Comparative transcriptomics of hemimetabolan and holometabolan metamorphosis

Guillem Ylla Bou

TESI DOCTORAL UPF / 2017

DIRECTORS DE LA TESI

Dr. Xavier Bellés Ros

Dra. Maria-Dolors Piulachs Bagà

Institut de Biologia Evolutiva

(CSIC - Universitat Pompeu Fabra)

A tots els que d'una forma o altra n'heu format part.

Agraïments

Vull començar donant les gràcies als meus directors de tesi, per l'oportunitat que em van brindar en permetre'm dur a terme aquest doctorat. Vull agrair-los-hi haver-me transmès la seva passió per la ciència, les seves ensenyances i la seva paciència. Als dos, Xavier Bellés i Maria Dolors Piulachs, moltes gràcies. També vull mostrar el meu agraïment a en José Luís Maestro, per les discussions científiques que hem tingut de les quals tant he après i tant útils han estat per a aquesta tesi.

Són molts els companys que han passat pel laboratori P64 en aquests 5 anys; Alba H. Alba V., Alberto, Ana, Aníbal, Carlos, Carol, Elena, Jaume, Jimena, Jesús, Mahboubeh, Moysés, Nashwa, Natalia, Paula, Raúl, Sarai, Sheila, Tim, Viviana i altres visitants de curta estada. A tots, moltes gràcies pel que m'heu ensenyat i per les estones que hem compartit. Vull agrair també l'ajuda dels col·laboradors que he tingut, en especial a en Bastian Fromm, qui em va ajudar a submergir-me en el món dels microRNAs i a en Mark Harrison, pels seus comentaris constructius sobre els meus manuscrits.

Aquesta tesi molts cops ha sobrepassat el camp professional, i els seus efectes sovint els han «patit» les persones del meu voltant. Thank you Aleksandra to be always there. Thank you for your support. Thank you very much for your patience with my grumpy days when results were not as expected. Thank you for your time

reading and checking spelling mistakes of this thesis. Dziękuję Ci bardzo.

Vull també recordar i donar les gràcies a tots els que heu fet que hagi pogut arribar fins aquí. Moltes gràcies a la meva família, especialment als pares, germà i àvia, per haver-vos tingut sempre a prop i haver-me ajudat sempre. En vosaltres sempre he tingut el suport necessari, sempre m'heu animat i recolzat, moltes gràcies.

A tots vosaltres, a tots els que ho heu fet possible, moltes gràcies.

Abstract

The evolutionary success of insects was particularly shaped by the innovation of the metamorphosis, especially by the transition from hemimetaboly to holometaboly. The mechanisms underlying this evolutionary transition represent an unsolved question, although different approaches have been used to study them. In the present thesis we followed a transcriptomic approach, comparing data on mRNA and miRNA expression in key developmental moments, comprising embryonic and postembryonic stages, in species representing the hemimetabolan and holometabolan modes. Most of the work has been carried out in the hemimetabolan species *Blattella germanica*, but we have used also other reference species for comparison, especially the holometabolan *Drosophila melanogaster* and *Tribolium castaneum*. The results show that there are not qualitative gene differences between holometabolan and hemimetabolan species, but differences in patterns of expression and potential networking of orthologous genes. Transcription factors, epigenetic modifiers, and miRNAs appear as important players in both developmental modes.

Resum

L'èxit evolutiu dels insectes ha estat marcat per la innovació de la metamorfosi i, en especial, per la transició de la metamorfosi hemimetàbola a holometàbola. Els mecanismes subjacents en aquesta transició evolutiva representen una qüestió no resolta. Per tal d'estudiar aquesta transició, en aquesta tesi hem utilitzat un enfocament transcriptomic comparant dades de mRNA i miRNA en estadis clau del desenvolupament, incloent-hi estadis embrionaris i post embrionaris en espècies representatives de metamorfosis hemimetàbola i holometàbola. La major part dels anàlisis s'han centrat en l'hemimetàbol *Blattella germanica*, tot i que s'han utilitzat dades d'altres espècies com a contrast, especialment dels holometàbols *Drosophila melanogaster* i *Tribolium castaneum*. Els resultats mostren que no hi ha diferències qualitatives en relació a gens dels hemimetàbols i holometàbols, en canvi les principals diferències consisteixen en els diferents perfils d'expressió de gens comuns i la seva xarxa de d'interacció. Els factors de transcripció, els modificadors epigenètics i els miRNAs emergeixen com a principals protagonistes dels mecanismes reguladors en ambdós models de desenvolupament.

Preface

Discovering the genetic mechanisms that triggered the origin and evolution of insect metamorphosis is the ambitious aim of the Bellés lab, while the objective of the Piulachs lab is to discover and understand the different reproductive strategies that insects adopted along evolution. Therefore, the general approach followed in this thesis is to gather a robust information on this subject in a phylogenetically basal insect (the German cockroach, *Blattella germanica* is the model used in the laboratory), which shows a gradual, hemimetabolan mode of metamorphosis, and compare this data with equivalent information available in phylogenetically distal species, showing a complete, holometabolan metamorphosis mode, which has been thoroughly studied (for example, the fly *Drosophila melanogaster*). In this context, the objective of the present thesis has been to add new light on the mechanisms that control insect metamorphosis and evolution following the approach of the comparative analysis of "omics" data in *B. germanica* and in other insect species. The starting point has been the analysis of twenty-two transcriptomes covering the most significant transitions in the ontogeny of *B. germanica*. In parallel, equivalent libraries of small RNAs were also prepared and analyzed, in order to study the microRNAs operating in each transition. The results of these analyses, and the comparison with equivalent data available in *D. melanogaster* and, occasionally, in the beetle *Tribolium castaneum*, constitute the essential basis of our work.

Table of Contents

1. INTRODUCTION

"… From so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved."

- Charles Darwin

With more than 1 million described species, insects represent about 90% of the total number of animal species (Grimaldi et al. 2005). Insects are one of the most diverse lineages on Earth in the number of species and morphological variety, which allowed them to colonize most of the terrestrial ecosystems (Engel et al. 2004). The long evolution of insects explains, in part, their success.

1.1. Insect evolution

According to recent phylogenetic studies (Misof et al. 2014), insects originated in the Early Ordovician period (Figure 1), around ~479 million years ago (Mya), although the first insect fossils available date from the Early Devonian (~412 Mya). It is believed that insects were among the first animals to colonize terrestrial and fresh water ecosystems, the first animals to fly and also the first ones to develop social behavior (Grimaldi et al. 2005).

Although some small branches of the insects phylogeny are still not well resolved, the evolutionary relationships between the main insect orders are well established and supported. Insect orders show an increase in organism complexity from the less modified Palaeoptera, to Polyneoptera, and Paraneoptera, until to the more

modified and diverse Endopterygota (Grimaldi et al. 2005).

With the origin of Endopterygota (which emerged ~345 Mya) occurred the most dramatic expansion of species and genera. This expansion was accompanied by the emergence of a new kind of metamorphosis, namely holometabolan metamorphosis. As a consequence of Endopterygota expansion, most insect species (ca. 83%) follow the holometabolan type of metamorphosis, exhibiting a spectacular range of morphologies and lifestyles. Hence, explaining how holometabolan metamorphosis originated is equivalent to explain how most of the present insect diversity emerged.

1.2. Insect metamorphosis. The transition from juvenile stages to the adult

Insect metamorphosis has fascinated humans from all civilizations across the centuries as can be observed in ancient art and manuscripts. In the ancient Egypt, insect metamorphosis was considered a representation of resurrection, and scarabs were represented in Pharaohs' tombs. Aristotle, in the ancient Greece, wrote that the "caterpillar is nothing more than a soft egg", thus hypothesizing that the larva is still part of the embryogenesis.

Darwin also wrote about insects' metamorphosis in his book "The Voyage of the Beagle" highlighting that in those times, insect metamorphosis was considered a kind of witchcraft in some cultures.

Modern science, from the 19th century onwards, left behind part

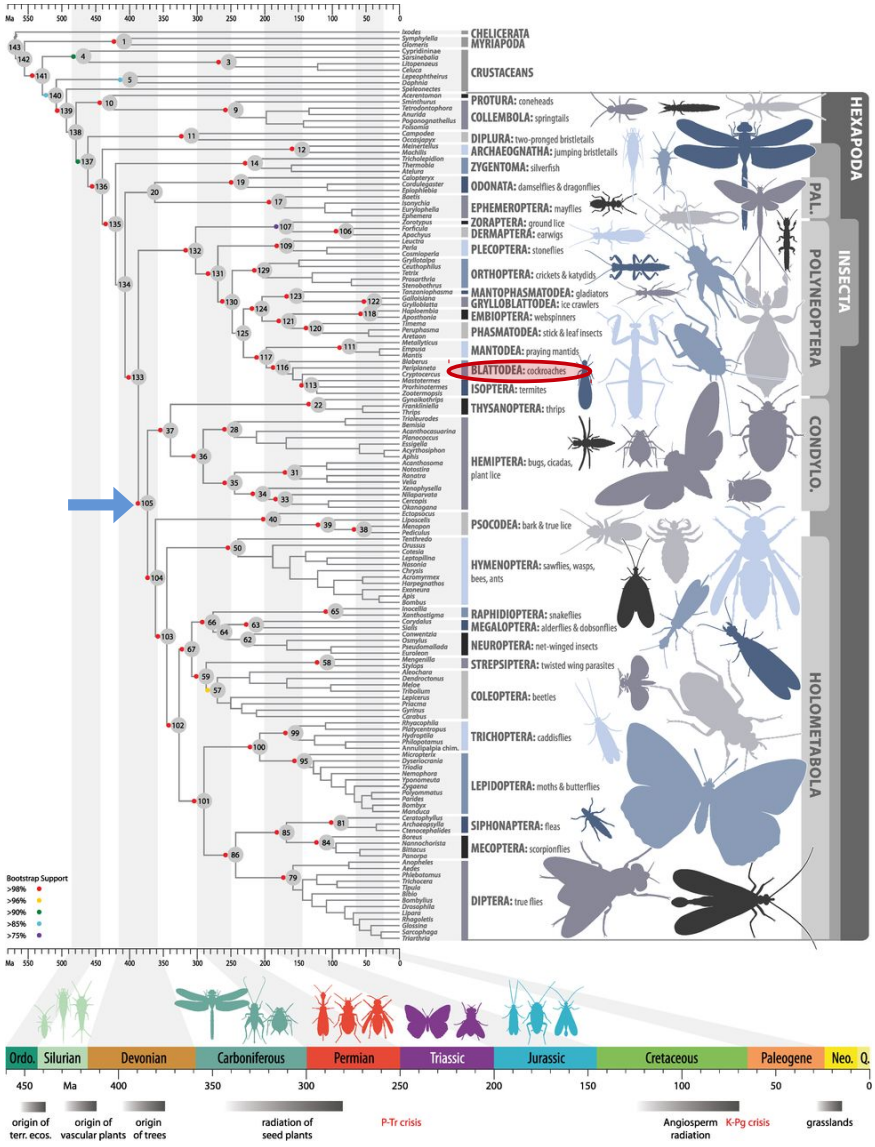

Figure 1: Phylogenetic tree of insects modified from Misof et al. (2014). The Blattodea branch, which contains *Blattella germanica*, is highlighted in red and the monophyletic origin of holometabolan metamorphosis is indicated with a blue arrow.

of the mystery associated with the insects metamorphosis. Nowadays, the general regulatory elements of insect metamorphosis, especially the hormonal factors, have been elucidated. Nevertheless, how the metamorphosis is regulated at a gene expression level is not yet fully understood.

Around 90% of the presently living insect species show some kind of metamorphosis along postembryonic development. The metamorphosis, from the Greek "transformation", refers to a biological process in which an animal changes its body form. Depending on the degree of abruptness of the changes, insect metamorphosis has been classified into three types (Figure 2):

a) Ametabolan, by which the changes of body morphology are practically nonexistent, and adults and juvenile stages differ only on size, and on minor details, like the genitalia present in the adult. The adult continues molting after acquiring the reproductive capabilities.

b) Hemimetabolan, by which the juvenile stages, or nymphs, and the adult differ not only in size but also in few but significant morphological characters, like the presence of fully developed wings and genitalia in the adult. The adult, once formed after metamorphosis, does not molt anymore.

c) Holometabolan, which has 3 different stages with completely different morphology, such as larvae, pupa and adult. The larva shows a morphology very different from that of the adult and requires the intermediate pupal stage to

bridge the transition to the adult. The adult, once formed after metamorphosis, do not molt anymore.

From an endocrinological point of view, two main hormones regulate insect metamorphosis. From one side, the 20-hydroxyecdysone (Ecd) is responsible for promoting the successive molts of the insect life cycle. While the Ecd peak occurs in presence of Juvenile hormone (JH), the molt results in another juvenile stage, but when Ecd secretion occurs without the presence of JH, the insect molts to the adult form. Therefore, Ecd and JH can be considered as key endocrine regulators of metamorphosis.

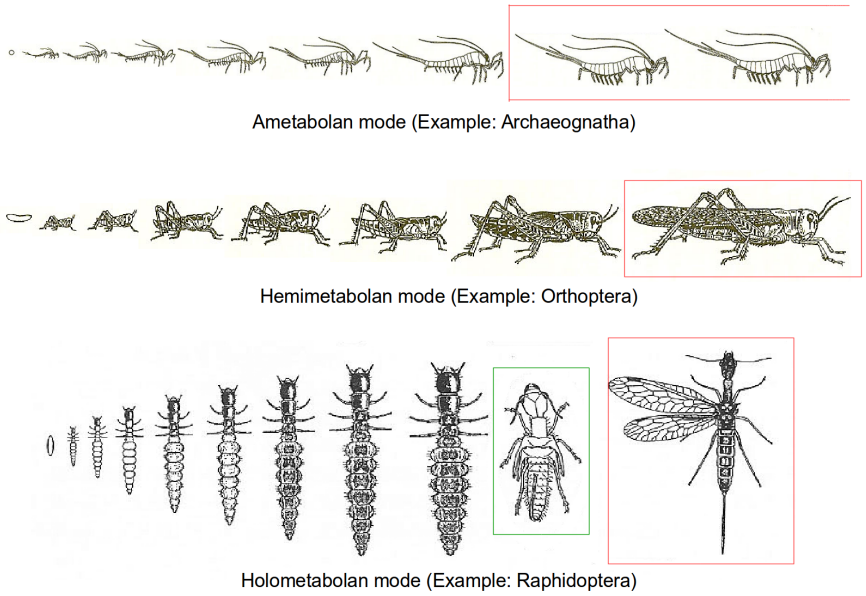

Figure 2*:* The main types of insect metamorphosis, ametabolan, hemimetabolan and holometabolan. Quiescent stage of holometabolan is framed in green. Adult, reproductively competent stages are showed in a red square. In the ametabolans, the red square is open because the adult continues molting. Modified from Belles (2011).

