

Diverse &ITCis&IT-Regulatory Mechanisms Contribute to Expression Evolution of Tandem Gene Duplicates

Journal Article**Author(s):**

Baudouin-Gonzalez, Luís; Santos, Marília A.; Tempesta, Camille; Sucena, Élio; Roch, Fernando; Tanaka, Kohtaro

Publication date:

2017-12

Permanent link:

<https://doi.org/10.3929/ethz-b-000220107>

Rights / license:

[Creative Commons Attribution-NonCommercial 4.0 International](#)

Originally published in:

Molecular Biology and Evolution 34(12), <https://doi.org/10.1093/molbev/msx237>

Diverse *Cis*-Regulatory Mechanisms Contribute to Expression Evolution of Tandem Gene Duplicates

Luís Baudouin-Gonzalez,^{‡,1,2} Marília A. Santos,^{§,1} Camille Tempesta,² Élio Sucena,^{1,3} Fernando Roch,^{†,2} and Kohtaro Tanaka^{*,†,1}

¹Instituto Gulbenkian de Ciência, Oeiras, Portugal

²Centre de Biologie du Développement, Centre de Biologie Intégrative, Université de Toulouse, CNRS, UPS, France

³Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

[‡]Present address: Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, United Kingdom

[§]Present address: Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: ktanaka@igc.gulbenkian.pt.

Associate editor: Patricia Wittkopp

Abstract

Pairs of duplicated genes generally display a combination of conserved expression patterns inherited from their unduplicated ancestor and newly acquired domains. However, how the *cis*-regulatory architecture of duplicated loci evolves to produce these expression patterns is poorly understood. We have directly examined the gene-regulatory evolution of two tandem duplicates, the *Drosophila* Ly6 genes CG9336 and CG9338, which arose at the base of the drosophilids between 40 and 60 Ma. Comparing the expression patterns of the two paralogs in four *Drosophila* species with that of the unduplicated ortholog in the tephritid *Ceratitis capitata*, we show that they diverged from each other as well as from the unduplicated ortholog. Moreover, the expression divergence appears to have occurred close to the duplication event and also more recently in a lineage-specific manner. The comparison of the tissue-specific *cis*-regulatory modules (CRMs) controlling the paralog expression in the four *Drosophila* species indicates that diverse *cis*-regulatory mechanisms, including the novel tissue-specific enhancers, differential inactivation, and enhancer sharing, contributed to the expression evolution. Our analysis also reveals a surprisingly variable *cis*-regulatory architecture, in which the CRMs driving conserved expression domains change in number, location, and specificity. Altogether, this study provides a detailed historical account that uncovers a highly dynamic picture of how the paralog expression patterns and their underlying *cis*-regulatory landscape evolve. We argue that our findings will encourage studying *cis*-regulatory evolution at the whole-locus level to understand how interactions between enhancers and other regulatory levels shape the evolution of gene expression.

Key words: *cis*-regulatory evolution, enhancer, *Drosophila*, gene duplication, gene regulation.

Introduction

Among the many mechanisms driving genome evolution and phenotypic change, gene duplication is arguably one of the most influential processes, as it provides raw material from which genes with diverse functions can evolve (Ohno 1970; Hahn 2009). Duplication events not only shape genomes at large evolutionary time scales but also have been shown to produce rapid adaptive changes at the population level (Perry et al. 2007; Schrider and Hahn 2010; Bass and Field 2011). However, in most cases, one of the duplicated copies becomes pseudogenized due to functional redundancy (Lynch and Conery 2003). Except in the rare cases where an increased dosage of the ancestral gene product is advantageous, the duplicates need to diverge rapidly in protein function and/or expression before one copy is lost through pseudogenization (Ohno 1970; Lynch and Conery 2003; Hahn 2009; Bass and Field 2011).

The process of expression divergence between paralogs, in particular, has received much attention from previous studies

(Force et al. 1999; Prince and Pickett 2002; Kassahn et al. 2009; Assis and Bachtrog 2013; Pegueroles et al. 2013). To explain the maintenance of duplicated copies, two main types of expression divergence have been proposed. First, each paralog can rapidly adopt distinct tissue, spatial or temporal specificities, thus making the two copies nonredundant (Hahn 2009). This process can occur through the acquisition of new expression (neofunctionalization) or through the partitioning of the original expression domains between the two copies (subfunctionalization) (Ohno 1970; Force et al. 1999; He and Zhang 2005; Hahn 2009). A classical model of subfunctionalization is the duplication–degeneration–complementation model, which postulates that *cis*-regulatory elements duplicated along with their coding sequences undergo complementary degeneration by neutral drift, eventually producing complementary expression patterns (Force et al. 1999). As this model requires the complete duplication of regulatory regions, it may be common in duplicate pairs resulting from whole-genome duplications or large-segmental

© The Author 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

duplications (Bruce et al. 2001; Jarinova et al. 2008; Kleinjan et al. 2008; Katju 2013).

Alternatively, changes in the expression levels of paralogs can lead to their retention (Force et al. 1999; Hahn 2009; Qian et al. 2010; Gout and Lynch 2015; Lan and Pritchard 2016; Thompson et al. 2016). According to the dosage sharing model (or quantitative subfunctionalization), the expression of the two copies in a given tissue is reduced to match the level originally produced by the unduplicated ortholog, thus making the contribution of both copies indispensable (Force et al. 1999; Lan and Pritchard 2016). As a corollary, paralogs can maintain the same expression domains for prolonged periods of time, although their relative expression levels may vary in different lineages. Eventually, one copy can lose expression in some tissues culminating in a subfunctionalization event (Gout and Lynch 2015; Lan and Pritchard 2016; Thompson et al. 2016). Interestingly, recent genome-wide studies in mammals and *Drosophila yakuba* populations have provided evidence indicating that dosage sharing may be a prevalent mechanism for the maintenance of tandem gene duplicates, which are more likely to be coregulated by their shared *cis*-regulatory environment (Lan and Pritchard 2016; Rogers et al. 2017).

Although many of the theoretical studies and experimental work examining the expression divergence between gene duplicates invoke different *cis*-regulatory mechanisms, few studies have directly investigated *cis*-regulatory evolution in duplicated loci (Hittinger and Carroll 2007; Jarinova et al. 2008; Kleinjan et al. 2008; Loehlin and Carroll 2016). Thus, there is a conspicuous lack of empirical data relating changes in *cis*-regulatory landscape to paralog expression divergence.

Drosophila melanogaster and its related species provide a solid phylogenetic framework to study *cis*-regulatory evolution and have been explored to unveil the nature of molecular changes underlying differences in gene expression between species (Ludwig et al. 1998; Kalay and Wittkopp 2010; Frankel et al. 2011; Arnoult et al. 2013; Arnold et al. 2014; Barriere and Ruvinsky 2014; Wunderlich et al. 2016). Here, we used this system to investigate the expression divergence of a pair of tandem duplicates and its underlying *cis*-regulatory bases. For this purpose, we chose the CG9336 and CG9338 paralogs, which arose from a single duplication event that occurred between 60 and 40 Ma and are found in all sequenced *Drosophila* species (Tanaka et al. 2015). Although the biological function of these genes has not yet been described, these duplicates belong to the insect Ly6 gene family, whose small protein products function as membrane ligands capable of binding to a wide range of targets in different biological contexts (Galat et al. 2008; Hijazi et al. 2009; Nilton et al. 2010; Wu et al. 2010; Kim and Marqués 2012; Chaudhari et al. 2013). Previous comparison of the embryonic expression patterns of CG9336 and CG9338 in *D. melanogaster* showed that these genes retain some of the tissue specificities of the unduplicated ortholog found in the tephritid *Ceratitis capitata* (Tanaka et al. 2015). However, they have also diverged from each other and the unduplicated ortholog, both by partitioning of the ancestral expression pattern and by acquisition of novel expression domains (Tanaka et al. 2015).

In this study, we have first characterized the tissue specificities of CG9336 and CG9338 expression in three additional *Drosophila* species to understand how and when the expression divergence arose. We then compared the enhancer activities of the entire locus in all four species to elucidate the *cis*-regulatory mechanisms underlying the expression evolution of the duplicates. We found that expression divergence, encompassing both sub- and neo-functionalization occurred both close to the time of duplication and more recently in a lineage-specific manner. Comparative analysis of the locus-wide *cis*-regulatory landscapes uncovered the presence of *cis*-regulatory modules (CRMs) associated with both the subfunctionalization events and the emergence of new expression domains underlying neofunctionalization. Surprisingly, some conserved tissue expression also appears to be driven by highly variable *cis*-regulatory architecture, highlighting the dynamic *cis*-regulatory basis of paralog expression evolution.

Results

To gain new insights into the mechanisms underlying paralog evolution, we chose to analyze in detail the process of expression divergence undergone by two *Drosophila* duplicates, CG9336 and CG9338 (Tanaka et al. 2015). A survey of multiple *Drosophila* species and two tephritid genomes indicates that these tandem duplicates arose in a common ancestor of drosophilids (fig. 1A). Their divergence has therefore accompanied the radiation of this group for more than 40 My. Comparison of the genomic annotations reveals that the two paralogs and their intron–exon structures have been preserved across *Drosophila* species (fig. 1A and B). In contrast, the adjacent and intervening noncoding regions have undergone alterations in different lineages, namely the insertion of different coding sequences in neighboring positions. Moreover, CG9336 and CG9338 have experienced further rounds of duplication in some lineages, producing additional species-specific paralogs (fig. 1A).

The Protein Products of CG9336 and CG9338 Exhibit Both Conserved and Divergent Features

The coding sequences of the two paralogs have diverged during drosophilid evolution (supplementary fig. S1A, Supplementary Material online), but their respective protein products display all the typical features of Ly6 family members: an N-terminal signal peptide, a three-fingers Ly6 domain with 10 conserved cysteines in stereotypical positions, and a short hydrophobic C-terminus that is cleaved for glycosylphosphatidylinositol-anchor addition during protein maturation (fig. 1B) (Hijazi et al. 2009; Nilton et al. 2010). We have previously shown that the paralog coding sequences have been maintained under purifying selection, indicating that the two proteins could perform distinct functions (Tanaka et al. 2015). To explore this possibility, we generated versions of both proteins with a fluorescent tag (mCherry) under the control of inducible promoters and compared their subcellular localizations in different tissues (supplementary fig. S1C–E and G–I, Supplementary Material online). We observed that the two gene products localize to the cell

