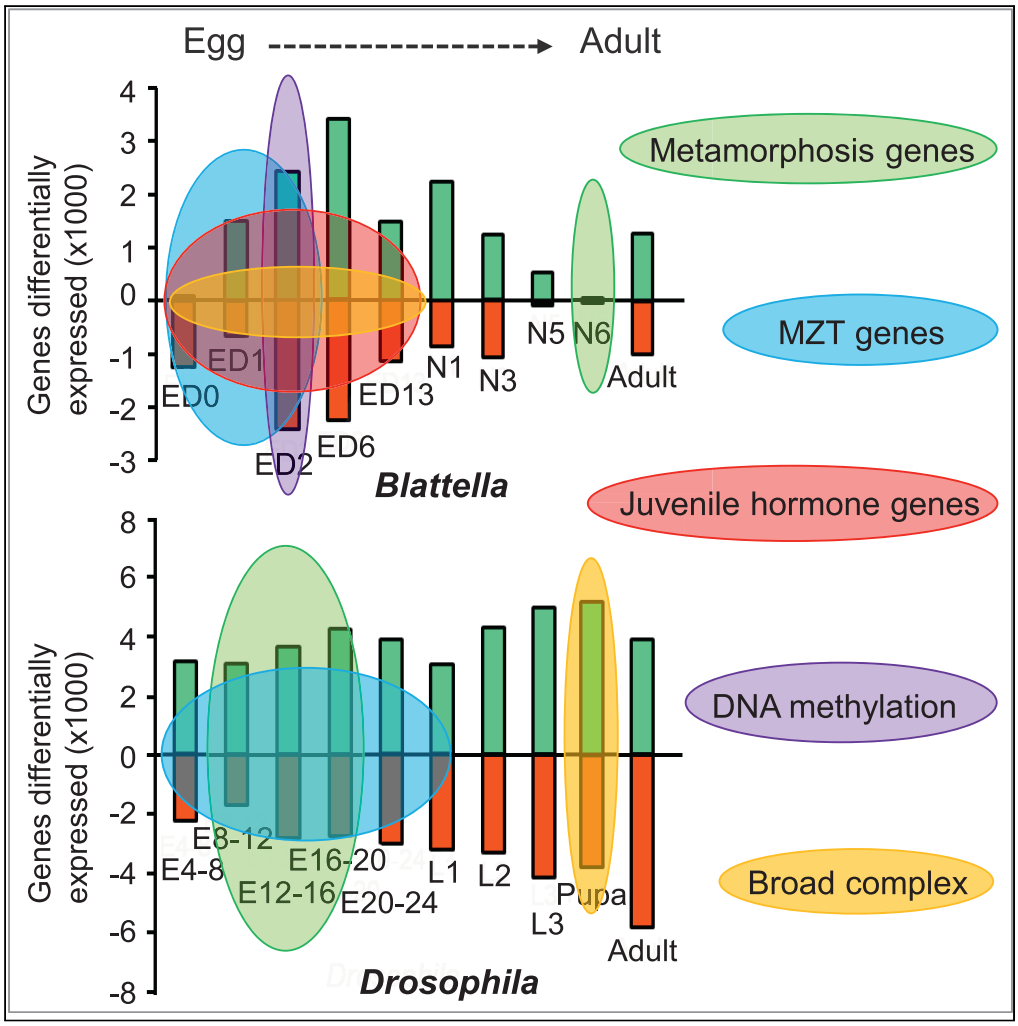
iScience

CellPress

Article

# Comparative Transcriptomics in Two Extreme Neopterans Reveals General Trends in the Evolution of Modern Insects




Guillem Ylla, Maria-Dolors Piulachs, Xavier Belles

xavier.belles@ibe.upf-csic.es

## HIGHLIGHTS

Transcriptomes of cockroaches and flies show key differences along development

Cockroaches and flies express metamorphosis factors with distinct timings in ontogeny

Cockroaches methylate DNA in early embryogenesis, whereas flies do not

MZT is limited to the early embryo in cockroaches, but it extends until hatching in flies

Ylla et al., iScience 4, 164–179 June 29, 2018 
https://doi.org/10.1016/j.isci.2018.05.017

iScience

CellPress

Article

# Comparative Transcriptomics in Two Extreme Neopterans Reveals General Trends in the Evolution of Modern Insects

Guillem Ylla,[1,2] Maria-Dolors Piulachs,[1] and Xavier Belles[1,3,*]

## SUMMARY

**The success of neopteran insects, with 1 million species described, is associated with developmental innovations such as holometaboly and the evolution from short to long germband embryogenesis. To unveil the mechanisms underlining these innovations, we compared gene expression during the ontogeny of two extreme neopterans, the cockroach *Blattella germanica* (polyneopteran, hemimetabolan, and short germband species) and the fly *Drosophila melanogaster* (endopterygote, holometabolan, and long germband species). Results revealed that genes associated with metamorphosis are predominantly expressed in late nymphal stages in *B. germanica* and in the early-mid embryo in *D. melanogaster*. In *B. germanica* the maternal to zygotic transition (MZT) concentrates early in embryogenesis, when juvenile hormone factors are significantly expressed. In *D. melanogaster*, the MZT extends throughout embryogenesis, during which time juvenile hormone factors appear to be unimportant. These differences possibly reflect broad trends in the evolution of development within neopterans, related to the germband type and the metamorphosis mode.**

## INTRODUCTION

With around 1 million species described, insects are the most diverse animal lineage on Earth. The extraordinary success of insects is due, at least in part, to their long evolutionary history, as they emerged some 450 million years ago (Mya) (Wang et al., 2016). This gave them enough time to evolve a series of key morpho-functional innovations that acted as drivers of expansion and diversification. A crucial innovation was the acquisition of wings, which took place about 410 Mya, with the emergence of the pterygote insects (Misof et al., 2014; Wang et al., 2016). A subsequent innovation was wing flexion over the dorsal body side (thus allowing a more efficient flight), which was achieved by neopteran insects some 380 Mya (Misof et al., 2014; Wang et al., 2016). Neopterans, or modern insects, represent more than 90% of the present insect species and have colonized all major terrestrial and freshwater habitats and exploited almost every organic resource, from dead plant and animal matter to all parts of green plants, and even to other kinds of animals, as predators or parasitoids. These ecological specializations have involved the corresponding adaptations, which has led to a formidable diversity in terms of morphology, physiology, and life cycles (Grimaldi and Engel, 2005).

From a developmental point of view, another key innovation that took place within neopteran evolution was metamorphosis (Nicholson et al., 2014), by which the individual acquires characteristic adult features and stops molting during postembryonic development. The ancestral metamorphosis mode was hemimetaboly, characterized by an embryogenesis that develops a first instar nymph displaying the essential adult body structure. The nymphs grow gradually, and the final molt to the adult stage completes the formation of functional wings and genitalia (Belles, 2011). From hemimetaboly emerged a metamorphosis mode known as holometaboly, in which the embryogenesis gives rise to a larva with a body structure considerably divergent from that of the adult, often more or less vermiform. The larva grows through various stages until molting to the pupal stage, which bridges the gap between the divergent larval morphology and that of the winged and reproductively competent adult (Belles, 2011). The holometabolan mode of metamorphosis was a successful innovation, as it was accompanied by an extraordinary radiation of the insect lineage (Misof et al., 2014). Indeed, more than 80% of currently known insect species follow the holometabolan metamorphosis (Condamine et al., 2016; Grimaldi and Engel, 2005).

Parallel innovations within the neopteran history occurred in embryogenesis, such as the evolution from short to long germband development. In long germband embryogenesis, the complete body segments

[1]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Passeig Marítim 37, 08003 Barcelona, Spain

[2]Present address: Department of Microbiology and Cell Science, Institute for Food and Agricultural Sciences, Genetics Institute, University of Florida, Gainesville, USA

[3]Lead Contact

*Correspondence: xavier.belles@ibe.upf-csic.es

https://doi.org/10.1016/j.isci.2018.05.017

(head, thoracic, and abdominal segments) are configured at the blastoderm stage. In short germband embryogenesis, the head lobes and the most anterior trunk segments are configured first and new segments are subsequently added from the posterior terminus. Less modified neopteran groups, mainly polyneopterans and paraneopterans, follow the short germband embryogenesis, whereas the more modified endopterygotes follow the long germband embryogenesis, in general (Chipman, 2015; Liu and Kaufman, 2005). Another process that evolved along neopteran history is blastokinesis, the movement of the embryo into the yolk mass that usually results in a partial revolution of the embryonic body (Panfilio, 2008). Blastokinesis occurs around mid-embryogenesis and is typical of short germband, hemimetabolan insects, whereas similar movements in long germband, holometabolan species are oversimplified or practically absent (Panfilio, 2008).

Most of the information regarding the detailed mechanisms regulating development has been described in the fruit fly *Drosophila melanogaster*, the model *par excellence* for genetic studies since the end of the 19$^{th}$ century (Markow, 2015). *D. melanogaster* is a singularly modified, endopterygote, holometabolan species, which shows long germband embryogenesis, practically without blastokinesis (Campos-Ortega and Hartenstein, 1985). The genome of *D. melanogaster* was the first to be sequenced among insects (Adams et al., 2000), but the availability of insect genomes has notably increased in recent years (I5K-Consortium, 2013). This allows entire genome comparisons, which may help to understand the genetic basis of given developmental innovations (see, for example, Harrison et al., 2018). However, developmental innovations largely evolve by altering the expression of functionally conserved genes, not by the emergence of new genes (Carroll, 2008). Therefore, comparative transcriptomics appears to be the most suitable approach to analyze the origin and evolution of developmental innovations. Again, the champion model concerning transcriptomic information is *D. melanogaster*, for which abundant high-throughput sequencing data are available, such as those generated in the modENCODE project (Celniker et al., 2009; modENCODE Consortium et al., 2010).

In contrast, transcriptomic data available in other insects are much less abundant and dispersed in a few species. Obviously, the heavy focus on *D. melanogaster* is a serious drawback if we aim at understanding the general trends of the evolution of development in insects through comparative transcriptomics. To partially fill this gap, we have produced extensive transcriptomic data along the ontogeny of the German cockroach, *Blattella germanica*, a polyneopteran hemimetabolan species, which shows short germband embryogenesis practically without blastokinesis (Tanaka, 1976), whose genome has recently been sequenced (Harrison et al., 2018). We have produced and sequenced 22 mRNA libraries from 11 developmental stages (two replicates each) covering the entire ontogeny: embryogenesis, nymphal stages, and the adult female. In total, we obtained 193,014,748 read pairs, which are now available to the scientific community. The study of these transcriptomes in *B. germanica* allowed describing the molecular basis of the main developmental transitions in this species. Then, searching in public databases, we found a most comprehensive RNA-seq dataset of *D. melanogaster* that comprises 22 libraries from 11 developmental stages (two replicates each) covering the entire ontogeny: embryogenesis, larval stages, the pupa, and the adult female (Celniker et al., 2009; modENCODE Consortium et al., 2010), with 129,507,378 read pairs in total (available at GEO: GSE18068). Then, we compared the respective ontogenetic sets of transcriptomes of *B. germanica* and *D. melanogaster*, with the idea of identifying differences among these two phylogenetically distant species that could illuminate broad trends in the evolution of development in neopteran insects.

## RESULTS AND DISCUSSION

### General Transcriptomic and Genomic Data

In *B. germanica*, the analyses were based on 22 mRNA libraries that were prepared in our laboratory, representing the following 11 stages (two replicates each): non-fertilized egg (NFE); 8, 24, 48, 144, and 312 hr after oviposition (ED0, ED1, ED2, ED6, and ED13); first, third, fifth, and sixth (last) nymphal instars (N1, N3, N5, and N6); and adult female (Table S1). In total, we obtained 198,970,437 read pairs (data from the 22 libraries accessible at GEO: GSE99785). After removing the adapters, filtering low-quality reads with FastQC (version 0.11.4) (Andrews, 2010), and merging read pairs, we obtained 193,014,748 read pairs (corresponding to 97.0% of the total sequenced read pairs) (Table 1), 66.8% of which mapped to the *B. germanica* genome.

The RNA-seq dataset of *D. melanogaster* used in the comparisons (GEO: GSE18068) comprises 22 libraries from 11 developmental stages (two replicates each) covering the entire embryo development

| Library | Raw Reads x2 | Clean Reads x2 | % | Mapped Read x2 | % |
|---|---|---|---|---|---|
| NFE | 14,413,472 | 13,736,963 | 95.31 | 11,055,040 | 80.48 |
| NFE_2 | 12,845,235 | 12,715,450 | 98.99 | 9,327,382 | 73.35 |
| ED0 | 4,664,861 | 4,349,066 | 93.23 | 3,348,444 | 76.99 |
| ED0_2 | 13,374,354 | 13,192,165 | 98.64 | 9,187,091 | 69.64 |
| ED1 | 2,530,147 | 2,339,451 | 92.46 | 1,705,713 | 72.91 |
| ED1_2 | 9,471,555 | 9,027,804 | 95.31 | 6,237,668 | 69.09 |
| ED2 | 8,023,009 | 7,652,728 | 95.38 | 5,603,324 | 73.22 |
| ED2_2 | 21,825,389 | 20,729,119 | 94.98 | 10,263,577 | 49.51 |
| ED6 | 8,659,285 | 8,427,098 | 97.32 | 6,894,844 | 81.82 |
| ED6_2 | 10,904,515 | 10,667,152 | 97.82 | 6,069,177 | 56.90 |
| ED13 | 10,364,701 | 10,015,047 | 96.63 | 7,892,059 | 78.80 |
| ED13_2 | 8,523,247 | 8,031,716 | 94.23 | 5,656,165 | 70.42 |
| N1 | 6,418,772 | 6,372,912 | 99.29 | 4,392,741 | 68.93 |
| N1_2 | 6,429,421 | 6,292,351 | 97.87 | 4,062,721 | 64.57 |
| N3 | 7,207,614 | 7,129,383 | 98.91 | 4,537,197 | 63.64 |
| N3_2 | 7,564,063 | 7,458,733 | 98.61 | 5,230,704 | 70.13 |
| N5 | 5,403,000 | 5,369,336 | 99.38 | 3,264,330 | 60.80 |
| N5_2 | 7,655,560 | 7,509,961 | 98.10 | 5,012,886 | 66.75 |
| N6 | 9,037,587 | 8,967,871 | 99.23 | 5,313,344 | 59.25 |
| N6_2 | 7,237,541 | 7,007,056 | 96.82 | 4,024,529 | 57.44 |
| Adult | 8,484,768 | 8,412,450 | 99.15 | 5,319,340 | 63.23 |
| Adult_2 | 7,932,341 | 7,610,936 | 95.95 | 4,609,493 | 60.56 |
| TOTAL | 198,970,437 | 193,014,748 | 96.98 | 129,007,769 | 67.66 |

**Table 1. Summary of the Reads Obtained from the Sequenced RNA-seq Libraries of *Blattella germanica***

For each library we show the number of read pairs sequenced, the number and percentage of reads after cleaning low-quality reads with Trimmomatic, and the number and percentage of clean reads mapped to the *B. germanica* genome (PRJNA427252).

(six sequential stages: 0–4 hr, 4–6 hr, 6–12 hr, 12–16 hr, 16–20 hr, 20–24 hr), the three larval stages (L1, L2, L3), the pupa, and the adult female. In postembryonic stages, we followed the correspondence *B. germanica* pre-last nymphal instars with *D. melanogaster* larvae, the last nymphal instar with the pupa (Belles and Santos, 2014), and the respective adult female stages. Correspondences between the embryo stages of *D. melanogaster* and *B. germanica* are summarized in Table S2. The analysis of the above-mentioned *D. melanogaster* libraries gave 129,507,378 read pairs, 95.2% of which mapped to the *D. melanogaster* genome (Table 2).