1.3. Insect embryo development

The embryo development of hemimetabolan insects results in a nymphal morphology, which already exhibits the essential adult body structure although some organs, such as wings and sexual organs, will not be developed until the adult stage. In contrast, holometabolan embryogenesis results in a larval morphology, usually vermiform, thus dramatically divergent from the adult body structure. Due to these differences in the output of embryogenesis in both metamorphosis types, we focused part of our analyses on the embryo development.

The oocytes developed by the adult female contain a single copy of the haploid genome plus nutrients and other molecules needed for embryo development. Especially important among these molecules are the maternal mRNAs, which are crucial to start the development of the embryo in their most early stages.

The egg fertilization occurs when the spermatozoon nucleus fuse with the oocyte nucleus resulting in a diploid cell. Then, the first important process that proceeds is the so-called maternal to zygotic transition (MZT). The MZT is characterized by the depletion of the maternal mRNAs and the transcription activation of the zygotic genome (Schier 2007; Tadros et al. 2009).

Regarding the MZT, the main factors involved in maternal mRNA degradation in insects are the protein Smaug and the miRNAs belonging to the Mir-309 family (Bushati et al. 2008; Chen et al. 2014). By contrast, the protein Zelda (also known as Vielfaltig), has been shown to play a crucial role in the activation of

the zygotic genome in *Drosophila melanogaster* (Schulz et al. 2015; Foo et al. 2014; Sun et al. 2015; Nien et al. 2011; Laver et al. 2015).

After the MZT, the orientation of the embryo is defined on a three-dimensional space where the segmentation pattern is determined. The entire process, from egg polarity to embryo segmentation, is determined by the successive expression of gap genes, pair-rule genes and segmentation genes. During this cascade of gene expression, the germ-band is defined and the body segments are determined. In the case of insects, there are two different kinds of germ-band (Liu et al. 2005): long and short germ-band (Figure 3).

Short germ-band development is typical of basal insects, such as locusts and cockroaches, whereas more derived species like most holometabolan insects predominantly follow the long germ-band development (Peel et al. 2005). The differences between these two developmental extremes are caused by a different expression of the patterning genes (Liu et al. 2005; Lynch et al. 2012). Long germ-band insects, such as the fruit fly *D. melanogaster* form the body segments simultaneously at the blastoderm stage, i.e., before gastrulation. In contrast, short germ-band species start gastrulation with just head and thoracic segments, and perhaps a few abdominal segments, and then progressively add new abdominal segments from a growth zone located at the distal end of the embryo (Chipman 2015).

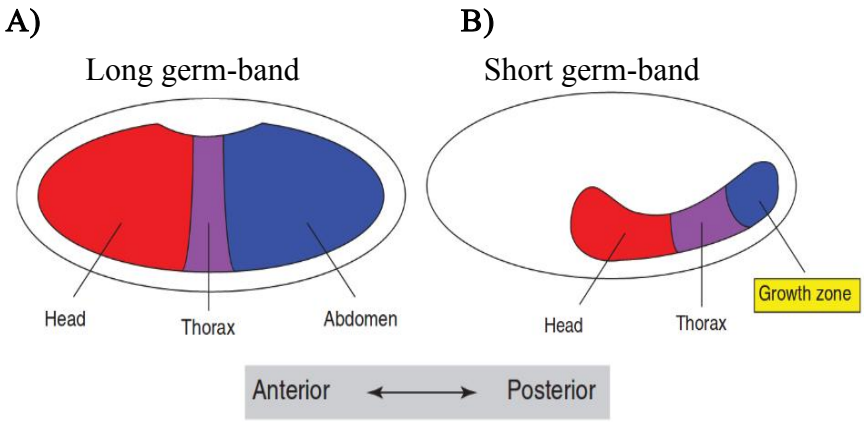

Figure 3: Germ-band types in insects. A) Long germ-band, which occupies all the oocyte length and all the segments are formed simultaneously at the blastoderm stage. B) short germ-band, which is formed on one extreme of the egg and constitutes the head and the thoracic segments; in addition, in the distal part of the germ-band successive abdominal segments will be formed from a growth zone. Image modified from Kimelman and Martin (2012).

After the germ-band formation, segmentation and eventually abdomen elongation, the body plan of the embryo is basically established. Then, the Hox genes are the main specifiers of the specific body structure (Hrycaj et al. 2016; Averof et al. 1995). Hox genes constitute a family that contains eight genes coding for homeobox-domain transcription factors. These genes usually cluster together in the insects genomes, and this synteny is conserved across distant species, such as insects and mammals, although the later have four copies of the Hox genes cluster, whereas insects have only one (Pace et al. 2016). The identity of the different segments is defined by a differential expression gradient of each Hox gene along the embryo. Interestingly, this mechanism for providing the segment identity through an expression gradient of Hox genes is also conserved across distant species (Figure 4).

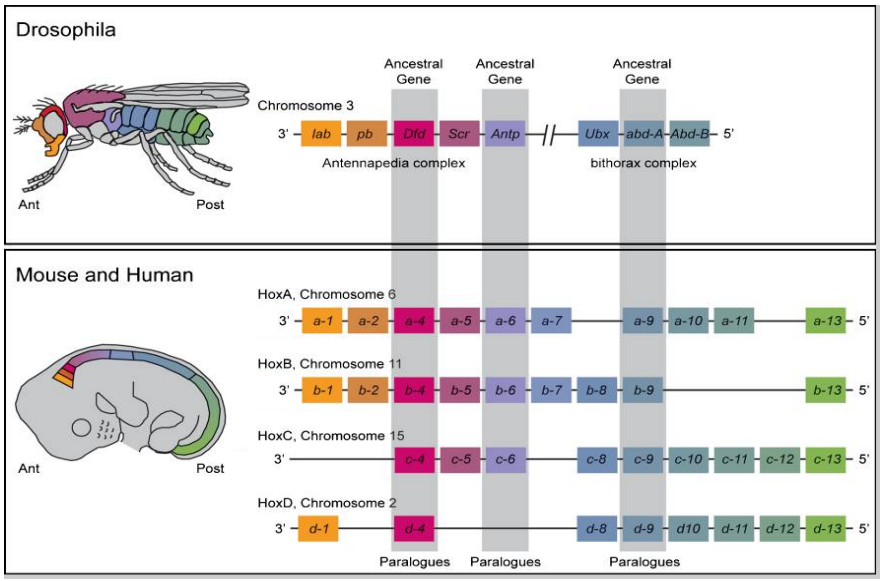

Figure 4: The Hox gene family synteny on chromosome 3 of *Drosophila melanogaster* and the segments where each Hox gene is expressed. The same Hox genes are conserved on mammals, which had four duplications of the Hox cluster. In both cases, each Hox gene is expressed on a delimited region of the embryo giving its identity to this region. Image modified from Pang and Thompson (2011).

1.4. Insects as experimental models in developmental studies

The use of animals for experimental purposes has been documented since the ancient Greece. Among the first reports of experiments with living animals there are those of Aristotle (384–322 BC) and Erasistratus (304–258 BC) (Fox et al. 2015). However, it was not until the 18th and 19th centuries when animals became a common model for experimentation. In particular, the use of insects as experimental models became popular after Thomas Hunt Morgan's (1866-1945) works with *D. melanogaster* in his famous Fly Room.

1.4.1. *Drosophila melanogaster*

Thomas Hunt Morgan, awarded with the Nobel prize in 1933, used the fruit fly *D. melanogaster* to study the heritability of phenotypic traits. He was a pioneer in describing the heritable units, afterward named chromosomes, and with this finding he set the first milestone of modern genetics.

Since then, *D. melanogaster* became the model par excellence for most of the genetic studies. Hence, nowadays we have a deep understanding of *D. melanogaster* life cycle and embryo development. The knowledge of the genetics and genomics of *D. melanogaster* is extensive, especially since its genome was sequenced and published in 2000 (Adams et al. 2000). The sequencing of the *D. melanogaster* genome represented the first available insect genome, and since that moment the amount of high-throughput sequencing (HTS) data has become almost overwhelming. For example, the modEncode project (Celniker et al. 2009) has already produced more than 1,400 HTS datasets on different *Drosophila* species, with the objective of identifying all of the sequence-based functional elements of different species of this fly genus.

D. melanogaster is an endopterygote insect that follows the holometaboloan mode of metamorphosis, and has a relatively short life cycle. Embryo development takes around 24 hours, counting from egg fertilization until larva hatching, at 25ºC. The larva molts 2 times during the next 8 days, thus having 3 larval stages before the pupal stage. The pupal stage lasts around 6 days, during which the

adult insect is formed. Finally, the adult fly emerges from the pupa exhibiting all the typically adult morphological characters, such as wings, compound eyes and genitalia.

1.4.2. *Tribolium castaneum*

The red flour beetle *Tribolium castaneum*, a worldwide pest of stored products (Grünwald et al. 2013), had been largely used as a representative model of the large insect order of Coleoptera (Wang et al. 2007).

The *T. castaneum* genome sequence was published in 2008 (Richards et al. 2008). The genome sequences and the gene annotations were hosted by the BeetleBase (Wang et al. 2007), which later became a centralized database for different kinds of sequencing data of *T. castaneum* (Kim et al. 2010).

An interesting feature of *T. castaneum* is that the embryo follows short germ-band development, although the species presents the holometabolan mode of metamorphosis. Even though the long germ-band emerged earlier than the holometabolan metamorphosis, and most of the holometabolans undergo the long germ-band embryogenesis, some coleopterans like *T. castaneum*, are short germ-band. This feature made the *T. castaneum* the model of choice of several researchers for studying the evolution of embryo development in insects.

At ~27ºC, the embryo of *T. castaneum* develops during 5 days until it hatches as a first instar larva. The larval life lasts for ca. 27

days and involve between 6 and 7 molts. Then the larva transforms into a pupa and remain in this stage for 5 days, until the imaginal molt.

1.4.3. *Blattella germanica*

In contrast with holometabolan insects, there is much less data and studies based on hemimetabolan insects. Among them, the German cockroach *Blattella germanica* is one of the most well studied at a molecular level, in a great measure owing to the work carried out in the laboratory since the 1980's.

Publicly available data about *B. germanica* was composed by a dozen of RNA sequencing datasets obtained in the laboratory in the recent years (Ylla et al. 2015; Cristino et al. 2011; Rubio et al. 2012). Fortunately, since 2016 we had access to the first assembly of the *B. germanica* genome in which we have later collaborated in the annotation process (Harrison et al. 2017; see Appendix 2).

The Blattodea order, the cockroaches, is composed by more than 4,500 described species living in a wide range of environments (Beccaloni et al. 2013). From fossils and amber inclusions, we know that cockroaches have remained practically unchanged morphologically for hundreds of millions of years. Considering that the morphology is a partial reflection of the genome, we can assume that, genetically, the cockroach genome has not significantly changed for millions of years. Thus, Blattodea species are usually defined as “basal” or close to “ancestral” insects and can be used as baseline for insect evolutionary studies. As an ancestral model,

cockroaches kept both basal metamorphosis (hemimetabolan) and basal germ-band (short).

For all the given reasons, it is important to increase knowledge about *B. germanica*, which has been, consequently, the main model insect used in this thesis.

1.4.4. Termites

Termites also belong to the Blattodea order and are closely related with cockroaches and, thus, with *B. germanica*. In spite of their close phylogenetic distance, as cockroaches and termites diverged 150 Mya (Bourguignon et al. 2015; Legendre et al. 2015), termites exhibit a dramatically derived life story and behavior. While cockroaches are solitary or simply gregarious, termites have developed a complex social organization, including the production of individuals belonging to morphologically different castes.

Regarding termite genomes, in 2014 those of two different species were released. These are the genomes of the *Zootermopsis nevadensis* (Terrapon et al. 2014) and *Macrotermes natalensis* (Poulsen et al. 2014).

1.4.5. Other relevant insect models

In the context of our analyses we used genomic and transcriptomic data of other insect models, which will sporadically appear in different sections of the present thesis.

One of these insects, which has been thoroughly studied due to its economic potential, is the silkworm *Bombyx mori*. This holometabolan species was domesticated in ancient times and more recently has been genetically selected for improving silk production. The sequencing of the *B. mori* genome by two independent research groups in 2004 (Mita et al. 2004; Xia et al. 2004), represented the first published lepidopteran genome. Later, in 2008, both groups collaborated and merged their genomic data for releasing and improved assembly and annotation of the genome (The International Silkworm Genome 2008).

Another holometabolan insect, the honey bee *Apis mellifera*, has been widely studied for its role on crop pollination and honey production, as well as for studying eusociality in holometabolan insects. Its genome was sequenced in 2006 (Honeybee Genome Sequencing Consortium 2006), and additional HTS datasets are publicly available for this insect, including extensive data on miRNAs (Macedo et al. 2016).

Regarding the hemimetabolan species, in addition to *B. germanica*, other relevant insects species are the locust *Locusta migratoria*, with a large genome (6.5 Gb) published in 2014 (Wang et al. 2014), and the pea aphid *Acyrtosiphon pisum* with the genome available since 2010 (International Aphid Genomics Consortium 2010). These two insects are dangerous threads for crop production, and thus, have usually been studied in the context of pest control.

1.5. The “omics” data era

We are currently immersed in the so called “omics era”, based on managing quantitative data related to the genome and their products. Thus, it is the era of genomics, transcriptomics, epigenomics, integromics, etc. After the publication of the first human genome in 2001 (Venter et al. 2001; Lander et al. 2001), the sequencing methodologies experienced a revolution with the release of the “High-Throughput Sequencing” (HTS) technologies (Metzker 2010). Among other things, these technologies dramatically decreased the price and timing of sequencing DNA and RNA molecules. Never before had been so effortless to obtain so much data from biological systems; the challenge now is to make sense out of so much data (Sboner et al. 2011).

1.5.1. Genomics

Consequently, thousands of genomes from different animal and plant species are being released from international projects such as 1,000 human genomes projects (Auton et al. 2015), the 5,000 arthropod genomes project (aka i5k) (I5K-Consortium 2013) or the 1,000 plant genomes project (1KP) (https://sites.google.com/a/ualberta.ca/onekp).

With regard to insects, and as stated above, the first sequenced genome was that of *D. melanogaster* in 2000 (Adams et al. 2000). After that, insect genomes were slowly being sequenced (Figure 5). Then, in 2013, the i5K genome project, accelerated the release of

new insect genomes and consequently, at the end of 2016 there were already 138 insect genomes available (Yin et al. 2016).