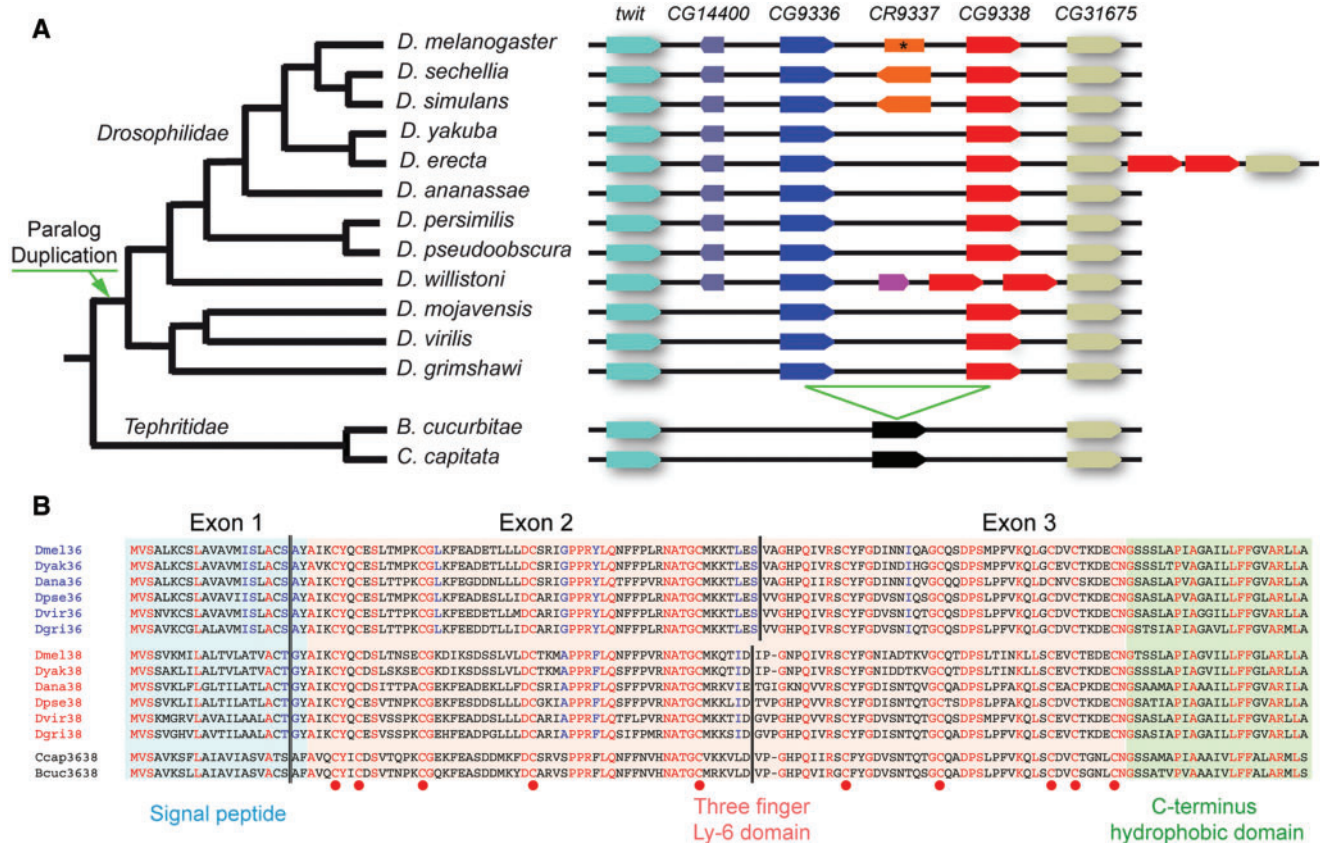


FIG. 1. Evolution of the CG9336–CG9338 genomic region after gene duplication. (A) Organization of the CG9336–CG9338 genomic region in different drosophilid and tephritid species. The phylogeny of the 14 species analyzed is on the left. Ortholog coding sequences are labeled using the same color code. The two paralogs originated from a duplication event in a common ancestor of the drosophilids. Several genomic insertions containing coding sequences for unrelated genes are found in some lineages. In addition, the region experienced additional duplications involving one or more coding units in *D. erecta* and *D. willistoni*. In *D. melanogaster*, the CR9337 coding region has been pseudogenized (asterisk). (B) Sequence alignment of protein products encoded by CG9336 (blue), CG9338 (red) and the unduplicated ortholog (black) in different drosophilid and tephritid species. Residues conserved in all three proteins are shown in red, whereas conserved amino acids specific to each paralog appear in blue. The exon structure and the different protein functional domains are also indicated. Highly conserved cysteine residues diagnostic of the Ly6 protein family are labeled with red circles.

membrane, and their overall distributions are indistinguishable (supplementary fig. S1C–E and G–I, Supplementary Material online). Thus, despite some degree of divergence at the levels of primary sequence and tissue-specific expression (see below), both proteins act in the same subcellular compartments. It is therefore possible that the two paralogs play redundant roles in tissues where they are coexpressed, although we cannot rule out that they may have acquired distinct targets.

Expression Divergence of CG9336 and CG9338 in *D. melanogaster* and *C. capitata*

Previous analyses showed that CG9336 and CG9338 are expressed in *D. melanogaster* embryos in both common and nonoverlapping domains, indicating that their tissue specificities have diverged after duplication (Tanaka et al. 2015). To gain a more complete picture of this divergence, we conducted an exhaustive analysis of their expression patterns not only during embryogenesis but also in third instar larvae. We examined their expression using both paralog-specific riboprobes and yellow fluorescent protein (YFP) reporters for each

gene, which faithfully reproduce the endogenous patterns and allow a higher cellular resolution and a more accurate assessment of expression levels (Lowe et al. 2014; Tanaka et al. 2015).

Both duplicates are expressed in glial tissues throughout development, but their relative levels vary in different glial subtypes, adopting in some cases complementary patterns (fig. 2). For instance, although CG9336 levels are higher, both paralogs are expressed in the embryonic glia associated with the peripheral nervous system (hereafter referred to as “axonal glia”; fig. 2A, B, G, and H) in the wrapping glia of the larval eye disc (hereafter referred to as “eye wrapping glia”; fig. 2M and S) and in the midline glia (fig. 2A, B, G, H, N, and T). In the surface glia of the embryonic central nervous system (CNS), however, only CG9336 is expressed (fig. 2A and B). In turn, CG9338 levels are higher in the larval CNS surface glia, and only this paralog is expressed in the eye carpet glia and the larval axonal glia (fig. 2S–V).

We observe a similar trend in other tissues. The embryonic Bolwig’s organ (fig. 2E and K), the heart (fig. 2F and L), and a ring of epidermal cells surrounding the embryonic anal plate (fig. 2D and J) express both genes, although CG9338 is seen at

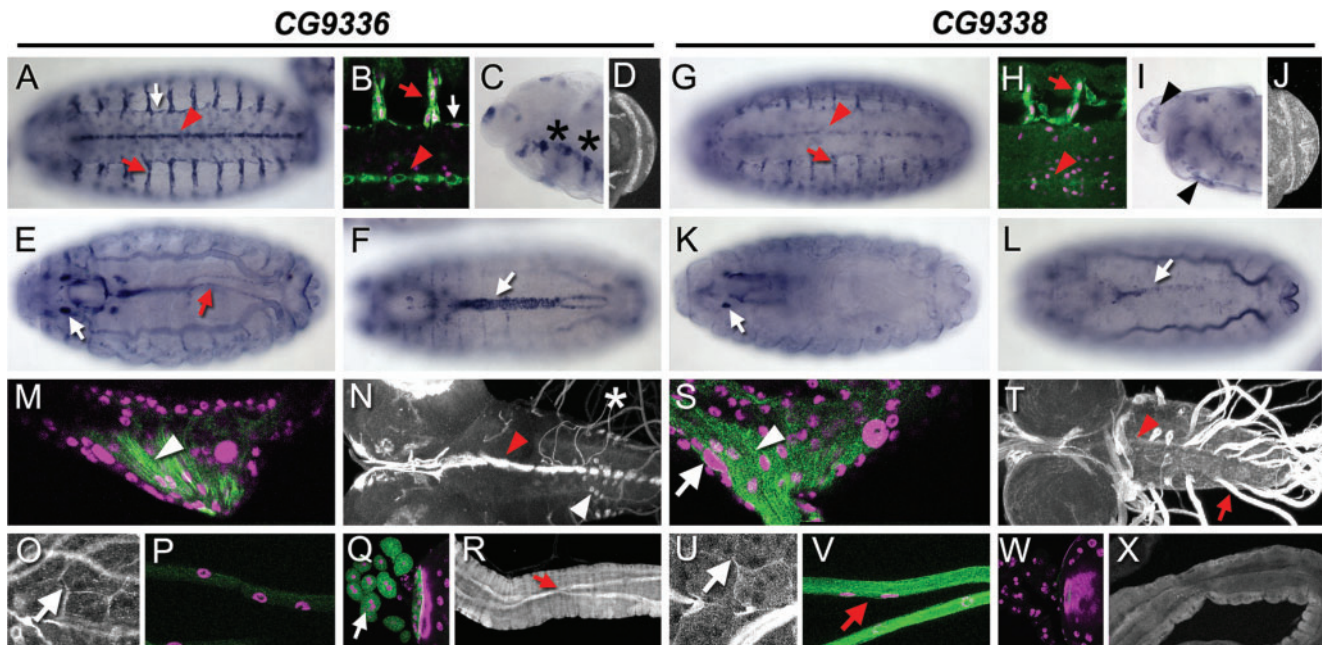


Fig. 2. Embryonic and larval expression patterns of CG9336 and CG9338 in *D. melanogaster*. (A–F) Expression of CG9336 mRNA (A, C, E, F) and the YFP-tagged protein product (B, D) in the embryo. In situ hybridization showing the CG9336 mRNA distribution in stage 16 (A, F), late stage 17 (E), and stage 15 embryos (C). The transcript is strongly expressed in the peripheral axon glia (red arrow, A), CNS surface glia (white arrow, A), and the midline glia (red arrowhead, A), the Bolwig’s organ (white arrow, E), hindgut boundary cells (red arrow, E), and the heart (white arrow, F). (B, D) Projections of confocal stacks showing the stage 16 embryos expressing CG9336-YFP and stained with a GFP antibody (green in B, b/w in D). (B) Expression is detected in glial cells (axonal glia, red arrow; CNS surface glia, white arrow) coexpressing the glial marker Repo (magenta) and in the midline glia (red arrowhead). (C) No expression is detected in the hemocytes. Asterisks, midline glia. (D) Expression is also seen in a ring of cells around the anal plate in the most posterior part of the embryo. (G–L) Expression of CG9338 mRNA (G, I, K, L) and YFP-tagged protein product (H, J) in the embryo. In situ hybridization showing the CG9338 mRNA distribution in stage 16 (G, L), late stage 17 (K), and stage 15 embryos (I). The transcript signal is detected in the peripheral glia (red arrow, G), the midline glia (red arrowhead, G), the Bolwig’s organ (white arrow, K), the heart (white arrow, L), and the hemocytes (black arrowheads, I). The signal was not detected in the hindgut boundary cells (K). (H, J) Projections of confocal stacks showing the stage 16 embryos expressing CG9338-YFP and stained with a GFP antibody (green in H, b/w in J). Expression is seen in axonal glial cells (red arrow, H) coexpressing Repo (magenta in H) and, at lower levels, in the midline glia (red arrowhead, H) and the anal plate (J). (M–R) Expression of the CG9336 YFP-tagged protein product revealed with a GFP antibody (green in M, P, Q, b/w in N, O, R) in third instar larvae. The CG9336-YFP-tagged protein distribution in the eye disc (M) reveals that the expression is in wrapping glial cells (white arrowhead). (N) In the larval CNS, high levels of expression in the midline glia (red arrowhead) and in the CAPA neurons (white arrowhead) are detected. Asterisk indicates the tracheal branches. We also detect expression in the large flat cells of the CNS surface glia (O, white arrow), but the signal is absent in the axonal glia (P, labeled in magenta with the Repo marker). The YFP-tagged protein is also detected in the nephrocytic garland cells (white arrow in Q; magenta shows TOPRO nuclear staining) and in the hindgut boundary cells (red arrow, R). (S–X) Expression of the CG9338 YFP-tagged protein product revealed with a GFP antibody (green in S, V, W, b/w in T, U, X) in third instar larvae. (S) In the larval eye disc, the CG9338-YFP protein trap is expressed in the wrapping glia (white arrowhead) and in the carpet cells (white arrow). In the larval CNS, a high level of expression is detected in the axonal glia (red arrow in T and V; Repo marker is in magenta) and the surface glia (U, white arrow). CG9338-YFP is detected neither in the nephrocytic garland cells (W, magenta shows TOPRO nuclear staining) nor in the hindgut (X).

much lower levels in the latter two tissues. Meanwhile, only CG9336 is detected in two rows of hindgut epithelial cells known as boundary cells (fig. 2E and R) (Iwaki and Lengyel 2002), in the nephrocytic garland cells (fig. 2Q), and in the CAPA-producing neurons of the CNS (fig. 2N) (Santos et al. 2007). In turn, CG9338 is the only paralog expressed in the embryonic hemocytes (fig. 2I).