We detected expression (>1 FPKM) for 90.1% of the annotated genes of *B. germanica* (25,643 out of 28,471) and 97.3% of *D. melanogaster* (17,004 out of 17,471). To facilitate comparisons, we obtained the set of orthologous genes shared by the two species. We retrieved the protein sequences from the 28,471 annotated genes of *B. germanica* and 17,471 annotated genes of *D. melanogaster* and identified 7,169 orthologous genes common to *B. germanica* and *D. melanogaster* following the best blast reciprocal hit approach. These 7,169 orthologues correspond to 25.2% of the *B. germanica* genes and 41.0% of those from *D. melanogaster*.

| Accession number | Name | Reads | Reads Mapping to the Genome | % |
|---|---|---|---|---|
| SRR030232 | E0-4_1 | 3,433,652 | 3,292,953 | 95.90 |
| SRR030233 | E0-4_2 | 4,093,252 | 3,989,272 | 97.46 |
| SRR030238 | E4-8_1 | 2,822,374 | 2,658,024 | 94.18 |
| SRR030239 | E4-8_2 | 3,800,699 | 3,689,337 | 97.07 |
| SRR030236 | E8-12_1 | 5,197,055 | 4,947,643 | 95.20 |
| SRR030237 | E8-12_2 | 5,146,028 | 4,958,807 | 96.36 |
| SRR030226 | E12-16_1 | 4,908,119 | 4,665,060 | 95.05 |
| SRR030227 | E12-16_2 | 3,829,586 | 3,705,026 | 96.75 |
| SRR030234 | E16-20_1 | 9,322,851 | 8,456,435 | 90.71 |
| SRR030235 | E16-20_2 | 6,222,965 | 5,676,767 | 91.22 |
| SRR030240 | E20-24_1 | 3,488,824 | 3,236,901 | 92.78 |
| SRR030241 | E20-24_2 | 5,442,437 | 5,200,659 | 95.56 |
| SRR030242 | L1_1 | 9,611,846 | 8,992,150 | 93.55 |
| SRR030243 | L1_2 | 6,504,722 | 6,110,616 | 93.94 |
| SRR030248 | L2_1 | 9,327,073 | 8,888,182 | 95.29 |
| SRR030249 | L2_2 | 11,399,101 | 10,962,849 | 96.17 |
| SRR030244 | L3_1 | 3,667,132 | 3,529,505 | 96.25 |
| SRR030245 | L3_2 | 9,330,126 | 9,143,635 | 98.00 |
| SRR030246 | Pupae_1 | 2,616,980 | 2,501,962 | 95.60 |
| SRR030247 | Pupae_2 | 8,983,652 | 8,704,123 | 96.89 |
| SRR030230 | Adult-Female_1 | 2,143,390 | 2,038,572 | 95.11 |
| SRR030231 | Adult-Female_2 | 8,215,514 | 7,927,427 | 96.49 |
| TOTAL | | 129,507,378 | 123,275,905 | 95.25 |

**Table 2. Summary of the *Drosophila melanogaster* RNA-seq Libraries Obtained from GEO: GSE18068**
For each library we show the number of reads and the number and percentage of reads mapped to the *D. melanogaster* genome from Flybase (version dmel_r6.12).

## General Gene Expression

The expression of all genes (Figure 1A) suggests that the duplicates of each stage-library behave similarly in *B. germanica* and in *D. melanogaster*. Moreover, principal component analysis (PCA) of the expression data of all libraries indicates that the replicates of each library group together (Figure 1B), which led us to represent the replicates joined in further figures. The PCA shows that all stages are well separated from each other, except N5 and N6 in *B. germanica*, and E0-4 and E4-8 in *D. melanogaster*, which are closely related. In *B. germanica*, the general expression (Figure 1A) indicates that many genes are more abundantly expressed during embryogenesis, whereas only a relatively small set is expressed at significant amounts in postembryonic stages. In *D. melanogaster*, the distinction between embryonic and postembryonic stages in terms of the abundance of gene expression is more diffuse. Characteristically, quite a high number of genes are highly expressed in the pupa and the adult (Figure 1A).

In *B. germanica*, the differential expression analysis between stages reveals that the most dynamic changes occur during embryogenesis (Figure 2A). In contrast, the number of gene expression changes is maintained

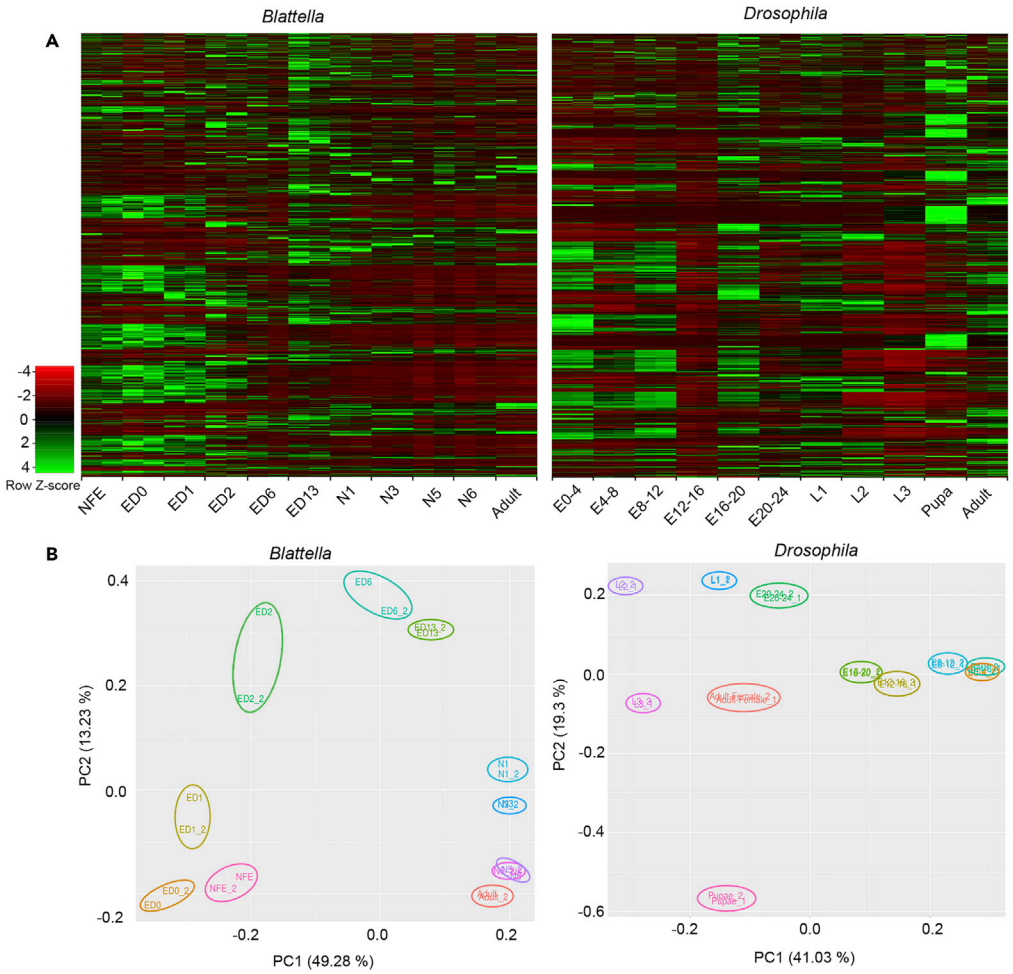



**Figure 1. Overall Gene Expression in the Stage-Libraries of *Blattella germanica* and *Drosophila melanogaster***

(A) Heatmap showing the expression of all genes (FPKM) in each of the stage-libraries.

(B) Principal component analysis plot showing the distribution of the two replicates of the stage-libraries.

at similar levels in all transitions in *D. melanogaster* (Figure 2A). These differences may be related to metamorphosis, given that in the hemimetabolan mode (*B. germanica*) the basic adult body structure is formed during embryogenesis. In contrast, in holometabolan species (*D. melanogaster*) the adult morphology is completed in postembryonic stages, around the pupal stage.

The GO-terms enrichment analyses of the expressed genes reveal different biological functions at different stages within the same species and general differences between *B. germanica* and *D. melanogaster* (Figures 2B and S1). In the embryo stages, the results indicate that both species are enriched in functions related to "cell cycle control," "gene expression," and "translation," suggesting an active transcriptional activity and cell proliferation, as expected in this developmental period. Functions associated with epigenetic control, such as "chromatin organization," are also enriched, but "DNA methylation" is enriched in the early *B. germanica* embryo but not in *D. melanogaster*. Functions related to adult morphogenesis, such as "metamorphosis," "imaginal disc development," and "pupal development," are enriched in the early-mid embryo in *D. melanogaster* and in late nymphal instars in *B. germanica*. This is consistent with the respective holometabolan and hemimetabolan metamorphosis mode of these species. In postembryonic development, we observed a clear enrichment in genes related to "cuticle development" in *D. melanogaster* pupae and *B. germanica* nymphs. In both species, the adult stage is enriched in genes related to homeostasis, such as metabolism, catabolism, and immune defense functions (Figure 2B).

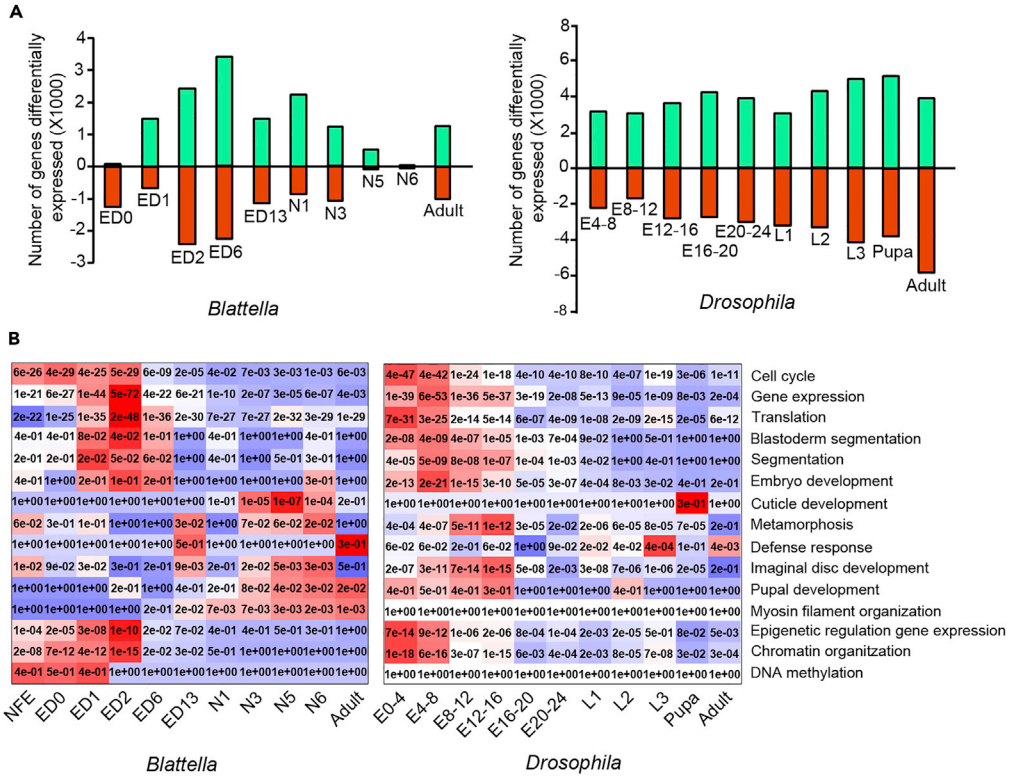



**Figure 2. Differential Expression Analysis and GO Terms from Enrichment Analysis**

(A) Number of genes significantly ($p < 0.05$) upregulated (green) and downregulated (red) according to the differential expression analysis between consecutive stage-libraries of *Blattella germanica* and *Drosophila melanogaster*.

(B) Selection of GO terms of biological processes from the enrichment analysis performed with the expressed genes at each stage in *B. germanica* and *D. melanogaster*; for each selected GO term the p value of the hypergeometric test is shown, and the color scale goes from red (low p value) to blue (high p value) normalized in each row.

The GO-enrichment analysis (which could suffer a bias because in *B. germanica* the GO terms are assigned on the best hit in *D. melanogaster*) is in agreement with the less informative but bias-free Pfam motifs enrichment analysis (Figure S2). Characteristically, Pfam motifs involved in metabolism and catabolism, such as those associated with peptidases, amylase, and hydrolases, and Pfam motifs related to immune defense response, such as "Defensin_2," are enriched in genes expressed in the adult. In contrast, embryos express genes with motifs associated with the regulation of gene expression, such as Zn-finger or Homeobox genes.

## Genes Mainly Associated with Embryogenesis

We paid special attention to maternally loaded transcripts, genes involved in the maternal to zygotic transition (MZT), genes driving the early embryo patterning, and Hox genes.

The NFE libraries of *B. germanica* contain maternally loaded mRNAs enriched for functions related to "cell cycle" and "embryo development" (Figures 2B and S1), as could be expected. They are also enriched for epigenetic functions ("epigenetic regulation of gene expression," "DNA methylation," and "chromatin organization"), but these GO terms do not appear in the earliest stage-libraries of *D. melanogaster* (Figures 2B and S1). Especially intriguing are the genes with the GO terms "metamorphosis" and "wing disc development" occurring in the NFE library of *B. germanica*. These include genes involved in the formation of bristles (*hairless*, *spineless*), legs (*croocked legs*, *rotund*, *spineless*, *vulcan*), antennae (*rotund*, *spineless*), and compound eyes (*Tartan*, *Hyperplastic discs*, *eyes absent*, *rotund*). The function of maternal transcripts with these GO terms is enigmatic but might be related to the hemimetabolan metamorphosis of *B. germanica*.

In the MZT transition of *D. melanogaster*, important genes are *smaug* (*smg*), associated with the elimination of maternal transcripts (Benoit et al., 2009; Chen et al., 2014; Tadros et al., 2007), and *zelda* (*zld*),

involved in the activation of the zygotic genome (Foo et al., 2014; Liang et al., 2008; Nien et al., 2011; Schulz et al., 2015; Sun et al., 2015). Moreover, Zelda promotes the expression of Mir-309 microRNAs (Fu et al., 2014) that, in turn, contribute to eliminating maternal mRNAs (Bushati et al., 2008). In *B. germanica*, *smg* shows an expression peak in ED0, whereas *zld* peaks in ED1, in both cases followed by an abrupt expression decrease, keeping low values during the remaining ontogeny. In contrast, *smg* and *zld* are consistently expressed during all embryogenesis and in the first larval instar of *D. melanogaster* (Figure 3A). In *D. melanogaster*, *smg* expression has been studied in terms of protein by western blot along the embryogenesis by Smibert et al. (1999), who observed a signal only in the first 3 hr of embryo development. However, the signal shown is very tenuous, which casts doubts about the possibility that a higher protein load would have allowed detecting signal at late embryogenesis. Subsequent works present *smg* western blot analyses only for the first 3–4 hr of embryogenesis (Benoit et al., 2009; Dahanukar et al., 1999). Regarding *zld* in *D. melanogaster*, northern blot analyses had shown that expression appears to be quite high in the embryo, L1 and L2, then decreases in L3 and the pupa, and slightly increases in the adult (Staudt et al., 2006), which is fairly coincident with the reads-based pattern obtained by us (Figure 3A). In *B. germanica* we sought to validate the reads-based pattern with qRT-PCR measurements. Interestingly, the obtained qRT-PCR profile showed a strong and significant correlation with the reads-based pattern (Pearson correlation of 0.904 with a p value = 0.00013) (Figure 3B).