Among the first genome drafts released by the i5K was that of *B. germanica*. We have actively participated in the *B. germanica* genome annotation consortium, which annotated this genome and that of the termite *Cryptotermes secundus*, which allowed the comparative genomics analysis between non-social cockroaches and eusocial termites. This collaboration resulted in the manuscript included in this thesis as Appendix 2 (Harrison et al. 2017), which is currently under revision, and available from the preprint server bioRxiv.

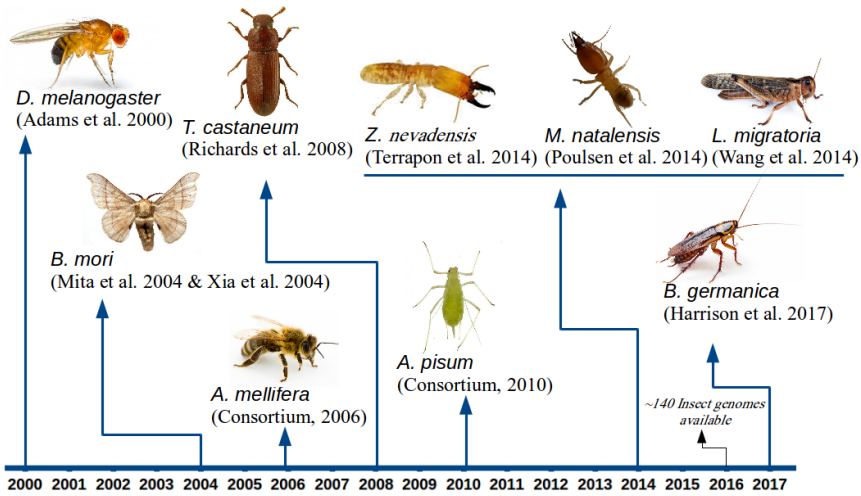

Figure 5: Chronology of the published insect genomes that are relevant for this thesis, since the first insect genome in 2000.

1.5.2. Transcriptomics

The word transcriptomic, refers to the study of the molecules transcribed from DNA, namely RNA. RNA molecules can also be sequenced taking advantage of the HTS technologies. There are especial databases and consortia producing and storing RNA sequencing data (RNA-seq). In the case of insects, the most relevant is the modEncode project (Celniker et al. 2009), which focuses on transcriptomics of different *Drosophila* species. In addition, the transcriptomic data repository "gene expression omnibus" (Edgar et al. 2002) contains RNA-seq data from different species, including a number of insects, uploaded by different laboratories.

In this context, data on hemimetabolan insects is scarce. Considering this unbalance, we obtained and analyzed different kinds of transcriptomic data from our model organism, *B. germanica*, during the work of the present thesis. These data, compared to those available in holometabolan species, gave us a number of clues on the evolution of gene expression regulation in insects.

The two kinds of data used in this thesis are mRNA and small RNA sequencing data. The mRNA data contains quantitative information about the product of coding genes. This is achieved by selecting the RNAs that contain a Poly-A tail prior to the sequencing step. The libraries of small RNAs are obtained by sequencing the RNA fraction containing molecules shorter than 50 nucleotides. This fraction of small RNAs contains regulators of gene expression, such as microRNAs and Piwi-interacting RNAs.

1.6. Gene expression regulatory networks

Comparative genomics has been useful to define phylogenetic relationships among species (Misof et al. 2014), however, it is not so efficient to explain phenotypic variation between close-related species. Even at a species level, genomic analysis cannot fully explain how a single genome produces a diversity of cell types, tissues and organs. In the case of holometabolan species, a single genome contains the information for producing extremely different body morphologies.

Phenotypic differences cannot be explained only on the basis of genomic information because the different ways that cells interpret the genomic information. The first level of "genomic interpretation" consists in the gene expression control. Each cell type expresses a particular set of genes, and modulate their expression in a specific way during development, thus producing a specialized cell type.

A precise regulation of gene expression is essential for the viability of any living organism. Therefore, a number of elements control gene expression at different levels. In the end, these regulatory elements ensure that the exact amounts of the gene products (proteins) are produced, thus determining the phenotype.

The integration of the different elements involved in the regulation of gene expression are commonly called gene regulatory networks (GRNs). Theoretically, it is possible to integrate all possible regulatory elements in a computational GRN model and simulate the functioning of a living cell or even a complete organism. However, nowadays we are still far from this possibility

as our knowledge of the regulatory elements and their precise mechanisms is still limited. Nevertheless, taking advantage of the expansion of the "omics" data, we can start to reconstruct simple GRNs that might help to understand the associated specific phenotypes.

In the context of this thesis, we focus on the following regulatory elements.

1.6.1. Transcription factors

Transcription factors (TFs) are proteins characterized by their function as necessary elements for gene transcription. TFs bind to the promoter region of target genes and recruit the necessary elements for its expression. Consequently, TFs must have at least one domain that can recognize and bind DNA (de Mendoza et al. 2013). Depending on this domain, the TF recognizes a more or less specific DNA sequence, which will determine the genes that can be regulated by it.

In a previous work, we unveiled the importance of TFs on the tergal gland morphogenesis (Ylla et al. 2015) (included in this thesis as Appendix 1) which suggested that they might be also important in the regulation of metamorphosis. Therefore, in this thesis we have intensively analyzed the evolution and expression of Tfs.

1.6.2. microRNAs

Since its discovery in 1993 (Lee et al. 1993; Wightman et al. 1993), microRNAs (miRNAs) have emerged as key regulators of gene expression at post-transcriptional level. miRNAs consists in a single strand RNA of 22-26 nucleotides resulting from the miRNA biogenesis pathway (Figure 6).

The functional, mature miRNA is loaded into the RISC complex and serves as a template for identifying a partially complementary sequence in the target mRNA. The miRNA-RISC complex binds the target mRNA and blocks their translation into protein.

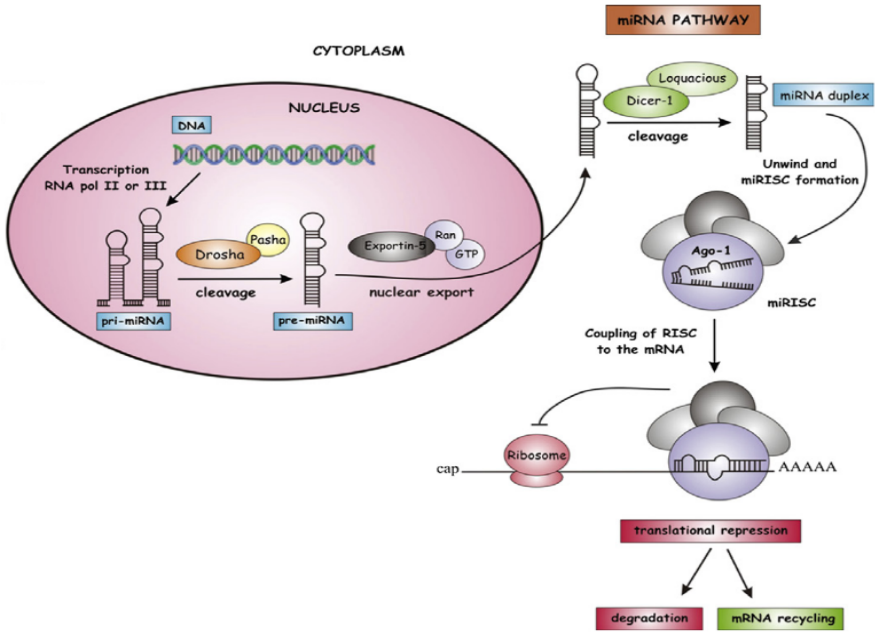

Figure 6: Scheme of the miRNA biogenesis process from Belles et al. 2012. The miRNA gene is transcribed in the nucleus as a miRNA primary transcript (pri-miRNA). Then, the pri-miRNA is processed by Drosha and Pasha enzymes producing the precursor miRNA (pre-miRNA). The pre-miRNA is then exported to the cytoplasm, where it is recognized by Dicer – in insects by Dicer-1 – which gives the mature 22-26 double stranded RNA sequence, which is then loaded to the RISC complex. The RISC complex will retain one of the two RNA strands (the functional, mature miRNA).

1.6.3. Epigenetic factors

Epigenetic modifications are heritable changes on the genetic material that do not directly alter the DNA nucleotide sequence. They form an additional regulatory layer of gene expression, usually acting by exposing or protecting particular DNA stretches for the accessibility of TFs. Epigenetic factors can directly affect the DNA strand, such as the DNA methylators, or affect chromatin elements, like histone modifiers.

DNA methylation is one of the most studied epigenetic processes in animals. It consists on adding methyl groups at the cytosine bases of the DNA, especially, on cytosines followed by guanines, what is known as “CpG” regions. In mammals, methylation has been associated with several biological functions, especially during embryogenesis (Cedar et al. 2012).

In insects, the role of DNA methylation is unclear. It has been reported that DNA methylation may play a role in cast differentiation in eusocial insects (Elango et al. 2009; Bewick et al. 2016). Herein, we show that that DNA methylation may also play an important roles on the regulation of gene expression in very early embryo stages.

2. OBJECTIVES

In the context of the antecedents mentioned in the Introduction, the following specific objectives were designed for the present thesis.

1. The first was to identify the complete catalog of miRNAs in *Blattella germanica*. Then, it could be used as a reference for studying the miRNAs across other insect lineages and derive evolutionary inferences.

2. After the first step of miRNA identification, the next goal planned was to analyze their expression along development of *B. germanica*, using 22 small RNA libraries that cover the entire ontogeny, in order to have information about their possible roles in the different stages of development. In addition, comparing the miRNA expression profiles in different insect species would help to explain the mechanisms accounting for the phenotypic differences found in these species.

3. At the same stages from which we obtained the libraries of small RNAs, we planned to obtain mRNA libraries. The objective was to study the expression of coding genes along the development of *B. germanica*, using the 22 mRNA libraries that cover the entire ontogeny. This would allow comparing the expression of coding genes along the development in our hemimetabolan model with

equivalent publicly available data from holometabolan models. The idea was to find differences in patterns of gene expression that could explain the developmental differences showed by the two metamorphosis modes, not only in postembryonic stages, but also during embryogenesis.

3. RESULTS

The Results chapter of this thesis is composed by three sections, each section corresponding to a scientific article resulting from the research carried out during the doctoral thesis work. Therefore, this thesis contains three scientific articles that the PhD candidate signs as the first author, two of them already published in scientific journals (Scientific Reports and BMC Genomics) while the third one has recently been submitted for publication.

The three articles appear in chronological order of publication. This order should allow the reader to follow and understand the storyline of the thesis. In brief, the story starts with the identification of the *Blattella germanica* miRNAs which led to an analysis of miRNA evolution across insect lineages. This was followed by the study of miRNA expression along the development of *B. germanica*, and the comparison of the miRNA expression in comparable stages of other insect models. Subsequently, we focused on the effectors of the miRNAs, that is, the mRNAs. In this context, we studied the expression of coding genes along the ontogeny of two insects with different metamorphosis modes: the hemimetabolan *B. germanica* and the holometabolan *Drosophila melanogaster*.

3.1. The microRNA toolkit of insects

Previous studies reported the importance of miRNAs in the metamorphosis of *B. germanica* (Lozano et al. 2015). In this article, we took advantage of the *B. germanica* genome and small RNA-seq data for annotating the miRNA complement of *B. germanica*. Then, we used this miRNA complement as a baseline for studying the miRNA evolution across the insect lineages.

In the context of this article, we achieved three major goals; **1)** We obtained a high-quality annotation of the *B. germanica* miRNAs. **2)** We developed a bioinformatic tool, the mirPlot software, which was later applied to re-annotate the miRNA complement of 6 other insects. **3)** We inferred the evolutionary history of miRNAs in insects, which show low rates of gains and losses of miRNA families.

The microRNA toolkit of insects

Guillem Ylla[1]*, Bastian Fromm[2]*, Maria-Dolors Piulachs[1]**, Xavier Belles[1]**

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

2 Department of Tumor Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Nydalen, N-0424, Oslo, Norway.

* These authors contributed equally to this work.

** Correspondence: X.B. (xavier.belles@ibe.upf-csic.es) or M.-D.P. (mdolors.piulachs@ibe.upf-csic.es)

Ylla G, Fromm B, Piulachs M.D., Belles X. The microRNA toolkit of insects. Scientific Reports (2016): 10.1038

3.2. Comparative analysis of miRNA expression during the development of insects of different metamorphosis modes and germ-band types

In the previous section, we defined the conserved miRNA toolkit of insects. We found that major transitions in insect evolution were not accompanied by dramatic gains/looses of miRNAs. The next goal was to study whether the expression pattern of miRNAs along the embryonic and postembryonic development could contribute to determine the type of embryogenesis, and the mode of metamorphosis.

In this article, we analyzed the miRNA expression along the *Blattella germanica* development, paying special attention to embryonic stages. In addition, in order to unveil possible functions related with the germ-band definition and insect metamorphosis, we compared our data on *B. germanica* with public data from other insect models with different developmental features. These models were *Tribolium castaneum,* a holometabolan insect that undergoes the short germ-band embryogenesis, and two different *Drosophila* species, *Drosophila melanogaster* and *Drosophila virilis*, also holometabolan but with a long germ-band embryo development.

Comparative analysis of miRNA expression during the development of insects of different metamorphosis modes and germ-band types

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

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

* Correspondence: X.B. (xavier.belles@ibe.upf-csic.es) or M.-D.P. (mdolors.piulachs@ibe.upf-csic.es)

Ylla G, Piulachs M.D., Belles X. Comparative analysis of miRNA expression during the development of insects of different metamorphosis modes and germ-band types. BMC Genomics (2017): 10.1186

3.3. Clues on the evolution of metamorphosis revealed by comparative transcriptomics of hemimetabolan and holometabolan insects

The following section corresponds to an article recently submitted for publication in a scientific journal. In this article, we analyze mRNA-seq data from the same eleven stages of *Blattella germanica* where we collected small RNA-seq data on the previous work and that cover the entire ontogeny.

We analyzed the developmental transcriptome of the hemimetabolan *B. germanica, which was compared* with that of the holometabolan *Drosophila melanogaster*. We found categorical differences in the expression of a number of genes acting in the maternal to zygotic transition, in the first steps of embryo patterning, in the formation of the nymphal or larval body structures during embryogenesis, and in postembryonic stages.