Comparing the expression patterns of CG9336 and CG9338 with that of their unduplicated ortholog (*a-36/38*) in *C. capitata* (fig. 3A–H), we could assess for each expression domain whether it was inherited by both *Drosophila* paralogs (conservation), by only one of them (subfunctionalization) or whether it constituted a novel acquisition (neofunctionalization) (fig. 3I). We infer that both paralogs inherited expression in the Bolwig’s organ, the embryonic axonal glia, and the eye

disc glia (fig. 3B, D, G, and I). In contrast, the expression in the hindgut and the embryonic CNS surface glia appears to have been retained only by CG9336 (fig. 3B, E, and I). Finally, the expression in the heart, the hemocytes, and the midline glia are likely novel domains, although we cannot exclude the possibility that they were secondarily lost in *C. capitata* (fig. 3A–C and I). Thus, our study case contains all the major evolutionary outcomes predicted for paralog expression divergence.

Interspecific Comparisons Reveal Paths to Sub- and Neo-Functionalization

The observed expression divergence of *D. melanogaster* CG9336 and CG9338 may have evolved soon after the duplication event and remained fixed throughout the evolution of

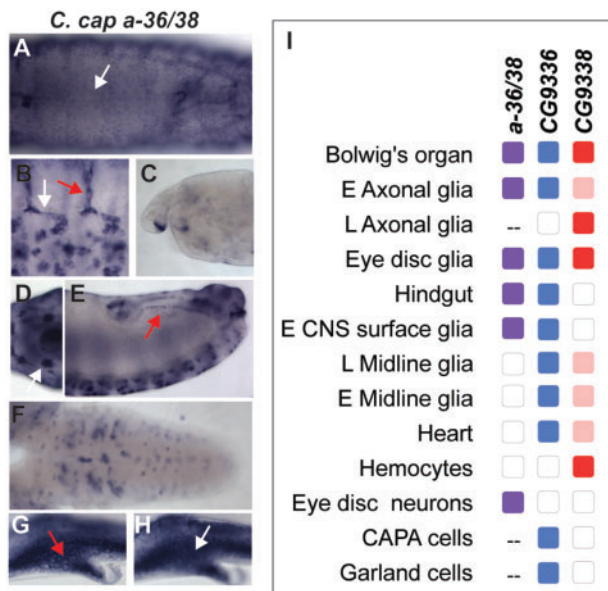


Fig. 3. Tissue-specific expression of the unduplicated ortholog. (A–E) mRNA localization of the unduplicated ortholog *a-36/38* in the *C. capitata* embryo. (A) There is no detectable expression in the heart (white arrow). The expression is detected in the axonal glia (B, red arrow) and the CNS surface glia (B, white arrow), the Bolwig's organ (D, white arrow), and the hindgut boundary cells (E, red arrow). No signal is detected in the midline glia (B) or the hemocytes (C). (F–H) *a-36/38* transcript localization in the tissues of the third instar larval tissues. (F) In the CNS, the expression was detected in a subset of neurons. In the eye disc, both the glia (G) and the photoreceptor neurons (H) showed expression. (I) Diagram summarizing the tissue-specific expression of *a-36/38*, CG9336, and CG9338. Lighter color denotes weak expression. The mRNA expression in the larval axonal glia of *C. capitata* could not be visualized by our in situ hybridization protocol. CAPA cells and garland cells could not be identified in *C. capitata*. L, larval; E, embryonic. See the main text for details.

Drosophila genus. Alternatively, the divergence could have taken place at different time points in a lineage-specific manner, yielding diverse but related expression patterns in different lineages. To gain further insights into the trajectories of expression divergence followed by the two paralogs after the duplication event, we characterized the embryonic and larval expression patterns of both genes in *D. ananassae*, *D. pseudoobscura*, and *D. virilis* by in situ hybridization (fig. 4 and supplementary fig. S2, Supplementary Material online). The results are summarized in figure 4A.

Both paralogs are expressed in the Bolwig's organ (fig. 4K, K', M, M', O, and O') and the embryonic axonal glia (fig. 4E, E', G, G', I, and I') in all the species examined. Another fully conserved feature concerns the hindgut boundary cells, which exclusively express CG9336 in all four species (figs. 2E and 4L, 4N, and 4P), suggesting that this domain was asymmetrically inherited from the unduplicated ortholog (fig. 3E) before the drosophilid radiation. In contrast, other complementary patterns appear to have arisen by inactivation of one of the paralogs in specific lineages. For example, although both genes are expressed in the eye disc glia of *D. melanogaster* and *D. virilis* (fig. 2M and S and supplementary fig. S2M and P, Supplementary Material online) only CG9336 or

CG9338 are detected in *D. ananassae* or *D. pseudoobscura* eye discs, respectively (supplementary fig. S2A, D, G, and J, Supplementary Material online). Similarly, in the embryonic CNS surface glia of *D. melanogaster* and *D. virilis*, only CG9336 is present (figs. 2A, 2B, and 4I), whereas CG9338 alone is expressed in this tissue in *D. ananassae* (fig. 4E').

The neofunctionalization events identified above in *D. melanogaster* also appear to have occurred at different time points in the paralog divergence process (fig. 4A). In the case of the midline glia, the expression of both paralogs is conserved in all the species in both embryonic and larval stages (figs. 2A, 2G, 2N, 2T, 4A, 4E, 4E', 4G, 4G', and supplementary fig. S2B, E, H, and K, Supplementary Material online) with the exception of *D. virilis*, where the expression is only present in the larva (fig. 4A and supplementary fig. S2N and Q, Supplementary Material online). Thus, this tissue specificity could have been acquired close to the duplication event, at least in the larva. Similarly, the embryonic heart expresses one or both paralogs in all the species examined (fig. 4A–D'). However, the extent of this expression is very variable, ranging from strong CG9336 levels in the whole organ, as seen in *D. melanogaster* (fig. 2F and L) to only a subset of cells, as observed in *D. pseudoobscura* (fig. 4C). Thus, this tissue-specific expression could have also originated soon after the duplication and subsequently undergone spatial modifications in different lineages.

In contrast, other novel expression domains appear to have originated within specific lineages. For instance, the expression of CG9338 in the embryonic hemocytes is only seen in *D. melanogaster* and *D. ananassae* (figs. 2I and 4F'). Further examples include the CG9336 expression in the garland cells of *D. melanogaster* (fig. 2Q) and in the eye photoreceptors of *D. pseudoobscura* (fig. 4A and supplementary fig. S2G', Supplementary Material online). Finally, we detected expression of CG9336 (fig. 2N) in the CAPA neurons of *D. melanogaster* and of both paralogs in *D. virilis* (fig. 4A and supplementary fig. S2N' and Q', Supplementary Material online).

In summary, our interspecific comparisons reveal that the present-day expression patterns of the two paralogs have been shaped by a series of lineage-specific events involving both the asymmetric inheritance of ancestral patterns and the acquisition of novel transcriptional profiles.

Characterization of the Cis-Regulatory Landscape Controlling Paralog Expression in *D. melanogaster*

We next sought to dissect the cis-regulatory regions of the CG9336-CG9338 locus to determine the genetic basis of their shared and divergent expression patterns. For this purpose, we built a series of *D. melanogaster* GAL4 reporter constructs containing all the intronic and flanking intergenic regions of the two paralogs, with the exception of the short 58 bp intron 1 of CG9338 (fig. 5). For some regions, we built additional constructs with smaller overlapping fragments to better define the position of particular CRMs (fig. 5). Comparing the tissue-specific activity of these reporters with the endogenous expression patterns of the two paralogs, we have uncovered a set of CRMs capable of driving expression in most of the native domains (figs. 5 and 7, and supplementary fig. S7, Supplementary Material online). Our results reveal a complex

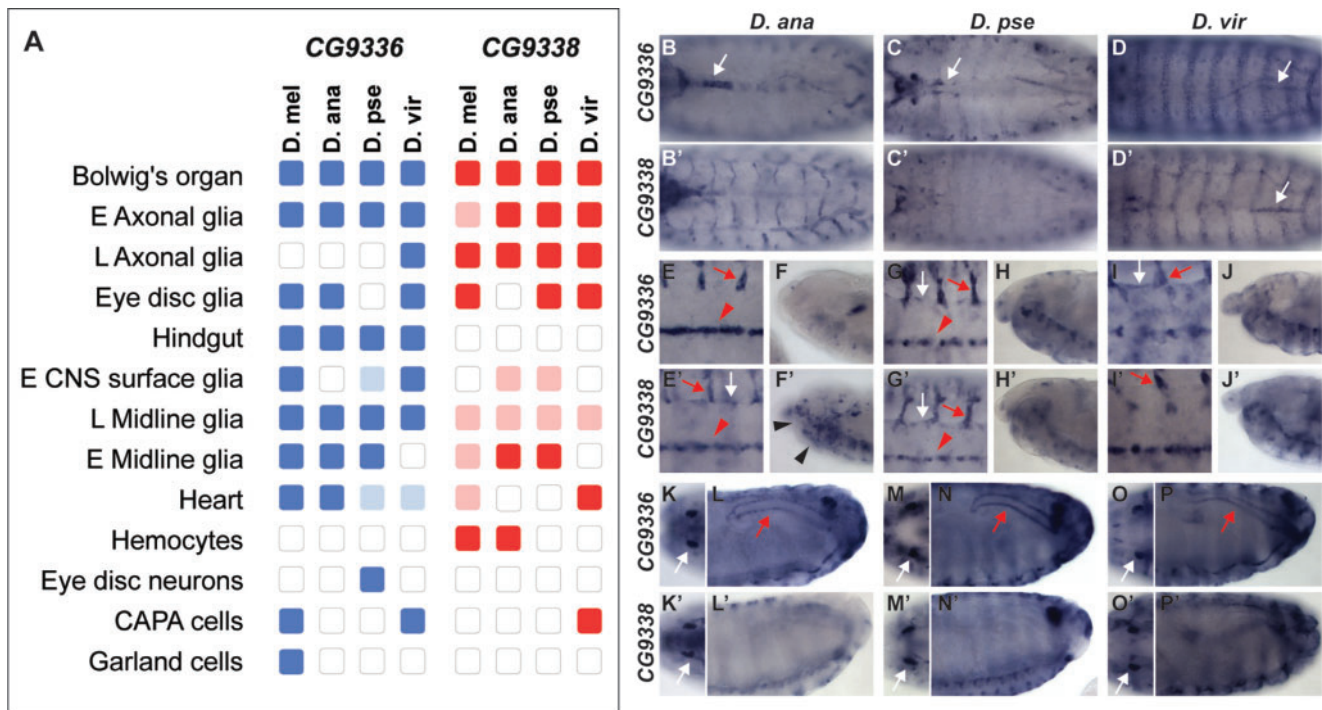


Fig. 4. Comparative analysis of the embryonic expression patterns of CG9336 and CG9338 orthologs. (A) Summary of tissue-specific expression of CG9336 (blue) and CG9338 (red) in four *Drosophila* species. Lighter colors denote weak expression and empty squares represent lack of expression. L, larval; E, embryonic. (B–D') In situ hybridization showing mRNA expression in the heart. In *D. ananassae* (B, B'), CG9336 is expressed most prominently in the cardiac cells in the anterior portion of the heart (B, arrow), with weaker expression in more posterior regions, whereas CG9338 is not detected (B'). In *D. pseudoobscura* (C, C'), CG9336 mRNA is detected in few pericardial cells in the anterior heart (C, arrow), whereas CG9338 mRNA is absent (C'). In *D. virilis*, both paralogs are detected in the pericardial cells in the posterior heart (D, D', arrows), with CG9338 being expressed more prominently (D'). (E, E', G, G', I, I') Expression of the paralogs in the glia. In *D. ananassae* (E, E'), both paralogs are expressed in the axonal glia of the PNS (red arrows) and the midline glia of the VNC (arrowheads). In the CNS surface glia, only CG9338 is expressed (E', white arrow). In *D. pseudoobscura* (G, G'), both paralogs are detected in the axonal glia (red arrows), in the midline glia (arrowheads) and, weakly, in the surface glia (white arrows). In *D. virilis* (I, I'), both paralogs are expressed in the axonal glia of the PNS (red arrows), whereas only CG9336 is detected the CNS surface glia (I, white arrow). Neither paralogs are expressed in the midline glia (black arrows). (F, F', H, H', J, J') Expression of the paralogs in the hemocytes, as shown in the embryonic head. The only paralog detected in these cells is *D. ananassae* CG9338 (F', arrowhead). Both paralogs are expressed in the Bolwig's organ in all species (K, K', M, M', O, O'). In the boundary cells of the hindgut, CG9336 is expressed in all species (L, N, P; red arrows), whereas CG9338 is absent in all species (L', N', P').

cis-regulatory architecture in which an array of modular enhancers distributed throughout the locus controls the tissue-specific expression of the two paralogs (fig. 5).

