.QVACTPFL.PSA.HG.S.SASVGGNT.ATPSSLPSD SSNSS.AGFINHSHFAAAAAAA O.TASPSIAASSNTNPVALOVRESGENSGEFL.MITHYGSR.SASVGGNT.ATPSSLPSD SSNSGPP.PPPPISITHPHF AAAAAA AAAAAFNVALOSSASTASNTPFSABSGYRA.QVACTPFL.MITHYGSR.SASVGGNT.ATPSSLPSD SSNSSAFP.SSGPSD.PPPISITHPHF AAAAAA O.TASPSIAASSNTNPVALOVRESGENSGEFL.MITHYGSR.SASVGGNT.ATPSSLPSD SSNSGPP.PPPISITHPHF AAAAAA SANTAFYA.ERSLESSL.MITSKS.SASVGGNT.ATPSSLPSD SSNSGPP.PPISITHPHF AAAAAA SANTYFF.SLERSGESS.MITSSASST.PVALOVRESGENSGEFP.NVALTAAA SAANAFSR.KERSLESSL.MITAGASSASSNSSAFP.SSSAFP.PPISITHPHF AAAAAAYA SAANTAFF.SASSSTRA.QVACTPFL.AAAFTAAA SAANTAFF.SASSST.PVALOVRESGENSGEFP.NVALTASSASSAFP.NVALTASSASSAFP.PPISITHPHF AAAAAAA SANTYFF.SLERSGESST.MITSSASSAFP.NVALTAFS.LENGSSAFP.NVALTASSASSAFP.NVALTASSA SAANTYFF.SLERSGESST.MITSSASSAFP.NVALTAFAA SAANTYFF.BERSLESSL.MITSSASSAFP.SSSAFP.PPISITHPHF AAAAAAYA SAANTYFF.BERSLESSL.MITSSASSAFP.SSSAFP.NVALTASSA SAANTYFF.BERSLESSL.MITSSASSAFP.SSSAFP.NVALTASSA SAANTYFF.BERSLESSL.MITSSASSAFP.SSSAFP.NVALTHSSAT.POISTHPHF SAAAAYA SAANTYFF.BERSLESSL.MITSSASSAFP.NVALTASSAT.POISTHPHF SAAAAYA SAANTYFF.BERSLESSL.MITSSASSAFP.SSSAFP.NVALTHSSAT.POISTHPHF SAAAAYA SAANTYFF.BERSLESSL.MITSSASSAFP.NYAAASSAT.POISTHPHF SAAAAYA SAANTYFF.BERSLESSL.MITSSASSAFP.NYAAASSAT.POISTHPHF SAAASSA SVTLFFRS.MITSCSSSL.AAAAHTSR.SVAA SAANTYFF.BERSLESSL.MITSSASSAFP.NYAAAASSATASSAFP.SSGAFF.NYAAASSAFP.SSGAFF.Y. SVTLFFRS.MITSCSSSL.AAAAAAAAFYAAAAFYAAAAFAAAAAAAAAAAAAAAAA

BTB
Bab2 sequence conservation among muscomorphans (Part2)

Dmel2 Dpse2 Dvir2 Blac2 Ccap2 Tmin2 Scal2 Lcup2 Freg2 Chom2 Pmac2 Gbre2 Gbre2 Gbre2 Edim2 Mequ2 Edim2	488 540 537 621 639 673 550 601 584 523 531 530 511 530 511 414 393 345	AAHHAAAAYPQQAPPP .AMH.HHAAAAAAQQLAAQ.HQLAHSH. AMASALAA AAAGGAGAGGAGGGGGGGGGGGGGGGGGGGGGG	
Dme12 Dpse2 Dvir2 Blac2 Ccap2 Tmin2 Tda12 Mdom2 Scal2 Lcup2 Preg2 Chom2 Preg2 Gbre2 Gbre2 Gbre2 Gbre2 Mequ2 Edim2 Mabd2	578 637 631 714 730 751 637 683 665 606 612 614 613 599 599 599 599 440 419 426	SEAR MTECC GEG CM GAA AAATG AASVAAD Y QY CLOCM W QKCMTNQOL DEFERGPLESM DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL BERAK MTESC GEG W GAAAAAT GASVAAD Y QY CLOCM W KKWTNQON LYO DEFERGPLESM DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL BERAK MTETS GEG GG W GAAAAT GASVAAD Y QY CLOCM W KKWTNQON LYO DEFERGPLESM DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL BERAK MTETS GEG GG W GAAAAT GASVAAD Y QY CLOCM W KKWTNQON LYO DEFERGPLESM DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL BERAK MTETS GE GG GG W GAGAST SSVTAD Y QY CLOCM W KKWTNQON LYO DEFERGPLESM DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GE GG W GW GAAST SSVTAD Y QY CLOCM W KKWNTNQON LYO DEFERGPLESM DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GE GH DW M GAAST SSVTAD Y QY CLOCM W KKWNTNQON LYO DEFERGPLESM DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GEH DW M GAAST SSVTAD Y QY CLOCM W KKWNTNQON LYO DEFERGPLESM DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GEH DW M GG ASS SSVTAD Y QY CLOCM W KKWNTNQON LYO DEFERGPLESM DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GEH DW M GG ASS SSVTAD Y QY CLOCM W KKWNTNQON LYO DEFERGPLESM DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GEH DW M GG ASS SSVTAD Y QY CLOCM W KKWNTNQON LYO DEFERGPLESW DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GH DW M GAAST SSVTAD Y QY CLOCM W KKWNTNQON LYO DEFERGPLESW DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GH DW M GAAST SSVTAD Y QY CLOCM W KKWNTNQON LYO DEFERGPLESW DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GH DW M GAAST SSVTAD SY CLOCM W KKWNTNQON LYO DEFERGPLESW DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GH DW M GAAST SSVTAD SY CLOCM W KKWNTNQON LYO DEFERGPLESW DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GH DW M GAST SSVTAD SY CLOCM W KKWNTNQON LYO DEFERGPLESW DE MAEATSVLK GLSS QAARK D I PYPT V LYAN RVENNL SERAK MTES GH DW M GAST SSVTAD SY CLOCM W KKWNTNQON LYO DEFERGPLESW DE MAEATSVLK GLSS QAAR	BabCD
Dmel2 Dpse2 Dvir2 Blac2 Ccap2 Tmin2 Tdal2 Mdom2 Scal2 Lcup2 Preg2 Chom2 Pmac2 Gbre2 Gbre2 Gbre2 Gbre2 Mequ2 Edim2 Mabd2	689 746 738 818 835 855 742 782 742 742 742 742 710 710 710 710 716 717 703 703 543 523 538	G DE LO GA DE BEXAR GEP OR ILLGNW PERLY EVYLAVVER DYRE TE DE SALVA NHO GHGTY GGT, TENGYHSAAARIAAONAAL. GYSLD GG DE BEXAR GEP OR ILLGNW PELYE SVI KAVVERDYRE TE DE SALVAN AN O'N HOOGHGTY GGTY GGTY GGTY GGT GPSLD GG TD REPKAR GEP OR ILLGNW PELYE SVI KAVVERDYRE TE DE SALVAN AN O'N HOOGHGAH GGAH GGAN G'HSAAARIAAONAAL. GPSLD GG TD REPKAR GEP OR ILLGNW PELYE SVI KAVVERDYRE TE DIAHV.YAR O AHGAH GGAH GGAN G'HSAAARNAAAONAAL GPSLD GG TD REPKAR GEP OR ILLGNW PELYE SVI KAVVERDYRE TE BUJAHV.YAR O MQAGSY GSAGKNAVAAA. AAN O'N HSAARNAAAONAAL GPSLD GG TD REPKAR GEP OR ILLGNW PELYE SVI KAVVERDYRE TE BUJAHV.YAR O MQAGSY GSAGKNAVAAA. AAN O'N HSAARNAAAAONAAL GSSLD GG TD REPKAR GEP OR ILLGNW PELYE SVI KAVVERDYRE TE BUJAHV.YAR O MQAGSY GSAGKNAVAAA. AAN O'N HNAARNAAAAONAAL GSSLD GG TD REPKAR GEP OR ILLGNW PELYE SVI KAVVERDYRE TE BUGH.H.S.YAR O GGHSGY GSAGKNAVAAA. AAN O'N HNAARNAAAAAANAAL GSSLD GG TD REPKAR GEP OR ILLGNW PELYE SVI KAVVERDYRE TE BUGH.H.S.YAR O GHSSF GGGAAGGES SASA. AAN O'N HNAARNAAAAAAAA GSSLD GG TD REPKAR GEP OR ILLGNW PELYE SVI KAVVERDYRE TE BUGGIF FANG O'AHA	
Dmel2 Dpse2 Dvir2 Blac2 Ccap2 Tnin2 Tdal2 Scal2 Lcup2 Chom2 Preg2 Chom2 Pmac2 Gbre2 Gbre2 Gbre2 Magu2 Edim2 Mabd2	778 835 822 9129 959 830 885 864 801 807 808 808 793 793 793 793 637 620 633	A P DA S P 16 S M T T L RO I LSO O CURO HUCO AHHOO O P S HUO O S HAO O P S HUO O S HAO Y K S P X LO NO E I E O Y S A A Y A A A A K HOO O	
Dme12 Dpse2 Dvir2 Blac2 Ccap2 Tmin2 Tda12 Sca12 Lcup2 Preg2 Chom2 Pmac2 Gbre2 Gbre2 Gbre2 Mequ2 Edim2 Mabd2	876 923 909 981 998 1052 900 967 941 868 874 868 874 875 862 862 862 862 865 687 696	E D S A 1. C LM CLP CL NVMP R . C R C Y C R P D S A 1. C LM CLP CL NV 1P S . C A G A A P S . A A S A R L S R E R E R D R E R E R A C Y C R P D S A 1. C LM CLP CL NV 1P S . C A G A G S A G C G G G G G G G G G G G G G G G G G	
Dmel2 Dpse2 Dvir2 Blac2 Ccap2 Tmin2 Tdal2 Scal2 Lcup2 Chom2 Preg2 Ghre2 Ghre2 Ghre2 Ghre2 Mequ2 Mabd2	939 989 993 1044 1081 1106 983 1060 943 946 953 947 926 926 926 763 744 740	0 B 1 C S 2 C C S 2 C C S 2 C C S 2 C C S 2 C C S 2 C C C S 2 C C C S 2 C C C C	
Dme12 Dpse2 Dvir2 Blac2 Ccap2 Train2 Train2 Train2 Sca12 Lcup2 Chom2 Preg2 Ghre2 Ghre2 Ghre2 Edim2 Edim2 Mabd2	1005 1072 1066 1115 1157 1168 1051 1108 1005 1012 1005 1012 1005 995 838 826 779	. AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	