Clues on the evolution of metamorphosis revealed by comparative transcriptomics of hemimetabolan and holometabolan insects

Guillem Ylla, Maria-Dolors Piulachs, Xavier Belles*

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

* Correspondence: X.B. (xavier.belles@ibe.upf-csic.es)

Ylla G, Piulachs M.D., Belles X. **Comparative transcriptomics of hemimetabolan and holometabolan insects along ontogeny reveal clues on the evolution of metamorphosis.** Submitted

Clues about the evolution of insect metamorphosis revealed by comparative transcriptomics of hemimetabolan and holometabolan species

Guillem Ylla, Maria-Dolors Piulachs, Xavier Belles*

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

Abstract

We obtained and sequenced 22 transcriptomes covering 11 key stages of the ontogeny of the cockroach *Blattella germanica*, a hemimetabolan model, and compared them with an equivalent transcriptomic set from *Drosophila melanogaster*, a holometabolan model. The analysis revealed the following important differences between the two species. In *B. germanica* the most diverse changes in gene expression occur during embryogenesis, whereas *D. melanogaster* maintains a similar level of changes throughout its ontogeny. In *B. germanica*, Smaug and Zelda are acutely expressed in the maternal to zygotic transition, whereas in *D. melanogaster* they maintain high expression during all embryogenesis. DNA methylation appears to operate in *B. germanica* embryonic development, but not in *D. melanogaster*. The expression of gap, pair-rule and segment polarity genes reveals differences between *B. germanica* and *D. melanogaster*; for example hairy is highly expressed in mid-late embryonic stages of *B. germanica* but not in

D. melanogaster. With regard to Hox genes, Abd-B is highly expressed in mid-late embryonic stages of *D. melanogaster* but not *B. germanica*. The expression of transcription factors shows many quantitative and qualitative differences between the two species. Finally, the expression of genes related to apoptosis is high in early to mid-embryonic stages of *B. germanica*, whereas in *D. melanogaster*, it is high in practically in all embryonic and postembryonic stages. These differences may account for the different metamorphosis modes, hemimetabolan in *B. germanica* and holometabolan in *D. melanogaster*, and bring new light on the mechanisms underlying the evolution from hemimetaboly to holometaboly.

Introduction

With around 1 million species described, insects are the most diverse animal lineage on Earth (I5K-Consortium, 2013), having colonized practically all terrestrial and freshwater ecosystems on all continents. Part of this evolutionary success is due to the innovation of metamorphosis (Nicholson et al. 2014), by which the individual acquires adult features and becomes reproductively competent during postembryonic development. The first innovation in this sense was hemimetaboly, characterized by a type of embryogenesis that generates a first instar nymph displaying the essential adult body structure, and which grows gradually during the juvenile stages. The final molt into an adult completes the formation of the wings and functional genital structures. From an ancestral hemimetabolan group, some 350 Mya, holometaboly emerged. This is a type of metamorphosis in which embryogenesis gives rise to a

larva with a body structure considerably divergent from that of the adult. The larva then 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). Holometabolan metamorphosis was a very successful innovation that was accompanied by a large radiation of the insect lineage (Misof et al. 2014), demonstrated by the fact that, today, more than 80% of currently known insects undergo this kind of metamorphosis (Grimaldi et al. 2005).

Despite the importance of insect metamorphosis, the mechanisms underlying it are far from fully understood. The holometabolan mode has been studied in greater depth, particularly thanks to the intensive research carried out on the fruit fly, *Drosophila melanogaster*, the model *par excellence* for genetic studies since the beginning of the 19th century. For example, the modEncode project (Celniker et al. 2009), has made extensive high throughput sequencing data available from this and other Drosophilid species. In contrast, hemimetabolan metamorphosis has received much less attention, and this is a serious drawback as good knowledge of this process is essential for understanding the evolution of the more successful holometaboly. In order to partially cover this gap, we have produced extensive transcriptomic data along the ontogeny of the German cockroach, *Blattella germanica*, a well-known hemimetabolan model, whose genome has been recently made available (https://www.hgsc.bcm.edu/arthropods/german-cockroach-genomeproject). In particular, we have produced 22 transcriptomes covering 11 different stages along the ontogeny of *B. germanica*, six

of them corresponding to embryo development. This emphasis in embryogenesis is important as we are convinced that the comparative study of hemimetabolan and holometabolan embryo development can afford important clues regarding the mechanisms differentiating both metamorphosis modes.

Embryo development begins with the maternal to zygotic transition (MZT), during which maternal mRNAs are cleared and the zygotic genome is activated (Schier 2007; Tadros et al. 2009). In insects, the main factors contributing to maternal mRNA clearance are the protein Smaug and the MIR-309 family of miRNAs (Tadros et al. 2007; Ylla et al. 2016; Bushati et al. 2008; Chen et al. 2014). In contrast, the protein Zelda (=Vielfaltig), plays a crucial role in zygotic genome activation, at least in *D. melanogaster (Schulz et al. 2015; Foo et al. 2014; Sun et al. 2015; Nien et al. 2011; Laver et al. 2015)*. A process that could be important in the MZT is DNA methylation. Given that methylated cytosines tend to experience spontaneous deamination, DNA methylation can be studied indirectly given that genomic CpG depletion (CpGo/e) strongly correlates with highly methylated DNA regions (Bewick et al. 2016; Glastad et al. 2013; Park et al. 2011). During early patterning, gap genes are expressed under the control of the maternal mRNAs and typically encode transcription factors that control the expression of pair-rule genes. In turn, pair-rule genes activate the expression of the segment polarity genes, thus determining the general polarity of the embryo (Peel et al. 2005). Subsequently, the Hox genes are expressed and play key roles in morphogenesis and body structure shaping (Averof et al. 1995).

Embryo development results in a nymph or larvae in hemimetabolan and holometabolan species, respectively. Subsequently, postembryonic development up to the adult stage is mainly regulated by ecdysteroids, which promote the successive molts, and juvenile hormone (JH), which essentially represses the adult molt until a critical size is reached (Belles 2011). Ecdysteroids and JH are synthesized through a number of enzymatically regulated steps (Niwa et al. 2014; Bellés et al. 2005), whose expression can be modulated by different factors, like myoglianin, which represses the expression of JH acid *O*-methyl transferase, the last and key enzyme of JH biosynthesis (Ishimaru et al. 2016). In turn, the action of ecdysone and JH is mediated by transcription factors that transduce the hormonal signal (King-Jones et al. 2005; Jindra et al. 2013). Recent studies have shown that the molecular action of JH on metamorphosis is based on the MEKRE93 pathway (Belles et al. 2014; Jindra et al. 2015). In pre-last juvenile instars, the MEKRE93 pathway starts with JH interactig with its receptor, Methoprene-tolerant (Met) and stimulating the expression of the transcription factor, Krüppel homolog 1 (Kr-h1); this represses another transcription factor, E93, which triggers adult morphogenesis. Thus, when JH vanishes in the last juvenile instar, E93 becomes de-repressed and metamorphosis proceeds (Belles et al. 2014). Many other transcription factors are crucial for regulating developmental processes. As they bind to specific DNA regions, they contain a characteristic set of DNA binding motifs (de Mendoza et al. 2013; Ylla et al. 2015), which serve for identification and comparison. Finally, apoptosis is also important in the context

of metamorphosis to eliminate old structures that must be replaced by new ones (Suzanne et al. 2013; Accorsi et al. 2015).

All these aspects were analyzed in the transcriptomes we obtained covering the entire ontogeny of *B. germanica*. Subsequently, the same analysis was carried out on a publicly available equivalent transcriptomic set of *D. melanogaster*. Comparing the results of the two analyses highlighted key differences in regulatory mechanisms that could account for the different modes of metamorphosis utilized by these two species, and which could shed light on the basis of the evolutionary transition from hemimetaboly to holometaboly.

Results

General transcriptomic and genomic data

The analyses were based on 22 mRNA libraries of *B. germanica*, prepared in our laboratory, representing the following 11 selected stages (2 replicates each): non-fertilized egg (NFE), 8, 24, 48, 144 and 312 h after fecundation (ED0, ED1, ED2, ED6 and ED13), first, third, fifth and sixth (last) nymphal instars (N1, N3, N5 and N6), and adult female (Table 1). From the 22 libraries, we obtained a total of 198,970,437 read pairs (the data from the 22 libraries is accessible at GEO: PRJNA382128, GSE99785). After removing the adapters, filtering low quality reads, and merging read pairs, we obtained a total of 193,014,748 read pairs (corresponding to 97.01% of the total sequenced read pairs) (Supplemental Table S1), 66.84% of which mapped to the *B. germanica* genome. On the basis of normalized gene expression (Fig. 1A), we obtained the hierarchical clustering of samples, which demonstrated that the two replicates of

each stage-library grouped together (Supplemental Fig. S1A). The only exception was N5, which clustered as a sister group of N5-2+N6. Given the high similarity between the two library replicates of the same stage, we joined these for further gene expression analyses.

Table 1. Biological data corresponding to the 11 key stages in which mRNA libraries were obtained in the present work.

Library	Developmental period	Age after oviposition (AO)	% Embryo development	Embryo stage	Hormonal context
NFE	Egg	Day 8 of the first gonadotrophic cycle (in preoviposition)	--	--	Not determined within the egg. High levels of 20E and JH in the surrounding haemolymph.
ED0	Embryo	8 h AO, when the ootheca is still vertical	2%.	Only yolk granules observed.	No detectable levels of 20E and JH
ED1	Embryo	24 h AO	6%	Energids at low density spread among the yolk granules. Tanaka stage 1	No detectable levels of 20E and JH
ED2	Embryo	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	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	312 h AO	72%	Eyes well colored, antennae and legs reaching the 5th abdominal segment. Tanaka stage 15	Peak of 20E. High levels of JH
N1	1st nymphal instar	1-2 days old	--	--	High levels of 20E and JH
N3	3rd nymphal instar	2-4 days old	--	--	High levels of 20E and JH
N5	5th nymphal instar	3-5 days old	--	--	High levels of 20E and JH
N6	6th nymphal instar	5-7 days old	--	--	High levels of 20E, no JH
Adult	Adult (female)	5 days old	--	--	Low (ovarian) levels of 20E and high JH levels

A similar RNA-seq dataset of *D. melanogaster* that is available at GEO GSE18068 was used for comparative purposes. It comprises 22 libraries from 11 developmental stages (2 replicates each) covering the entire embryo development (six chronologically sequential stages: 0-4h, 4-6h, 6-12h, 12-16h, 16-20h, 20-24h), the three larval stages (L1, L2, L3), the pupa, and the adult male and female. In the case of the adult stage, we chose the female library as this was comparable with the equivalent library of *B. germanica*. For the other postembryonic stages, pre-last nymphal instars of *B. germanica* can be compared to *D. melanogaster* larvae, and the last nymphal instar can be compared to pupae. Equivalencies between

embryo stages are not direct, thus a detailed examination based on developmental and molecular data led us to propose the correspondences between the transcriptomes of *B. germanica* and *D. melanogaster* summarized in Supplemental Table S2. When analyzed with the procedure used on the libraries of *B. germanica*, a total of 129,507,378 RNA-seq fragments were obtained from the 22 libraries, 95.19% of which mapped to the *D. melanogaster* genome (Supplemental Table S3). The expression of all genes in each library is shown in Figure 1A, and the results of the associated clustering analysis are summarized in Supplemental Figure S1B. This shows that the two replicates of each stage-library clustered together, a fact that led us to join the two replicates in further analyses. Among all the RNA-seq libraries from each insect, we detected expression evidence (>1 FPKM) for 90.07% of the annotated genes of *B. germanica* (25,643 out of 28,471), and 97.33% of *D. melanogaster* (17,004 out of 17,471).

In order for the datasets to be comparable from the point of view of functional annotation, we annotated the *B. germanica* genes with reference to orthologous *D. melanogaster* genes. We retrieved the protein sequences from the 28,471 annotated genes of *B. germanica* and those of the 17,471 annotated genes of *D. melanogaster*. Following the best blast reciprocal hit (BBRH) approach, we identified 7,169 orthologous genes common to *B. germanica* and *D. melanogaster*. This corresponds to 25.18% of the *B. germanica* genes and 41.03% of those from *D. melanogaster*. The gene names used are those currently applied to *D. melanogaster*.

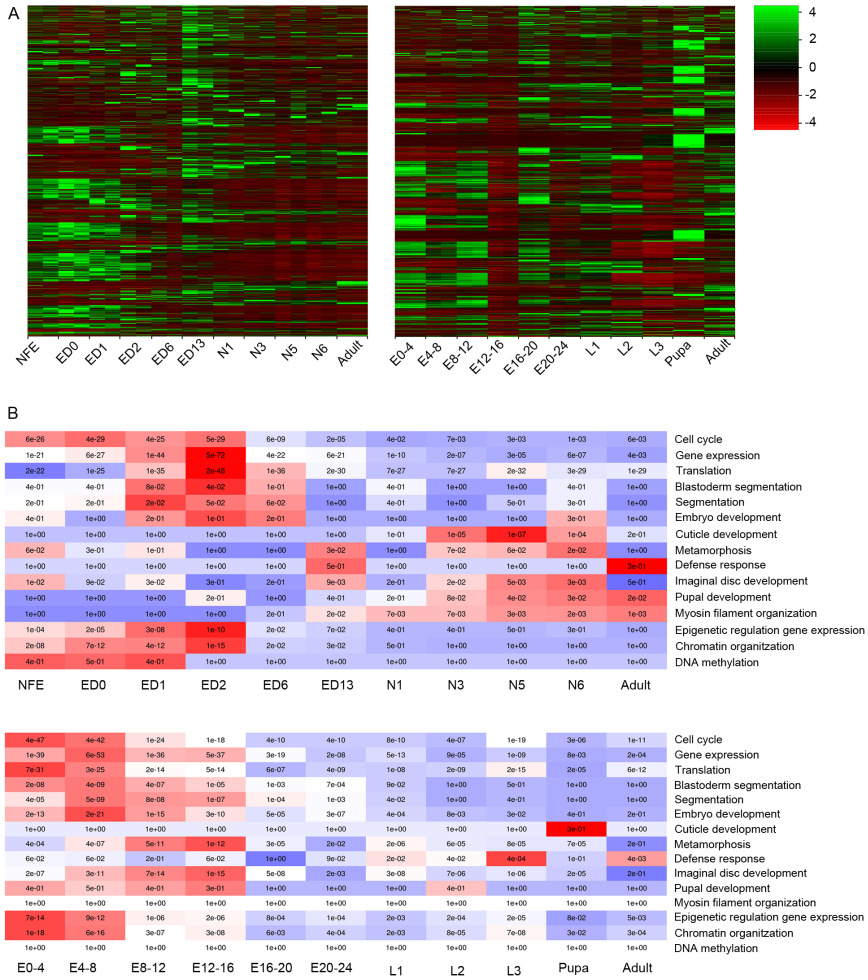

Figure 1: Overall gene expression along ontogeny, and selected GO terms from enrichment analysis. **A)** Heatmap showing the expression of all genes in each of the libraries studied of *Blattella germanica* (left) and *Drosophila melanogaster* (right). **B)** Selection of GO terms of biological process from the enrichment analysis performed with the expressed genes at each stage in *B. germanica* (upper panel) and *D. melanogaster* (lower panel). 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.