Specifically, we identified CRMs active in paralog-specific domains. CG9336 expression in the CAPA neurons and the garland cells could be driven by modules present in the 5' intergenic region of CG9336 (Dmel A construct; fig. 5), whereas the expression in the hindgut could be regulated by three CRMs (Dmel A, C, and E constructs, fig. 5). Similarly, two modules present in intron 2 of CG9338 and the 3' intergenic region of CG9338 (Dmel Fc 5' Hemo and Dmel G constructs, fig. 5) likely drive the expression of this gene in the hemocytes.

We also identified CRMs for domains shared by the two paralogs, such as the glial cells. We found two separate regions displaying enhancer activity in this tissue. One is located in the intergenic region and drives expression in all glial cell types that express the two paralogs (Dmel D construct, fig. 5). Through enhancer bashing, we have delimited the glial enhancer activity to a 908-bp region immediately downstream of the CG9336 coding region (Dmel Da construct).

Further bashing of this construct indicates that this region contains at least three parts, driving both common and distinct expressions. Whereas the 5' end harbors the midline glia activity, the central region (Dmel Dc5' Peak construct) and the 3' end (Dmel Dc3' Peak construct) can independently drive expression in the rest of the glial domains (fig. 5 and supplementary fig. S7, Supplementary Material online). In fact, the only difference in the activities of the latter two is in the wrapping glia in the eye disc, where only the Dmel Dc 5' Peak fragment is active (fig. 5 and supplementary fig. S7, Supplementary Material online). The other glial CRM identified resides in the second intron of CG9338 and drives expression in all glial domains, with the exception of the midline glia (Dmel F construct, fig. 5 and supplementary fig. S7, Supplementary Material online). This configuration thus suggests that the intergenic midline glia CRM could be shared by both paralogs, but each duplicate could have a dedicated CRM driving transcription in the other glial subtypes (figs. 2 and 3).

Other CRMs potentially shared by the two paralogs are the embryonic heart CRM, which overlaps extensively with the

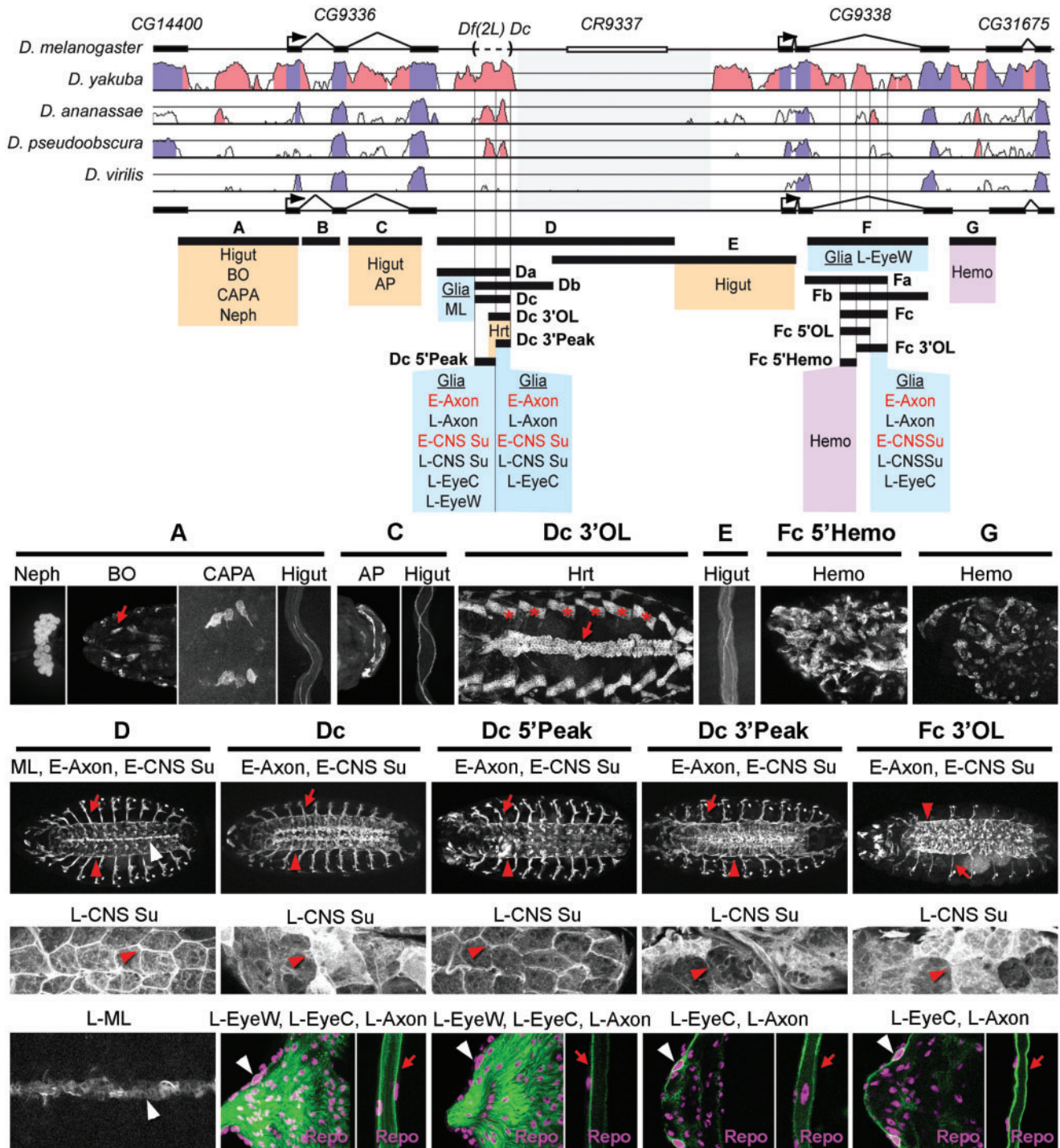


FIG. 5. Locations of CRMs controlling the CG9336 and CG9338 expression in *D. melanogaster*. The diagram on the top shows VISTA plots displaying sequence conservation between the *D. melanogaster* CG9336–CG9338 genomic region and homologous regions in different drosophilid species. Highly conserved regions between two species are represented as red (noncoding DNA) or purple (coding DNA) peaks. The coding regions of the two *D. melanogaster* paralogs and the adjacent genes are indicated above the plots. The region deleted in the *Df(2L)Dc* mutant is shown between brackets. The *D. melanogaster*-specific insertion containing the CR9337 pseudogene (white bar) is shaded in grey. We did not find any relevant enhancer activity associated with this region. Below the alignments, the genomic fragments included in different GAL4:VP16 reporter constructs are represented by black bars. Construct names are indicated with bold face capital letters. The inferred locations of different tissue-specific CRMs, narrowed down to minimal regions by comparison of overlapping constructs, are indicated in colored boxes (blue, glia; purple, hemocyte; orange, other). Higut, hindgut boundary cells; BO, Bolwig’s organ; CAPA, CAPA-peptide abdominal neurons; Neph, nephrocytes (garland cells); AP, anal plate ring cells; Hrt, embryonic heart; Hemo, embryonic hemocytes; ML, embryonic and larval midline glia; E-Axon and L-Axon, embryonic and larval axonal glia; E-CNS Su and L-CNS Su, embryonic and larval CNS surface glia; L-EyeW, eye wrapping glia; L-EyeC, eye carpet glia. The confocal images are stack projections of embryonic and larval tissues showing mCD8-GFP (embryos, b/w) or mCD8-mCherry (larvae, b/w or green) reporter expression, driven by the different GAL4:VP16 constructs. The A construct drives expression in the embryonic Bolwig’s organs (red arrow), the

intergenic glial CRM (Dmel Dc3' OL construct, fig. 5), the anal plate CRM located in the CG9336 Intron 2 (Dmel C construct, fig. 5), and the Bolwig's organ CRM located in the 5' upstream region of CG9336 (Dmel A construct, fig. 5). As we did not find other CRMs active in these tissues, any of these modules could be responsible for the expression of both CG9336 and CG9338 in these domains. However, we cannot rule out that additional CRMs present in the coding regions or outside the regions examined regulate the two genes.

Endogenous Deletion of the Dc Region Reveals the Activity of Shared Modules

To determine if the paralog coexpression in the embryonic glia and the heart is due to sharing of the intergenic enhancers (fig. 5), we used the CRISPR/Cas9 technology (Gratz et al. 2014) to delete the genomic region corresponding to the Dmel Dc construct (figs. 5 and 6). We note that this deficiency, *Df(2L)Dc*, which is fully viable, deletes a genomic region particularly well conserved among *Drosophila* species (fig. 5). In *Df(2L)Dc* homozygous embryos, the expression of both paralogs disappears in the heart and in all the glial tissues, including the midline (fig. 6C, C', D, G, G', and H). Thus, the intergenic CRMs appear to be shared and are essential for the embryonic glial and heart expression of both paralogs (fig. 6W). This result also suggests that although the CG9338 intron2 glial CRM is active in the embryo in the transgenic assay, it is not capable of driving the embryonic glial expression in its endogenous genomic context. Finally, our observations also indicate that the Dc region is necessary but not sufficient for the embryonic midline glia expression, as this fragment on its own does not display activity in this tissue (fig. 5 and supplementary fig. S7, Supplementary Material online).

In the *Df(2L)Dc* larvae, CG9336 expression is also abolished in the eye wrapping glia and the CNS surface glia but is only reduced in the midline glia (fig. 6I–N). In contrast, the CG9338 glial expression remains unaffected except in the midline, where it becomes undetectable (fig. 6O–V). Thus, the Dc region is strictly required for CG9336 midline expression in the embryo but not in the larva. We also observe that the *Df(2L)Dc* deletion alters paralog expression in tissues a priori dependent on the activity of other CRMs. In particular, both paralogs are clearly up-regulated in a series of dorsal epidermal stripes and in the anal plate of mutant embryos (fig. 6A'–D' and E'–H'). In contrast, in the larval stage, we did not observe paralog up-regulation or ectopic expression in the

tissues examined (supplementary fig. S3, Supplementary Material online).

In summary, our results reveal the complex nature of the regulatory logic operating in this locus and suggest the contribution of stage-specific enhancer–promoter and enhancer–enhancer interactions to the regulation of the two paralogs.

Interspecific Comparison of Cis-Regulatory Architectures

To compare the cis-regulatory landscapes of the four drosophilids, we built for the other three species a series of GAL4 constructs containing the regions homologous to the *D. melanogaster* reporter constructs. We then assessed their activities in *D. melanogaster* transgenic hosts (fig. 7 and supplementary figs. S4–S7, Supplementary Material online). Our data indicate that several aspects of the cis-regulatory architecture observed in *D. melanogaster* are preserved in the other species. To begin with, the positions of the two glial CRMs always coincide (fig. 7). In addition, the location of the enhancers active in the Bolwig's organ (CG9336 5' region), the anal plate ring (CG9336 Intron 2), and the heart (intergenic region) is conserved, as it is also the case for most of the multiple modules driving expression in the hindgut boundary cells (CG9336 5', CG9336 intron 2, and intergenic region, fig. 7).