In *B. germanica*, the expression of *smg* is compatible with the role of eliminating maternal transcripts and that of *zld* is compatible with a stimulatory role on Mir-309 microRNAs expression, which, according to Ylla et al. (2017), peaks on ED2. In *B. germanica*, *smg*, *zld*, and Mir-309 show a narrow window of expression between ED0 and ED2, framing the MZT within the first 12% of embryogenesis. In *D. melanogaster*, *smg* and *zld* maintain quite high levels of expression throughout embryogenesis and even the first larval instar (Figure 3A). This continued expression of *smg* and *zld* in *D. melanogaster* that look like an "extended" MZT, is consistent with the stable expression changes along all embryogenesis (Figure 2A), and might be related to the formation of the morphologically divergent holometabolan larva.

The functional enrichment analysis (Figure 2B) suggested that DNA methylation operates during the MZT in *B. germanica* embryos, whereas this is not the case with *D. melanogaster*. Thus, we examined the CpG depletion (CpGo/e, observed versus expected number of CpGs), which is a reliable predictor of DNA methylation (Bewick et al., 2016). The comparison of CpGo/e with gene expression in the 11 stage-libraries of *B. germanica* revealed a significant negative correlation between both parameters in ED0, ED1, and ED2 stages. The genes overexpressed in these stages, covering the MZT, had the lowest levels of CpGo/e (Figure 3C). Moreover, *Dnmt1*, a gene coding a DNA methyltransferase (Lyko, 2018), is predominantly expressed in these same stages (Figure 3D). Interestingly, the expression of *Dnmt2*, whose gene product catalyzes tRNA methylation (Goll et al., 2006), also peaks in very early embryo development (Figure 3D). In *D. melanogaster*, the expression of *Dnmt2* shows a peak around mid-embryogenesis and then a significant expression is kept all along ontogeny (Figure 3D), a pattern that is in agreement with previous northern and western blot studies (Kunert et al., 2003; Lyko et al., 2000). *D. melanogaster* does not have *Dnmt1*, which is consistent with data suggesting that DNA methylation is quantitatively irrelevant in dipterans (Marhold et al., 2004), although limited DNA methylation has been observed to occur in short motifs, independent of Dnmt2 (Takayama et al., 2014).

Our data suggest that a discrete wave of DNA methylation promotes temporal expression of a set of genes during the MZT of *B. germanica* that might be necessary for the zygotic activation. DNA methylation is currently associated with a repressed chromatin state and inhibition of gene expression, although in some instances it can also have an activating effect (Siegfried and Simon, 2010). In insects, levels of DNA methylation are much higher in the hemimetabolan than in the holometabolan species (Bewick et al., 2016), although DNA methylation appears to be important for caste differentiation in holometabolan, social bees (Elango et al., 2009), and ants (Kay et al., 2018), as well as in hemimetabolan termites (Harrison et al., 2018). Among the holometabolan groups, DNA methylation is limited and atypical in *D. melanogaster* (Dnmt1- and Dnmt3-independent) (Elango et al., 2009; Marhold et al., 2004; Takayama et al., 2014). In the beetle *Tribolium castaneum*, which possesses Dnmt1 and Dnmt2, but not Dnmt3, two types of DNA methylation exist: conserved CpG methylation catalyzed by Dnmt1 and non-CpG methylation, which shows high similarity to *D. melanogaster* methylation and would be catalyzed by still unknown methyltransferases (Feliciello et al., 2013; Song et al., 2017). The contrast between the high and low levels of DNA methylation in hemimetabolan and holometabolan species, respectively (Bewick

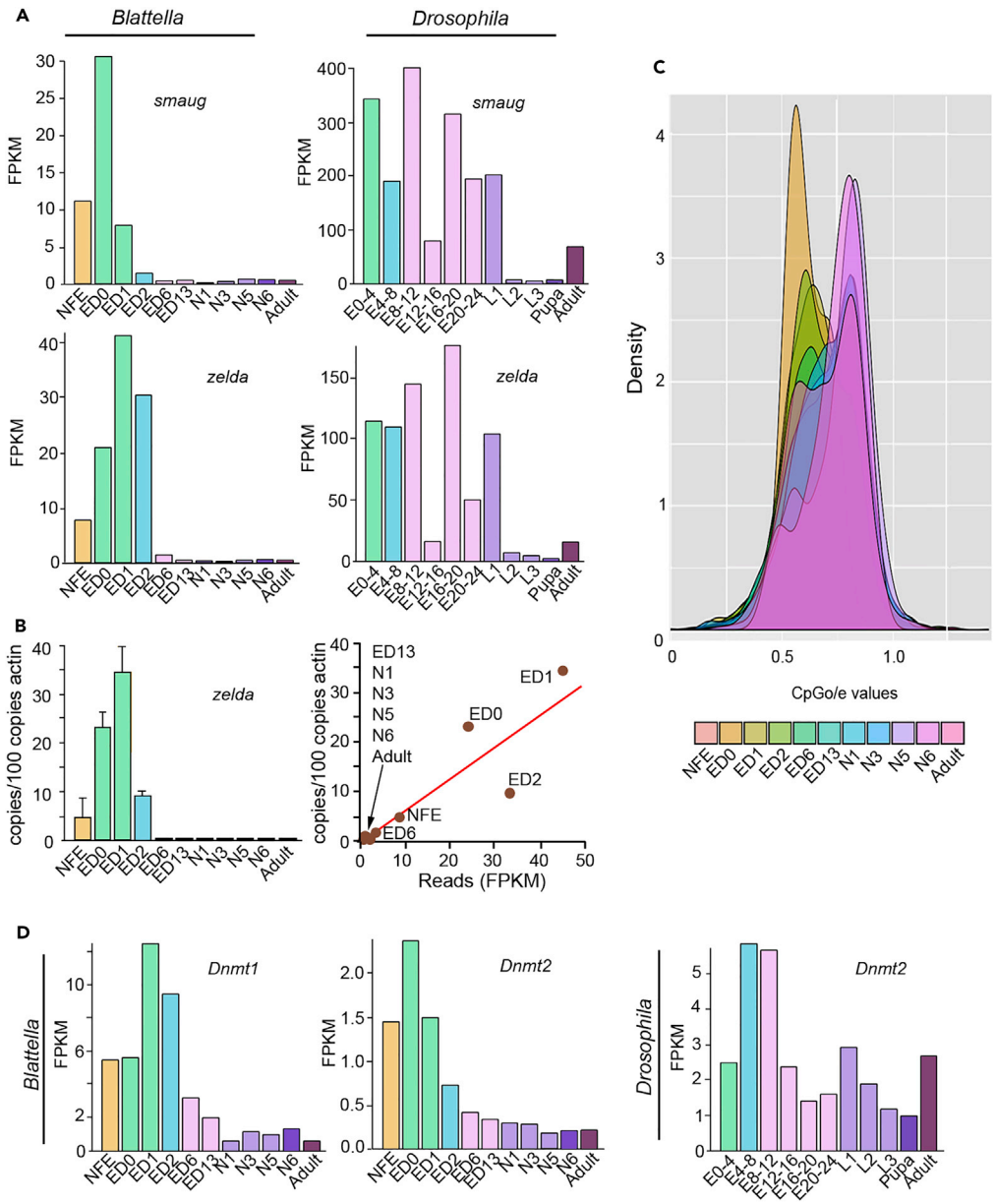



**Figure 3. Gene Expression and Methylation during the Maternal to Zygotic Transition in *Blattella germanica* and *Drosophila melanogaster***

(A) Reads-based expression of *smaug* and *zelda* along the different stage-libraries.

(B) Left: qRT-PCR-based expression of *zelda* along the same stages in *B. germanica*; each value represents three biological replicates and it is represented as copies of *zelda* mRNA per 100 copies of BgActin-5c mRNA (mean ± SEM). Right: Representation of the FPKM and qRT-PCR values of expression of *zelda* in the stages studied and the regression line obtained.

(C) The CpGo/e distribution of the differentially expressed genes in each stage-library of *B. germanica*.

(D) Expression of DNA methyltransferase *Dnmt1* and the tRNA methyltransferase *Dnmt2*, along the different stage-libraries. In *A*, *B*, and *D*, identical bar colors indicate equivalent developmental periods, according to the criteria summarized in Table S2.

et al., 2016), and our observations in *B. germanica* suggest that DNA methylation operates in early embryo development of hemimetabolan species, contributing to the zygote gene activation in the MZT. We propose that this is an ancestral feature in neopteran insects, whose functional relevance may have been progressively lost in holometabolan species (see also Bewick et al., 2016).

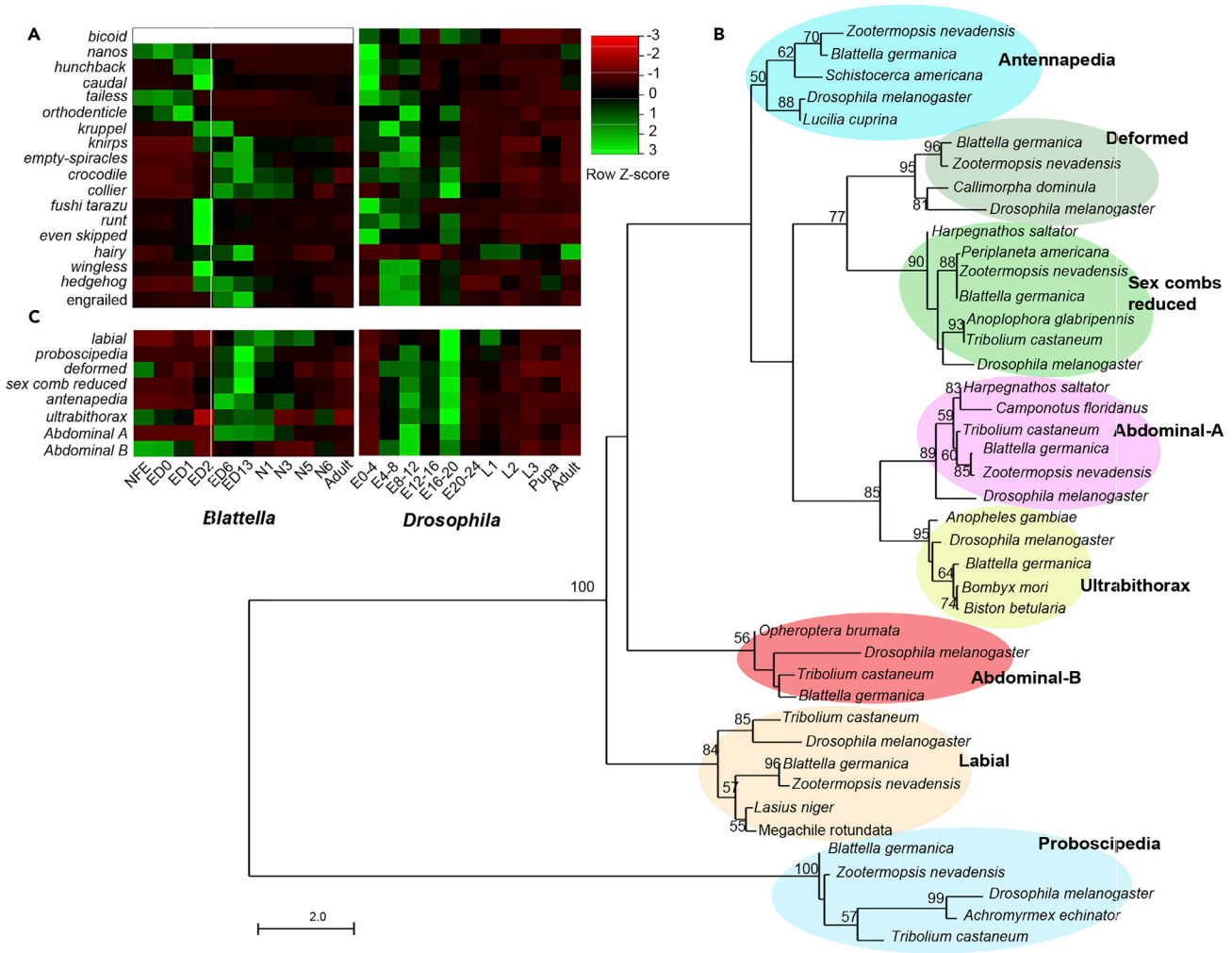



**Figure 4. Expression of Early Patterning Genes, Hox Genes, Transcription Factors (TFs), and Genes Related to Hormonal Biosynthesis and Signaling along the Different Stage-Libraries of *Blattella germanica* and *Drosophila melanogaster***

(A) Heatmap showing the expression of maternal, gap, pair-rule, and segmentation genes; *bicoid* has no orthologs in *B. germanica*.

(B) Phylogenetic relationships of the Hox proteins of *Blattella germanica* with those of other insect species. Bootstrap values > 50 are indicated in the corresponding nodes. Scale bar indicates the number of substitutions per site.

(C) Heatmap showing the expression of Hox genes. In (A) and (C), the expression is indicated in FPKM.

Concerning early embryo patterning, we examined the expression of the most representative gap, pair-rule, and segment polarity genes, which determine the general polarity of the embryo (Peel et al., 2005). The most obvious difference between *B. germanica* and *D. melanogaster* in very early embryogenesis is the absence of *bicoid* in the former species, as this gene is exclusive to higher dipterans (Schröder, 2003). In *B. germanica*, the gap-gene cascade is initiated by maternal *tailless*, followed by *orthodenticle*, *huckbein*, and *Krüppel* (Figure 4A). In general, the expression patterns are similar in both species, showing approximately the cascade of maternal, gap, pair-rule, and segment polarity genes. Only *hairy* (*h*) exhibits a neatly divergent pattern, being predominantly expressed in mid-late embryogenesis in *B. germanica* and in postembryonic stages in *D. melanogaster* (Figure 4A). In *D. melanogaster*, *h* acts as a pair-rule in early embryo development, whereas in larvae, by binding to the protein Achaetae, regulates the patterning of sensory organs in the developing wings and legs (Fisher and Caudy, 1998). Through other mechanisms, *h* might also contribute to regulating the morphogenetic furrow in the developing eye (Bhattacharya and Baker, 2012). The latter functions explain the expression that we observed in *D. melanogaster* larvae, and we speculate that the high level of expression in the mid-late embryo of *B. germanica* might be due to the formation of nymphal structures, such as proper chaetotaxy and compound eyes.

Subsequently, we examined the Hox genes, which play key roles in morphogenesis and body structure shaping (Averof and Akam, 1995). We identified the eight canonical Hox genes in the *B. germanica* genome (Figure 4B) and observed that most of them are fully expressed in the mid-late embryo, such as in *D. melanogaster*, when dorsal closure occurs (Figure 4C). The main difference between the two species is *Abdominal-B* (*Abd-B*), which, in *B. germanica*, shows the highest transcript levels in NFE and ED0. The function of the high maternal load of *Abd-B* is enigmatic, but the low expression levels in the mid-late embryo might have to do with dorsal closure. In *D. melanogaster*, mixer cell remodeling regulates tension along the leading edge during dorsal closure. *Abdominal-A* (*abd-A*) is a pro-mixing factor in the first five abdominal segments, whereas *Abd-B* represses mixing in posterior segments. At late closure in the central segments, the tension increases and *abd-A* is not repressed by *Abd-B* in these segments (Roumengous et al., 2017). If *abd-A* and *Abd-B* played the same role in *B. germanica*, then the low expression of *Abd-B* would suggest that the pro-mixing action of *abd-A* is needed all along the leading edge during dorsal closure.