General expression

Expression in each stage-library

The expression of all the genes in each library of *B. germanica* shows that many of them are more expressed in embryonic stages, whereas a relatively small set of genes are specific to postembryonic stages (Fig. 1A). In *D. melanogaster*, many genes are well expressed in embryonic and larval stages, whereas a smaller set of genes is highly expressed in pupal and adult stages (Fig. 1A). To analyze how the different stages of *B. germanica* compare with those of *D. melanogaster*, we carried out a differential expression analysis of the 7,169 orthologous genes with each possible library pair of the two species. From each comparison we obtained the number of genes which are differently expressed at a significant level (adjusted p-value < 0.01). Within each species, the results show that *B. germanica* stage-libraries form more or less well defined transcriptomic similarity groups in the very early and early embryonic stages. Conversely, no large groups are observed in *D. melanogaster* (Supplemental Fig. S2). Considering interspecific relationships, the analysis revealed clear transcriptomic similarities between *B. germanica* nymphs and *D. melanogaster* larvae, but the pupal stage (*D. melanogaster*) appears quite unique. The lowest number of differences of the pupa was found with the adult and the last larval instar (L3) of *D. melanogaster*, and with the last nymphal instar (N6) of *B. germanica* (Supplemental Fig. S2, inset).

Functional enrichment at each stage

GO-terms and enrichment analyses of the expressed genes in each library revealed different biological functions at different stages

within the same species, and general differences between *B. germanica* and *D. melanogaster* (Fig. 1B, & Supplemental Figs. S3, S4). In the embryonic stages, the results show that both species are enriched with respect to functions relating to "cell cycle control", "gene expression", and "translation", suggesting active transcriptional activity and cell proliferation (Fig. 1B, & Supplemental Figs. S3, S4). Functions relating to epigenetic control such as "chromatin organization" are also enriched but, intriguingly, "DNA methylation" functions are enriched in the early *B. germanica* embryo, but not in *D. melanogaster*. In postembryonic development, we observed a clear enrichment in genes relating to "cuticle development" in *D. melanogaster* pupae and *B. germanica* nymphs. In both species, the adult stage is enriched in genes relating to homeostasis, such as metabolism, catabolism, and immune defense functions (Supplemental Figs. S3, S4).

Differential expression between stages

In *B. germanica*, differential expression changes between successive stages reveal that embryonic stages experience the most dynamic changes, whereas changes are much less conspicuous during postembryonic life, especially between nymphal stages (Fig. 2). Thus, while the strongest difference is between ED2 and ED6, with 5.694 differentially expressed genes, the lowest number of changes is observed between N5 and N6 with only 69 genes significantly changing their expression. In contrast, the differential expression analysis of *D. melanogaster* revealed that number of changes is similar in all transitions (Fig. 2).

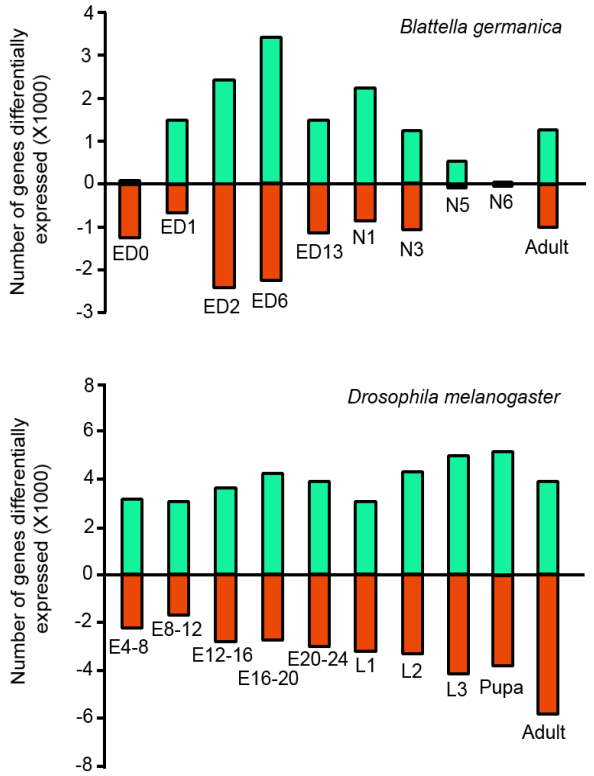

Figure 2: Differential expression analysis between consecutive stage-libraries. Number of genes significantly ($p < 0.05$) upregulated (green) and downregulated (red) according to the differential expression analysis between consecutive libraries of *Blattella germanica* and *Drosophila melanogaster*.

Maternal to zygotic transition

Maternal transcripts

The NFE libraries of *B. germanica* contain maternally loaded mRNAs that, unsurprisingly, are enriched for functions relating to "cell cycle" and "embryo development" (Fig. 1B & Supplemental Fig. S3). Interestingly, they are also enriched for genes linked to epigenetic regulation, like those with the GO-terms "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* (Fig. 1B Supplemental Fig. S4). Especially intriguing are the genes with the GO-terms "metamorphosis" and "wing disc development", which are well represented in the NFE library of *B. germanica* (284 and 360 genes, respectively) but do not figure in the 0-4h embryo library of *D. melanogaster* (Fig. 1B). A number of genes corresponding to these GO-terms, are typical of adult morphogenesis, 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).Their occurrence as maternal transcripts in *B. germanica* is enigmatic as these processes develop later in embryogenesis.

Smaug, Zelda and DNA methylation

In *B. germanica*, Smaug shows a peak in ED0, whereas Zelda peaks in ED1, in both cases followed by an abrupt decrease, so that the levels are very low in mid and late embryonic and in postembryonic stages (Fig. 3A). In contrast, Smaug and Zelda are well expressed during all the embryogenesis stages in *D. melanogaster*, as well as in the first larval instar (Fig. 3A). Moreover, the functional enrichment analysis (Fig. 1B) suggests DNA methylation operates in early *B. germanica* embryos during the MZT (NFE, ED0 and ED1 libraries), whereas this is not the case with *D. melanogaster*. Thus, we studied DNA methylation in *B. germanica* using the indirect approach of examining the relationship between CpGo/e and expression in each one of the 11 stage-libraries of *B. germanica*. The results revealed a significant negative correlation between gene

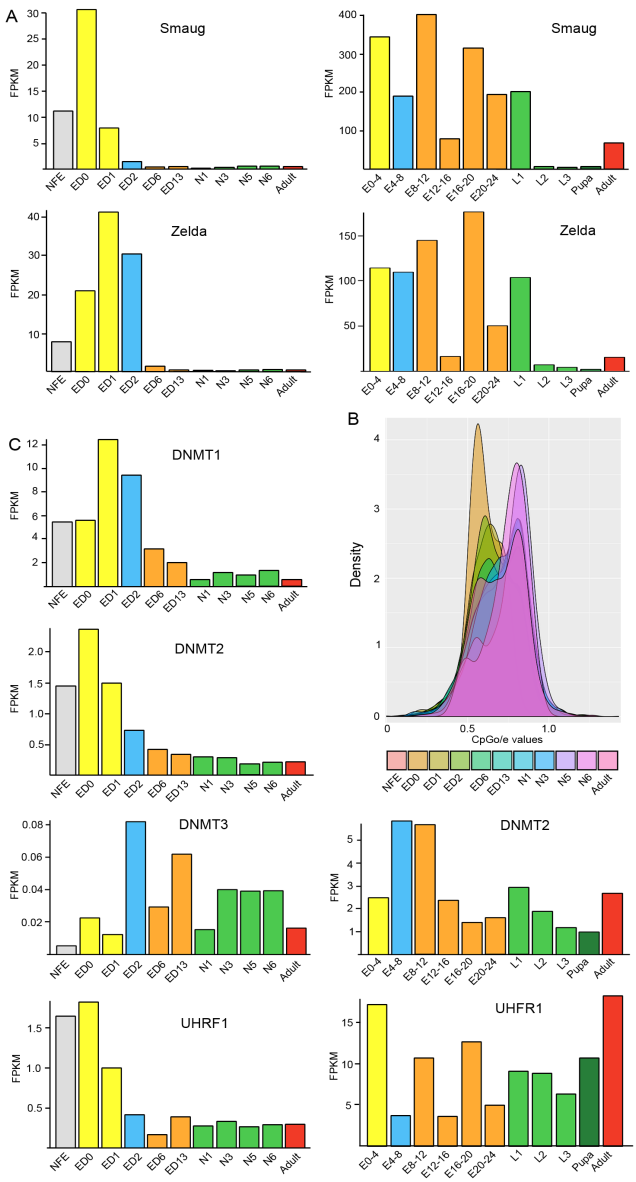

Figure 3: Gene expression and methylation during the maternal to zygotic transition in Blattella germanica and Drosophila melanogaster. A) Expression of Smaug and Zelda along the different stage-libraries. **B)** The CpGo/e distribution of the differentially expressed genes in each stage-library of *B. germanica*. **C)** Expression of DNA methylation enzymes, DNMT and UHRF, along the different stage-libraries. In A and C, the left diagrams correspond *to B. germanica* and the right diagrams to *D. melanogaster*; and identical bar colors indicate equivalent developmental periods according to the criteria summarized in Supplemental Table S2.

expression and CpGo/e at ED0, ED1 and ED2. The strongest and most significant correlation coefficient between CpG depletion and expression was observed at ED0. Indeed, the genes overexpressed at ED0 show the lowest levels of CpGo/e (Fig. 3B). Moreover, the genes coding for enzymes that catalyze DNA methylation, like DNMT (DNA methyltransferase) 1, 2 and 3, and UHRF1 (Ubiquitin like with PHD and ring finger domains 1), are well expressed in the early embryo. DNMT1 and 2, and UHRF1 peak at ED0 and ED1, whereas DNMT 3 shows lower expression levels and peaks at ED2 (Fig. 3C). In *D. melanogaster*, methylation enzymes DNMT2 and UHRF1 show moderate to low, and relatively sustained levels of expression throughout ontogeny (Fig. 3C), which is consistent with data suggesting that DNA methylation is irrelevant in dipterans (Marhold et al. 2004).

From gap to hox genes

Early patterning

The expression of the most representative gap, pair-rule and segment polarity genes is shown in Figure 4A. In *B. germanica*, nanos appears to be maternal but continues to be expressed in ED0 and ED1, whereas hunchback and caudal show an expression peak in ED2. The gap-gene cascade is initiated by maternal tailless, followed by orthodenticle, huckbein, and Küppel. In contrast, the pair-rule genes show a delimited peak at ED2, except hairy which has a delayed expression, with a peak at ED13. Segment polarity genes are expressed from ED2 to ED13. *D. melanogaster* libraries do not show the cascade of maternal, gap, pair rule and segment polarity genes very well, possibly due to the arbitrary staging of

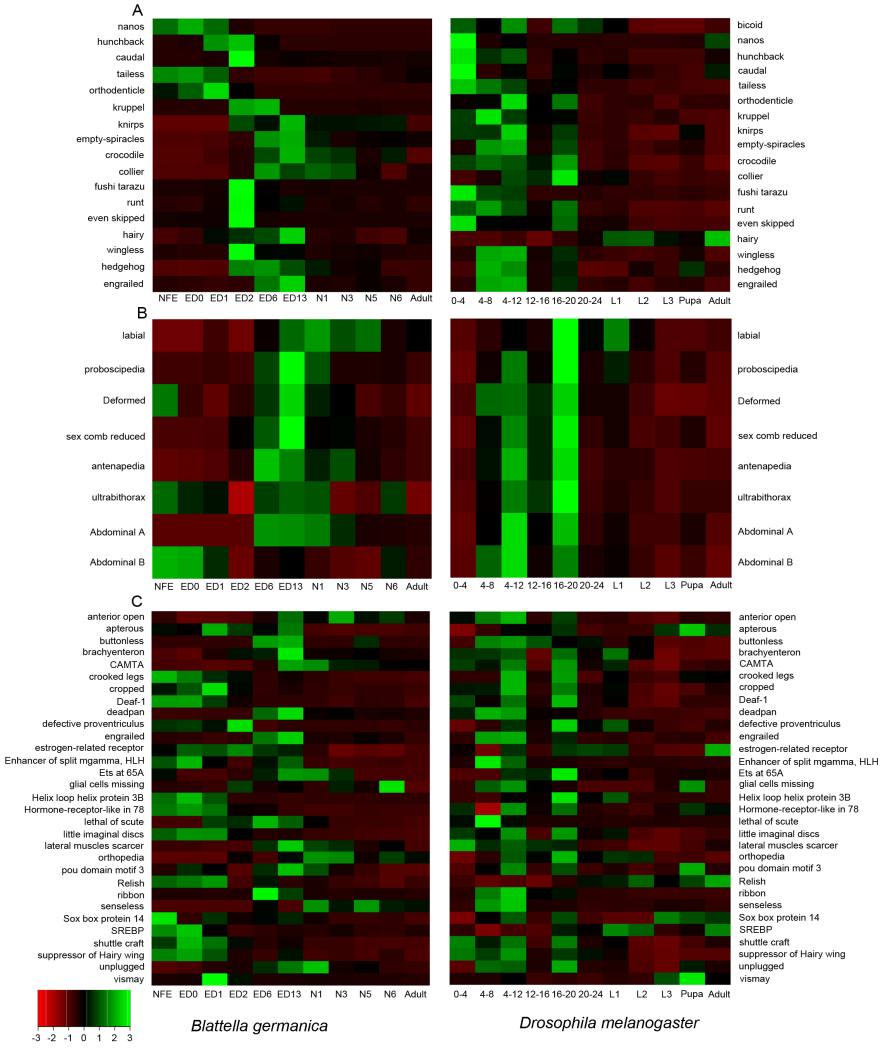

Figure 4: Expression of maternal and early embryo genes, hox genes and transcription factors. A) Expression of maternal, gap, pair-rule and segmentation genes along the different stage-libraries. **B)** Expression of Hox genes. **C)** Expression of a selection of 34 orthologous transcription factors common to *Blattella germanica* and *Drosophila melanogaster*.

these libraries. In any case, bicoid, nanos, hunchback, and caudal, which are described as maternal genes in *D. melanogaster*, show the highest mRNA abundance in the earliest stage available (E0-4h) (Fig. 4A). The other genes appear in the mid-late embryo stage, and

only hairy has an atypical pattern, being predominantly expressed in postembryonic stages.