Despite the conserved location of the two glial CRMs in the different species, fine-scale analysis reveals varying activities in the different glial subtypes (fig. 7 and supplementary figs. S4–S7, Supplementary Material online). Importantly, this variation correlates with the endogenous expression of the paralogs in each species (fig. 4A). For instance, both CRMs are active in the eye disc glia of *D. melanogaster* and *D. virilis*, where both paralogs are expressed (fig. 4A). In comparison, CG9338 is not detected in this tissue in *D. ananassae* (fig. 4A), reflecting the lack of activity of its CG9338 intron 2 CRM (fig. 7; Dana F, Fb and Fc, supplementary fig. S4, Supplementary Material online). Similarly, the intergenic module of *D. pseudoobscura* is not active in the eye wrapping glia (although it is active in the thin carpet glial cells, which are not reliably stained with in situ hybridization; fig. 7; Dpse D, supplementary fig. S5, Supplementary Material online), providing a potential explanation for the lack of CG9336 expression observed in the eye glia of this species (fig. 4 and supplementary fig. S2G, Supplementary Material online).

Based on these observations, we reasoned that in all species a dedicated CRM drives the transcription of each paralog in the eye disc glia. However, in other tissues, the regulatory logic controlling paralog expression seems to vary among

Fig. 5 Continued

larval hindgut boundary cells, CAPA neurons, and nephrocytes. The C construct elicits expression in the embryonic anal plate and the larval hindgut boundary cells. Strong expression in both the heart (red arrow) and in dorsal muscles (red asterisks) is observed for the Dc 3'OL construct. The E construct drives the larval hindgut boundary cell expression. Both Fc 5'Hemo, and G constructs drive expression in embryonic hemocytes, visible here in stage 15 embryos. The D, Dc, Dc 5'Peak, Dc 3'Peak and Fc 3'OL constructs are active in the axonal glia in both stage 17 embryos (red arrows) and in third instar larvae (visible in green, red arrows). They also drive expression in the embryonic (arrowhead) and larval CNS surface glia (red arrowheads). The D construct, in addition, is active in the midline glia at both stages (white arrowheads). In the larval eye disc, Dc, Dc 5'Peak, Dc 3'Peak and Fc 3'OL drive expression in the carpet cells (white arrowheads), but only Dc and Dc 5'Peak drive expression in the eye wrapping glia (broad GFP signals in the eye disc).

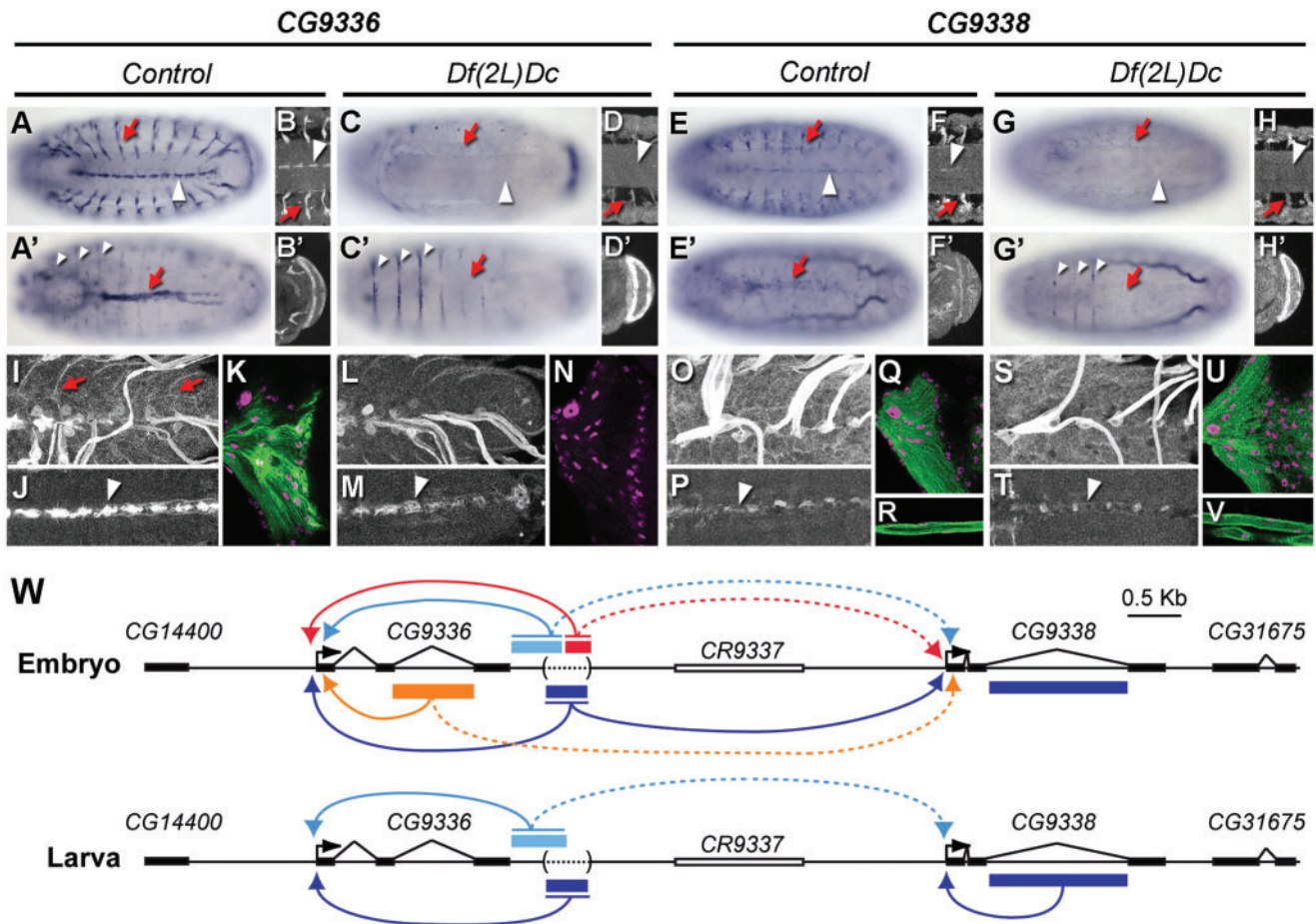


Fig. 6. Functional analysis of the Dc region in *D. melanogaster*. (A–H') Expression of CG9336 and CG9338 in wild-type and *Df(2L)Dc* embryos (stages 16 and 17) revealed by in situ hybridization (A, A', C, C', E, E', G and G'), anti-GFP staining in embryos expressing CG9336-YFP (B, B', D, D') or CG9338-YFP (F, F', H, H'). (A–H) Ventral views showing expression of the two paralogs in the axonal glia (red arrows) and the midline glia (white arrowheads) of the CNS. In the mutants (C, D, G, H), the signal is absent in these two tissues. (A', C', E', G') Dorsal views showing paralogs expression in the heart (red arrows) and in a series of anterior epidermal stripes (white arrowheads). Heart expression is lost in the mutants (C', G'), but both paralogs are up-regulated in the epidermal stripes. (B', D', F', H') Confocal stack projections of the anal plate region showing up-regulation of both paralogs in *Df(2L)Dc* mutants. (I–V) Confocal stack projections of third instar larvae carrying the YFP-tagged proteins visualized with anti-GFP (b/w and green). Magenta shows Repo staining. In the larval CNS surface glia (I and L, red arrows) and the eye wrapping glia (K and N), the CG9336-YFP expression is lost in the mutant. The midline glia expression (white arrowheads) is reduced in the mutant larva (compare J and M). CG9338-YFP expression is not affected in the CNS glial cells (O and S), eye disc glia (Q and U), or axonal glia (R and V), but is lost in the midline glia (white arrowheads, compare P and T). (W) Schematic summarizing the inferred activity of different CRMs (colored rectangles) on the expression of the two paralogs in the embryonic and larval stages. The *Df(2L)Dc* deletion is indicated by brackets. Directional activities of different CRMs on the two paralogs' promoters are illustrated with arrows (plain lines, high activity levels; dotted lines, low levels). In the embryo, single CRMs for the anal plate (orange), the midline glia (pale blue), and the heart (red) drive CG9336 and, at lower levels, CG9338 expression in these tissues. A single shared intergenic CRM within the *Df(2L)Dc* deletion is responsible for glial expression of both paralogs (dark blue). In the larvae, the shared midline glia CRM (light blue) also drives expression of both paralogs, but its deletion does not completely abolish the expression. The intergenic glial CRMs are responsible for CG9336 activation in the CNS surface glia and the eye disc wrapping glia. CG9338 glial expression is under the control of the second glial CRM located in its intron.

species. As an example, our analysis of the *Df(2L)Dc* mutants shows that the intergenic enhancer is the sole element controlling CG9336 expression in the CNS surface glia of *D. melanogaster*. However, in *D. ananassae*, this CRM is not active in the embryonic surface glia (fig. 7; Dana DE construct, supplementary fig. S4, Supplementary Material online) and, accordingly, CG9336 is not expressed (fig. 4A and E). In contrast, the CG9338 intron 2 CRM is active in this tissue (fig. 7; Dana F, Fb and Fc, supplementary fig. S4, Supplementary Material online)

and CG9338 is expressed (fig. 4E'). Thus, whereas a single CRM drives the embryonic glial expression of both paralogs in *D. melanogaster*, both CRMs appear to contribute to their expression in *D. ananassae*.

Overall, our results are consistent with the idea that most of the interspecific variations observed in the endogenous expression patterns are due to *cis*-regulatory divergence rather than *trans*-regulatory changes. This is further illustrated by the presence of CRMs active in the novel lineage-specific

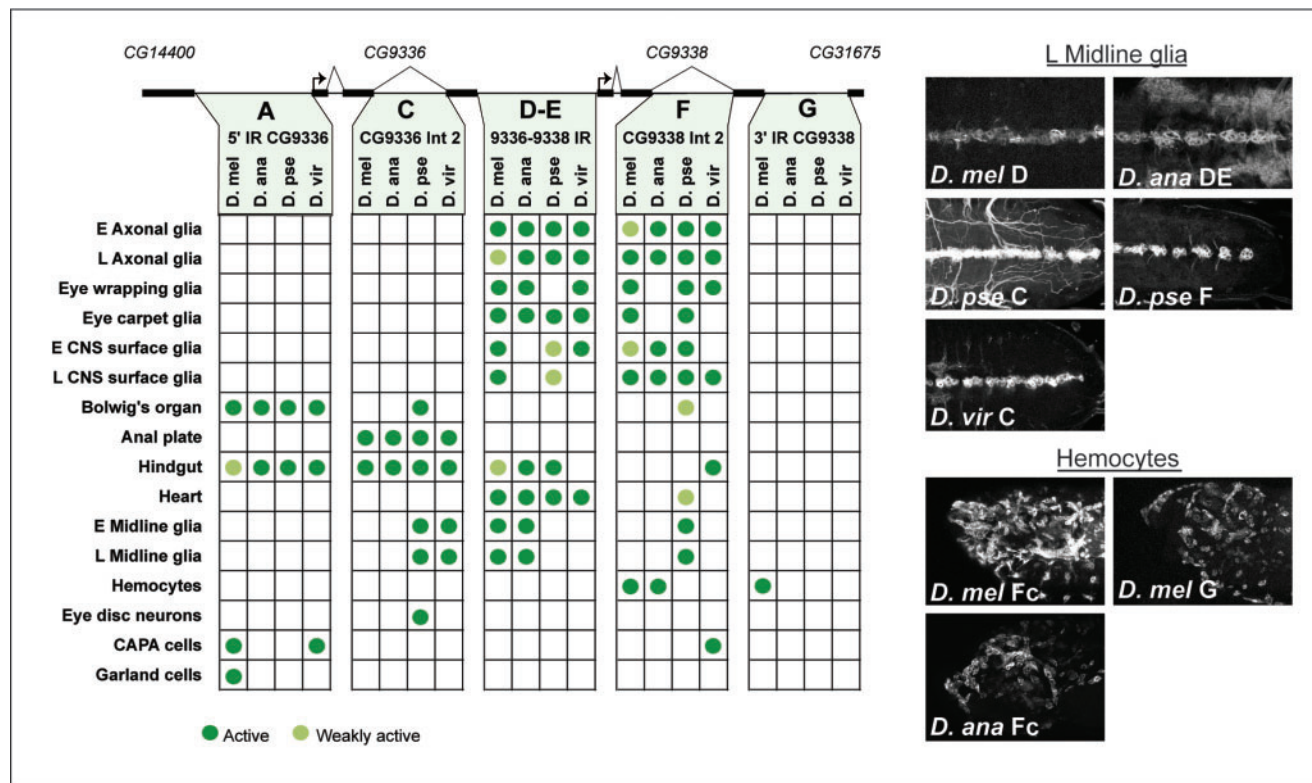


Fig. 7. Locations of tissue-specific CRMs in four *Drosophila* species. (Left) The top scheme shows the noncoding regions in the locus corresponding to the constructs A, C, DE (or D and E in *D. melanogaster*), F, and G, whose enhancer activities were tested. Green circles represent the tissue-specific CRMs found in each region. “E” and “L” before the tissue names indicate the embryonic and the larval stages, respectively. The lighter green circles indicate weak activities. (Right) Confocal images are stack projections showing the activities of the larval midline glia and the embryonic hemocyte CRMs from different species. For both tissues, the number and the location of CRMs vary among different species. A full description of the activities of all constructs tested is available in supplementary figure S7, Supplementary Material online.

domains, including the hemocyte expression in *D. melanogaster* and *D. ananassae* (fig. 7; in the constructs Dmel F and Dmel G, fig. 5, and in Dana F, supplementary fig. S4, Supplementary Material online), the *D. melanogaster* CG9336 expression in the garland cells (fig. 7; in Dmel A, fig. 5), and the *D. pseudoobscura* CG9336 expression in the eye photoreceptors (fig. 7; in Dpse C, supplementary fig. S5, Supplementary Material online).