Sequence conservation between Bab1/2 paralogs (Part1)

BTB

Sequence conservation between Bab1/2 paralogs (Part2)

I

Dmell Dprel Dvirl Ccapi Tmin1 Tdall Moon1 Scall Preg1 Chom1 Preg1 Chom2 Dvir2 Blac2 Dvir2 Blac2 Ccap2 Tdal2 Ccap2 Tdal2 Dvir2 Blac2 Ccap2 Tdal2 Ccap2 Tdal2 Ccap2 Tdal2 Ccap2 Tdal2 Ccap2	155 163 229 228 224 195 212 212 212 193 164 177 179 177 179 250 2250 2250 2250 2252 223 248 239 2174 172	<pre>LIAE TP CONP IV MRDVN C DIKAT VEFWYRGII WS DDOL DIL RINEW KVRGIA DV RIMEAA. TAARA. AAASS</pre>
Dmell Dpwsl Dvirl Riati Ccapi Triall Triall Naoni Lcupi Lcupi Dres Ediml Mabdi Dvir2 Riac2 Ccap2 Blac2 Ccap2 Blac2 Ccap2 Blac2 Ccap2 Lcup2 Dvir2 Blac2 Ccap3 Ccap3	2466 2568 3109 3144 2911 2934 2111 2934 2112 215 2688 275 2688 275 2688 275 2688 275 2688 275 2688 275 3437 344 344 3363 352 337 2357 2357 2357 2357 2357 2357	<pre>NUMBER ADDITION ADDITION ADDITIONAL ADD</pre>
Dmell Dpwsl Dvirl Ccapl Tminl Tdall Scall Lcupl Pregl Choml Pregl Choml Pmacl Gbrel Gmorl Ediml Dmel2 Dvir2 Blac2 Ccap2 Tdal2 Ccap2 Tdal2 Ccap2 Tdal2 Ccap2 Tdal2 Ccap2	300 326 314 412 413 380 357 358 363 357 358 363 357 363 357 363 363 357 363 363 357 364 450 478 408 419 391 391 392 392 392 392 392 376 290 290 263	FNPFAP.PSFLSSLIAAREMMELE OKERE ROADC PNPFA SPLSSLIAAREMELE OKERE ROADC PNPFA SPLSSLIAAREMELE OKERE ROADC PONFA SPLSSLIAAREMENDER RERENSE ROADC PONFF SPLSSLIAAREMENDER RERENSE ROADC PNPF SPLSSLIAAREMENDER RERENSE ROADC PNNFF SPLSSLIAAREMENDER RERENSE RENSE PNNFF SPLSSLIAAREMENDER RERENSE RESENDER PNNFF SPLSSLIAAREMENDER RESENDERARE KARKIT FNNFF SPLSSLIAAREMENDER RESENDERARE KARKIT FNNFF SPLSSLIAAREMENDERARE RESENDERARE KARKIT FNNFF SPLSSLIAAREMENDERARE RESENDERARE KARKIT FNNFF SPLSSLIAAREMENDERARE RESENDERARE KARKIT FNNFF SPLSSLIAA
Dmell Dprel Dvirl Ccapi Taini Madoml Lcupi Lcupi Pregl Chomi Pmaci Gbrel Ediml Mabdi Dmel2 Djar2 Blac2 Ccap2 Blac2 Ccap2 Blac2 Ccap2 Cap2 Cap2 Blac2 Ccap2 Cap2 Cap2 Cap2 Blac2 Ccap2 Cap2 Cap2 Blac2 Ccap2 Cap2 Cap2 Blac2 Ccap2 Cap2 Cap2 Cap2 Blac2 Ccap2 Cap2 Cap3 Cap3 Cap3 Cap3 Cap3 Cap3 Cap3 Cap3	340 364 464 460 473 425 407 393 394 402 407 393 394 402 407 393 394 402 407 393 394 402 407 402 407 402 407 407 402 407 402 407 402 403 404 407 402 403 404 405 405 405 405 405 405 405 405 405	PD MSSGS TVVA TRELETATH ALDMESPAATEGPLS RSS_RFES_10SPC00

Sequence conservation between Bab1/2 paralogs (Part3)