Hox genes

We identified the eight canonical Hox genes in the *B. germanica* genome, which are well conserved with respect to other species (Supplemental Fig. S5). Most of them are fully expressed in ED6 and ED13. Abd-A and lab continue to be expressed in the first nymphal stages. Intriguingly, maximal mRNA levels of Abd-B occur in NFE and ED0. Ubx is interesting because it is also present in NFE and is additionally expressed in N6 (Fig. 4B). In *D. melanogaster*, Hox genes are expressed between E8-12 and E16-20h. Only lab shows signs of expression beyond the embryo, in L1 (Fig. 4B). In the equivalent embryo libraries of *B. germanica*, ED6 and ED13, Hox genes are also abundantly expressed, with the exception of Abd-B, which is, in contrast, expressed in N6 (Fig. 4B).

General development and metamorphosis

Transcription factors

As a first step in studying transcription factors, we performed a PfamScan search among all annotated *B. germanica* and *D. melanogaster* proteins. This gave 17,196 PFAM-A motifs (4,280 unique) associated to 12,789 *B. germanica* genes, and 15,5475 PFAM-A motifs (4,339 unique) associated to 10,759 *D. melanogaster* genes. Among these, we identified 600 *B. germanica* genes and 458 *D. melanogaster* genes containing at least one Pfam

motif unequivocally linked to a transcription factor function (de Mendoza et al. 2013; Ylla et al. 2015).

Most of these transcription factors are more highly expressed in the embryo stages of both species (Supplemental Fig. S6). Characteristically, many of them are also highly expressed in the pupal and adult stages of *D. melanogaster*. In any case, there are stage-specificities characterized by the high or low expression of a given set of transcription factors. In order to see comparable qualitative differences between the two species, we retrieved the subset of orthologous transcription factor genes common to *B. germanica* and *D. melanogaster*. In total we found 297 genes common to both species (Supplemental Table S4). The expression profiles in *B. germanica* and *D. melanogaster* (Supplemental Figs. S7, S8) are reminiscent of those observed when studying all genes (Fig. 1A). A selection of those that display a greater contrast in expression between species and stages is shown in Figure 4C. 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*, whereas the E0-4h stages in *D. melanogaster* are characterized by a high level of expression of lateral muscles scarcer, hinge 1 and calmodulin-binding transcription activator. In mid-late embryo stages, lethal of scute is characteristically highly expressed in ED6 to ED13 in *B. germanica*, whereas shuttle craft, rotund, orthopedia, little imaginal discs, *HR78* (Hormone-receptor-like in 78), ELL-associated factor (*Deaf*), cropped, and crooked legs are highly expressed in *D. melanogaster* in E8-20h. With respect to postembryonic stages, unplugged,

senseless, sortilin, *Ets65A*, and orthopedia are typically highly expressed in *B. germanica* nymphs, whereas *SREBP*, *Sox14*, relish, and hairy are typical of *D. melanogaster* larvae. The pupa of *D. melanogaster* continue to express high levels *Sox14* and relish, which are also expressed in the adult, and have characteristic expression peaks for vismay, apterous, and glial cells missing. Interestingly, glial cells missing shows also an expression peak in L6 of *B. germanica*.

Apoptosis

Among the genes involved in programmed cell death, we studied the expression of Dredd (coding for caspase 8), Dronc (caspase 9), Drice (caspase 3), and Dark (contributing, with Dronc, to the formation of the apoptosome). In *B. germanica*, the expression of these 6 genes is generally high in early to mid-embryo stages, but then decreases substantially in late embryonic stages, nymphal stages, and adults (Fig. 5). In *D. melanogaster*, their expression is generally also high in early embryonic stages, but this remains high in late embryogenesis, as well as in larva, pupa and adult (Fig. 5).

Ecdysone and JH biosynthesis

In *B. germanica*, the genes involved in ecdysone biosynthesis are more highly expressed in all embryonic stages than in nymphs or adults (Figs. 6A). In *D. melanogaster*, however, they are mostly expressed in very early embryos and late postembryonic stages (Figs. 6A). The genes involved in JH biosynthesis are expressed throughout the ontogeny of *B. germanica*, although there is a

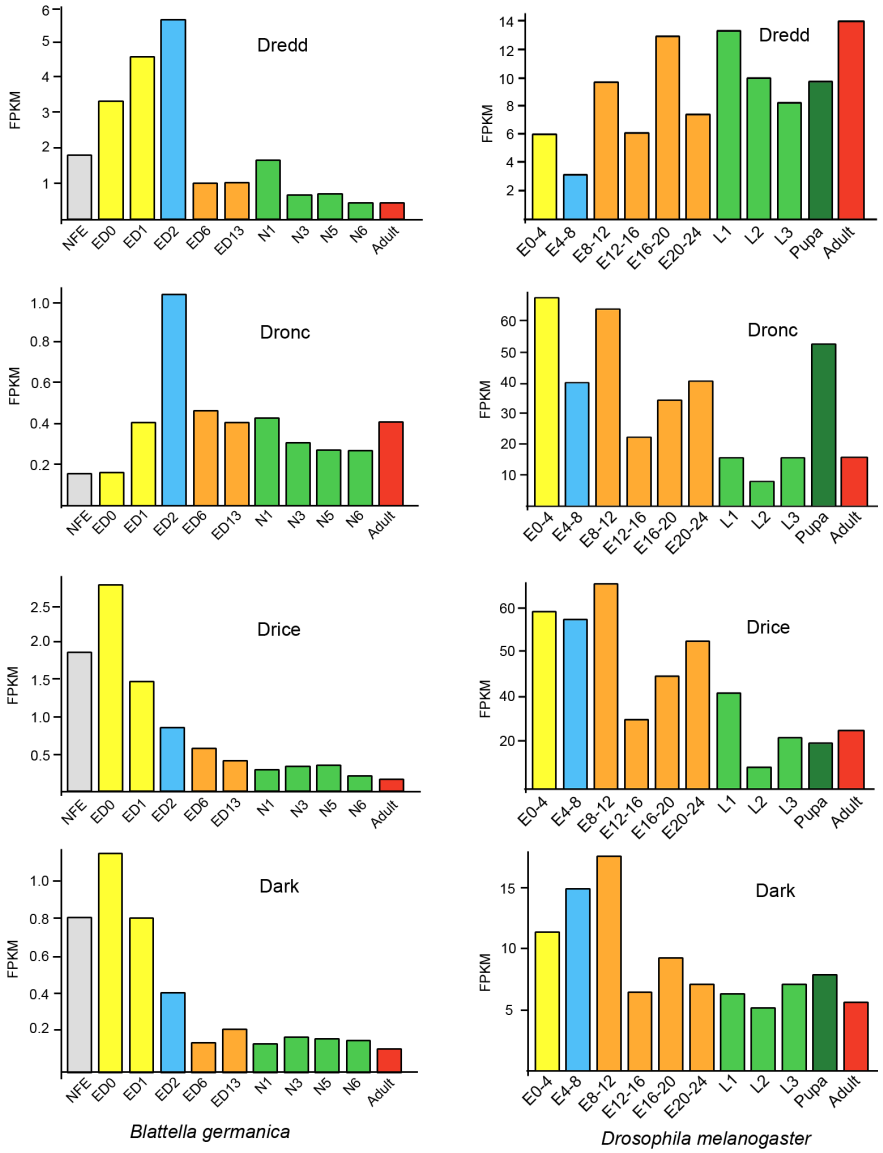

Figure 5: Expression of genes involved in programmed cell dead in *Blattella germanica* and *Drosophila melanogaster*. The genes selected were Dredd (coding for caspase 8), Dronc (caspase 9), Drice (caspase 3) and Dark (contributing, with Dronc, to the formation of the apoptosome). Identical bar colors indicate equivalent developmental periods according to the criteria summarized in Supplemental Table S2.

tendency for the genes coding for the mevalonate branch to be more highly expressed in the embryonic stages, whereas those coding for enzymes specific to the JH branch are, in general, more highly expressed in the postembryonic stages (Figs. 6B). A similar situation, but with a more marked difference between the embryonic and postembryonic stages is observed in *D. melanogaster*. The genes coding for the mevalonate branch show a window of high expression in the late embryonic stage (E16-20h), whereas the expression of those coding for enzymes specific to the JH branch is predominantly seen in the postembryonic stages (Figs. 6B). Interestingly, there is no conspicuous expression of any enzymes in early embryonic stages of *D. melanogaster*, in contrast to that seen in *B. germanica*.

Ecdysone and JH signaling

The main genes regulating metamorphosis include those belonging to the signal transducer pathways of ecdysone and JH. In *B. germanica*, the expression of typical transducers of the ecdysone signal, like ecdysone receptor (*EcR*), ultraspiracle (*USP*), *E75*, *HR3* and *HR4*, are distributed over the embryonic and postembryonic stages. Interestingly, Fushi tarazu factor 1 (*Ftz-f1*) shows a clear expression peak in N5. There is comparatively higher expression of typical JH transducers, like *Met*, Taiman (*Tai*) and *Kr-h1* in embryonic than in postembryonic stages. Interestingly, this is also the case for Broad-complex (*BR-C*) expression. Ecdysone-induced protein 93F (*E93*) peaks are seen on N6, as expected. Myoglianin shows a clear peak on N5 (Fig. 6C & Supplemental Fig. S7). In *D.melanogaster*, the expression of most of the ecdysone and JH

signal transducers shows the highest levels in the late embryonic stage (E16-20h), and then pupa and adult. *BR-C* is practically not expressed in the early embryo, *E93* expression concentrates in the pupa and adult, and myoglianin is more expressed in embryonic than postembryonic stages (Fig. 6C & Supplemental Fig. S7).

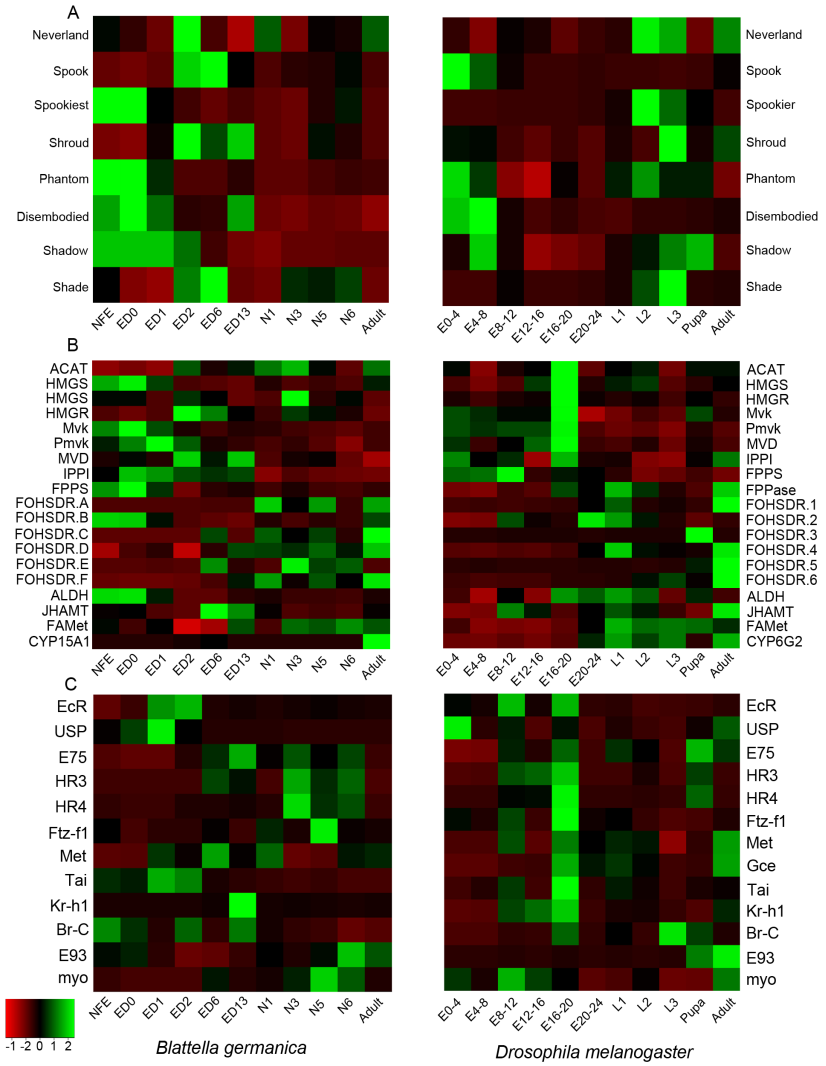

Figure 6: Expression of genes related with hormonal biosynthesis and signaling. A) Expression of genes coding for enzymes of the ecdysone synthesis along the different stages. **B)** Expression of genes coding for juvenile hormone (JH) synthesis enzymes. **C)** Expression of genes coding for key transcription factors that transduce the ecdysone and JH signals. In all cases, the left panel corresponds to *Blattella germanica* and the right to *Drosophila melanogaster*.

Discussion

Cockroach and fly gene expression. General features

Gene expression in ontogeny reveals that while *B. germanica* has the highest diversity of genes expressed during embryogenesis, the genes expressed in *D. melanogaster* are rather diverse in the embryonic and larval stages, and show a peak of intense expression in the pupal stage (Fig. 1A). Similarly, the results of the differential expression analysis between successive stages show that the most dynamic changes in *B. germanica* occur during embryogenesis, whereas in *D. melanogaster* quite dynamic changes are seen all through ontogeny (Fig. 2). These general differences may be related to the fact that in the hemimetabolan mode the essential adult body structure is built during embryogenesis, while postembryonic stages essentially focus on growth. In contrast, in holometabolan species the adult morphology is completed in postembryonic stages, mainly in the transitions from the last instar larvae, to pupa and then adult.

The maternal to zygotic transition

Intriguingly, the pool of maternal mRNAs of *B. germanica* includes transcripts enriched for genes with the GO-term "metamorphosis", which is not the case in the early embryogenesis of *D. melanogaster* (Fig. 1B). Specifically, in *B. germanica*, the maternal material includes significant amounts of *BR-C* transcripts (Fig. 6C), which is a factor typically promoting morphogenesis, notably the formation of the holometabolan pupa (Zhou et al. 2002). The occurrence of *BR-C* and other transcripts currently involved in morphogenesis in the pool of maternal mRNAs in *B. germanica* could be related to the hemimetabolan mode of metamorphosis. Alternatively, these types

of transcripts in the very early embryo might be related to the different germ-band types: short in *B. germanica* and long in *D. melanogaster*.