Taken together our data indicate that most of the lineage-specific sub- and neo-functionalization events under study result from changes in the activities of conserved CRMs or from the emergence of new modules.

Conserved Expression: Compensatory and Redundant Enhancers

Although expression in several conserved domains appears to be driven by homologous CRMs, our interspecific comparisons reveal cases in which expression could be regulated by CRMs that exist in variable numbers and occupy different locations. For instance, whereas *D. melanogaster* and *D. ananassae* have a single midline glia CRM in the intergenic region, this activity resides in the second intron of CG9336 in *D. pseudoobscura* and *D. virilis* (fig. 7; constructs Dpse C and Dvir C, supplementary figs. S5 and S6, Supplementary Material online). Moreover, *D. pseudoobscura* has an

additional midline CRM in the CG9338 intron 2 (fig. 7; Dpse F construct, supplementary fig. S5, Supplementary Material online). Thus, expression in this tissue depends on CRMs that have changed position during drosophilid evolution and appear to function as compensatory enhancers (also referred to as nomadic enhancers) (Kalay and Wittkopp 2010; Arnold et al. 2014).

Another such example is provided by the CRMs driving expression in the hindgut boundary cells. All four species have multiple CRMs active in this tissue but not always in homologous regions (fig. 7). We also found that *D. pseudoobscura*, in comparison to the other species, has additional CRMs for the Bolwig's organ (fig. 7; CG9336 Intron 2, Dpse C, supplementary fig. S5, Supplementary Material online) and the heart (CG9338 Intron 2, Dpse F, supplementary fig. S5, Supplementary Material online). Similarly, *D. virilis* has two CAPA neuron CRMs (Dvir A and Dvir F, supplementary fig. S6, Supplementary Material online). However, given that in this species both genes are expressed in these cells (fig. 4A and supplementary fig. S2N' and Q, Supplementary Material online), each module could be dedicated to a specific paralog, as it could also be the case for the two *D. pseudoobscura* midline CRMs (fig. 7; Dpse C and F, supplementary fig. S5, Supplementary Material online). Finally, *D. melanogaster* has two CRMs driving expression in the embryonic hemocytes

(fig. 7; Dmel Fc 5' Hemo and Dmel G, fig. 5). Thus, these modules could potentially control CG9338 expression in these cells in a redundant manner.

In summary, our locus-wide inventory of enhancer activities uncovers that conserved expression features can also rely on a highly evolvable *cis*-regulatory architecture.

Discussion

Lineage-Specific Changes Shaped Evolving Expression Patterns during Paralog Divergence

Few empirical studies have directly examined the contribution of *cis*-regulatory evolution to the functional divergence of gene paralogs (Hittinger and Carroll 2007; Jarinova et al. 2008; Kleinjan et al. 2008; Loehlin and Carroll 2016). In this work, we have characterized the expression patterns of a pair of tandem duplicates and its underlying *cis*-regulatory basis in four drosophilid species. We have surveyed in this way a diversification process of 40 My, the estimated age of the last common ancestor of the species analyzed (Russo et al. 1995; Obbard et al. 2012). Based on the phylogeny, we could infer that many of the present-day expression divergence between CG9336 and CG9338 have been shaped by lineage-specific sub- and neo-functionalization events after fixation of the two paralogs (fig. 4). This and two additional observations argue against the initial preservation of the two paralogs through the rapid acquisition of complementary patterns, which evolved immediately after the duplication event (Ohno 1970; Force et al. 1999). First, paralog coexpression is still prevalent in many tissues (figs. 3I and 4A) and, second, at least in the two developmental stages examined, we could not detect ancient domains unique to CG9338, which would have made its initial retention necessary. Thus, alternative processes such as sharing of the ancestral gene dose between the two paralogs could have been determinant for their initial maintenance in the genome, which was then followed by subfunctionalization in certain lineages (Lan and Pritchard 2016; Thompson et al. 2016).

The dosage-sharing mechanism, in fact, has been reported as prevalent among tandem duplicates, both in mammals and in different populations of *D. yakuba* (Lan and Pritchard 2016; Rogers et al. 2017). In tandem configurations, presumably, the shared *cis*-regulatory environment allows the maintenance of the overlapping expression domains for prolonged periods, whereas other mechanisms could alter the total level of transcription from the duplicated coding units. As detailed below, our study provides a clear example where shared enhancers contribute to the coregulation of the two paralogs in an ancestral expression domain. However, our results also illustrate ways in which *cis*-regulatory evolution can decouple paralog regulation. Our data thus provide empirical support to the notion that the dynamic properties of CRMs facilitate the emergence of divergent paralog expressions and functional specialization in specific cellular contexts.

Cis-Regulatory Evolution and the Emergence of Novel Tissue-Specific Enhancers

Our results show that most of the evolutionary changes generating expression pattern divergence of CG9336 and CG9338 have a *cis*-regulatory basis. Our assay of tissue-specific enhancer activities relied on a heterologous expression system, as all the activities of CRMs were monitored in the *D. melanogaster* host. The fidelity of this approach thus depends on a high degree of conservation in the *trans*-regulatory networks operating among the species examined, a feature of our experimental system that has been extensively tested by many previous studies (Ludwig et al. 1998; Kalay and Wittkopp 2010; Frankel et al. 2011; Rebeiz et al. 2011; Arnoult et al. 2013; Arnold et al. 2014; Wunderlich et al. 2016). Although we cannot completely rule out that some of the expression differences observed are due to evolutionary changes in upstream regulatory networks, most of the reporters analyzed respond to the *trans*-regulatory landscape of *D. melanogaster* and collectively recapitulate the tissue-specific patterns observed in their respective endogenous species. For instance, this is clearly the case for the five neofunctionalization events identified in our data set, where for each novel expression domain we have detected the concomitant emergence of tissue-specific CRMs (fig. 7; heart, midline glia, hemocytes, eye photoreceptors, and garland cells). As some of these CRMs are only found in single lineages, it indicates that the CRM with novel tissue-specific activities appeared in relatively short evolutionary timescale contributing to the rapid expression divergence of the paralogs.

Conserved Glial CRMs and Paralog Expression Divergence

The most conserved *cis*-regulatory feature of the CG9336–CG9338 locus is the presence of two glial CRMs, which collectively control the expression of the two paralogs in these tissues (fig. 7). These modules are located in homologous regions in all the species examined, and the intergenic glial CRM is indeed the only regulatory element in the locus displaying significant sequence conservation (fig. 5). Our data show that despite many similarities, the activities of these two CRMs are not equivalent (fig. 7) resulting in the divergent expression of the two paralogs in this tissue.

First, our analysis of the reporter constructs showed that the two CRMs display different enhancer activities within the glial subtypes indicating that they integrate common developmental cues, hence their redundancy in some glial subtypes, but also distinct inputs, allowing, for instance, differential expression in the eye carpet and wrapping glia (fig. 7). Second, by deleting the intergenic glial CRM in *D. melanogaster*, we demonstrated that the two glial CRMs have distinct temporal requirement as well as different capacities to interact with the promoters of each paralog (fig. 6). Together, these differences contribute to the partially overlapping glial expression of the two paralogs (fig. 7).

Evolutionary changes in the capacities of the two CRMs also appear to underlie the interspecific variation. We have observed in the different species several examples of

subfunctionalization in glial subtypes that correlate with changes in the activity of the two glial CRMs (figs. 4 and 7). However, whether the promoter preferences of these CRMs have also changed during evolution is difficult to establish in absence of functional data describing their endogenous activity. It is nevertheless tempting to speculate that in coregulated loci, expression divergence may not exclusively depend on the inactivation of specific enhancers but also on the dynamic modulation of enhancer–promoter interactions, which, as we have shown, can contribute to gene regulation integrating both temporal and spatial developmental cues.

Redundant Enhancers and the Evolution of Cis-Regulatory Architecture

Previous studies have demonstrated the pervasive presence of redundant CRMs (enhancers with overlapping tissue activities) in the genome of *D. melanogaster* (Frankel et al. 2010; Cheng et al. 2014; Ross et al. 2015; Cannavò et al. 2016; Wunderlich et al. 2016). Our locus is not an exception to this trend, and the glial CRMs are not the only redundant modules we identified (fig. 7). Interestingly, we found that the number and location of these redundant enhancers vary among species, suggesting that they can undergo lineage-specific births and losses (fig. 7). As proposed by Kalay and Wittkopp (2010), such a dynamic pattern of enhancer evolution is likely to underlie the existence of compensatory enhancers.

Specifically, we found in different species two and even three different CRMs driving expression in the hindgut boundary cells. Although it could be argued that some of these modules may not be functional in their endogenous context, these CRMs appear to interact specifically with the CG9336 promoter, as only this paralog is expressed in this tissue. This suggests that selective pressures acting on the promoter preference of these CRMs could prevent paralog coexpression in this tissue. This example demonstrates that variations in both CRM number and position do not necessarily translate into interspecific differences in the expression of the two paralogs. However, this underlying variation could potentially lead to operational changes in the cis-regulatory logic. For example, both CG9336 and CG9338 are expressed in the midline glia in all species (fig. 4), and a single midline glial CRM drives their expression in *D. melanogaster*, *D. ananassae*, and *D. virilis* (fig. 7). However, in *D. pseudoobscura*, there are two midline CRMs (fig. 7). This raises the possibility that in this species, a paralog-specific enhancer could control the midline expression of each gene. If this was the case, the emergence of a redundant module could have promoted the transition from the state of coregulation to that of independent regulation. Thus, the recurrent appearance of redundant enhancers could present opportunities to decouple paralog regulation in specific tissues, a process that may play a recurrent role during paralog divergence.

The Role of Endogenous Genomic Context in Paralog-Specific Regulation

Our observations reveal that the developmental regulation of the two paralogs requires the contribution of dynamic

enhancer–promoter interactions. Two additional observations further highlight the role played by the endogenous genomic context in the transcriptional regulation of this locus. First, we have shown in the *Df(2L)Dc* mutants that the deletion of a CRM can perturb the activity of other CRMs, suggesting that enhancer–enhancer interactions may also contribute to paralog regulation (fig. 6). Second, we observed that the activities of some of our reporters do not recapitulate the stage- and tissue-specific variations observed in the expression of the endogenous transcripts (figs. 2, 4, 7, and supplementary fig. S2, Supplementary Material online). For instance, CG9336 expression in the axonal glia is restricted to embryonic stages in all the species except in *D. virilis*. However, all the reporter constructs containing the axonal glial CRMs are active in both embryonic and larval stages (figs. 4 and 7).