Dmell Dpsel Dvirl Stall Ccapl Tminl Scall Scall Lcupl Pregl Choml Pregl Choml Dmse2 Dvir2 Blac2 Dvir2 Blac2 Ccap2 Tmin2 Scall Dmse2 Dvir2 Blac2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2 Ccap2 Tmin2 Scal2	$\begin{array}{c} 401\\ 436\\ 450\\ 537\\ 533\\ 494\\ 458\\ 479\\ 472\\ 458\\ 461\\ 554\\ 461\\ 554\\ 461\\ 554\\ 604\\ 604\\ 604\\ 604\\ 554\\ 458\\ 504\\ 494\\ 458\\ 503\\ 557\\ 494\\ 488\\ 836\\ 734\\ 735\\ 347\\ 353\\ \end{array}$	LHP	<pre>H HAS PAPHPS QTAGS</pre>	AHHF ASPAG SQH6VSSAG ASPAG SQH6VSSAG ASPAG PTHPSSLIS.GO, LTATTVGSGVS PTHSSLIS.GO, LTATTVGSGVS PTHSSLIS.GO, LTATTVGSGVS ASPAG THRSSLIS.GO, LTATTVGSGVS ASPAG THRSSLSGOLTATTVGGVGVS ASPAG THRSSLSGOLTATVGGVGVS ASPAG THRSSLSGOLTATVGVGVS GGV THRSSLSGOLTATVGVGVSVS GGV THRSSLSSGOLGGGGGAGAGGS GGV THRSSLSSR FLTPSRSVSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
Dmell Dpsel Dvirl Rlat1 Ccapl Tidll Scall Ndoml Scall Pregl Choml Pregl Choml Pregl Choml Dmel2 Dpse2 Dvir2 Blac2 Dvir2 Blac2 Ccap2 Tidl2 Dise Dise Ccap2 Tidl2 Dise Dise Ccap2 Tidl2 Ccap2 Tidl2 Ccap2 Tidl2 Ccap2 Tidl2 Ccap2 Tidl2 Ccap2 Tidl2 Ccap2 Tidl2 Ccap2 Tidl2 Ccap2 Tidl2 Ccap2 Tidl2 Ccap2 Ccap2 Tidl2 Ccap2 Ccap2 Ccap2 Ccap2 Tidl2 Ccap2 Ccap2 Cidl2 Ccap2 Ccap2 Ccap2 Ccap2 Cidl2 Ccap2 Cidl2 Ccap2 Cidl2 Cidl2 Cidl2 Cidl2 Cidl2 Cidl2 Cidl2 Cidl2 Ccap2 Cidl2 Cidl2 Cidl2 Cidl2 Cidl2 Cidl2 Cidl2 Ccap2 Cidl	429 468 4506 589 589 580 582 514 495 522 514 498 563 563 564 498 563 564 498 563 564 565 571 615 537 543 543 545 553 543 543 564 498 563 564 498 563 564 498 564 564 498 563 564 498 564 564 498 564 564 498 564 564 498 564 564 498 564 564 498 566 564 498 566 566 572 572 574 498 566 566 572 574 574 575 574 498 566 575 575 575 575 575 575 575 575 575	DERFFLGP. AAAM. AAAMELS. VHTAAAESRFPLGP. AAAM. AAAMELS. VHTAAESRFPLGP. TAAM. AAAMELS. GRSEDSOGGANDEFALGS.OAAM. AAA VAAOMDLG. GRSEFSCAADERFALGS.OAAM. AAA VAAOMDLG. SERFEASADERFALGS.OAAAMAAAAAAAAAHNLSGHHTHGHP. RGYESHADESFPLGPAOAAMIGAA. AASOMELS. GEGGQMAGD.SHRFJNGPAOAAM. AAA.AAAAAHNLG. EGGQGNGD.SHRFJNGPAOAAM. AAA.AAAAMIDLG. EGGQGNGDSSHRFJNGPAOAAM. AAA.AAAAMILG. EGGQGNGDSSHRFJNGPAOAAM. AAA.AAAAMILG. EGGQGNGDSSHRFJNGPAOAAM. AAA.AAAAMILG. SESSHHFSNGSVOAAAMAAAA.AAAHIDLS. SESSHHFSNGSVOAAAMAAAA.AAAHIDLS. SESSHHFSNGSVOAAAMAAAA.AAAHIDLS. AAAAQOLAAO.HOL. AHSHAAM. ASALAAASI AAAAQOLAAO.HOL. AHSHAAM. ASALAAASI AAAAQOLAAO.HOL. AHSHAAM. ASALGAS AAAAQOLAAO.HOL. AHSHAAM. ASALGAS AAAAQOLAAO.HOL.AHSHAAM. ASALGAS AAAAQOLAAO.HOLAAHSHAM. ASALGAS AAAAQOLAAO.QOLHAAAHSHAM. ASALGAS AAAAQOLAAO.QOLHAAAHSHAM. ASALGAS AAAAQOLAAO.QOLHAAAHSHAM. ASALGAS AAAAQOLAAO.QOLHAAAHSHAM. ASALGAS AAAAQOLAAO.QOLHAAAHSHAM. ASALGAS AAAAQOLAAO.QOLHAAAHSHAM. ASALGAS AAAAQOFAAO.QOLHAAAHSHAM. ASALGAS	CLG. 	Р G P S A. E
Dmel1 Dpsel Dvir1 Blat1 Ccapl Tidal1 Scal1 Lcupl Preg1 Chomi Breg1 Chomi Breg1 Chomi Breg2 Dvir2 Cmac Dvir2 Capl Drac Comi Breg2 Dvir2 Capl Drac Chomi Breg2 Dvir2 Capl Dvir2 Capl Drac Chomi Breg2 Dvir2 Capl Dvir2 Capl Drac Capl Drac Capl Drac Chomi Breg2 Dvir2 Capl Dvir2 Capl Drac Capl Drac Capl Drac Capl Drac Capl Drac Capl Drac Capl Drac Capl Drac Capl Drac Capl Drac Capl Chomi Drac Capl Drac Capl Drac Capl Drac Capl Chomi Drac Capl Drac Capl Capl Capl Capl Capl Capl Capl Capl	$\begin{array}{c} 457\\ 495\\ 495\\ 640\\ 587\\ 674\\ 587\\ 561\\ 550\\ 551\\ 550\\ 552\\ 256\\ 605\\ 605\\ 605\\ 605\\ 605\\ 605\\ 605\\ 6$	PRLP DPPEHHHGGGGVGVGGGGVGGGGAGVVGSGGGSLADDUGLSK PRLP DPPHHHGSQGVGVGGGGGGVGGGGVGSQVVGF,SGGSLADDUGSK PRLP DPPHHHGSQF TGGSGVGGGG,GGVGGGGVGGGGGVGSQVVGF,SGGSLADDUGSK PTHAN,DSQVGGGG,GGVGGGGGVGGGGGVGGGGGGGVGGGGGGGGGG		M G A T I G . S T L A A D S Y Q Y Q D D S M W Q K C W T I M G L M H Z M G A T A A D S Y Q Y Q D D S M W Q K C W T I M G L M H Z M M M M G A M T A A D S Y Q Y Q D D S M W Q K C W T I M G M H Z M M M M M M M M M M M M M M M M M
Dmell Dpsel Dvirl Blat1 Ccapl Tidall Scall Lcup1 Preg1 Chom1 Preg1 Chom1 Preg1 Chom1 Dmac2 Com2 Tid12 Disc2 Com2 Tid12 Disc2 Ccap2 Tid12 Disc2 Ccap2 Tid12 Disc2 Ccap2 Tid12 Chom1 Dmel2 Disc2 Ccap2 Tid12 Chom1 Disc2 Disc2 Ccap2 Tid12 Chom2 Disc2 Ccap2 Tid12 Chom1 Disc2 Disc2 Ccap2 Tid12 Chom1 Disc2 Disc2 Ccap2 Disc2 Ccap2 Cin2 Cin2 Cin2 Cin2 Disc2 Cin2 Cin2 Disc2 Cin2 Cin2 Cin2 Cin2 Cin2 Cin2 Cin2 Cin	559 595 7199 7133 6666 641 633 667 650 641 634 635 635 635 635 635 635 635 635 635 635			

BabCD

Sequence conservation between Bab1/2 paralogs (Part4)