### Transcription Factors

Important players in gene regulatory networks are transcription factors (TFs) (de Mendoza et al., 2013). To study them, we performed a PfamScan search in annotated proteins, which gave 17,196 PFAM-A motifs (4,280 unique) associated with 12,789 *B. germanica* genes and 15,475 (4,339 unique) associated with 10,759 *D. melanogaster* genes. Among these, we identified 600 genes in *B. germanica* and 458 in *D. melanogaster* that contained at least one Pfam motif associated with a TF function (de Mendoza et al., 2013; Ylla and Belles, 2015). Most of these TFs are differentially expressed during embryogenesis of both species, and many of them are also highly expressed in the pupal and adult stages of *D. melanogaster* (Figure 5A). To identify comparable differences between the two species, we retrieved the subset of orthologous TF genes common to *B. germanica* and *D. melanogaster*, obtaining 297 genes shared by the two species (Data S1). The expression of these 297 genes in *B. germanica* and *D. melanogaster* (Figure S3) reminds that observed when representing all genes (Figure 1A) or all TFs (Figure 5A), with many TF genes more abundantly expressed during embryogenesis in *B. germanica*, whereas in *D. melanogaster* the diversity of expression is more similar in embryonic and postembryonic stages.

A selection of TF genes that display a greater contrast in expression between species and stages is shown in Figure 5B. We can see, for example, that *vismay*, *SREBP* (sterol regulatory element binding protein), and *HLH3B* (helix loop helix protein 3B) are specifically highly expressed in the very early embryonic stages (ED0 and ED1) of *B. germanica*. High expression of *SREBP* suggests that lipogenesis and lipid homeostasis (Shao and Espenshade, 2012) are important in these stages in *B. germanica*. Conversely, *SREBP* appears to be not as relevant in *D. melanogaster*, but others, such as *lateral muscles scarcer*, involved in the development of embryonic lateral transverse muscles (Müller et al., 2010), are highly expressed in early embryo development. In mid-late embryogenesis, *lethal of scute*, a gene involved in the neurogenesis and specification of sensory organs (Negre and Simpson, 2015), is highly expressed in *B. germanica*, whereas *shuttle craft*, required to maintain the proper morphology of motoneuronal axon nerve routes (Stroumbakis et al., 1996), is highly expressed in *D. melanogaster*. Also typical of late embryo stages of *D. melanogaster* is the high level of expression of *little imaginal discs*, a histone demethylase that specifically removes H3K4me3, a mark associated with active transcription (Li et al., 2010), and *cropped*, a gene essential for embryonic tracheal terminal branching (Wong et al., 2015). The aforementioned expression divergences refer to genes with no relevant roles in general patterning or organogenesis and could respond to circuitries specific of *B. germanica* and *D. melanogaster*, rather than being considered as reflecting general trends of evolution of development.

With respect to postembryonic stages, *unplugged* (*unpg*), required for the formation of specific tracheal branches (Chiang et al., 1995), and *senseless* (*sens*), crucial for the peripheral nervous system development (Nolo et al., 2000), are typically highly expressed in *B. germanica* nymphs, whereas in *D. melanogaster*, expression of these genes appears to be insignificant. Conversely, *Sox14,* required for 20E signaling at the onset of metamorphosis (Ritter and Beckstead, 2010), and *Relish* (*Rel*), which promotes the transcription of innate immune response genes (Petersen et al., 2013), are characteristically expressed in *D. melanogaster* larvae. The pupa of *D. melanogaster* continues expressing *Sox14* and *Rel* at high levels. These expression divergences may reflect the different development of cockroach nymphs and fly larvae and pupae. Compared with fly larval growth, the development of cockroach nymphs involves a considerable increase in size; thus, growth of the tracheal and the peripheral nervous systems promoted by *unpg* and *sens* makes sense in this context. The expression of *Sox14* in larvae and pupae might be related to

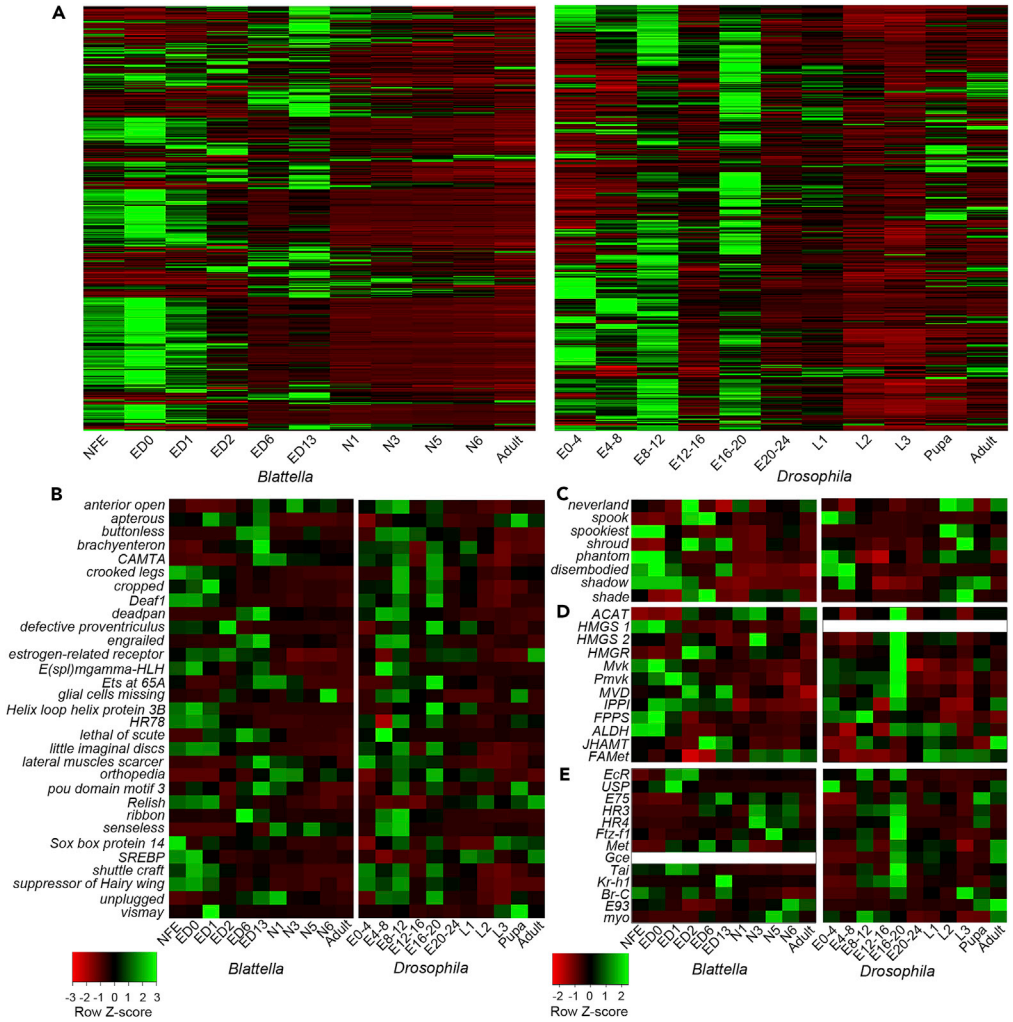



**Figure 5. Expression of Transcription Factors (TFs) and Genes Related to Hormonal Biosynthesis and Signaling along the Different Stage-Libraries of *Blattella germanica* and *Drosophila melanogaster***

(A) Expression of genes containing at least one Pfam motif unequivocally linked to a transcription factor function in *B. germanica* (600 genes identified) and in *D. melanogaster* (458 genes identified).

(B) Selection of 34 orthologous TFs common to *B. germanica* and *D. melanogaster* with conspicuously divergent expression.

(C) Genes coding for enzymes for ecdysone (20E) synthesis.

(D) Genes coding for enzymes for juvenile hormone (JH) synthesis; *HMGS1* has no orthologs in *D. melanogaster*.

(E) Genes coding for key TFs that transduce the 20E and JH signals; *gce* and *Met* have only one ortholog in *B. germanica*.

The expression is shown in FPKM in all cases. The left color scale refers to panels (A) and (B), and the right one to panels (C), (D) and (E).

the complex ecdysone signaling that regulates the holometabolan postembryonic development, which requires precise increases and decreases of hormonal signaling in narrow temporal windows (Riddiford et al., 2003). The continued expression of *Rel* in the pupa must be associated with the vulnerability to infections of this immobile stage.

## Genes Associated with Metamorphosis

We have considered genes related to the two main hormones regulating metamorphosis, the juvenile hormone (JH) and the ecdysone, or more properly 20-hydroxyecdysone (20E), which is the most well-known active form. During the juvenile postembryonic life, JH levels are high, but in the pre-adult stage, they fall dramatically until being practically undetectable. JH has a repressor role upon metamorphosis, and its

absence determines the metamorphosis. 20E has an ecdysteroidal structure, and during juvenile stages, it is synthesized by the prothoracic glands. The most important role of 20E is to promote molting, and in the pre-adult stage, in the absence of JH, it promotes the metamorphic molt (Belles, 2011; Nijhout, 1994).

In general, the genes involved in 20E biosynthesis are more highly expressed in all embryonic stages than in nymphs or adults in *B. germanica*, whereas in *D. melanogaster* they are mostly expressed in very early embryos and pre-adult stages (Figure 5C). The genes involved in JH biosynthesis are expressed throughout the ontogeny of *B. germanica* and *D. melanogaster* in a relatively similar way, although in the very early embryogenesis expression is high in *B. germanica* and low in *D. melanogaster* (Figure 5D), whereas only in late embryo stages, especially in the E16-20 stage, it becomes high in *D. melanogaster*. This appears to be the general trend: JH (and JH signaling) appears earlier in hemimetabolan than in holometabolan species (Truman and Riddiford, 1999). Moreover, the high expression of JH genes in early embryogenesis observed in *B. germanica* may be typical of the hemimetabolan species. Indeed, JH genes have been shown to have important functions in early embryogenesis of *B. germanica* (Fernandez-Nicolas and Belles, 2017), which does not appear to be the case in the holometabolan silkworm, *Bombyx mori* (Daimon et al., 2015).

In *D. melanogaster*, the expression of typical transducers of the 20E signal (King-Jones and Thummel, 2005), such as *ecdysone receptor* (*EcR*), *ultraspiracle* (*USP*), *E75*, *HR3*, and *HR4*, appears to be more concentrated in the E16-20 stage, whereas in *B. germanica* it spreads in earlier embryo stages. Singularly, *Fushi tarazu factor 1* (*Ftz-f1*) exhibits a predominant expression in N5 in *B. germanica*, whereas in *D. melanogaster* it is mainly expressed in mid-embryogenesis (Figure 5E). The characteristic expression in *B. germanica* may suggest that *Ftz-f1* plays important roles in the penultimate nymphal instar, when it is defined the genetic program of the last nymph (in which metamorphosis is determined). We have reported previously that *Ftz-f1* has critical functions during the last nymphal molts in *B. germanica* (Cruz et al., 2008). Concerning JH transducers (Jindra et al., 2015), there are not great differences of expression patterns of *Methoprene-tolerant* (*Met*), *Taiman* (*Tai*), and *Krüppel homolog 1* (*Kr-h1*) between the two species studied (Figure 5E). *Broad-complex* (*BR-C*) is interesting, as its expression shows a divergent pattern in *D. melanogaster*, where it is concentrated in the last larval instar and the pupa. In *B. germanica* maternal *BR-C* transcripts are abundant, and the gene is significantly expressed during embryogenesis (Figure 5E). This is consistent with the important functions of *BR-C* in embryo development of this species (Piulachs et al., 2010), whereas in postembryonic development *BR-C* is involved in promoting wing pad growth (Huang et al., 2013). Conversely, *BR-C* has a key function in pupal morphogenesis in *D. melanogaster* and in holometabolan insects, in general (Zhou and Riddiford, 2002). The occurrence of significant amounts of *BR-C* transcripts in the maternal load of *B. germanica* could be associated with the formation of the short germband type of this species, whereas their expression in mid and late embryogenesis might be related to the formation of a first instar nymph with basic adult features, typical of the hemimetabolan mode of metamorphosis. Other differences in the expression of JH-associated genes between *B. germanica* and *D. melanogaster* during postembryonic development may be simply idiosyncratic, as JH does not completely repress metamorphosis in higher flies, as occurs generally in insects, including cockroaches (Riddiford and Ashburner, 1991).

Other genes related to metamorphosis are *E93*, which triggers adult morphogenesis in hemimetabolan and holometabolan species (Belles and Santos, 2014; Ureña et al., 2014), and *myoglianin* (*myo*), which in the cricket *Gryllus bimaculatus* regulates the JH decrease that occurs in the last nymphal instar, which triggers metamorphosis (Ishimaru et al., 2016). Concentrated *E93* expression in pre-adult and adult stages in both *B. germanica* and *D. melanogaster* (Figure 5E) is consistent with its role of adult specifier. The high expression of *myo* in the pre-adult stages of *B. germanica* (Figure 5E) is in agreement with the inhibitory role on JH production described in *G. bimaculatus*. This role could be, therefore, conserved in hemimetabolan species, but not in holometabolan species, such as *D. melanogaster*, where *myo* expression is practically absent in pre-adult stages but concentrates in mid embryogenesis (Figure 5E), which is consistent with its role in the formation of embryo glial cells and myoblasts (Lo and Frasch, 1999).

## Conclusions

A significant part of the transcriptomic differences observed appears to be specific of cockroaches or flies. This must be the case of expression divergences in many TFs, which probably reflect differences in the expression and circuitry in functionally similar genetic networks. However, the differences underlined later

might reflect broad trends in the evolution of basic processes within neopterans, such as the development of the germband type or the metamorphosis mode.

*B. germanica* exhibits the most dynamic gene expression changes during embryogenesis. In contrast, *D. melanogaster* keeps a similar level of expression changes throughout ontogeny. This may be related to the different types of metamorphosis: hemimetabolan in cockroaches, where the adult body structure is shaped during embryogenesis, and holometabolan in flies, which shapes the adult morphology in post-embryonic stages. Genes associated with adult morphogenesis ("metamorphosis," "imaginal disc development," "pupal development") are predominantly expressed in late nymph stages in *B. germanica* and in the early-mid embryo in *D. melanogaster*. Again, this reflects a basic difference between hemimetabolan and holometabolan metamorphosis.

In *B. germanica*, the expression of *smg* and *zld*, which are important players in the MZT (see Liang et al., 2008, and Nien et al., 2011, for functional studies), concentrates in early embryogenesis (from 0% to 12% development), whereas in *D. melanogaster* there is significant expression throughout the entire embryogenesis. This sort of "extended" MZT might be related to the evolutionarily derived embryo morphogenesis and to the hemimetabolan mode of metamorphosis.

DNA methylation in early embryogenesis, which possibly promotes the expression of genes involved in the zygotic activation, is detected in *B. germanica* but not in *D. melanogaster*. This is consistent with the fact that hemimetabolan species have high levels of DNA methylation, in general, whereas they are much lower in holometabolans (Bewick et al., 2016). Thus, progressive loss of DNA methylation, in this case in the embryo, may have been a mechanism driving the evolution from hemimetabolan polyneopterans and paraneopterans to holometabolan endopterygotes.