Altogether, these findings are consistent with the notion that the coordinated expression of tandem duplicates depends on the complex interaction of CRMs with other encoded elements, including other enhancers, tethering elements, repressors, insulators, and local chromatin regulators, which contribute to restrict CRM activity to specific paralogs, developmental stages, or cell types (Calhoun et al. 2002; Carvajal et al. 2008; Jarinova et al. 2008; Tsujimura et al. 2010; Kvon et al. 2014; Long et al. 2016; Maeso and Tena 2016). These interactions can thus become potential targets of evolutionary changes during paralog divergence, just as the individual activities of each CRM.

This work is one of the first empirical analyses describing the expression evolution of a duplicated gene pair in closely related species. The rapidly evolving nature of the paralogs has permitted an unprecedented level of detail in the dissection of the mechanisms shaping their expression evolution. This gene pair and other loci with equivalent properties can thus constitute promising study systems, which can help us gain deeper understanding of gene regulatory evolution. With the advent of genome editing technology, which can be applied to non-model drosophilids (Ding et al. 2016; Karageorgi et al. 2017; Stern et al. 2017), the inquiry into cis-regulatory evolution can progress beyond the study of individual enhancers and to the functions of other layers of cis-regulation.

Materials and Methods

Sequences and Phylogeny

The genomic sequences covering the CG9336–CG9338 region were obtained from the *D. melanogaster* R6.13, *D. yakuba* R1.05, *D. ananassae* R1.05, *D. pseudoobscura* R3.04, and *D. virilis* R1.06 genome releases and were pairwise aligned using the mVISTA tool (Frazer et al. 2004). Protein coding and amino acid sequences were aligned using the Clustal Omega software with the default settings (<http://www.ebi.ac.uk/Tools/msa/clustalo/>; last accessed September 13, 2017). The GenBank sequence accession numbers are: *Dmel36*, NM_136225; *Dyak36*, XM_002090241; *Dana36*, XM_001965237; *Dpse36*, DR121923; *Dvir36*, XM_002052027; *Dgri36*, XM_001988362; *Dmel38*, NM_136227; *Dyak38*,

XM_002090240; *Dana38*, XM_001965236; *Dpse38*, XM_002132762; *Dvir38*, XM_002052026; *Dgri38*, XM_001988361; *Ccap3638*, XM_004524746; and *Bcuc3638*, XM_011189182. The sequences of the coding regions were used to compute the gene tree (supplementary fig. S1, Supplementary Material online) with the Maximum Likelihood method in MEGA6 package (Tamura et al. 2013), using the Tamura–Nei model.

Animal Husbandry and Fly Stocks

Wild-type strains of *D. melanogaster*, *D. ananassae*, *D. pseudoobscura*, and *D. virilis* came from the Drosophila Species Stock Center (San Diego, USA). The YFP-reporter lines used were the YFP protein trap lines, CG9336^{CPT1001654} (DGRC #115180) and CG9338^{CPT1100000} (DGRC #115071). In these lines, the YFP is incorporated into the endogenous products, which remain under the control of their native *cis*-regulatory regions (Lowe et al. 2014). Other *D. melanogaster* strains used include the reporters *UASmCD8-GFP* (Bloomington#5137) and *UASmCD8-mCherry* (B#27392), and the GAL4 drivers *w*; *breathless-GAL4*⁵², *UASActin5C-GFP* (B#8807), and *P(GawB)Mz97-GAL4* (Ito et al. 1995). We also used the integration platform *y sc v P(nos-phiC31); P(CaryP)attP2* (B#25710); and the *y w vasa-Cas9* stock (B#51323). Balancers combinations used were *w*; *TM3/TM6B*, *w*; *amos^{Tf}/CyO wg-lacZ*, *w*; *wg^{Sp1}/CyO tub-PBac T* (B#8285), *w*; *wg^{Gla1}/CyO twist-GAL4 UAS-GFP* (B#6662), and *w*; *sna^{ScO}/CyO, PsChFP2* (B#35523). The drosophilid cultures were raised at 25 °C in standard cornmeal medium. The *C. capitata* culture (kindly provided by Dr A. Jessup, IAEA Seibersdorf, Austria) was maintained at 25 °C on a diet of sugar and hydrolyzed yeast protein for the adults and on a *Drosophila* food medium for the larvae.

mCherry-Tagged Proteins and Live Imaging

For the generation of full length CG9336 and CG9338 tagged with mCherry, the mCherry coding sequence, flanked on each side by a single L residue, was introduced in frame between the predicted signal peptide (conserved residue Y22) and their respective three finger domains (see fig. 1B). We used the RE67340 (CG9336) and GH07967 (CG9338) EST clones (DGRC) as cDNA templates. Constructs were then sequenced and subcloned into the pUAST vector for the generation of transgenic strains permitting ectopic expression of each protein under the control of the GAL4 system (Brand and Perrimon 1993). For imaging, third instar larvae were dissected and immediately mounted in Schneider S2 medium (Gibco) between a slide and a coverslip separated by thin spacers. Tissues were imaged within 20 min, using a Zeiss 710 confocal microscope.

Reporter Constructs

The intronic, intergenic, and flanking noncoding regions of CG9336 and CG9338 from the four drosophilid species in this study were PCR-amplified from genomic DNA, using the Phusion High-Fidelity DNA polymerase (NEB). Primer combinations used are described in the Supplementary Material online. PCR products were cloned into the Gateway

pENTR1A Dual Selection Vector (Invitrogen) or Gateway pENTR/D-TOPO Vector (Invitrogen), then sequenced to verify the identity and orientation of the inserts (the 5′ intergenic regions were inserted in the 5′–3′ orientation relative to the promoter, whereas the rest was inserted in the 3′–5′ orientation). Using the Gateway LR Clonase II system (Invitrogen), the inserts were transferred to the pBPGAL4.2: VP16Uw vector (Pfeiffer et al. 2010), which contains a mini-white marker and an attB sequence for site-specific integration. Each construct was injected at a concentration of 0.5–1 μg/μl into *y sc v P(nos-phiC31); P(CaryP)attP2* embryos expressing the phiC31 site-specific integrase. Emerged adults were crossed to *y w* flies, and the progeny was screened for *w*⁺ insertions.

In Situ Hybridization

To synthesize paralog-specific riboprobes for each species, 3′ or 5′ untranslated regions were cloned from embryonic cDNA libraries and used as templates. The sequences of the cloning primers and the probes are listed in the Supplementary Material online. Embryos were dechorionated and fixed according to Tautz and Pfeifle (1989). In situ hybridization was carried out as in Panganiban et al. (1995) based on Tautz and Pfeifle (1989) with the following modifications: *C. capitata* embryos were incubated for 3 min in 4 μg/ml proteinase K at 37 °C, and the hybridization buffer included heparin instead of glycogen. Hybridization was carried out at 60 °C. Embryos were mounted in 70% glycerol in phosphate-buffered saline (PBS) and observed under the Leica DM LB2 upright microscope.

Third instar wandering larvae were dissected in PBS, fixed for 30 min in 4% formaldehyde in PBT (PBS, 0.1% Tween-20) and dehydrated in 100% methanol. Samples were rehydrated in 1:1 methanol/5% formaldehyde in PBT (5 min), postfixed in 5% formaldehyde in PBT (30 min), and washed three times for 10 min in PBS-Triton (PBS, 0.1% Triton X-100). After incubation for 5 min in 50 μg/ml proteinase K at room temperature in PBS-Triton, tissues were postfixed for 30 min in 5% formaldehyde in PBS-Triton. Hybridization was carried out overnight at 55 °C. Tissues were mounted in 60% glycerol and imaged in a Nikon Eclipse 80i microscope equipped with a DXM1200C digital camera.

Analysis of Reporter Expression and Immunohistochemistry

For the analysis of embryonic reporter activities, males from each line were crossed with *UASmCD8-GFP* virgin females. Embryos were fixed as indicated above and blocked in 5% normal goat serum in PBT for 30 min. Primary and secondary antibody incubations and washing steps were carried out in PBT. All samples were imaged on a Leica SP5 inverted confocal microscope. Reporter activities in the larva were analyzed in the third instar stage in the progeny of a cross between males of each reporter strain and CG9336^{CPT1001654}; *UASmCD8-mCherry* or CG9338^{CPT1100000}; *UASmCD8-mCherry* virgin females to assess the match with the endogenous expression. Larval tissues were fixed as described above for 20 min. Subsequent blocking, 4 °C overnight antibody incubations and washes were carried out in PBS-Triton with 0.1%

BSA. All samples were mounted in VECTASHIELD (Vector Laboratories) and imaged on a Leica SP8 upright confocal microscope. The antibody concentrations used were 1:1000 rabbit anti-GFP (Molecular Probes), 1:50 mouse anti-Repo (8D12, DSHB), 1:1000 Alexa488 anti-rabbit, and 1:1000 Alexa546 anti-Mouse (Invitrogen). All images were processed using Fiji software (Schindelin et al. 2012) and Adobe Photoshop (Adobe Systems).

CRISPR/Cas9-Mediated Genome Engineering

The *Df(2L)Dc* deletion lines were generated by first replacing the Dc region with a DsRed marker via homology directed repair, followed by the removal of the marker. The single-guide RNA (sgRNA) target sites were searched using the fly CRISPR Optimal Target Finder online tool (Gratz et al. 2014). Oligonucleotides corresponding to these sites were cloned into the pCFD3-dU6:3gRNA vector to make the sgRNA plasmids (Gratz et al. 2013). For building the donor plasmid containing the DsRed marker, ~1 kb regions flanking the Dc region (homology arms) were PCR-amplified from genomic DNA of the *y w vasa-Cas9* line with the Phusion High-Fidelity DNA polymerase (NEB), then cloned into the pHD-ScarlessDsRed vector (obtained from DGRC), using the In-Fusion HD cloning kit (Clontech). Sequences of the cloning primers for the homology arms and sgRNAs are available in the Supplementary Material online. A mix of two sgRNA plasmids and the donor plasmid (100 and 500 ng/μl, respectively) was injected into embryos of three different strains: *y w vasa-Cas9*, *y w vasa-Cas9*; CG9336^{CPT1001654}, and *y w vasa-Cas9*; CG9338^{CPT1100000}.

Emerging males were crossed with *w; amos^{Tf}/CyO wg-lacZ* flies and the F1 male progeny was screened for DsRed-positive individuals. Stocks carrying the insertions were established, using the *w; wg^{Gla1}/CyO twist-GAL4 UAS-GFP*, and *w; sna^{ScO}/CyO, PsChFP2* balancers. Removal of the DsRed cassette (which is flanked by PBac transposon ends) was accomplished using the *w; wg^{SP1}/CyO tub-PBac T* strain as a Piggy-Bac transposase source. Males carrying both the DsRed insertions and the transposase chromosome were crossed to *w; amos^{Tf}/CyO wg-lacZ* virgin females, and their progeny was screened for the loss of DsRed signal. Deletions were verified by PCR analysis of genomic DNA using flanking primers.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Patrícia Beldade for the critical reading of the manuscript and acknowledge Diogo Manoel, the Instituto Gulbenkian de Ciência Imaging Facility and the Toulouse RIO Imaging platform for their assistance in transgenics and imaging, Luisa Vasconcelos, Christen Mirth, and Ivo Telley for sharing reagents and species stocks. We also wish to thank the Bloomington Drosophila Stock Center (Indiana, USA), the Developmental Studies Hybridoma Bank (Iowa, USA), Drosophila Genomics Resource Center (Indiana, USA;

supported by NIH grant 2P0OD010949) for making fly stocks, antibodies, and vectors available. C.T. was supported by a contract of the Agence Nationale de la Recherche (ANR, France) ANR-13-ISV7-0001-01. K.T. was supported by Fundação para a Ciência e a Tecnologia (FCT, Portugal) post-doctoral fellowship (SFRH/BPD/75139/2010). This work was supported by Fundação Calouste Gulbenkian/Instituto Gulbenkian de Ciência, ANR (France) and FCT (Portugal) through ANR-13-ISV7-0001-01 to F.R., FCT-ANR/BIA-ANM/0003/2013 to E.S., and FCT-EXPL/BEX-GMG/2197/2013 to K.T.