Dmell Dpsel Dyirl Blatl Thinl Scall Midoml Scall Pregl Loupi Pregl Choml Pregl Choml Pregl Blac2 Scal2 Dpse2 Scal2 Preg2 Scal2 Preg2 Scal2 Preg2 Ccap2 Clom2 Pregl Clom2 Comp Comp Comp Comp Comp Comp Comp Comp	673 709 695 826 826 826 826 827 738 746 751 746 751 747 746 751 747 746 751 748 805 739 748 805 739 748 803 813 813 813 813 813 813 813 814 815 815 815 815 815 815 815 815 815 815	HSPAFPLO.DLPLS., YPGASGALAGAPSSMACPN HSDMTLS., YPGASGGLTGPGAASLSALACPN HSVPFPO.DMTLS., YPGASGGTGPGAASLSALACPN HSPWFPFO.DCFLN.YPGASG HSPMFPFO.DCFLN.YPGAGG HSPMFPFO.DCFLN.YPGAGG HSSMFPFO.DCFLN.YPGMGG HSSMFPFO.DCFLN.YPGMGG HSSMFPFO.DCFLN.YPGMGG HSSMFPFO.DCFLN.YPGMGG HSSMFPFO.DCFLN.YPGMGG HSSMFPFO.DSPLS.YPGVGG HSSLPFFO.DSPLS.YPGVGG HSSVFFCO.EGLS.YPGVGG HSVYFFO.EGLS.YPGVGG HSVYFFO.EGLS.YPGVGG HSVYFFO.EGLS.YPGVGG HSVYFFO.DNILNYYGGPSS HSVYFFO.DNILNYYFG.GG HSVYFFO.DNILNYYFG.APP HSVFFFO.DNILNYYFF HSNAAAA MQAGSPFC.AAC HSAAA HSAAFG HSAAFG HSAAAA HSAAAAA HSAAFG HSAAAAA HSAAAAA HSAAAAA HSAAAAA HSAAAAAA HSAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	S G P C T	.000. 000000H0HHSHH 2.0AHH000PS. 2.0
Denell Dpsell Dyir1 Blat1 Trainl Trainl Trainl Trainl Trainl Trainl Trainl Drain Pregl Drain Blac2 Cap2 Cap2 Cap2 Cap2 Cap2 Cap2 Cap2 Cap	745 737 786 898 898 893 893 893 893 893 893 893 893	GPP. VVP PG. FNL PP H GPP. VVP PG. FNL PP H GPP. VVP PG. LPNL PP H GPP. GPF. GPG. LPNL PP H GPG. GPG. GPG. LPNL PP H HGD. GPG. GGAGTGGGGYVP.GPG. LPNL PP H HGD. GPG. GGAGTGGGGYVP.GPG. LPNL PP H HGD. GPG. GGAGTGGGGYVP.GPG. LPNL PP H HGD. GPG. GFG. GGAG. LPNL PP H GPG. GGAGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	. GVGCGGVGNVPGAAG	AA 0.000000000000000000000000000000
Dmell Dpsel Dvirl Blatl Ccapl Trainl Mdoml Lcupl Choml Dreiz Crapz Dvirz Crapz Dvirz Crapz Dvirz Crapz Dvirz Crapz Dvirz Crapz Dvirz Crapz Dvirz Crapz Dvirz Crapz Dvirz Crapz Dvirz Crap	783 824 954 948 913 951 889 889 889 889 889 886 880 850 850 850 914 901 973 990 8867 8862 8867 8867 8867 8854 8667 8654 8654 8654	G P R H A P S P C G P . A C L . P N P P S	MAVALEH	H H H M Q Q H H H L Q Q H H L Q Q H H Q Q Q H Q Q P Q H Q Q Q Q Q Q Q Q H Q Q Q Q Q Q Q Q H Q Q Q Q Q Q H Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q
Dmell Dysel Dysel Dysel Statl Ccapl Tmin1 Tdall Kdoml Chomi Chomi Chomi Chomi Chomi Chomi Chomi Dyse2 Ccap2 Train2 Scal1 Dyse2 Ccap2 Train2 Scal2 Ccap2 Train2 Scal2 Ccap2 Ccap2 Train2 Scal2 Ccap3 Ccap2 Cc	824 883 8871 1007 10061 10061 948 975 948 976 922 907 922 907 8900 524 925 970 8900 524 974 1062 1087 974 934 934 934 934 934 934 934 934 934 93	LH LQ QCQ AH L HH HCQ QCQ CQ QC HH HG . LH LQ QCQ A QC HH LH QC QV AA	. GRQVAHKSGFGAB	00

Sequence conservation between Bab1/2 paralogs (Part5)