The expression of TFs reveals many differences between *B. germanica* and *D. melanogaster* in embryonic and postembryonic stages. Many of them appear to be specific, but some observed in JH and 20E transducers could be representative of the type of embryogenesis and/or metamorphosis. For example, transcripts of JH transducers that are present at significant amounts in very early embryo stages of *B. germanica*, but not in *D. melanogaster*, may reflect a different regulation of the blastoderm formation related to the germband type, short (*B. germanica*) or long (*D. melanogaster*) (see Fernandez-Nicolas and Belles, 2017, for functional studies). It is plausible that loss of these JH transducers in the very early embryo has been one of the drivers of evolution from short to long germband. Another difference relates to the expression of these hormonal transducers in the mid and late embryo, which is quantitatively and functionally important in hemimetabolan species (Fernandez-Nicolas and Belles, 2017; Piulachs et al., 2010) but not in holometabolans (Daimon et al., 2015). Thus, the declining influence of JH in the embryo may have been another factor driving the morphological divergence of juvenile stages in the holometabolan last common ancestor and the evolution of metamorphosis toward holometaboly.

Comparisons also highlighted *BR-C* as a particularly important TF. In *D. melanogaster*, *BR-C* expression concentrates in prepupal and pupal stages, which is consistent with its key role in pupal morphogenesis of holometabolan insects (Zhou and Riddiford, 2002). In turn, the low expression during embryogenesis fits with the practically dispensable role of *BR-C* in embryo development in holometabolan insects (Daimon et al., 2015). In *B. germanica*, in contrast, the highest expression of *BR-C* is observed along embryogenesis, which is in agreement with its important morphogenetic roles (Piulachs et al., 2010), which would be characteristic of embryogenesis in hemimetabolan species. Comparatively, the expression of *BR-C* in nymphal stages is low, which corresponds to its limited role of sustaining the growth of wing pads (Huang et al., 2013). As proposed by Huang et al. (2013), a fundamental innovation in postembryonic development in holometabolans has been an expansion of *BR-C* functions, from one specialized in wing development to a larger array of morphogenetic functions that culminated with the pupal specifier role. Conversely, in hemimetabolans, *BR-C* would have important morphogenetic roles in embryo development, and its loss may have been an important factor in the evolution of holometaboly from hemimetaboly (see also Fernandez-Nicolas and Belles, 2017).

## METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, three figures, three tables, and four data files and can be found with this article online at https://doi.org/10.1016/j.isci.2018.05.017.

## ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Economy and Competitiveness (grants CGL2012-36251 and CGL2015-64727-P to X.B., including FEDER funds), Spanish Ministry of Science and Innovation (grants BFU2011–22404 and CGL2016–76011-R to M.-D.P., including FEDER funds), and Catalan Government (grants 2014 SGR 619 and 2017 SGR 1030). Thanks are also due to Alba Ventos-Alfonso, who carried out the qRT-PCR measurements.

## AUTHOR CONTRIBUTIONS

X.B. and M.-D.P. designed the research; X.B., M.-D.P., and G.Y. performed the research; GY analyzed data; X.B., M.-D.P., and G.Y. discussed and interpreted the analyses; X.B., M.-D.P., and G.Y. wrote the paper.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 15, 2018
Revised: April 13, 2018
Accepted: May 23, 2018
Published: June 29, 2018

## REFERENCES

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. (2000). The genome sequence of *Drosophila melanogaster*. Science *287*, 2185–2195.

Andrews, S.. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Averof, M., and Akam, M. (1995). Hox genes and the diversification of insect and crustacean body plans. Nature *376*, 420–423.

Belles, X. (2011). Origin and evolution of insect metamorphosis. In Encyclopedia of Life Sciences (ELS) (John Wiley & Sons, Ltd.), pp. 1–11.

Belles, X., and Santos, C.G. (2014). The MEKRE93 (Methoprene tolerant-Krüppel homolog 1-E93) pathway in the regulation of insect metamorphosis, and the homology of the pupal stage. Insect Biochem. Mol. Biol. *52*, 60–68.

Benoit, B., He, C.H., Zhang, F., Votruba, S.M., Tadros, W., Westwood, J.T., Smibert, C.A., Lipshitz, H.D., and Theurkauf, W.E. (2009). An essential role for the RNA-binding protein Smaug during the *Drosophila* maternal-to-zygotic transition. Development *136*, 923–932.

Bewick, A.J., Vogel, K.J., Moore, A.J., and Schmitz, R.J. (2016). Evolution of DNA methylation across insects. Mol. Biol. Evol. *34*, 654–665.

Bhattacharya, A., and Baker, N.E. (2012). The role of the bHLH protein hairy in morphogenetic furrow progression in the developing *Drosophila* eye. PLoS One *7*, e47503.

Bushati, N., Stark, A., Brennecke, J., and Cohen, S.M. (2008). Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in *Drosophila*. Curr. Biol. *18*, 501–506.

Campos-Ortega, J.A., and Hartenstein, V. (1985). The Embryonic Development of *Drosophila melanogaster* (Springer).

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell *134*, 25–36.

Celniker, S.E., Dillon, L.A., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M., et al. (2009). Unlocking the secrets of the genome. Nature *459*, 927–930.

Chen, L., Dumelie, J.G., Li, X., Cheng, M.H., Yang, Z., Laver, J.D., Siddiqui, N.U., Westwood, J.T., Morris, Q., Lipshitz, H.D., et al. (2014). Global regulation of mRNA translation and stability in the early *Drosophila* embryo by the Smaug RNA-binding protein. Genome Biol. *15*, R4.

Chiang, C., Young, K.E., and Beachy, P.A. (1995). Control of *Drosophila* tracheal branching by the novel homeodomain gene *unplugged*, a regulatory target for genes of the *bithorax* complex. Development *121*, 3901–3912.

Chipman, A.D. (2015). Hexapoda: comparative aspects of early development. In Evolutionary Developmental Biology of Invertebrates 5, A. Wanninger, ed. (Springer), pp. 93–110.

Condamine, F.L., Clapham, M.E., and Kergoat, G.J. (2016). Global patterns of insect diversification: towards a reconciliation of fossil and molecular evidence? Sci. Rep. *6*, 19208.

Cruz, J., Nieva, C., Mané-Padrós, D., Martín, D., and Belles, X. (2008). Nuclear receptor BgFTZ-F1 regulates molting and the timing of ecdysteroid production during nymphal development in the hemimetabolous insect *Blattella germanica*. Dev. Dyn. *237*, 3179–3191.

Dahanukar, A., Walker, J.A., and Wharton, R.P. (1999). Smaug, a novel RNA-binding protein that operates a translational switch in *Drosophila*. Mol. Cell *4*, 209–218.

Daimon, T., Uchibori, M., Nakao, H., Sezutsu, H., and Shinoda, T. (2015). Knockout silkworms reveal a dispensable role for juvenile hormones in holometabolous life cycle. Proc. Natl. Acad. Sci. USA *112*, E4226–E4235.

Elango, N., Hunt, B.G., Goodisman, M.A.D., and Yi, S.V. (2009). DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. Proc. Natl. Acad. Sci. USA *106*, 11206–11211.

Feliciello, I., Parazajder, J., Akrap, I., and Ugarković, Đ. (2013). First evidence of DNA methylation in insect *Tribolium castaneum*. Epigenetics *8*, 534–541.

Fernandez-Nicolas, A., and Belles, X. (2017). Juvenile hormone signaling in short germ-band hemimetabolan embryos. Development *144*, 4637–4644.

Fisher, A., and Caudy, M. (1998). The function of hairy-related bHLH repressor proteins in cell fate decisions. Bioessays *20*, 298–306.

Foo, S.M., Sun, Y., Lim, B., Ziukaite, R., O'Brien, K., Nien, C.Y., Kirov, N., Shvartsman, S.Y., and Rushlow, C.A. (2014). Zelda potentiates

morphogen activity by increasing chromatin accessibility. Curr. Biol. *24*, 1341–1346.

Fu, S., Nien, C.-Y., Liang, H.-L., and Rushlow, C. (2014). Co-activation of microRNAs by Zelda is essential for early *Drosophila* development. Development *141*, 2108–2118.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.-L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science *311*, 395–398.

Grimaldi, D., and Engel, M.S. (2005). Evolution of the Insects (Cambridge University Press).

Harrison, M.C., Jongepier, E., Robertson, H.M., Arning, N., Bitard-Feildel, T., Chao, H., Childers, C.P., Dinh, H., Doddapaneni, H., Dugan, S., et al. (2018). Hemimetabolous genomes reveal molecular basis of termite eusociality. Nat. Ecol. Evol. *2*, 557–566.

Huang, J.-H., Lozano, J., and Belles, X. (2013). Broad-complex functions in postembryonic development of the cockroach *Blattella germanica* shed new light on the evolution of insect metamorphosis. Biochim. Biophys. Acta *1830*, 2178–2187.

I5K-Consortium (2013). The i5K initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. J. Hered. *104*, 595–600.

Ishimaru, Y., Tomonari, S., Matsuoka, Y., Watanabe, T., Miyawaki, K., Bando, T., Tomioka, K., Ohuchi, H., Noji, S., and Mito, T. (2016). TGF-β signaling in insects regulates metamorphosis via juvenile hormone biosynthesis. Proc. Natl. Acad. Sci. USA *113*, 5634–5639.

Jindra, M., Belles, X., and Shinoda, T. (2015). Molecular basis of juvenile hormone signaling. Curr. Opin. Insect Sci. *11*, 39–46.

Kay, S., Skowronski, D., and Hunt, B.G. (2018). Developmental DNA methyltransferase expression in the fire ant *Solenopsis invicta*. Insect Sci. *25*, 57–65.

King-Jones, K., and Thummel, C.S. (2005). Nuclear receptors-a perspective from *Drosophila*. Nat. Rev. Genet. *6*, 311–323.

Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in *Drosophila*. Development *130*, 5083–5090.

Li, L., Greer, C., Eisenman, R.N., and Secombe, J. (2010). Essential functions of the histone demethylase Lid. PLoS Genet. *6*, e1001221.

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. Nature *456*, 400–403.

Liu, P.Z., and Kaufman, T.C. (2005). Short and long germ segmentation: unanswered questions in the evolution of a developmental mode. Evol. Dev. *7*, 629–646.

Lo, P.C., and Frasch, M. (1999). Sequence and expression of *myoglianin*, a novel *Drosophila* gene of the TGF-beta superfamily. Mech. Dev. *86*, 171–175.

Lyko, F. (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. Nat. Rev. Genet. *19*, 81–92.

Lyko, F., Whittaker, A.J., Orr-Weaver, T.L., and Jaenisch, R. (2000). The putative *Drosophila* methyltransferase gene *dDnmt2* is contained in a transposon-like element and is expressed specifically in ovaries. Mech. Dev. *95*, 215–217.

Marhold, J., Rothe, N., Pauli, A., Mund, C., Kuehle, K., Brueckner, B., and Lyko, F. (2004). Conservation of DNA methylation in dipteran insects. Insect Mol. Biol. *13*, 117–123.

Markow, T.A. (2015). The secret lives of *Drosophila* flies. Elife *4*, e06793.

de Mendoza, A., Sebé-Pedrós, A., Sestak, M.S., Matejcic, M., Torruella, G., Domazet-Loso, T., and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. USA *110*, E4858–E4866.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. Science *346*, 763–767.

modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F., et al. (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. Science *330*, 1787–1797.

Müller, D., Jagla, T., Bodart, L.M., Jährling, N., Dodt, H.-U., Jagla, K., and Frasch, M. (2010). Regulation and functions of the *lms* homeobox gene during development of embryonic lateral transverse muscles and direct flight muscles in *Drosophila*. PLoS One *5*, e14323.

Negre, B., and Simpson, P. (2015). The *achaete - scute* complex in Diptera: patterns of noncoding sequence evolution. J. Evol. Biol. *28*, 1770–1781.

Nicholson, D.B., Ross, A.J., and Mayhew, P.J. (2014). Fossil evidence for key innovations in the evolution of insect diversity. Proc. Biol. Sci. B *281*, https://doi.org/10.1098/rspb.2014.1823.

Nien, C.Y., Liang, H.L., Butcher, S., Sun, Y., Fu, S., Gocha, T., Kirov, N., Manak, J.R., and Rushlow, C. (2011). Temporal coordination of gene networks by Zelda in the early *Drosophila* embryo. PLoS Genet. *7*, e1002339.

Nijhout, H.F. (1994). Insect Hormones (Princeton University Press).

Nolo, R., Abbott, L.A., and Bellen, H.J. (2000). Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in *Drosophila*. Cell *102*, 349–362.

Panfilio, K.A. (2008). Extraembryonic development in insects and the acrobatics of blastokinesis. Dev. Biol. *313*, 471–491.

Peel, A.D., Chipman, A.D., and Akam, M. (2005). Arthropod segmentation: beyond the *Drosophila* paradigm. Nat. Rev. Genet. *6*, 905–916.

Petersen, A.J., Katzenberger, R.J., and Wassarman, D.A. (2013). The innate immune response transcription factor Relish is necessary for neurodegeneration in a *Drosophila* model of ataxia-telangiectasia. Genetics *194*, 133–142.

Piulachs, M.-D., Pagone, V., and Belles, X. (2010). Key roles of the *Broad-Complex* gene in insect embryogenesis. Insect Biochem. Mol. Biol. *40*, 468–475.

Riddiford, L.M., and Ashburner, M. (1991). Effects of juvenile hormone mimics on larval development and metamorphosis of *Drosophila melanogaster*. Gen. Comp. Endocrinol. *82*, 172–183.

Riddiford, L.M., Hiruma, K., Zhou, X., and Nelson, C.A. (2003). Insights into the molecular basis of the hormonal control of molting and metamorphosis from *Manduca sexta* and *Drosophila melanogaster*. Insect Biochem. Mol. Biol. *33*, 1327–1338.

Ritter, A.R., and Beckstead, R.B. (2010). Sox14 is required for transcriptional and developmental responses to 20-hydroxyecdysone at the onset of *Drosophila* metamorphosis. Dev. Dyn. *239*, 2685–2694.

Roumengous, S., Rousset, R., and Noselli, S. (2017). Polycomb and Hox genes control JNK-induced remodeling of the segment boundary during *Drosophila* morphogenesis. Cell Rep. *19*, 60–71.

Schröder, R. (2003). The genes orthodenticle and hunchback substitute for bicoid in the beetle *Tribolium*. Nature *422*, 621–625.

Schulz, K.N., Bondra, E.R., Moshe, A., Villalta, J.E., Lieb, J.D., Kaplan, T., McKay, D.J., and Harrison, M.M. (2015). Zelda is differentially required for chromatin accessibility, transcription-factor binding and gene expression in the early *Drosophila* embryo. Genome Res. *25*, 1715–1726.

Shao, W., and Espenshade, P.J. (2012). Expanding roles for SREBP in metabolism. Cell Metab. *16*, 414–419.

Siegfried, Z., and Simon, I. (2010). DNA methylation and gene expression. Wiley Interdiscip. Rev. Syst. Biol. Med. *2*, 362–371.

Smibert, C.A., Lie, Y.S., Shillinglaw, W., Henzel, W.J., and Macdonald, P.M. (1999). Smaug, a novel and conserved protein, contributes to repression of nanos mRNA translation in vitro. RNA *5*, 1535–1547.

Song, X., Huang, F., Liu, J., Li, C., Gao, S., Wu, W., Zhai, M., Yu, X., Xiong, W., Xie, J., et al. (2017). Genome-wide DNA methylomes from discrete developmental stages reveal the predominance of non-CpG methylation in *Tribolium castaneum*. DNA Res. *24*, 445–457.

Staudt, N., Fellert, S., Chung, H.-R., Jäckle, H., and Vorbrüggen, G. (2006). Mutations of the *Drosophila* zinc finger-encoding gene vielfältig impair mitotic cell divisions and cause improper chromosome segregation. Mol. Biol. Cell *17*, 2356–2365.