Two important players in the MZT are Smaug, which contributes to the clearance of maternal mRNA (Tadros et al. 2007; Chen et al. 2014), and Zelda, which activates the zygotic gene expression (Schulz et al. 2015; Foo et al. 2014; Sun et al. 2015). In *B. germanica* ontogeny, Smaug and Zelda show respective expression peaks in ED0 and ED1 (Fig. 3A), which suggests that the MZT takes place between these two stages. Additionally, we recently reported (Ylla et al. 2017) that microRNAs of the MIR-309 family, which in *D. melanogaster* contribute to clearing maternally loaded mRNAs (Bushati et al. 2008), show an acute expression peak at ED2. If, as suggested (Ylla et al. 2017), the MIR-309 microRNAs play the role of maternal mRNA scavengers in *B. germanica*, as occurs in *D. melanogaster*, then their expression in ED2 may characterize a final step of the MZT. In *D. melanogaster*, Smaug and Zelda maintain quite high levels of expression throughout embryogenesis and even the first larval instar (Fig. 3A). This continued expression of Smaug and Zelda may be related to the evolutionarily derived embryogenesis in *D. melanogaster* and, by extension, of holometabolan species.

A process that may be relevant in very early ontogenetic stages of *B. germanica*, but not *D. melanogaster*, is DNA methylation (Fig. 1B). DNA methylation is associated with a repressed chromatin state and inhibition of gene expression (Siegfried et al. 2010), and is generally conserved among insects (Zemach et al. 2010; Bewick et al. 2016). DNA methylation is

important in the caste differentiation of social insects (Elango et al. 2009; Wang et al. 2006), but does not appear relevant in dipterans (Marhold et al. 2004). In *B. germanica*, DNA regions which are prone to methylation (with a low CpGo/e) show the highest levels of expression in early embryogenesis, from day 0 to day 2 (Fig. 3B). They then vanish in subsequent embryonic and postembryonic stages. In parallel, the DNMT enzymes also show high levels of expression from day 0 to day 2 (Fig. 3C). Different DNMTs have been associated to different methylation roles. DNMT3 is considered to methylate *de novo*, whereas DNMT1 and UHRF1 maintain the established methylation profiles (Bostick et al. 2007; Bestor et al. 2015). DNMT2 was first reported as an enzyme methylating DNA, but it was later discovered that it methylates tRNAs (Goll et al. 2006). This data taken as a whole leads us to propose that in the first moments when the zygotic program is activated, between ED0 and ED1, DNA is poorly or not at all methylated, allowing the expression of those genes prone to methylation. Thereafter, and considering the expression patterns of DNMTs, especially DNMT1 and DNMT3, a wave of DNA methylation would silence the expression of the above genes for the remaining developmental stages. In any case, our data show that DNA methylation appears to play a role in the regulation of gene expression in the *B. germanica* embryo, which might be a feature of hemimetabolan embryogenesis.

Gene expression in the embryo

The most obvious difference in the expression of gap, pair-rule and segment polarity genes between *B. germanica* and *D. melanogaster*

is the absence of bicoid in the former species (Fig. 4A), as this gene is exclusive to higher dipterans (Schröder 2003). In *D. melanogaster*, bicoid organizes the patterning of the anterior region by activating zygotic hunchback and Krüppel (Driever et al. 1989; Hülskamp et al. 1990). In short germ-band hemimetabolan species, like the cricket *Gryllus bimaculatus*, the role of bicoid appears to be replaced by caudal, which triggers a slightly different genetic cascade of genes (Shinmyo et al. 2005). The expression of hairy is especially interesting, as in mid-late embryo stages its expression is intense in *B. germanica* but very weak in *D. melanogaster* (Fig. 4A). In *D. melanogaster*, hairy acts as a pair-rule in early embryo development, whereas in larvae, by binding to the protein achaetae, it allows the proper patterning of sensory organs in the developing wings and legs (Fisher et al. 1998). Through other mechanisms, hairy might also contribute to regulating the progression of the morphogenetic furrow in the developing eye (Bhattacharya et al. 2012). The latter functions explain the expression that we observed in *D. melanogaster* larvae (Fig. 4A), and we speculate that the high level of expression in the mid-late embryo of *B. germanica* (Fig. 4A) might be due to the generation of nymphal structures like the formation of proper chaetotaxy or compound eyes.

Among the Hox genes, Abd-A is well expressed in the mid-late embryo in *B. germanica*, when dorsal closure occurs, but Abd-B is not. Conversely both Abd-A and Abd-B are abundantly expressed in the equivalent stages of *D. melanogaster* (Fig. 4B). In *D. melanogaster*, mixer cell remodeling regulates tension along the leading edge during dorsal closure. Abd-A is a pro-mixing factor in the first five abdominal segments, whereas Abd-B represses mixing

in posterior segments. At closure in the central segments, which close last, the tension increases; thus in these segments Abd-A is not repressed by Abd-B (Roumengous et al. 2017). If Abd-A and Abd-B play the same role in *B. germanica*, then the low level of expression of Abd-B would suggest that the pro-mixing action of Abd-A is needed all along the leading edge during dorsal closure.

The study of transcription factors highlighted a number of genes which are specifically important in the embryonic development of *B. germanica*, compared with *D. melanogaster* (Fig. 4C). For example, high expression of SREBP in ED0 and ED1 of *B. germanica* suggests that lipogenesis and lipid homeostasis (Shao et al. 2012) is important in early embryogenesis of the cockroach. Conversely, in *D. melanogaster* SREBP appears to be not as relevant, but others, like 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 embryonic stages, lethal of scute, a gene involved in the neurogenesis and specification of sensory organs *(Negre et al. 2015)* is highly expressed in in *B. germanica*, whereas shuttle craft, which is required to maintain the proper morphology of motoneuronal axon nerve routes (Stroumbakis et al. 1996) is highly expressed in *D. melanogaster*. Also typical of the late embryonic 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).

A gene showing high expression levels in *B. germanica* embryogenesis, but not *D. melanogaster*, is *BR-C* (Fig. 6C), which is consistent with its important functions in cockroach embryo development (Piulachs et al. 2010). Conversely, a gene that is highly expressed in *D. melanogaster* embryogenesis, but whose expression is not relevant in *B. germanica* (Fig. 6C), is myoglianin, which is involved in the formation of glial cells and myoblasts (Lo et al. 1999). Finally, a difference that could be functionally significant in late embryo stages is the relatively low expression levels of genes related to apoptosis in *B. germanica*, in contrast with *D. melanogaster* (Fig. 5), suggesting that the formation of derived vermiform body morphology requires extensive processes of programmed cell death.

Postembryonic development

With respect to transcription factors (Fig. 4C), unplugged, required for the formation of specific tracheal branches, specifically the cerebral branch (Chiang et al. 1995), and senseless, which is crucial for the development of the peripheral nervous system *(Nolo et al. 2000)*, are typically highly expressed in *B. germanica* nymphs. In *D. melanogaster* these genes are not highly expressed. Conversely, *Sox14,* required for 20E signaling at the onset of metamorphosis *(Ritter et al. 2010)*, relish, which promotes the transcription of innate immune response genes (Petersen et al. 2013) and hairy, which is involved in sensory organ patterning (Fisher et al. 1998), are characteristically expressed in *D. melanogaster* larvae. The pupa of *D. melanogaster* continues to express *Sox14* and relish at high levels, and shows a characteristic expression of apterous, a gene

involved in wing morphogenesis (Michel et al. 2016) (which, in *B. germanica* is well expressed during embryo development), and glial cells missing, a gene that controls the determination of glial versus neuronal fate (Egger et al. 2002). Interestingly, glial cells missing also shows a peak of expression in L6 of *B. germanica*, suggesting that it has a conserved neuromorphogenetic role in the pre-adult stage of both hemimetabolan and holometabolan species.

The higher expression of apoptosis-related genes in *D. melanogaster* with respect to *B. germanica* (Fig. 5) might indicate that extensive processes of programmed cell death are required in the latter species. This observation stands not only for the pupal stage, which experiences an obvious process of disintegration and construction of new body structures, but also in larvae, despite the apparent morphological conservation during the three larval stages. Related to ecdysone and JH, an important difference between *B. germanica* and *D. melanogaster* is the expression of *BR-C* in L3 and the pupa of *D. melanogaster*, whereas in the postembryonic stages of *B. germanica* this transcription factor is expressed only at low levels (Figs.6C). These differences are due to the fact that *BR-C* has a key function in pupal morphogenesis in *D. melanogaster* (Zhou et al. 2002), whereas in *B. germanica* it plays a minor role in wing pad growth (Huang et al. 2013). Also of interest is the characteristic expression peak of *Ftz-f1* and myoglianin in N5 of *B. germanica* (Figs.6C), suggesting that these factors play important roles in the penultimate nymphal instar, possibly related to the genetic program that is installed in the last nymphal instar, where the 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), and a recent report on the cricket *G. bimaculatus* reveals the crucial role of myoglianin in the regulation of the JH decrease that occurs in the last nymphal instar (Ishimaru et al. 2016). Expression of *E93* in pre-adult stages in both species is consistent with its role as adult specifier in hemimetabolan and holometabolan species (Urena et al. 2014; Belles et al. 2014).

Conclusions

Gene expression in ontogeny reveals that *B. germanica* shows the most dynamic and diverse changes during embryogenesis. In contrast, *D. melanogaster* maintains a similar level of gene expression throughout ontogeny. This general difference may be related to the different types of metamorphosis: hemimetabolan in *B. germanica*, where the adult body structure is shaped during embryogenesis, and holometabolan in *D. melanogaster*, which completes the adult morphogenesis in postembryonic stages.

In *B. germanica*, Smaug and Zelda are acutely expressed in the MZT, whereas in *D. melanogaster* quite high expression levels are maintained throughout embryogenesis. This continued expression of Smaug and Zelda may be related to the evolutionarily derived embryo morphogenesis in *D. melanogaster* and, by extension, of holometabolan species.

DNA methylation appears relevant in the early embryonic development of *B. germanica*, but not *D. melanogaster*. We suggest that a wave of DNA methylation in the early embryogenesis of *B. germanica* silences the expression of a set of earlier genes. If so, then this mechanism has been lost in *D. melanogaster,* as well as perhaps in other holometabolan species.

The expression of gap, pair-rule, and segment polarity genes shows differences between *B. germanica* and *D. melanogaster*. For example, hairy is highly expressed in the mid-late embryonic stages of *B. germanica* but not *D. melanogaster*. This stage-specific expression of hairy in *B. germanica* might be related to the formation of nymphal structures, like the formation of proper chaetotaxy or compound eyes.

Concerning Hox genes, Abd-B is highly expressed in the mid-late embryonic stage of *D. melanogaster* but not *B. germanica*, which could be related to different mechanisms for achieving dorsal closure in the two species.

The expression of transcription factors reveals many quantitative and qualitative differences between *B. germanica* and *D. melanogaster* in embryonic and postembryonic stages. For example, *BR-C* appears to be important in embryo morphogenesis in *B. germanica* whereas it is crucial for pupal morphogenesis in *D. melanogaster*.

In *B. germanica*, the expression of genes related to apoptosis is high in early to mid-embryonic stages of development, whereas in *D. melanogaster* it is high in practically in all embryonic and postembryonic stages. This indicates a differential use of apoptosis to shape morphologies, which is more extensive in *D. melanogaster*.

Methods

Insect colony

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 1). Data on juvenile hormone (JH) and ecdysteroids (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 transcriptomes. 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 validated by 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 codes PRJNA382128 & GSE99785.

Analysis of the RNA-seq data

In the *B. germanica* libraries, we removed the adapters and trimmed the low quality bases on the reads extremes using Trimmomatic (Bolger et al. 2014) with the parameters "2:30:10:8:TRUE SLIDINGWINDOW:4:15". RNA-seq data along the development of *D. melanogaster* was retrieved from Gene Expression Omnibus (Edgar et al. 2002) under the accession GSE18068. All the RNA-seq datasets, were mapped to their correspondent insect genome (dmel_r6.12 and Bger_0.5.3) using the STAR software (Dobin et al. 2013) and the table of counts obtained with the R implementation of featureCounts (Liao et al. 2014) using the correspondent gene annotation of each insect. For clustering purposes reads were normalized with the "varianceStabilizingTransformation" function implemented at DESeq2 R package (Love et al. 2014), for gene expression profiles and visualization we used the FPKMs normalization.

Functional annotation of genes

Functional annotation of genes based on their protein sequence was obtained using PfamScan (Li et al. 2015) together with the PFAM-A database version 30.0 (Bateman et al. 2004). To obtain the transcription factor genes, we selected those genes with a Pfam motifs unequivocally related to transcription factor activity (de Mendoza et al. 2013; Ylla et al. 2015). Another level of functional

annotations are the GO-terms, which we retrieved for the *D. melanogaster* genes using the Annotation Forge package (Carlson et al. 2016), and used for the corresponding *B. germanica* orthologs. Gene orthology between *B. germanica* and *D. melanogaster* where obtained by using Blastp (version 2.5.0+) (Camacho et al. 2009) reciprocal best hits (BRBHs) strategy (Rivera et al. 1998). In the case of the Hox genes, we aligned the candidate of *B. germanica* protein sequences with the eight canonical Hox genes (Negre et al. 2007) of different insect species with ClustalX (Larkin et al. 2007) and we carried out a phylogenetic reconstruction with RAxML (Stamatakis 2014). For all the genes whose expression profile is shown, their accession code in both insects is shown at Supplemental Table S4. The enrichments analysis tests for PFAM motifs and GO-terms was performed using the expressed genes at each stage (>1FPKM) and the hypergeometrical test applied with the GOstats package (Falcon et al. 2007).

Differential expression analysis

The differential expression (DE) analysis tests, were performed with 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 < 0.01. When comparing gene expression between two insect species, we used only the subset of orthologous genes common to both species.

Methylation

The CpG depletion it is a robust indicator of with DNA methylation (Upadhyay et al. 2013; Xiang et al. 2010; Bewick et al. 2016). Consequently, following the method described by (Elango et al. 2009), we calculated the ratio between the observed frequency of CpG and the expected frequency for each annotated gene.

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.

Data access

All sequencing data from the 22 transcriptomes of *Blattella germanica* analysed in the present work is accessible at NCBI Gene Expression Omnibus (GEO: PRJNA382128, GSE99785).

Acknowledgements

The work was supported by the Spanish Ministry of Economy and Competitiveness (grants CGL2012-36251 and CGL2015-64727-P to X.B), Spanish Ministry of Science and Innovation (grants BFU2011–22404 and CGL2016–76011-R to M.D.P.) and Catalan Government (2014 SGR 619). It also received financial assistance from the European Fund for Economic and Regional Development (FEDER funds to X.B. and M.D.P.).

Disclosure declaration

The authors declare no competing or financial interests.

References

Accorsi A, Zibaee A, Malagoli D. 2015. The multifaceted activity of insect caspases. *J Insect Physiol* **76**: 17–23.