	0.77						
Dnse1	938	· · · · · · · · · · · · · · · · · · ·	ISHSHPHSHSHSHG EKOO	KKG.	SPL	RSEIPRL.HSPLGDLG	LDMA
Dvir1	941		DKSK	S KG	SPL	RSETFRL.HSPLAELG	LELGY
Blat1	1083	Η	HQSASVSATATKAKCPS	P	SLLE	RRSSPSGNSHL.HSTLTELG	LDMGY
Ccap1	1082	HOOOHOHOOOHOKOOHOOOH	HQSVNLGGTASKMKCPS	P ,	SSLLE	QRPSPSVDSHL.HSTLTELG	LDMGE
Tdal1	1026	HOH1	TSAVNTSSSSIIAKPKS	VKD	SPT	ORSSPRHETPL.HSPFTELG	LEIGY
Mdom1	1009		.HHTTAITTASSIKSSS	STV	S FL	RRTSPPSVSPLTELG	LEMAF
Loupl	1076	·····LC	HPSSIATITASSIKSSS HPSLOKSTTATST KPHS	STS	SPD	RSSPSHSVSPLTELG	LEMAE
Preg1	972		HQSMQKTTTASSIKPPS	STS	SPS	RRTSPHSLTPL.HSPLTELG	LEMSF
Choml	963	MI	HQTMQKSASASSTKPLS	STS	S PS	RRNSPHSLTPL.HSPLTELG	LEMSF
Ghrel	968		OHSTATTISADOT KPSS	131 997	SPSTP P	CRUSTE CNTTL ESSLELG	DEMOR
Gmor1	953		QHSISTIISAPQTKFSS	SSI	SPSLER	QRQSTPGNTAL.RSSLTDLA	LDFGYS
Edim1	936	********************		T	SPSESAHDL	RMNSPEERLL.NSPL.QMS	LEPSVNLAVGVSGMAY
Mabd1 Dme12	538				SPHDM	RMQSPEDTPI.SSPI.EMT	LEPSVNLAVGVSGLP.
Dpse2	1003			A	SPYSSHYAKOQOQOQOQOQOHS	KQQEQHPNYAY.NKRFLE.S	LFAGIDFEAIANGLLQ
Dvir2	1007			A	<mark>S</mark> FYSAHHSHGHYA	KHAKEQPGYAY.NKRFLE.S	LPAGIDFEAIANGLLQ
Blac2	1095		QKMPATATAAAUSA	A	AAYAHY	KNKEHA INYAF.NKRFAE.I	LPPGIDFEAIANGLLQ
Tmin2	1120		DSGA		PLYKQS	KQDAHQQYAAY.NKRFLE.S	LPPGIDFEAFANGLLQ
Tda12	993		QKQSLGPNNNTISGG	G	<mark>S</mark> NFQH	RNKEHPSYAY.NKRLLE.S	L PPGIDFESIANNLLQ
Mdom2	1075		DMQK SAGS	I	SPYTHQHM	KNNKDQMAVNYAY.NKRFLE.N	LPPGIDFEAIANGLLQ
Lcup2	953		QK	A	SPYSQ	KMKEHP	LPPGIDFEAIANGLLQ
Preg2	956		QK	A	<mark>S</mark> PYSQM	KMKEHP.,, NYAY.NKRFLD.S	LPPGIDFEAIANGLLQ
Chom2	963	******************	QKSIGS.	A	SPFSQ	KMKEHPNYAY.NKRFLD.S	LPPGIDFEAIANGLLQ
Gbre2	936		DGOKOVPPSCFSIGP	s	SPYFYF	KNKDHA IOYAY.NKKFLE.N	LPPGIDFEAIANGLFO
Gmor2	936		DGQKQAPPSCFSIGP	S	<mark>5</mark> РҮРҮF	KNKDHAIQYAY.NKKFLE.N	LPPGIDFEAIANGLFQ
Mequ2	777	******	PYGQQRGGNHQKSSS	T	SPYSMNY	KTKPETSAYKFPDKRMLE.G	L TPSIDFEAITNGLLQ
Mabd2	752			S	YP	AIRCOS	YPAGINFEALANGMLH
Dmel1	906	SYK	EFSPSRLFAEDLAE	LVGASVS	· · · · · · · · · · · · · · · · · · ·	SSAAAATAPPERS	
Dvir1	977	K	EFSPIRLFAEDLAE	LVGVAPA	· · · · · · · · · · · · · · · · · · ·	SATTPST	
Blat1	1135	K TT	S, SAYSPIRLFSEDLAA	LVGASEDSP	PGTT.ASSTN.VSS <mark>T</mark>	STVTTPAM	
Ccapl	1150	KTS	SSAYSPIRLFSEDLAA	LVGASEDSP	PDTA.ATSTI.VSRT	STVTTSAM	
Tdall	1078	KHV	GYSPSRLFTEDIAE	LMGAATA	PQIP.AVHIS.QPQVSAIS	SITSSSSASPATS	SI
Mdom1	1051	K P S	AFSPSRLFSDDIAD	IVGAAAAAA	AVAA.ASSTS.SSTCITTATVT <mark>T</mark>	ATITGSSSYSLPATS	SSSMMAMDAAGQQHQS
Scal1	1122	K P A	PFSPSRLFADDIAD	IVGAAAAAA	AAATNVSSTS.SSACTTTATMT <mark>T</mark>	ATITASSSSSTYGLPTAT	GAG AMDSSSSSALG
Pregl	1028	KPS	PFSPSRLFSDDISD	IVGVAASPL IVGVAASPL	RSPIISIHPUSSLAA.VII RSPITS.HPSSI.AA.VII	ATITTTSSINIA, TVRSSSSEPSGS	
Chom1	1012	KPS	PFSPSRLFSDDIS <mark>D</mark>	IVGVAASPL	RSPSTS.HFSTA.TA.VT <mark>T</mark>	ATITTTSSINFA.TVRSSSSEPSVS	AI
Pmac1	1018	KPS	SFSPSRMFSDDISE	IVGVAASPL	RSPSTS.YSSCSVTG.VST	ATITATSSINTTRASSLEPSVI	
Gbrei Gmorl	1002	KST.	SFSPSRLFPDDLAD	LVGTSPS		STITUTKSNTPSET	
Ediml	983	K P S	GYSSPRPEHLFQEDIA	LVGSASD		STANLENY	*************
Mabd1	580		GKPPEILFADDLSR	NCISSNI		TSA	
Dnse2	1002	KSV0	K. SPRFEDFFPGPG. QDMSE	FANFDA		AAAAA	
Dvir2	1063	KSV	KSPRFEDFFPGQDMSE	LFADAGIAA	GAGSAA	AAAAAAAA	
Blac2	1112	KSV	KSPRFEDFFPS	FGNAES		GAGA	****************
Ccap2 Tmin2	1154	KSV	K. SPRFEDFFPS	FGNADS		AAAAGVS SYPP	
Tda12	1048	KSV	KSPRFEDFFPGQDMS E	LFSNPEM		G	
Mdom2	1128	K SSAAAAAAAAAAASSTAA	K SPRFEDFFPG PDMSE	LIANAEA		ASAQFPA	
Lcup2	999	KST	K. SPRFEDFFOG ODMND	LMTGPES.	· · · · · · · · · · · · · · · · · · ·	AGA	*************
Preg2	1002	KST	KSPRFEDFFQGQDMND	LMSGPES		A.A	
Chom2	1009	KST	KSPRFEDFFQGQDMND	LMTGPES		A.T	
Gbre2	992	KSA	KSPRFEDLFSGODANE	LLANNET		ATAFPS	
Gmor2	992	KSA	KSPRFEDLFSGQDGNE	LLPINET		AT	
Mequ2	835	KSH	P SPRFEDFFPS PDVNE	LFAGNAA		SANFP	
Mabd2	776	KSGLSS	VSPRFEDFFONHOVVD	LLNSTAD	· · · · · · · · · · · · · · · · · · ·	AGIF	
						a de la competencia de mana a se	
Dmell Dmael	947		SAATGADAPSSSSS.	. GGIKVEPI	I. T TS E		
Dvirl	1013		SGATTIDICSGSSSG	GGGIKVEPI	T. T. S. S E		
Blat1	1185	sG	SHADVSITSSAASGNISS.	. SN iki epi	T. T <mark>SS</mark> E		
Ccapl	1200		SHADVSITSSATSGNISS.	- SNIKIEPI	I.ISSE		
Tdal1	1121	LSSLTSA	TGNVATAASAPTSSKISS.	SSIKIEPI	T. T. S. S E		
Mdom1	1130	SSIASSAGISPSHSSTSMAT	TTAAATVVTSASSSNISS.	. TG IKLE PI	T. T <mark>SS</mark> E		
Scall	1203	SVAISSATDAGGGMASSAAF	TISIATVVSSAASSSSSNISS.	TOTAL PI	T.TTSD		
Preg1	1086		SSAASVAIDSMAS.ASSNISS.	TGIKLEPI	T.T TS D		
Choml	1081		TSTASAAIDSTAS.TTSNISS.	. TG <mark>IKL</mark> EPI	T. T <mark>SS</mark> D		
Pmac1 Ghrel	1084		TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	NSVELEDT	T.TSSD		
Gmorl	1057	LTSVAC	TSTTTSTTSNKSSS.	NSVELEPI	T. T. S. S E		
Ediml	1025			. IN <mark>IKME</mark> PI	TEC <mark>RS</mark> E		
Mabd1	1043		CATE FOR	. GNIKLEPI	H ATRIPHED		
Dpse2	1117		VR.ESP	LMKIKLEOO	QHATELPHED		
Dvir2	1114			LMK <mark>IKLE</mark> QQ	Q.A <mark>AE</mark> LPHED		
Blac2	1145			LMKIKLEHO	A.SAEVPHED		
Tmin2	1207		VR.ESP	LMKIKLEOO	A.TAEMPHDE		
Tda12	1083		MR.DTQ	LMK IKLE Q.	AAEVPHED		
Mdom2	1180		QR.EYN	LMKIKLEQQ	P. TEVOHED		
Lcup2	1035		OR.ETN	LMKIKLEHO	PTELQHED		
Preg2	1037		QRGETN	LMK <mark>IKLE</mark> HQ	HTELQHED		
Chom2	1044	*****************	QR.ETN	LMKIKLEHQ	P. TELQHED		
Gbre2	1027		OR DSN	LMOIKLEOO	Q. ITEMONEG		
Gmor2	1027		QR.DSN	LMQIKLEQQ	Q. I <mark>TE</mark> MQNEG		
Mequ2 Edim2	870			TAKIKLEHO	H. HSADTHEE		
Mabd2	813		PTA	LTKIKMEP.	TVABNRED		

BER^{OCS1} sequence conservation among Drosophilidae (Part1)



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



BER^{OCS2} sequence conservation among Drosophilidae



BER^{OC53} sequence conservation among Drosophilidae (Part1)





BER^{OC53} sequence conservation among Drosophilidae (Part2)



BER^{OCS4} sequence conservation among Drosophilidae



BER^{OCS5} sequence conservation among Drosophilidae



36

BER^{OCS6} sequence conservation among Drosophilidae (Part1)



BER^{OCS6} sequence conservation among Drosophilidae (Part2)



BER^{OCS7} sequence conservation among Drosophilidae (Part1)



BER^{OCS7} sequence conservation among Drosophilidae (Part2)