Stroumbakis, N.D., Li, Z., and Tolias, P.P. (1996). A homolog of human transcription factor NF-X1

encoded by the *Drosophila shuttle craft* gene is required in the embryonic central nervous system. Mol. Cell. Biol. *16*, 192–201.

Sun, Y., Nien, C.-Y., Chen, K., Liu, H.-Y., Johnston, J., Zeitlinger, J., and Rushlow, C. (2015). Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation. Genome Res. *25*, 1703–1714.

Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). SMAUG is a major regulator of maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. Dev. Cell *12*, 143–155.

Takayama, S., Dhahbi, J., Roberts, A., Mao, G., Heo, S.-J., Pachter, L., Martin, D.I.K., and Boffelli, D. (2014). Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity. Genome Res. *24*, 821–830.

Tanaka, A. (1976). Stages in the embryonic development of the German cockroach, *Blattella germanica* Linné (Blattaria, Blattellidae). Kontyû, Tokyo *44*, 1703–1714.

Truman, J.W., and Riddiford, L.M. (1999). The origins of insect metamorphosis. Nature *401*, 447–452.

Ureña, E., Manjon, C., Franch-Marro, X., and Martin, D. (2014). Transcription factor E93 specifies adult metamorphosis in hemimetabolous and holometabolous insects. Proc. Natl. Acad. Sci. USA *111*, 7024–7029.

Wang, Y.-H., Engel, M.S., Rafael, J.A., Wu, H.-Y., Rédei, D., Xie, Q., Wang, G., Liu, X.-G., and Bu, W.-J. (2016). Fossil record of stem groups employed in evaluating the chronogram of insects (Arthropoda: Hexapoda). Sci. Rep. *6*, 38939.

Wong, M.M.-K., Liu, M.-F., and Chiu, S.K. (2015). Cropped, *Drosophila* transcription factor AP-4, controls tracheal terminal branching and cell growth. BMC Dev. Biol. *15*, 20.

Ylla, G., and Belles, X. (2015). Towards understanding the molecular basis of cockroach tergal gland morphogenesis. A transcriptomic approach. Insect Biochem. Mol. Biol. *63*, 104–112.

Ylla, G., Piulachs, M.-D., and Belles, X. (2017). Comparative analysis of miRNA expression during the development of insects of different metamorphosis modes and germ-band types. BMC Genomics *18*, 774.

Zhou, X., and Riddiford, L.M. (2002). Broad specifies pupal development and mediates the "status quo" action of juvenile hormone on the pupal-adult transformation in *Drosophila* and *Manduca*. Development *129*, 2259–2269.

**ISCI, Volume *4***

## Supplemental Information

# Comparative Transcriptomics in Two Extreme Neopterans Reveals General Trends in the Evolution of Modern Insects

**Guillem Ylla, Maria-Dolors Piulachs, and Xavier Belles**

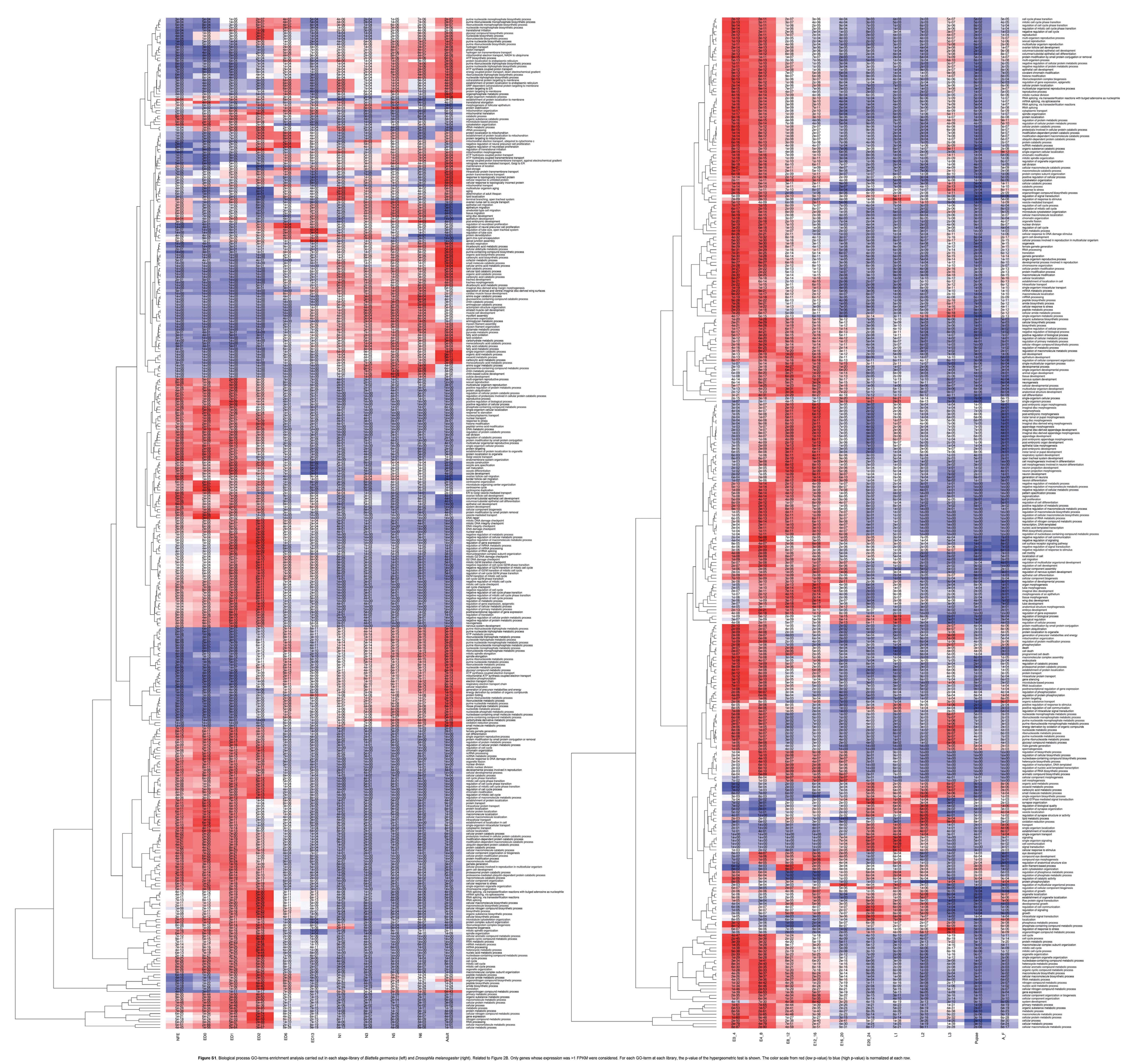
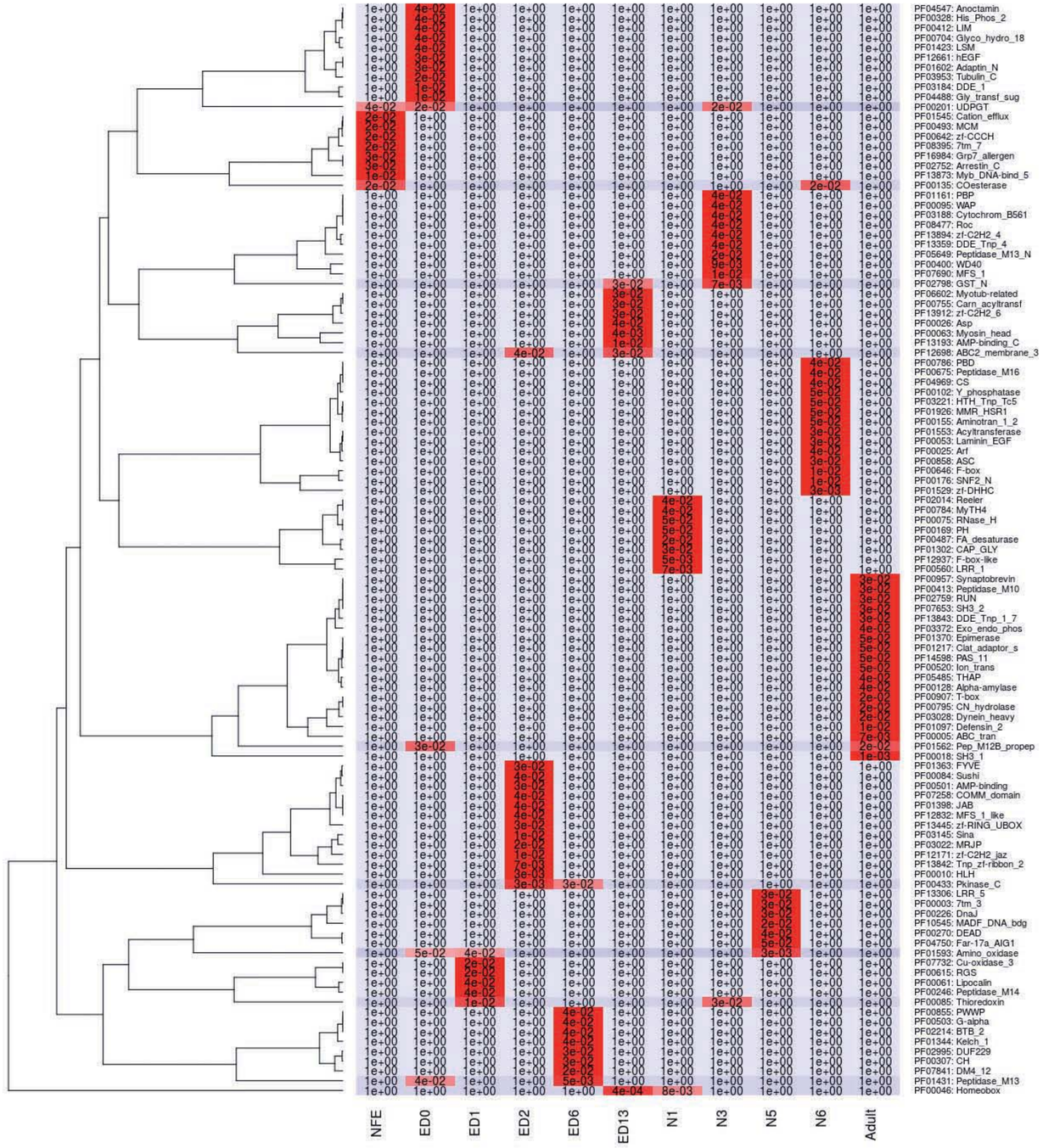



**Figure S1**. Biological process GO-terms enrichment analysis carried out in each stage-library of *Blattella germanica* (left) and *Drosophila melanogaster* (right). Related to Figure 2B. Only genes whose expression was >1 FPKM were considered. For each GO-term at each library, the p-value of the hypergeometric test is shown. The color scale from red (low p-value) to blue (high p-value) is normalized at each row.



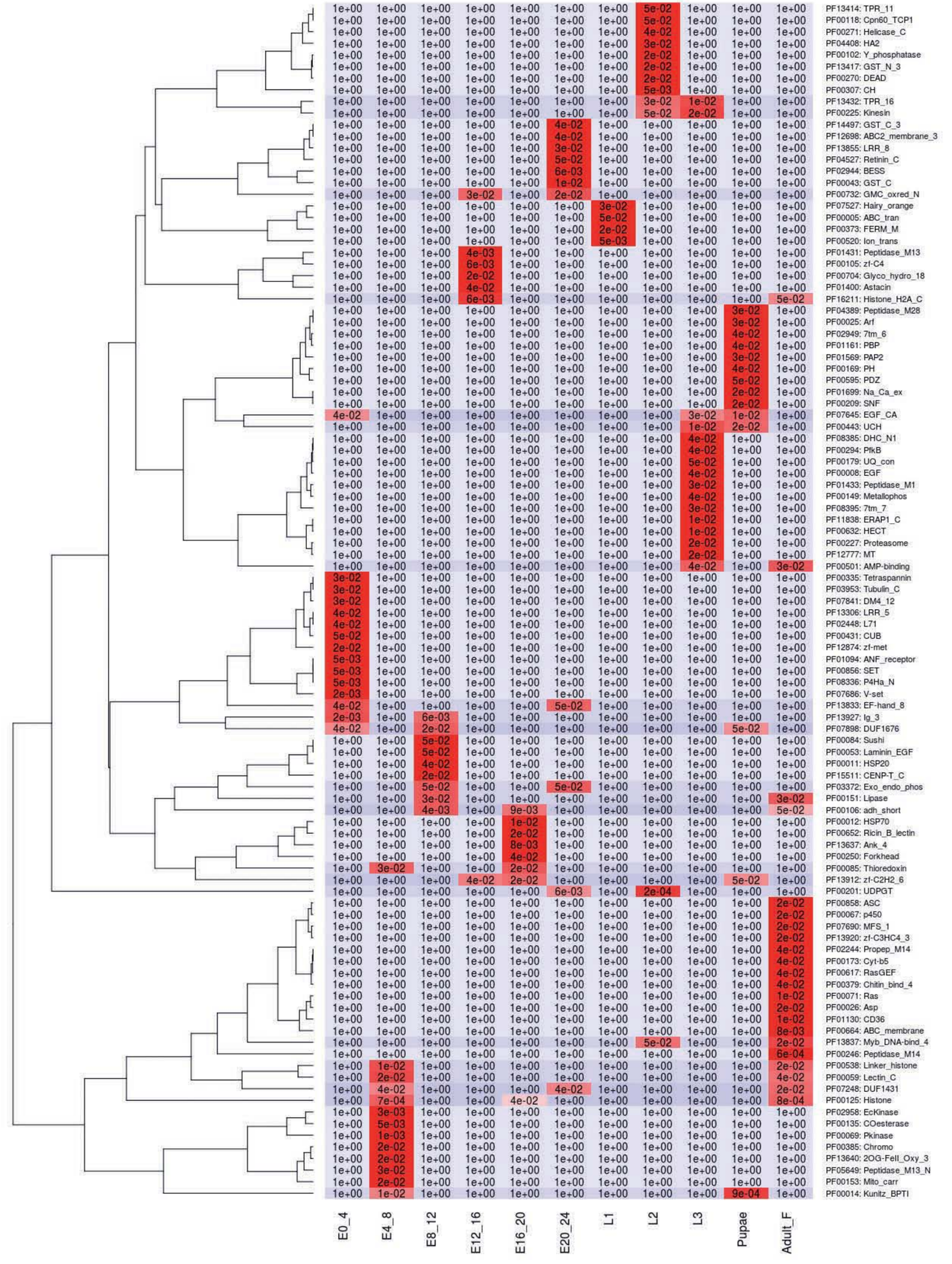




**Figure S2.** PFAM motifs enrichment analysis carried out in each stage of of *Blattella germanica* (left) and *Drosophila melanogaster* (right). Related to Figure 2B. Only genes whose expression was >1 FPKM were considered. The p-value of the hypergeometric test is shown, and only those PFAMs showing a significant enrichment (p-value < 0.05), in at least one stage, are shown. The color scale from red (low p-value) to blue (high p-value) is normalized at each row.