References

- Arnold CD, Gerlach D, Spies D, Matts JA, Sytnikova YA, Pagani M, Lau NC, Stark A. 2014. Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nat Genet.* 46(7): 685–692.
- Arnoult L, Su KFY, Manoel D, Minervino C, Magrina J, Compel N, Prud'homme B. 2013. Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. *Science* 339(6126): 1423–1426.
- Assis R, Bachtrög D. 2013. Neofunctionalization of young duplicate genes in *Drosophila*. *Proc Natl Acad Sci U S A.* 110(43): 17409–17414.
- Barriere A, Ruvinsky I. 2014. Pervasive divergence of transcriptional gene regulation in *Caenorhabditis* Nematodes. *PLoS Genet.* 10(6): e1004435.
- Bass C, Field LM. 2011. Gene amplification and insecticide resistance. *Pest Manag Sci.* 67(8): 886–890.
- Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118(2): 401–415.
- Bruce AEE, Oates AC, Prince VE, Ho RK. 2001. Additional hox clusters in the zebrafish: divergent expression patterns belie equivalent activities of duplicate *hoxB5* genes. *Evol Dev.* 3(3): 127–144.
- Calhoun VC, Stathopoulos A, Levine M. 2002. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila Antennapedia* complex. *Proc Natl Acad Sci U S A.* 99(14): 9243–9247.
- Cannavò E, Khoueiry P, Garfield DA, Gleeher P, Zichner T, Gustafson EH, Ciglar L, Korbel JO, Furlong EEM. 2016. Shadow enhancers are pervasive features of developmental regulatory networks. *Curr Biol.* 26(1): 38–51.
- Carvajal JJ, Keith A, Rigby PWJ. 2008. Global transcriptional regulation of the locus encoding the skeletal muscle determination genes *Mrf4* and *Myf5*. *Genes Dev.* 22(2): 265–276.
- Chaudhari SS, Arakane Y, Specht C. a, Moussian B, Kramer KJ, Muthukrishnan S, Beeman RW. 2013. Retroactive maintains cuticle integrity by promoting the trafficking of Knickkopf into the procuticle of *Tribolium castaneum*. *PLoS Genet.* 9(1): e1003268.
- Cheng Y, Brunner AL, Kremer S, DeVido SK, Stefaniuk CM, Kassis JA. 2014. Co-regulation of *invected* and *engrailed* by a complex array of regulatory sequences in *Drosophila*. *Dev Biol.* 395(1): 131–143.
- Ding Y, Berrocal A, Morita T, Longden KD, Stern DL. 2016. Natural courtship song variation caused by an intronic retroelement in an ion channel gene. *Nature* 536(7616): 329–332.
- Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151(4): 1531–1545.
- Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* 466(7305): 490–493.
- Frankel N, Erezylmaz DF, McGregor AP, Wang S, Payre F, Stern DL. 2011. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* 474(7353): 598–603.
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32(Web Server): W273–W279.

- Galat A, Gross G, Drevet P, Sato A, Ménez A. 2008. Conserved structural determinants in three-fingered protein domains. *FEBS J.* 275(12): 3207–3225.
- Gout J-F, Lynch M. 2015. Maintenance and loss of duplicated genes by dosage subfunctionalization. *Mol Biol Evol.* 32(8): 2141–2148.
- Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM. 2013. Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics* 194(4): 1029–1035.
- Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, O'Connor-Giles KM. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* 196(4): 961–971. 2014.
- Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered.* 100(5): 605–617.
- He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169(2): 1157–1164.
- Hijazi A, Masson W, Augé B, Waltzer L, Haenlin M, Roch F. 2009. *Boudin* is required for septate junction organisation in *Drosophila* and codes for a diffusible protein of the Ly6 superfamily. *Development* 136(13): 2199–2209.
- Hittinger CT, Carroll SB. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449(7163): 677–681.
- Ito K, Urban J, Technau GM. 1995. Distribution, classification, and development of *Drosophila* glial cells in the late embryonic and early larval ventral nerve cord. *Roux Arch Dev Biol.* 204(5): 284–307.
- Iwaki DD, Lengyel JA. 2002. A Delta-Notch signaling border regulated by engrailed/invented repression specifies boundary cells in the *Drosophila* hindgut. *Mech Dev.* 114(1-2): 71–84.
- Jarinova O, Hatch G, Poitras L, Prudhomme C, Grzyb M, Aubin J, Berube-Simard F-A, Jeannotte L, Ekker M. 2008. Functional resolution of duplicated *hoxb5* genes in teleosts. *Development* 135(21): 3543–3553.
- Kalay G, Wittkopp PJ. 2010. Nomadic enhancers: tissue-specific *cis*-regulatory elements of yellow have divergent genomic positions among *Drosophila* species. *PLoS Genet.* 6(11): e1001222.
- Karageorgi M, Bräcker LB, Lebreton S, Minervino C, Cavey M, Siju KP, Grunwald Kadow IC, Gompel N, Prud'homme B. 2017. Evolution of multiple sensory systems drives novel egg-laying behavior in the fruit pest *Drosophila suzukii*. *Curr Biol.* 27(6): 847–853.
- Kassahn KS, Dang VT, Wilkins SJ, Perkins AC, Ragan MA. 2009. Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. *Genome Res.* 19(8): 1404–1418. doi: 10.1101/gr.086827.108.
- Katju V. 2013. To the beat of a different drum: determinants implicated in the asymmetric sequence divergence of *Caenorhabditis elegans* paralogs. *BMC Evol Biol.* 13: 73.
- Kim NC, Marqués G. 2012. The Ly6 neurotoxin-like molecule target of wit regulates spontaneous neurotransmitter release at the developing neuromuscular junction in *Drosophila*. *Dev Neurobiol.* 72(12): 1541–1558.
- Kleinjan DA, Bancewicz RM, Gautier P, Dahm R, Schonthaler HB, Damante G, Seawright A, Hever AM, Yeyati PL, van Heyningen V, et al. 2008. Subfunctionalization of duplicated zebrafish *pax6* genes by *cis*-regulatory divergence. *PLoS Genet.* 4(2): e29.
- Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schemhuber K, Dickson BJ, Stark A. 2014. Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. *Nature* 512(7512): 91–95.
- Lan X, Pritchard JK. 2016. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science* 352(6288): 1009–1013.
- Loehlin DW, Carroll SB. 2016. Expression of tandem gene duplicates is often greater than twofold. *Proc Natl Acad Sci U S A.* 113(21): 5988–5992.
- Long HK, Prescott SL, Wysocka J. 2016. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* 167(5): 1170–1187.
- Lowe N, Rees JS, Roote J, Ryder E, Armean IM, Johnson G, Drummond E, Spriggs H, Drummond J, Magbanua JP, et al. 2014. Analysis of the expression patterns, subcellular localisations and interaction partners of *Drosophila* proteins using a pigP protein trap library. *Development* 141(20): 3994–4005.
- Ludwig MZ, Patel NH, Kreitman M. 1998. Functional analysis of *eve* stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* 125(5): 949–958.
- Lynch M, Conery JS. 2003. The evolutionary demography of duplicate genes. *J Struct Funct Genomics.* 3(1-4): 35–44.
- Maeso I, Tena JJ. 2016. Favorable genomic environments for *cis*-regulatory evolution: a novel theoretical framework. *Semin Cell Dev Biol.* 57: 2–10.
- Nilton A, Oshima K, Zare F, Byri S, Nannmark U, Nyberg KG, Fehon RG, Uv AE. 2010. Crooked, coiled and crimped are three Ly6-like proteins required for proper localization of septate junction components. *Development* 137(14): 2427–2437.
- Obbard DJ, Maclennan J, Kim K-W, Rambaut A, O'Grady PM, Jiggins FM. 2012. Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Mol Biol Evol.* 29(11): 3459–3473.
- Ohno S. 1970. Evolution by gene duplication. Berlin, Germany: Springer-Verlag.
- Panganiban G, Sebring A, Nagy L, Carroll SB. 1995. The development of crustacean limbs and the evolution of arthropods. *Science* 270(5240): 1363–1366.
- Pegueroles C, Laurie S, Albà MM. 2013. Accelerated evolution after gene duplication: a time-dependent process affecting just one copy. *Mol Biol Evol.* 30(8): 1830–1842.
- Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, et al. 2007. Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* 39(10): 1256–1260.
- Pfeiffer BD, Ngo T-TB, Hibbard KL, Murphy C, Jenett A, Truman JW, Rubin GM. 2010. Refinement of tools for targeted gene expression in *Drosophila*. *Genetics* 186(2): 735–755.
- Prince VE, Pickett FB. 2002. Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet.* 3(11): 827–837.
- Qian W, Liao B-Y, Chang A-Y-F, Zhang J. 2010. Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends Genet.* 26(10): 425–430.
- Rebeiz M, Jikomes N, Kassner VA, Carroll SB. 2011. Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proc Natl Acad Sci U S A.* 108(25): 10036–10043.
- Rogers RL, Shao L, Thornton KR. 2017. Tandem duplications lead to novel expression patterns through exon shuffling in *Drosophila yakuba*. *PLoS Genet.* 13(5): e1006795.
- Ross J, Kuzin A, Brody T, Odenwald WF. 2015. *cis*-regulatory analysis of the *Drosophila* PDM locus reveals a diversity of neural enhancers. *BMC Genomics* 16(1): 1–20.
- Russo CA, Takezaki N, Nei M. 1995. Molecular phylogeny and divergence times of drosophilid species. *Mol Biol Evol.* 12(3): 391–404.
- Santos JG, Vömel M, Struck R, Homberg U, Nässel DR, Wegener C. 2007. Neuroarchitecture of peptidergic systems in the larval ventral ganglion of *Drosophila melanogaster*. *PLoS One* 2(8): e695.
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 9(7): 676–682.
- Schrider DR, Hahn MW. 2010. Gene copy-number polymorphism in nature. *Proc R Soc B Biol Sci.* 277(1698): 3213–3221.
- Stern DL, Crocker J, Ding Y, Frankel N, Kappes G, Kim E, Kuzmickas R, Lemire A, Mast JD, Picard S. 2017. Genetic and transgenic reagents for *Drosophila simulans*, *D. mauritiana*, *D. yakuba*, *D. santomea*, and *D. virilis*. *G3 (Bethesda)* 7: 1339–1347.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 30(12): 2725–2729.

- Tanaka K, Diekmann Y, Hazbun A, Hijazi A, Vreede B, Roch F, Sucena É. 2015. Multispecies analysis of expression pattern diversification in the recently expanded insect *Ly6* gene family. *Mol Biol Evol.* 32(7): 1730–1747.
- Tautz D, Pfeifle C. 1989. A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene hunchback. *Chromosoma* 98(2): 81–85.
- Thompson A, Zakon HH, Kirkpatrick M. 2016. Compensatory drift and the evolutionary dynamics of dosage-sensitive duplicate genes. *Genetics* 202(2): 765–774.
- Tsujimura T, Hosoya T, Kawamura S. 2010. A single enhancer regulating the differential expression of duplicated red-sensitive opsin genes in zebrafish. *PLoS Genet.* 6(12): e1001245.
- Wu MN, Joiner WJ, Dean T, Yue Z, Smith CJ, Chen D, Hoshi T, Sehgal A, Koh K. 2010. SLEEPLESS, a *Ly-6/neurotoxin* family member, regulates the levels, localization and activity of Shaker. *Nat Neurosci.* 13(1): 69–75.
- Wunderlich Z, Bragdon MDJ, Vincent BJ, White JA, Estrada J, DePace AH. 2016. *Krüppel* expression levels are maintained through compensatory evolution of shadow enhancers. *Cell Rep.* 14(12): 3030.