BER^{OC58} sequence conservation among Drosophilidae (Part1)

145	1	10 20	ç 3 ç	40, 50,		6 Q
Dmel Dsim Dsec		CGCAGGCCAC.AGGA CGCAGGCCAC.AGGA CGCAGGCCAC.AGGA	ATTTCAATCACCTCTT ATTTCAATCACCTCTT ATTTCAATCACCTCTT	GGTCACCGAAGCGTTC CGTCGCCGAAGCGTTC GGTCGCCGAAGCGTTC	CGGTGC	CTGCTCCCCG CTGCTCCCCG CTGCTCCCCCA
Dyak Dere		CGCAGGCCAC . AGGA	ATTTCAATCACCTCTT ATTTCAATCACCTCTT	GTCACCGA GTCACCGA GCGTTC	CGGTGG	CTGCTCCCCG
Dana Dper		CGCAGGCCAC.AGGA	TTTCAATCACTTCTC CACAGGAATTTGC	AGTCACCAAACTGTTC AATCTTCGAGCCACAC	CAGTGA	GCAACCCCCA
Dyir	AAGGGGAAC	ACAAAACCAC.AAGGGC	CCACCAGGAAATTTGC ATGCCGAAAGACAATT ATGAATATTCATATTC	AATCTICGAGCCACAC GAGAATGAATGTGTTT A TAAATAAATGTATT	CCCTTGGCTGAT	TT AGTGTGTGTGCACAAG
Dmoj Dgri			CAATT	ACATTCA ATCTTCTCTCTC	ATATTTAACAATCAATT GTGCCTGTCTGTCTGTT	AATGT <mark>GCG</mark> GA C ATAA IGTGT <mark>GTGT</mark> A
Dmel	70 80 AGAACCC. AAGAATCGGTGG	90 100 ATCTGACCTGAGAAGO	GGTACTA		110 120 A <mark>ATACTATGTGGAG</mark>	130 CAAATCTCTCCGG
Dsim Dsec	AGAACCCGAGAATCGGGGG1 AGAACCCAAGTATCGGGGG1	ATCTGACCT <mark>GAGAAGO</mark> ATCTGACCT <mark>GAGAAGO</mark>	G <mark>G</mark>			CAAATCTCCGCGG CAAATCTCCGCGGG
Dyak Dere	AGAACCCCAGAATCGGGGGG AGAACCC.AAAGAATCGGGGGG AGAACCCCAATTCGAATCGGGGGG	ATCTGACCTGAGAATC ATCTGACCTGCGAATC	AGAAGIG		CTAAATATGTGGAG	CAAATCTCCGCGG CAAATCTCCGCGT
Dper	AACCCCC					CCGACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Dwil Dvir	TACGCCC <mark>A</mark> GA <mark>ACGGA1</mark> GAATTCCGAA <mark>TGGGA1</mark>	ACCTATGACAAGACGA ACCTACGACAAGGCGA	AAATGAA AAAATGT		T <mark>GTGGAACGCAGAG</mark> .	AG AA CTC <mark>C</mark> ATAAA.A ACG ATC G <mark>C</mark> GAGC <mark>G</mark> .A
Dmoj Dgri	GGAGGAAGCATCGCATGTAGAA GAAATCCAGAACGGAA	ACCTACGACAAGGCGA ACCTAC <mark>GACAAIA.</mark> AA	AAAATGT AAAA TGT			ACG ATC GCGAGCA CCC A AAG <mark>C</mark> AACAGCA
Dmel	···· <mark>·</mark> ················	<mark></mark>		GAAGAAAGAG	GGCCGAGGAGAGCA	
Dsec	· · · · · · · · · · · · · · · · · · ·			GAAGAAAGAGO	GGCCGAGGAGAGCA	
Dere Dana	GGGCAAGCAGCCCAAGCAGCCCCAA	CAGAGAGCGAG	AAGATATAG	<mark>GAAGAAGGAGG</mark>	CCCCCCACCACCACCACACCACACACACACACACACAC	
Dper Dpse	CACCAAAIGIGGCACACAC CACCAAAIGIGGCACACAC			AAAGAGAGAGAGA. AA <mark>AGAGAGAGAGAG</mark> Caacaa aaa ac aa caa	GGAGAGGGGAGAGGG	
Dvir Dmoj	AAAGACGTGGCCCGGCAACAAC GAAGACGTGG.CCGGCAACAA	CATTGAGT CAATGAACAGT.GGA/	ACAGAAGCAGCGGCCA	GGCAGCAGC IGGCA <mark>GAGGCA</mark> GAG G	AGCGGCAGCAGCAG	
Dgri	<mark>.</mark>	CA ACGAATGATCGCG <i>I</i>	AGCCAAGCAGCAGCAG	IAGCA <mark>GAAGCAGCAGC</mark>	CGGCGGCAAACACGTCGC	GACACGTTGCCAGCC
	170	180	190 2	00 210	220	230 240
Dmel Dsim	AACAATGGAAC	CGTGGAACGTGGAGCI CGTGGAACGTGGAGCI	CGTGAATGCTG	ICAATCATTCCG <mark>GGC</mark> A ICAATCATTCCG <mark>GGC</mark> A	ATGGCCACTCTGGC	. GGTGGGGT. GGGGA . GGTGGGGT. GGGGA
Dsec Dyak	AACAATGGTCGTGGAG	CGTGGAACGTGGAGC CGTGGAACGTGGAGC CATGGAACGTGGAGC	CGTGAATGCTG	ICAATCATTCCGGGCA ICAATCATTCCGGGCA	ATGGGCACTCTGGT	GGTGGGGGC.GAGGA GGTGGGGGT.GGGGGA
Dana Dper	AACAATAA	GCGACGTGGAGCA GAAGCGACGTGAAGCA	CGTGAATGCTG	ICAATCACTTGTGGA ICAATCATGAGAAGG	GAGGATACCCGGAG	GAGGGGGT.GG GACAAGCC.ACCCC