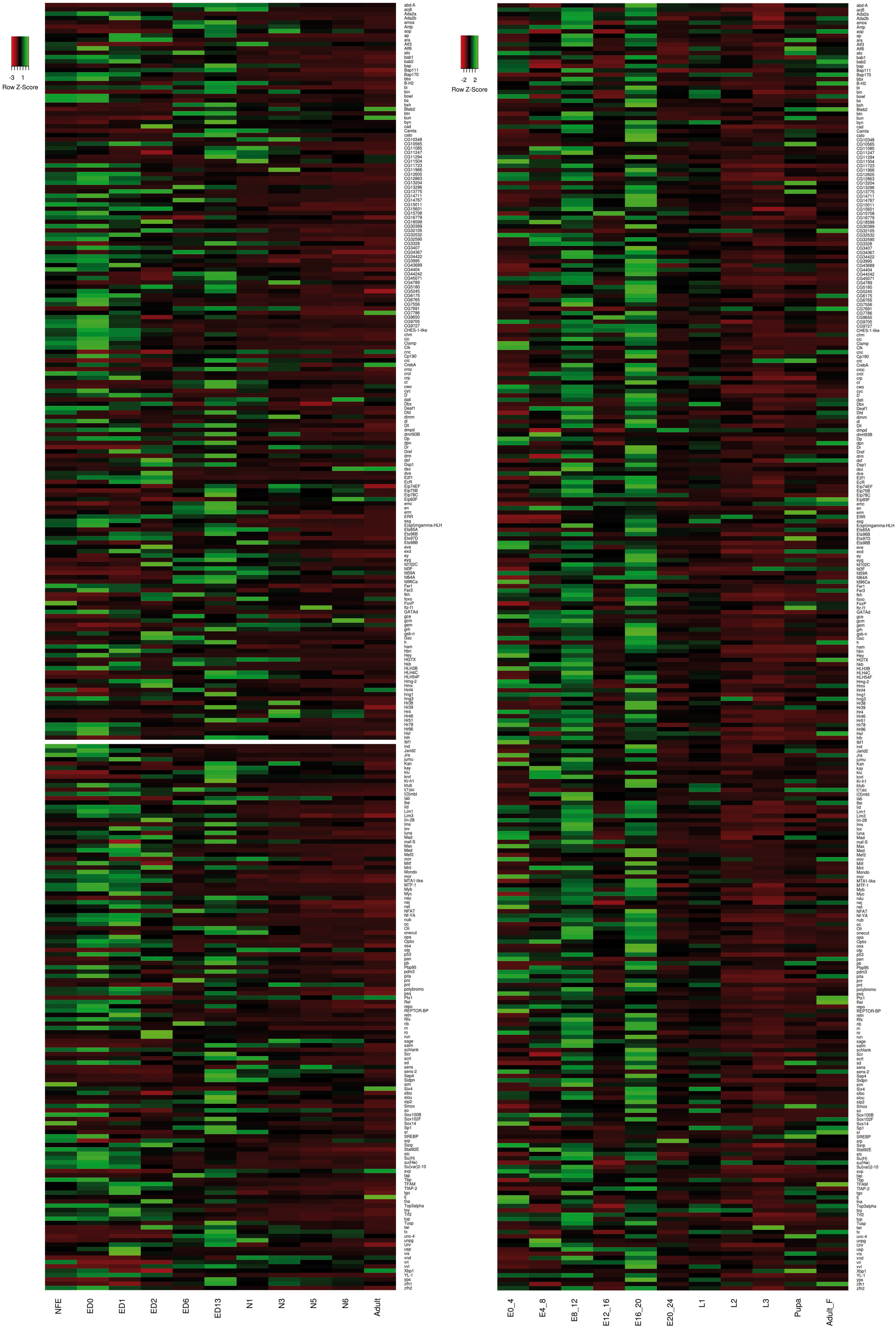
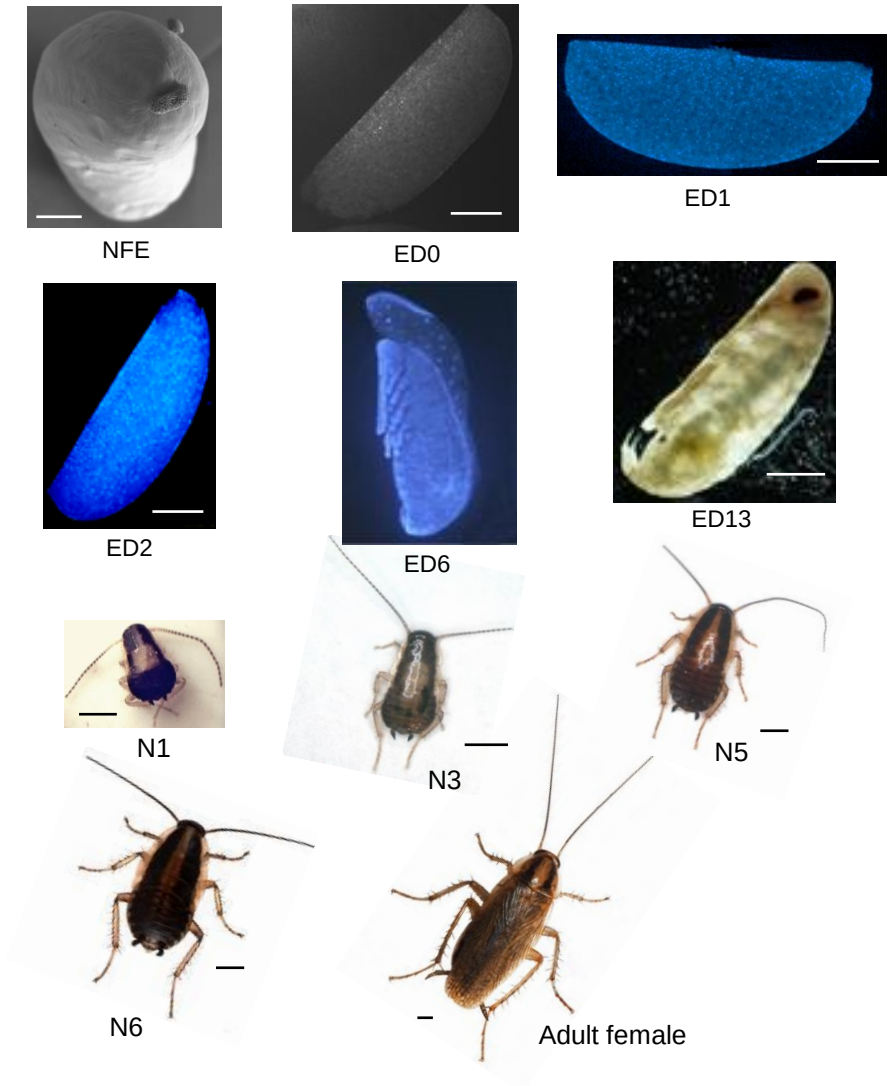



**Figure S3.** Expression in *Blattella germanica* (left) and *Drosophila melanogaster* (right) of the 297 transcription factors common to the two species. Related to Figure 5A.

**Table S1.** Biological data corresponding to the 11 transcriptomes studied in the present work. Related to Experimental Procedures. Data on juvenile hormone (JH) and ecdysteroids (20E) are from Treiblmayr et al. (2006) (JH in nymphal stages), Maestro et al. (2010) (JH in embryo stages), Cruz et al. (2003) (20E in nymphal stages), Piulachs et al. (2010) (20E in embryo stages). Stages are from Tanaka (1976). Scale bars equivalent to 500 µM (from NFE to ED13) or to 1 mm (from N1 to adult ).




| Library | Period development | Pooled individuals per sample (n = 2) | Age (AO: after oviposition) | % embryo development | Embryo development, Tanaka stage | Hormonal context |
|---|---|---|---|---|---|---|
| NFE | Egg | n 1 = 360 eggs from 24 females<br>n 2 = 270 eggs from 18 females | Day 8 of the first gonadotrophic cycle (in preoviposition) | -- | -- | Not measured in the egg. High levels of 20E and JH in the surrounding haemolymph. |
| ED0 | Embryo | n 1 = 45 oothecae<br>n 2 = 24 oothecae | 8 h AO, when the ootheca is still vertical | 2%. | Only yolk granules observed. | No detectable levels of 20E and JH |
| ED1 | Embryo | n 1 = 25 oothecae<br>n 2 = 24 oothecae | 24 h AO | 6% | Energids at low density spread in the yolk . Tanaka stage 1 | No detectable levels of 20E and JH |
| ED2 | Embryo | n 1 = 8 oothecae<br>n 2 = 8 oothecae | 48 h AO | 12% | Abundant energids, germ band anlage well delimited, slightly expanded at both sides. Tanaka stage 2 | Burst of 20E inferred from the expression of HR3 (a 20E-dependent gene). No detectable JH |
| ED6 | Embryo | n 1 = 4 oothecae<br>n 2 = 4 oothecae | 144 h AO | 33% | Pleuropodia well apparent, legs segmented, caudal space arises. Tanaka stage 8 | Peak of 20E. Very low levels of JH |
| ED13 | Embryo | n 1 = 4 oothecae<br>n 2 = 4 oothecae | 312 h AO | 72% | Eyes colored, antennae and legs reaching the 5th abdominal segment. Tanaka stage 15 | Peak of 20E. High levels of JH |
| N1 | 1st nymphal instar | n 1 = 5 individuals (indeterminate sex)<br>n 2 = 5 individuals (indeterminate sex) | 1-2 days old | -- | -- | High levels of 20E and JH |
| N3 | 3rd nymphal instar | n 1 = 5 individuals (indeterminate sex)<br>n 2 = 5 individuals (indeterminate sex) | 2-4 days old | -- | -- | High levels of 20E and JH |
| N5 | 5th nymphal instar | n 1 = 5 individuals (females)<br>n 2 = 5 individuals (females) | 3-5 days old | -- | -- | High levels of 20E and JH |
| N6 | 6th nymphal instar | n 1 = 5 individuals (females)<br>n 2 = 5 individuals (females) | 5-7 days old | -- | -- | High levels of 20E, no JH |
| Adult | Adult | n 1 = 5 females<br>n 2 = 5 females | 5 days old | -- | -- | Low (ovarian) levels of 20E and high JH levels |

**Table S2**. Caracteristics of the embryo mRNA libraries of *Blattella germanica* and *Drosophila melanogaster* used in this work. Related to Experimental Procedures. Same color indicates equivalent developmental periods.

| ***Blattella germanica*** | | ***Drosophila melanogaster*** | |
|---|---|---|---|
| **Library, hours and developmental/molecular features** | **% development (a)** | **Library, hours and developmental/molecular features** | **% development (b)** |
| NFE-Non-fecunded eggs | 0.00% | | |
| ED0 8h AO. Pre-blastoderm stage. T48 highly expressed. Maternal-gap genes (Nanos and tailless) well expressed. | 2.00% | Embryo 0-4h. Pre-bastoderm+blastoderm+gastrula stages. T48 highly expressed. Maternal-gap genes (Nanos, hunchback, caudal and tailless) well expressed. | 0.0% - 17% |
| ED1 24h AO. Early blastoderm. T48 still expressed. Gap genes (orthodenticle, huckbein) well expressed. Pair rule genes not expressed yet. | 6.00% | Embryo 4-8h. Early extended germ band+late extended germ band. Gap, pair rule and segmentation genes well expressed. | 17% - 34% |
| ED2 48h AO. Germ band being formed. Pair rule and segmentation genes well expressed. | 12.00% | Embryo 8-12h. Onset of dorsal closure. Gap, pair rule and segmentation genes well expressed. Hox genes expressed. | 34% - 50% |
| ED6 144h AO. Onset of dorsal closure. Segmentation genes still expressed. Hox genes begin to be expressed. | 33.00% | Embryo 12-16h. End of dorsal closure+beginning of late embryo stage. | 50% - 67% |
| ED13 312h AO. Dorsal closure finished. Hox genes well expressed. | 72.00% | Embryo_16-20h. Late embryo stage. Hox genes well expressed. | 67% - 83% |
| | | Embryo 20-24h. Late embryo stage. | 83% - 100% |

(a) 100% = 18 days
(b) 100% = 24 hours

**Table S3.** Names and the corresponding accession codes of genes whose expression profiles are showed. Related to Experimental Procedures. For *Drosophila melanogaster* the Flybase gene accession is used and for *Blattella germanica*, the codes were based on genome annotation available from NCBI bioproject under the accession code PRJNA427252.

| Gene name | FlyBase Accession | Bger code |
|---|---|---|
| *abdominal A* | FBgn0000014 | Bger_24605 |
| *Abdominal B* | FBgn0000015 | BgerTmpA028008-RA[a] |
| *ACAT* | FBgn0035203 | Bger_10151[b] |
| *ALDH* | FBgn0010548 | Bger_14856[c] |
| *antennapedia* | FBgn0260642 | Bger_28331 |
| *anterior open* | FBgn0000097 | Bger_12886 |
| *apterous* | FBgn0267978 | Bger_ M0006[a] |
| *bicoid* | FBgn0000166 | No orthologous in *B. germanica* |
| *brachyenteron* | FBgn0011723 | Bger_17804 |
| *Broad Complex (Core)* | FBgn0283451 | Bger_M0004[a] |
| *buttonless* | FBgn0014949 | Bger_04705 |
| *CAMTA* | FBgn0259234 | Bger_11477 |
| *caudal* | FBgn0000251 | Bger_M0002[a] |
| *collier* | FBgn0001319 | Bger_28042 |
| *crocodile* | FBgn0014143 | Bger_08908 |
| *crooked legs* | FBgn0020309 | Bger_24171 |
| *cropped* | FBgn0001994 | Bger_15672 |
| *deadpan* | FBgn0010109 | Bger_12374 |
| *Deaf1* | FBgn0013799 | Bger_03953 |
| *defective proventriculus* | FBgn0020307 | Bger_25497 |
| *Deformed* | FBgn0000439 | Bger_21960 |
| *Disembodied* | FBgn0000449 | Bger_24331 |
| *Dnmt1* | No orthologous in *D. melanogaster* | Bger_02650 |
| *Dnmt2* | FBgn0028707 | Bger_06901 |
| *E(spl)mgamma-HLH* | FBgn0002735 | Bger_05984 |
| *E75* | FBgn0000568 | Bger_26014 |
| *E93* | FBgn0264490 | Bger_M0005[a] |
| *EcR* | FBgn0000546 | Bger_08790 |
| *empty-spiracles* | FBgn0000576 | Bger_26557 |
| *engrailed* | FBgn0000577 | Bger_00720 |
| *engrailed* | FBgn0000577 | Bger_00720 |
| *estrogen-related receptor* | FBgn0035849 | Bger_11720 |
| *Ets at 65A* | FBgn0005658 | Bger_27146 |
| *even skipped* | FBgn0000606 | Bger_18070 |
| *FAMeT* | FBgn0034583 | Bger_20710[c] |
| *FPPS* | FBgn0025373 | Bger_01636[b] |
| *Ftz-f1* | FBgn0001078 | Bger_09989 |
| *fushitarazu* | FBgn0001077 | Bger_21966 |
| *gce* | FBgn0261703 | No orthologous in *B. germanica* |
| *glial cells missing* | FBgn0014179 | Bger_00054 |
| *grainy head* | FBgn0259211 | Bger_26553 |
| *hairy* | FBgn0001168 | Bger_26281 |
| *hedgehog* | FBgn0004644 | Bger_02233 |
| *Helix loop helix protein 3B* | FBgn0011276 | Bger_13620 |
| *HMGR* | FBgn0263782 | Bger_14007[b] |
| *HMGS1* | No orthologous in *D. melanogaster* | Bger_03932[b] |
| *HMGS2* | FBgn0010611 | Bger_02739[b] |
| *HR3* | FBgn0000448 | Bger_00728 |