Dpse Dwil	AGGGGCGACGACAATTO	GAAGCGACGTGGAGCA ACTACGACGTGTAGCA	ACGTGATGAATGCCTG	FCAATCA <mark>T</mark> GA <mark>G</mark> AAGG FCAATCATCAGAAAGG	CCTG	GCCGCAAT
Dvir Dmoj Dgri	ACAACAAAGAGAAGCACGACGCCGC	GCGACGACGTGAAGCA GCGACGACGTGCAGCA CAGACGACGTGAAGCA	ACGTGAATGCTG ACGTGAATGCTG ACGTGAATGCTG	ICAATCAAGAGCGCCC ICAATCAAGAGCGCGCA ICAATCA <mark>TT</mark> GGC <mark>G</mark> AGC	CGAGICAGCGACA CGA	ACAACAAC.AGCAA ACAACAAC
		-	CS1			
Dmel	250 G TTCCACAG AG <mark>C</mark> GCT G C.	260 270 GACCGTAGCTCCGC	280 ATTCCGCG	СТС	290 GACGTCAC	300 GACACGAGGCG
Dsim Dsec	G <mark>TTCCACAG</mark> AG <mark>C</mark> GCT <mark>G</mark> G. G <mark>TTCCACAG</mark> AG <mark>C</mark> GCT <mark>G</mark> G.	. G <mark>a</mark> c <mark>cggagctccgc2</mark> . G <mark>a</mark> c <mark>cggagctccgc2</mark>	ATTCCGCG	<mark>стд</mark> <mark>стд</mark>	GACGTCAC GACGT <mark>C</mark> AC	<mark>GACACGAGGCG</mark> <mark>GAC</mark> ACGAGGCG
Dyak Dere	GTTCCAAAGAGCGGGGG GTTCCGAAGAGCGGGGGG	TACCGGAGCTCCGCA	ATTCCGCG	CTG	GACGTCAC	GACACGAGGCG
Dper	CCACCACACCCCCCT CCACCACACCACCACCACCCCC	TTCTATACATTCCAC	CCAACGG	CAG	GACGTCACG	ACTGAGGGCAGC
Dwi1 Dvir	caacaacaacaacaaacaa	AC <mark>A</mark> TCAGCATTCCAC	GTAGCCG		GACGTCACAACGGCAGG	CGAG <mark>GTGAGCAAGCG</mark> <mark>GTCA</mark> CGCGGCG
Dmoj Dgri	ATTCCACGGTAGCTGGTC ATTCCATTCGAGCAAC	AGTCGCGA	IGACAAGC		TCTGTGAC	GTCACGCGGCG
	310 320			330	340	350
Dmel Dsim	AAATGCCGCCGGCAGAAT			TCGCAGCCGAZ	GATTGAAGATCGCC	GAAAGGCGCG
Dsec Dyak	AAATGCCG <mark>CCG</mark> <mark>GCAGAAT</mark> AAATGCCG <mark>CCG</mark> <mark>GCAGAAT</mark>			<mark>TCGCAGCC</mark> GAZ	GATTGAAGATTGCC	GAAAG <mark>TCG</mark> CG GAAAG <mark>GCG</mark> CG
Dere Dana	AAGAGCCGCCGGCAGAAT	r		TCGCAGCCGAA	GATTGAAGATCGCC	GAAAGAGCGCG
Dpse Dwil	AAATGCCGCCGGCAGAAA	GCAGTCAGTAGCAGC	AGCA	.GCAGTCGAATTCGGA	AAT.AAAAATCCCAACC	GAAAGGCGCG
Dvir Dmoj	AAATGCCGTCGGCAAAAG			GCAGCCGA	ACAAAATGAAGA	
Dgri	AAATGCCGACGGCAAAAAA	AAAGTAGCCAGCAGC	AGCAGCGGGGCAAAGT	CAATG <mark>CCGAAGCCGA</mark>	ACAAAATGAAGATCG <mark>C</mark> A.	ACAG <mark>TTCGCGCGCG</mark>
	360 370 3	380 390	400	410	Next to CS2 420	430
Dmel Dsim	CGAATCGTGAATCAAAACAAA CGAATCGTGAATCAAAACAAA CGAATCGTGAATCAAAACAAA	CCGAGGCAAAAGATCC CCGAGGCAAAAGATCC	GATAATGAAT GATAATGAAT	GGAGAGAGAGATAT		GAAGCAGCGCTGCG GGAAGCAGCGCTGCG
Dyak	CGAATCGTGAATCAAAACAAAA CGAATCGTGAATCA <mark>A</mark> AACAAAA CGAATCGTGAATCA <mark>A</mark> AACAAAA	CCGAGGCAAAAGATCC CCGAGGCAAAAGATCC	GATAATGAAT GATAATGAAT	GGAGAGAGATAT	AGAC	GAAGCAGCGCTGCA
Dana Dper	CGAATCGTGAATCAAAACAAA CGAAACGTGAATCAAAAAAAAAA	CCGGGGGAAACGATCC CGATCGAGAGAGCGAC	GAGAGCCATC	GAAAGAGAGCA <mark>AT</mark> GT. GAAAGAGAGCGAG		GGGGGCAGCACAGT IGAAGCAGCGCTGTG
Dwil Dvir	CGAAACGTGAATCAAAACAAA CGAAATATGAATCAAAACAAA CGAAACGTAAATCAAAACAAA	AAACCCCGAGACGAGGG ACAGAGAGAGAGAGACA	AGAGCGCGCCAAGC	GCAAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGAGTGA	TGCTTGACGCGTGAGCG	GCAAGCAACGCTGTG GCCAGCAGCGATGC.
Dmoj Dgri	C <mark>GAAACGTA</mark> AATCA <mark>A</mark> AACAAA C <mark>c</mark> aa <mark>acgta</mark> aatca <mark>t</mark> aacaaa	ACAGAGAGCG GCAGAGAGCGC		GCGATIG	CTITTGACGCGTG AGCG	TAAGCAAACGCTGC. CGAGAGAGCGCTGCT