| | | |
|---|---|---|
| *HR4* | FBgn0264562 | Bger_18448 |
| *hunchback* | FBgn0001180 | Bger_14305 |
| *IPPI* | FBgn0038876 | Bger_04562[b] |
| *JHAMT* | FBgn0028841 | Bger_04116[c] |
| *kirps* | FBgn0001320 | Bger_03424 |
| *krüppel* | FBgn0001325 | Bger_01853 |
| *Krüppel homolog 1* | FBgn0266450 | Bger_M0003[a] |
| *labial* | FBgn0002522 | Bger_17181 |
| *lateral muscle scarcer* | FBgn0002023 | Bger_02888 |
| *lethal of scute* | FBgn0002561 | Bger_14544 |
| *little imaginal disc* | FBgn0031759 | Bger_06256 |
| *Met* | FBgn0002723 | Bger_10267 |
| *MVD* | FBgn0030683 | Bger_12150[b] |
| *MvK* | FBgn0061359 | Bger_05361[b] |
| *myo* | FBgn0026199 | Bger_15600 |
| *nanos* | FBgn0002962 | Bger_23144 |
| *Neverland* | FBgn0259697 | Bger_22030 |
| *orthodenticale* | FBgn0004102 | Scaffold533:124865-197312[a] |
| *orthopedia* | FBgn0015524 | Bger_23610 |
| *Phantom* | FBgn0004959 | Bger_21433 |
| *Pmvk* | FBgn0032811 | Bger_06941[b] |
| *pou domain motif 3* | FBgn0261588 | Bger_09030 |
| *proboscipedia* | FBgn0051481 | Bger_17179 |
| *Relish* | FBgn0014018 | Bger_13050 |
| *ribbon* | FBgn0003254 | Bger_21758 |
| *runt* | FBgn0003300 | Bger_05999 |
| *senseless* | FBgn0051632 | Bger_16235 |
| *sex combs reduced* | FBgn0003339 | Bger_21963 |
| *Shade* | FBgn0003388 | Bger_13798 |
| *Shadow* | FBgn0003312 | Bger_09617 |
| *Shroud* | FBgn0262112 | Bger_07744 |
| *shuttle craft* | FBgn0001978 | Bger_06059 |
| *smaug* | FBgn0016070 | Bger_10865 |
| *Sox box protein 14* | FBgn0005612 | Bger_02065 |
| *Spook* | FBgn0003486 | Bger_25648 |
| *Spookiest* | No orthologous in *D. melanogaster* | Bger_04901 |
| *SREBP* | FBgn0261283 | Bger_16229 |
| *suppressor of Hairy wing* | FBgn0003567 | Bger_00591 |
| *tailless* | FBgn0003720 | Bger_19904 |
| *Taiman* | FBgn0041092 | Bger_21670 |
| *ultrabithorax* | FBgn0003944 | BgerTmpA017800-RA[a] |
| *unplugged* | FBgn0015561 | Bger_08368 |
| *USP* | FBgn0003964 | Bger_15477[b] |
| *vismay* | FBgn0033748 | Bger_07739 |
| *wingless* | FBgn0004009 | Bger_10842 |
| *zelda* | FBgn0259789 | Bger_M0001[a] |

[a] Genes manually annotated in *B. germanica* genome.

[b] Annotation based on Qu et al. (2017). MicroRNAs regulate the sesquiterpenoid hormonal pathway in *Drosophila* and other arthropods. Proceedings of the Royal Society B: Biological Sciences 284(1869). pii: 20171827.

[c] Annotation based on Harrison et al. (2018). Hemimetabolous genomes reveal molecular basis of termite eusociality. Nature Ecology and Evolution 2(3):557–566.

## Transparent Methods

### Insects

*B. germanica* specimens were obtained from a colony reared in the dark at 29 ± 1ºC and 60-70% relative humidity. All dissections and tissue sampling were carried out on carbon dioxide-anesthetized specimens. Tissues were frozen on liquid nitrogen and stored at -80ºC until use.

### Preparation and sequencing of mRNA libraries

We sequenced two biological replicates of each chosen stage along the ontogeny of the cockroach *B. germanica* (Table S1). Data on JH and 20E for the chosen stages are from Treiblmayr et al. (2006) (JH in nymphal stages), Maestro et al. (2010) (JH in embryo stages), Cruz et al. (2003) (20E in nymphal stages) and Piulachs et al. (2010) (20E in embryo stages). Tanaka stages are from Tanaka (1976). Total RNA was extracted using the GenElute Mammalian Total RNA kit (Sigma) following the manufacturer's protocol. Up to 10 μg of total RNA from pooled samples were used to prepare the libraries. The mRNAs were isolated by magnetic beads using the Dynabeads® Oligo (dT)25 (Invitrogen, Life Technologies) and following the manufacturer's protocol. Quality and quantity of mRNAs were assessed with a Bioanalyzer (Aligent Bioanalyzer® 2100). Libraries were prepared using NEBNext mRNA library Prep Master Mix Set for Illumina sequencing (New England Biolabs), and sequenced with 6 multiplexed runs of Illumina MiSeq. We did paired-end sequencing, with read length of 300 nucleotides. To avoid batch effects, replicates were never multiplexed together in the same run. We made all the datasets publicly available at Gene Expression Omnibus (Edgar et al. 2002) under the accession code GEO: GSE99785. For comparisons, we used an equivalent RNA-seq dataset of *D. melanogaster* (GEO: GSE18068) comprises 22 libraries from 11 developmental stages (2 replicates each) covering the entire embryo development, the three larval stages, the pupa, and the adult female. In postembryonic stages we followed the correspondence *B. germanica* pre-last nymphal instars with *D. melanogaster* larvae, the last nymphal instar with the pupa (Belles and Santos 2014), and the respective adult female stages. Correspondences between embryo stages of *D. melanogaster* and *B. germanica* are summarized in Table S2.

### Analysis of the RNA libraries

In the *B. germanica* libraries, we removed the adapters and trimmed the low quality bases on the reads extremes using Trimmomatic (version 0.32, relevant parameters: ILLUMINACLIP:"/TruSeq3-PE-2.fa":2:30:10:8:TRUE SLIDINGWINDOW:4:15) (Bolger et al. 2014). RNA-seq data along the development of *D. melanogaster* was retrieved from Gene Expression Omnibus under the accession GSE18068. All the RNA-seq datasets, were mapped to their correspondent insect genome using the STAR software (version 2.3.0, using default parameters) (Dobin et al. 2013) and the table of counts obtained with the R implementation of featureCounts (version 1.22.3, relevant parameters: allowMultiOverlap=T, countMultiMappingReads=T, useMetaFeatures=T) (Liao et al. 2014), using the corresponding gene annotation of each insect (Dataset S2 and S3). The genome assembly of *B. germanica* and corresponding gene annotations are available from NCBI bioproject under the accession code PRJNA427252. Regarding *D. melanogaster*, we used the genome assembly and gene annotation version "dmel_r6.12", available in Flybase (http://flybase.org/). For clustering purposes reads were normalized with the "varianceStabilizingTransformation" function implemented at DESeq2 R package (version 1.12.4) (Love et al. 2014), for gene expression profiles and visualization (e.g. heatmaps and bar plots) we used the FPKM normalization.

**Functional annotation of genes**

Using the protein sequence, functional annotation was obtained using PfamScan (version 1.5, Database Pfam-A, release 30.0) (Bateman et al. 2004; Li et al. 2015). Then, we selected those genes with a Pfam motifs unequivocally related to TF activity (de Mendoza et al. 2013; Ylla and Belles 2015). GO-terms were retrieved for the *D. melanogaster* genes with the AnnotationForge package (version 1.14.2) (Carlson and Pages 2016), and used for the corresponding *B. germanica* orthologues. Orthologous genes shared by *B. germanica* and *D. melanogaster* where obtained by following the Blastp (version 2.5.0+) (Camacho et al. 2009) reciprocal best hits (BRBHs) strategy (Rivera et al. 1998) (Dataset S4). For the Hox genes, we aligned the candidate of *B. germanica* protein sequences with the eight canonical Hox genes (Negre and Ruiz 2007) of different insect species with ClustalX (Larkin et al. 2007). Then, we performed a phylogenetic reconstruction with RAxML (within CIPRESScience Gateway, version V8.2X) (Stamatakis 2014), which is based on the maximum-likelihood principle, a JTT matrix, a gamma model of heterogeneity rate, and using empirical base frequencies and estimating proportions. The data was bootstrapped for 100 replicates. The accession

codes in both insects of these Hox genes, and that of other manually curated orthologous genes, are detailed in Table S3. The enrichments analysis tests for GO-terms was performed on the subset of expressed genes at each stage (>1FPKM) using the hypergeometrical test implemented in the GOstats package (version 2.38.1, relevant parameters: ontology = biologica process) (Falcon and Gentleman 2007), while the enrichment analysis for Pfam motifs was done on the same subset of genes with the hypergeometric test implemented in R (relevant parameters:phyper()) (R Development Core Team 2011).

**Differential expression analysis**

The differential expression analyses tests were performed with all genes using the DESeq2 package (Love et al. 2014). The obtained P-values were adjusted for multiple testing using the FDR (False Discovery Rate), and the threshold for significant expression change was set at an adjusted p-value < 0.05.

**DNA methylation**

We calculated the ratio between the observed frequency of CpG and the expected frequency in the gene body of each annotated gene (Elango et al. 2009).

$$\mathrm{CpGo/e} = \frac{f(CG)}{f(C) * f(G)}$$

The regression between $CpG_{o/e}$ of each gene and their expression level at each library was tested in R using the Pearson's product moment correlation coefficient.

**Quantification of mRNA levels of Zelda by qRT-PCR**

Quantitative real-time PCR (qRT-PCR) was carried out in an iQ5 Real-Time PCR Detection System (Bio-Lab Laboratories), using SYBR®Green (iTaq $^{TM}$Universal SYBR® Green Supermix; Applied Biosystems). Reactions were triplicate, and a template-free control was included in all batches. Primers used to detect Zelda mRNA levels were: TGTCCCAAACAGTTCAACCA (forward) and AAAGGGTTTCTCTCCCGTGT (reverse) designed on the sequence deposited in GenBank under the accession code LT717628.1. We validated the efficiency of each set

of primers by constructing a standard curve through three serial dilutions. In all cases, levels of mRNA were calculated relative to BgActin-5c mRNA levels,which were measured using the primers AGCTTCCTGATGGTCAGGTGA (forward) and TGTCGGCAATTCCAGGGTACATGGT (reverse), based on the sequence with the GenBank accession code AJ862721. Three biological replicates per point were measured and averaged, and results were calculated as copies of Zelda mRNA per 100 copies of BgActin-5c mRNA.

**Phylogenetic analysis of Hox proteins**

Sequences used were obtained by Blast from GenBank or from i5k project (https://i5k.nal.usda.gov/webapp/blast/)andfromFlybase(http://flybase.org/). Alignments were carried out with ClustalX (Larkin et al. 2007) and phylogenetic reconstruction with RAxML (Stamatakis 2014), based on the maximum-likelihood principle, a JTT matrix, a gamma model of heterogeneity rate, and using empirical base frequencies and estimating proportions. The data was bootstrapped for 100 replicates. The sequences used for comparison with those of *Blattella germanica* were: *Acromyrmex echinatior* EGI64564.1 (proboscipedia, pb); *Anopheles gambiae* XP_311623.2 (Ultrabithorax, Ubx); *Anoplophora glabripennis* XP_018562491 (Sex combs reduced, Scr); *Biston betularia* ADO33070.2 (Ubx); *Bombyx mori* NP_001107632.1 (Ubx); *Callimorpha dominula* AIB07881.1 (Deformed, Dfd); *Camponotus floridanus* EFN67233.1 (abdominal-A, abd-A); *Drosophila melanogaster* FBgn0000014 (abd-A),NP_524896 (Abdominal-B, Abd-B), FBgn0260642 (Antennapedia, Antp), FBgn0000439 (Dfd), FBgn0002522 (labial, lab),FBgn0051481(pb), FBgn0003339 (Scr), FBpp0082793 (Ubx); *Harpegnathos saltator* XP_011148886.1 (abd-A), EFN88927.1 (Scr); *Lasius niger* MQ93049.1 (lab); *Lucilia cuprina* KNC34760.1 (Antp); *Megachile rotundata* XP_012154280.1 (lab); *Operophtera brumata* KOB75113.1 (Abd-B); *Periplaneta americana* ADF35697.1 (Scr); *Schistocerca americana* AAB03236.1 (Antp); *Tribolium castaneum* AAB70263.1 (abd-A), NP_001034519.1 (Abd-B), NP_001107762.1 (lab), AAF03888.1 (pb), AAG13009.1 (Scr); *Zootermopsis nevadensis* KDR16991.1 (abd-A), KDR11585.1 (Antp), XP_021919824.1 (Dfd), KDR19418.1 (lab), KDR19417.1 (pb), KDR19415.1 (Scr).

**References**


Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., et al. (2004). The Pfam protein families database. Nucleic Acids Res. 32, D138-41.

Belles, X. and Santos, C. G. (2014). The MEKRE93 (Methoprene tolerant-Krüppel homolog 1-E93) pathway in the regulation of insect metamorphosis, and the homology of the pupal stage. Insect Biochem. Mol. Biol. 52, 60–68.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, btu170-.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009). BLAST+: architecture and applications. BMC Bioinformatics 10, 421.

Carlson, M. and Pages, H. (2016). AnnotationForge: Code for Building Annotation Database Packages. Available online at https://bioconductor.org/packages/release/bioc/html/AnnotationForge.html/.

Cruz, J., Martín, D., Pascual, N., Maestro, J. L., Piulachs, M. D. and Bellés, X. (2003). Quantity does matter. Juvenile hormone and the onset of vitellogenesis in the German cockroach. Insect Biochem. Mol. Biol. 33, 1219–1225.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Edgar, R., Domrachev, M. and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 30, 207–210.

Elango, N., Hunt, B. G., Goodisman, M. A. D. and Yi, S. V (2009). DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. Proc. Natl. Acad. Sci. U. S. A. 106, 11206–11211.

Falcon, S. and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. Bioinformatics 23, 257–258.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y. M., Buso, N. and Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. 43, W580-4.

Liao, Y., Smyth, G. K. and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930.

Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Maestro, J. L., Pascual, N., Treiblmayr, K., Lozano, J. and Belles, X. (2010). Juvenile hormone and allatostatins in the German cockroach embryo. Insect Biochem. Mol. Biol. 40, 660–665.

de Mendoza, A., Sebé-Pedrós, A., Sestak, M. S., Matejcic, M., Torruella, G., Domazet-Loso, T. and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. U. S. A. 110, E4858-4866.

Negre, B. and Ruiz, A. (2007). HOM-C evolution in *Drosophila*: is there a need for Hox gene clustering? Trends Genet. 23, 55-59.

Piulachs, M.-D., Pagone, V. and Belles, X. (2010). Key roles of the Broad-Complex gene in insect embryogenesis. Insect Biochem. Mol. Biol. 40, 468–475.

R Development Core Team (2011). R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available online at http://www.R-project.org/.

Rivera, M. C., Jain, R., Moore, J. E. and Lake, J. A. (1998). Genomic evidence for two functionally distinct gene classes. Proc. Natl. Acad. Sci. U. S. A. 95, 6239–6244.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Tanaka, A. (1976). Stages in the embryonic development of the German cockroach, *Blattella germanica* Linné (Blattaria, Blattellidae). Kontyû, Tokyo 44, 1703–1714.

Treiblmayr, K., Pascual, N., Piulachs, M.-D. M. D. M.-D., Keller, T. and Belles, X. (2006). Juvenile hormone titer versus juvenile hormone synthesis in female nymphs and adults of the German cockroach, *Blattella germanica*. J. Insect Sci. 6, 1–7.

Ylla, G. and Belles, X. (2015). Towards understanding the molecular basis of cockroach tergal gland morphogenesis. A transcriptomic approach. Insect Biochem. Mol. Biol. 63, 104–112.