BER^{OCS8} sequence conservation among Drosophilidae (Part2)





42

BER^{OC59} sequence conservation among Drosophilidae (Part1)





BER^{OCS9} sequence conservation among Drosophilidae (Part2)

BER^{OC59} sequence conservation among Drosophilidae (Part3)



BER^{OCS10} sequence conservation among Drosophilidae (Part1)



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



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)



P	Abdominal AE sequence conservation among Drosophilidae (Parti)															
	1, 10	G ۴	ata TCI Ţ	GATA	4 0	5 0					6 Q			7 Q	8 0	
Dmel Dsim Dsec Dyak Dana Dper Dpse Dwil Dvir Dmoj Dgri	GGACCACCACCA GGACCACCACCA GGACCACCACCACCA GGACCACCACCACCA GGACCACCACCACCA GGACCACCACCACCACCA GGACCACCACCACCACCACCACCACCACCACCACCACCAC	C T G A C A C T T X C T G A C A C T T X C T G A C A C T T X C T G A C A C T T X C T G A C A C T T X C T G A C A C T T X C T G A C A C T T X C T G A C A C T T X C T G A C A C T T X C T G A C A C T T X	NTCGCCTT NTCGCCTT NTCGCCTT NTCGCCTT NTCGCCTT NTCGCCTT NTCGCCTT NTCGCCTT NTCGCCTT NTCGCCTT	TATCTGCT TATCTGCT TATCTGCT TATCTGCT TATCTGCT TATCAGTGCT TATCAGTGCT TATCTGCT TATCTGCT TATCTGCT TATCTGCT TATCTGCT TATCTGCT	A TGGC A TGGC A TGGC A TGGC A TGGC C.T.TGGG C.T.TTATG C.T.TATG A AAAA A TGTG A TGTG A TGTG A TGTG	GGGAGC GGGAGC GGGAGC GGGAA GGGAAA GGGAA TGCATC TGCATC TGCATG TGCATG TGCATG TGCAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TCCAAA TC	SGAGGA SGAGGA SGAGGA SGAGGA SGAGGA SGAGGA SAAAAAA AAA	G. TATAAAT	GTATATA		T G G A G TA. . G G A G TA. . G G A G TA. A G G C A TA. A G G A A TA. A G A A A TA. A G A A A TA. A G A A A A A G A A A A A A G A A A A A	A GTC A GTC A GTC A GTC A GTC A A TCAA A A TCAA A A TCAA A A TCAA A A A TCAA A A A CAA A A A CAA G A A CAA G A GCGC G A GCAA G A GCAA G A GCAA	GAACCG AAATA AAATA AAATA AAATA AAATC AAATC AAATC AAATC AAATC AAA CTTG AGAGAAG	GAG GT GAG GT GAG GT GAG GT AAAG AT AAA AA AAT AC AAA AAT AAAAAT	GCAT C C GCAT C C GCAT C C GCAT T C GCAT T C GCAT T C ACAC A C A	ACACAA ACACAA ACACAA ACACAA ACACAA ACACAA ACACAC ACACAC GGAGAT AAATAA
Dmel Dsim Dsec Dyak Dana Dper Dyak Dwil Dwil Dvir Dmoj Dgri	9 ACA ACA ACATATAAA ACA ACA ACAC.ACACAACA ACAC.ACACAACA ACAC.ACACAACA ACAC.ACACAACA ACAC.ACACAACA ACAC.ACACAACA ACAC.ACACAACA ACAC.ACACAACA ACAC.ACACAACA ACAC.ACACAACA ACAC.ACACAACA ACAC.ACACAACAACA ACAC.ACACAACAACA ACAC.ACACAACAACAACAACAACAACAACAACAACAACA	9 TATAA TAA TATAA ATAT TATAA ATAT TATAA ATAT TATAA ATAT TATAAA AATAT CACACAA AAT G CAAAAAT G CAAAAAT G CAAAAT G CAAAAT	SCATAGCG SCATAGCG	AAGAGTGTC AAGAGTGTC	GCAAGTC GCAAGTC	GTACACT(GTACACT)	100 	ATGGAAG ATGGAAG ATGGAAG ATGGAAG ATGGAAG ATGGAAA ATGGAAA ATTGGAA CATAAA	ATT. AATT. AATT. AATT. SAATT. SAATT. SAATT. SAATT. SAATT. SAATT. SAATT. SCATT. SCATT. SCATT. STATT. SAATT.	ACGTGIG ACGTGIG	GATCGTT GATCGTT	GGACAGT/ GGACAGT/	AAATTCA AAGTTCA	TGAT.C TGATCC	ССАААА ССАААА	ATCCCT ATCCCT
Dmel Dsim Dsec Dyak Dana Dper Dpse Dwir Dvir Dmoj Dgri	CAAAATGITGGA CAAAATGITGGA		120 AATATTA AATATTA AATATTA AATATTA AATATTA AATATTA ATTTCTTA AATATTA AATATTA AATATTA AATATTA AATATTA	130 AAGAAGCT(AAGAAGCT(AAGAAGCT(AAGAAGCT(AAGAAGCT(AAGAAGCT(AAGAAGCT(AAGAAAGCT(AAGAAATTA AAGAAATTA AAGAAGAG GAA.	CACCGGA	TCAGG. TGGGAAT GGGAAT AGTC TATGGGA	ITITITIG ITITITIG ITIGGGTTA ATIGATIA . TIGATIA	CTTCGCA CTTCGCA CTATGCA CTATGCA CTTCGCA CTTCGCA	ICC AATTTTT AATTTTC GCTTATT IATAGAG I	ATTAAAA ATTAAAA GTCCGA. AGGGAA.	GATTTTT TATTTTT .TCTGTT	. GT GAAT, GT GAAT, GT GAAT, GT GAAT, GT GAAT, ATT TG TT AATT GT T GC AAT, CC AAT,	140 CGC TG CGC TG CGC TG CGCGC GG CGCGCG GG GGCGGCG GG GATTATA A. CTATA CTC GA	1 GCCGAG GCCGAG GCCGAG GCCGAG GCCGAG GITAGG GITAGG GITAGG GITAGG GAAGCG ACTGOG	50 GATGCA GATGCA GATGCA GATGCA GATGCA ATTAAA AATAAA AATAAA TATAAG GATGCC GATGCT	160 GCCCAT GCCCAT GCCCGT GCCCGT TICTAG TICTAG TICTAG ATACAA GCCTGT GGCTGT
Dmel Dsim Dsec Dere Dyak Dper Dpse Dwir Dvir Dmoj Dgri	CAATATTT CAATATTT CAATATTT CAATATTT CAATATTT CAATATTT AAAATTATAGAA AAAATTATAGAA AGGTATTCACAT CAATATTT CAATATTT	TTTCAGAAA TTTCAGAAAA TTT	AATTTAGT AATTTAGT AATTTAGT	TTTTTAAT TTTTTAAT TTCACACT	AAAGTIG AAAGTIG GAGCTCG	ТАААААА Т. АААААА Т. АААААА Т. ААААА	AAATICCP ATATICIP	GAAAATT GAAAATT	ITGGAAT ITAGAAT	TTCAGAA TTCAGAA	TTTTAAT	TTTTTGA1 TTTTTGA TTTTTGA	TTAAAGT TTAAAGT ACACGA TTCTCGA	TATACA TATACA CACACA GATATG	 AAAAAT AAAAAT CATGAT CTTAAG	TCTGG. TCTGG. TCTGG. AGAAGG. AAAAGG
Dmel Dsim Dsec Dere Dyak Dana Dper Dyse Dwil Dwir Dmoj Dgri	ТАСССАААААТС	GATTITTATC	AAAATTT AAAATTT AACGACT AATATTA	TIAAAITI TIAAAITI TAGAITA CGAAAAIG	AGAAAAT AGAAAAT CCTCTAA TTTCCTAT	TGTATTT TGCATTT AACGTAT ACGATAC GATGG	ITITGAI ITITGAI IGITGAI IGITGAI ITATACAI ITATACAI ITATGAIC	AAAGTTA AAAGTTA AAAGTTA AAGC GTTTATG	IGAAAAA IGAAAAA ATGGGAA	ATTATAG ATTATAG ATTATAA ATTATAA	AAAATIT AAAATIT CAGGTTA AAGGGTA	TAGAAGT TAGAAGT GCTCGAT TATTGGT	ITTAGAA ITTAGAA ICTCAAG IGTAGAG	ATTTTA ATTTTA ATTTTA G	GAATTT GAATTT	TAGAAA TAGAAA
Dmel Dsim Dsec Dere Dyak Dana Dper Dwil Dvir Dwil Dwir Dmoj Dgri	AATITATCTTTT AATITAACTTTT	TGATTAAAGI TGATTAAAGI	TAGGAAC TAGGAAC	AAATTCIGG	AAAATAT AAAATTT AACAGTT	TAAAATT TAAAATT GGCGACC	ICAGAAA ICAGAAA ICTGGCGG ICAAGAAT ICCATGI ICGTATGI	ATTTATG ATTTATG ACACAAG ATATAAA ATATAAA	TTTTTGA ITTTTGA ITTTTGA ITATTGA ZTCGCTC ZTGTTAT	TTAAAGC TTAAAGC CTCTAAA CTTAA CTCAATC	TATGAAA TATGAAA T TTAG	AAATTCT/ AAATTCT/	AG. AAAA AGAAAA 	TTTTAG TTTTAG TTTTAG TCCGAT TTTTAT	GIGCCT ATCATC	170 GAC GAC GAT GAT AAA AAA
Dmel Dsim Dsece Dyak Dana Dper Dwil Dwil Dvir Dwir Dpgri	180 TTTACACCTCG TTTACACCTCG TTTACACCTCG TTTACACCTCG TTTACACCTCC ATTTACACCTC TTTACACCTC TTTACACCTC TTTACACCTC TCTCACATGCAC	190 AAAGTGAAT AAAGTGAAT AGAAGTGAAT AGAAGTGCAT AGAAGGTGT AAGAGATTT AAAAGATTT AAAAAAATTT	ACCTGT. TCGTAT. TCGTAT. ATATGTA ATATGTA ATATGT ATATGTT ATATGTT	CGTATGTAC TGATTAAAC TGATTAAAC	CTAAATG TTAGAAG TTAAAAG CTAACAG	CAAATAC CAAATAC T	200	TCGAAAT TCGATAT TGATGCT TGATGCT TGATGCT TGAGAA TCAAGAA GCAAAA CCAAAC CCAACAA	210 3AC STG 3AC TTG 3AC TTG 3AC TTG CACTTG CACTTG CACTTG CACTTT CACTTT CACTTT CACTTT CACTTT CACTTT CACTTT CACTTT CACTTT CACTTT CACTTT CACTTT CACTTT CACTTT CACTTT CACTTT CACTT	GA G G AAAATTT AAAATTT AAAATTT CCCTACT TACCTTC	TAGAATT TAGAATA TAGAACA GAGGCTC CACACA. TGCAAG.	TTTTTTTT TTTTTTT TGCTTT 	LITITIT ITITITI AGCCCTTI GAATIG AGATIA	GATTAA GATTAA GATTAA GGTTCA GGTTCA A CATAAA		GAAAAA GAAAAA A AAAAAAC A.CTAGT

Abdominal AF sequence conservation among Drosonhilidae (Part1)

Abdominal AE sequence conservation among Drosophilidae (Part2)



Abdominal AE sequence conservation among Drosophilidae (Part3)



Abdominal AE sequence conservation among Drosophilidae (Part4)





Abdominal DE sequence conservation among Drosophilidae (Part2)