Averof M, Akam M. 1995. Hox genes and the diversification of insect and crustacean body plans. *Earth Atmos Aquat Sci Database pg* **376**.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, Studholme DJ, Yeats C, Eddy SR. 2004. The Pfam protein families database. *Nucleic Acids Res* **32**: D138-41.

Belles X. 2011. Origin and Evolution of Insect Metamorphosis. In *eLS*, John Wiley & Sons, Ltd, Chichester, UK.

Bellés X, Martín D, Piulachs M-D. 2005. The mevalonate pathway and the synthesis of juvenile hormone in insects. *Annu Rev Entomol* **50**: 181–199.

Belles X, Santos CG. 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.

Bestor TH, Edwards JR, Boulard M. 2015. Notes on the role of dynamic DNA methylation in mammalian development. *Proc Natl Acad Sci U S A* **112**: 6796–9.

Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. 2016. Evolution of DNA Methylation across Insects. *Mol Biol Evol* msw264.

Bhattacharya A, Baker NE. 2012. The role of the bHLH protein hairy in morphogenetic furrow progression in the developing Drosophila eye. ed. B.A. Hassan. *PLoS One* **7**: e47503.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: btu170-.

Bostick M, Kim JK, Estève P-O, Clark A, Pradhan S, Jacobsen SE. 2007. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science* **317**.

Bushati N, Stark A, Brennecke J, Cohen SM. 2008. Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in *Drosophila*. *Curr Biol* **18**: 501–6.

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

Carlson M, Pages H. 2016. AnnotationForge: Code for Building Annotation Database Packages.

Celniker SE, L Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD. 2009. Unlocking the secrets of the genome. *Nature June* **18**: 927–930.

Chen L, Dumelie JG, Li X, Cheng MH, Yang Z, Laver JD, Siddiqui NU, Westwood JT, Morris Q, Lipshitz HD, Smibert CA. 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 KE, Beachy PA. 1995. Control of Drosophila tracheal branching by the novel homeodomain gene unplugged, a regulatory target for genes of the bithorax complex. *Development* **121**.

Cruz J, Martín D, Pascual N, Maestro JL, Piulachs MD, Bellés X. 2003. Quantity does matter. Juvenile hormone and the onset of vitellogenesis in the German cockroach. *Insect Biochem Mol Biol* **33**: 1219–25.

Cruz J, Nieva C, Mané-Padrós D, Martín D, Bellés 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.

de Mendoza A, Sebé-Pedrós A, Sestak MS, Matejcic M, Torruella G, Domazet-Loso T, 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**: E4858-66.

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

Driever W, Nüsslein-Volhard C. 1989. The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. *Nature* **337**: 138–43.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**:

207–210.

Egger B, Leemans R, Loop T, Kammermeier L, Fan Y, Radimerski T, Strahm MC, Certa U, Reichert H. 2002. Gliogenesis in Drosophila: genome-wide analysis of downstream genes of glial cells missing in the embryonic nervous system. *Development* **129**: 3295–309.

Elango N, Hunt BG, Goodisman MAD, 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**: 11206–11.

Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**: 257–8.

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

Foo SM, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien CY, Kirov N, Shvartsman SY, Rushlow CA. 2014. Zelda potentiates morphogen activity by increasing chromatin accessibility. *Curr Biol* **24**: 1341–1346.

Glastad KM, Hunt BG, Goodisman MAD. 2013. Evidence of a conserved functional role for DNA methylation in termites. *Insect Mol Biol* **22**: 143–154.

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh C-L, Zhang X, Golic KG, Jacobsen SE, Bestor TH. 2006. Methylation of tRNAAsp by the DNA Methyltransferase Homolog Dnmt2. *Science* **311**: 395–398.

Grimaldi D, Engel MS. 2005. *Evolution of the Insects*. Cambridge University Press.

Huang J-H, Lozano J, 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 - Gen Subj* **1830**: 2178–2187.

Hülskamp M, Pfeifle C, Tautz D. 1990. A morphogenetic gradient of hunchback protein organizes the expression of the gap genes Krüppel and knirps in the early Drosophila embryo. *Nature* **346**: 577–80.

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, Mito T. 2016. TGF-$β$ signaling in insects regulates metamorphosis via juvenile hormone biosynthesis. *Proc Natl Acad Sci* **113**: 1600612113-.

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

Jindra M, Palli SR, Riddiford LM. 2013. The Juvenile Hormone Signaling Pathway in Insect Development. *Annu Rev Entomol* **58**: 181–204.

King-Jones K, Thummel CS. 2005. Nuclear receptors — a perspective from Drosophila. *Nat Rev Genet* **6**: 311–323.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–8.

Laver JD, Marsolais AJ, Smibert CA, Lipshitz HD. 2015. Regulation and Function of Maternal Gene Products During the Maternal-to-Zygotic Transition in *Drosophila*. In *Current Topics in Developmental Biology*, Vol. 113 pp. 43–84.

Li L, Greer C, Eisenman RN, Secombe J. 2010. Essential Functions of the Histone Demethylase Lid ed. A. Akhtar. *PLoS Genet* **6**: e1001221.

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

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–30.

Lo PC, Frasch M. 1999. Sequence and expression of myoglianin, a novel Drosophila gene of the TGF-beta superfamily. *Mech Dev* **86**: 171–5.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Maestro JL, Pascual N, Treiblmayr K, Lozano J, Bellés X. 2010. Juvenile hormone and allatostatins in the German cockroach embryo. *Insect Biochem Mol Biol* **40**: 660–665.

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

117–123.

Michel M, Aliee M, Rudolf K, Bialas L, Jülicher F, Dahmann C. 2016. The Selector Gene apterous and Notch Are Required to Locally Increase Mechanical Cell Bond Tension at the Drosophila Dorsoventral Compartment Boundary ed. A. Singh. *PLoS One* **11**: e0161668.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, Niehuis O, Petersen M, Izquierdo-Carrasco F, Wappler T, Rust J, Aberer AJ, Aspock U, Aspock H, Bartel D, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**: 763–767.

Müller D, Jagla T, Bodart LM, Jährling N, Dodt H-U, Jagla K, Frasch M. 2010. Regulation and Functions of the lms Homeobox Gene during Development of Embryonic Lateral Transverse Muscles and Direct Flight Muscles in Drosophila ed. P. Callaerts. *PLoS One* **5**: e14323.

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

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

Nicholson DB, Ross AJ, Mayhew PJ. 2014. Fossil evidence for key innovations in the evolution of insect diversity. *Proc R Soc B Biol Sci* **281**: 20141823–20141823.

Nien CY, Liang HL, Butcher S, Sun Y, Fu S, Gocha T, Kirov N, Manak JR, Rushlow C. 2011. Temporal coordination of gene networks by Zelda in the early *Drosophila embryo*. *PLoS Genet* **7**.

Niwa R, Niwa YS. 2014. Enzymes for ecdysteroid biosynthesis: Their biological functions in insects and beyond. *Biosci Biotechnol Biochem* **78**: 1283–1292.

Nolo R, Abbott LA, Bellen HJ. 2000. Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in Drosophila. *Cell* **102**: 349–62.

Park J, Peng Z, Zeng J, Elango N, Park T, Wheeler D, Werren JH, Yi S V. 2011. Comparative Analyses of DNA Methylation and Sequence Evolution Using *Nasonia* Genomes. *Mol Biol Evol* **28**: 3345–3354.

Peel AD, Chipman AD, Akam M. 2005. Arthropod Segmentation: beyond the

Drosophila paradigm. *Nat Rev Genet* **6**: 905–916.

Petersen AJ, Katzenberger RJ, Wassarman DA. 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, Bellés X. 2010. Key roles of the Broad-Complex gene in insect embryogenesis. *Insect Biochem Mol Biol* **40**: 468–475.

Ritter AR, Beckstead RB. 2010. Sox14 is required for transcriptional and developmental responses to 20-hydroxyecdysone at the onset of drosophila metamorphosis. *Dev Dyn* **239**: 2685–2694.

Rivera MC, Jain R, Moore JE, Lake JA. 1998. Genomic evidence for two functionally distinct gene classes. *ProcNatlAcadSciUSA* **95**: 6239–6244.

Roumengous S, Rousset R, Noselli S. 2017. Polycomb and Hox Genes Control JNK-Induced Remodeling of the Segment Boundary during Drosophila Morphogenesis. *Cell Rep* **19**: 60–71.

Schier AF. 2007. The Maternal-Zygotic Transition: Death and Birth of RNAs. *Science* **316**.

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

Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, McKay DJ, Harrison MM. 2015. Zelda is differentially required for chromatin accessibility, transcription-factor binding and gene expression in the early *Drosophila* embryo. *Genome Res* 1715–1726.

Shao W, Espenshade PJ. 2012. Expanding roles for SREBP in metabolism. *Cell Metab* **16**: 414–9.

Shinmyo Y, Mito T, Matsushita T, Sarashina I, Miyawaki K, Ohuchi H, Noji S. 2005. caudal is required for gnathal and thoracic patterning and for posterior elongation in the intermediate-germband cricket Gryllus bimaculatus. *Mech Dev* **122**: 231–9.

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

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

Stroumbakis ND, Li Z, Tolias PP. 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 CY, Chen K, Liu HY, Johnston J, Zeitlinger J, Rushlow C. 2015. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation. *Genome Res* **25**: 1703–1714.

Suzanne M, Steller H. 2013. Shaping organisms with apoptosis. *Cell Death Differ* **20**: 669–675.

Tadros W, Goldman AL, Babak T, Menzies F, Vardy L, Orr-Weaver T, Hughes TR, Westwood JT, Smibert CA, Lipshitz HD. 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.

Tadros W, Lipshitz HD. 2009. The maternal-to-zygotic transition: a play in two acts. *Development* **136**: 3033–42.

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, Keller T, 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.

Upadhyay M, Samal J, Kandpal M, Vasaikar S, Biswas B, Gomes J, Vivekanandan P. 2013. CpG dinucleotide frequencies reveal the role of host methylation capabilities in parvovirus evolution. *J Virol* **87**: 13816–24.

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

Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA, Robinson GE. 2006. Functional CpG Methylation System in a Social Insect. *Science* **314**.

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

Xiang H, Zhu J, Chen Q, Dai F, Li X, Li M, Zhang H, Zhang G, Li D, Dong Y, Zhao L, Lin Y, Cheng D, Yu J, Sun J, Zhou X, Ma K, He Y, Zhao Y, et al. 2010. Single base–resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat Biotechnol* **28**: 516–520.

Ylla G, Belles X. 2015. Towards understanding the molecular basis of cockroach tergal gland morphogenesis. A transcriptomic approach. *Insect Biochem Mol Biol* **63**: 104–12.

Ylla G, Fromm B, Piulachs M-DD, Belles X. 2016. The microRNA toolkit of insects. *Sci Rep* **6**: 37736.

Ylla G, Piulachs M-D, 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.

Zemach A, McDaniel IE, Silva P, Zilberman D, Li X, Li M, Zhang H, Zhang G, Li D, Dong Y. 2010. Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science* **328**: 916–9.

Zhou X, Riddiford LM. 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**: 2259–69.

4. GENERAL DISCUSSION

Each section of the Results chapter of this thesis corresponds to a research paper. Thus, each section has its own introduction, results and discussion. Nevertheless, the different sections should not be considered as isolated blocks of information, but a set of articles with the same ultimate objective of elucidating the mechanisms regulating insect metamorphosis and their evolution. Each article is a step forward towards the objective, and the purpose of this chapter is to give an overview of the work carried out, as well as a global integration and interpretation of the results obtained.

The chosen approach was the analysis of transcriptomic data, obtained particularly in our model species, the German cockroach *Blattella germanica*, and comparative transcriptomics with other species, especially with *Drosophila melanogaster* and *Tribolium castaneum*, for which a considerable amount of transcriptomic data is publicly available (Celniker et al. 2009; Kim et al. 2010). In total, we have used 150 libraries for RNA-seq, of small RNAs and mRNAs, from 7 insect species (including those mentioned above). A remarkable number of these datasets, 55 libraries, were obtained in our laboratory. From these 55 RNA-seq libraries, 48 were specifically prepared and sequenced in the context of the present thesis. This thesis represents the most extensive source of transcriptomic data on *B. germanica*, and, by extension, the most extensive source of transcriptomic data from a hemimetabolan insect.

Although most of our work is based on transcriptomic data, the

genome sequence is an important tool for our analyses, since a reference of the gene sequences allows using the RNA-seq reads to quantify the expression of annotated genes. While in our previous works (i.e. Appendix 1) the sequence of the *B. germanica* genome was not available, since 2016, thanks to the i5k project (I5K-Consortium 2013), we have had access to the *B. germanica* genome assembly. The first version of the gene models annotation had a rather poor quality, containing pseudo-genes annotated as genes, exons of the same gene annotated as independent genes, regions of bacterial contamination, etc. Nevertheless, the last gene model version (v0.6.2) produced in the context of an international collaboration (Appendix 2), was greatly improved. This updated annotation, which is the one used in the present thesis, was improved in a great measure thanks to our RNA-seq datasets, which allowed to better delimit the transcribed regions of the genome.

Besides the coding genes, the genome contains other genetic elements which also have an impact on the phenotype. Among the regulatory elements, the most relevant at a post-transcriptional level are the microRNAs (miRNAs). Therefore, we proceed to annotate them in the *B. germanica* genome (chapter 3.1) using 7 small RNA-seq from different tissues and developmental stages previously sequenced in the laboratory. The use of the small RNA-seq data together with stringent filtering criteria resulted in a high confidence miRNA catalog for *B. germanica.* In our article (chapter 3.1), we considerably improved the previously published list of miRNA annotations of *B. germanica* (Cristino et al. 2011). This improvement was mainly due to newly acquired small RNA-seq datasets, the availability of the genome assembly and the

implementation of stringent biogenesis criteria. Therefore, we were able to report 40 miRNA families and about 50 miRNA genes previously not identified.

Then, in order to put our miRNA annotations in an evolutionary context, we needed to have comparable miRNA catalogs of similar quality from other insects. Consequently, we proceeded to gather public small RNA-seq data from the following six insect species: *D. melanogaster*, *Locusta migratoria*, *Acyrtosiphon pisum*, *Apis mellifera, T. castaneum* and *Bombyx mori*. Then, we re-annotated their miRNA complement following the stringent biogenesis criteria used in the case of *B. germanica*. This procedure allowed us to obtain comparable miRNA complements from each of the six insects. The comparisons of the miRNA complement between insect species revealed low ratios of loses/gains of miRNA families along insect evolution and a strongly conserved minimal insect miRNA toolkit, composed of a subset of 62 miRNA families. The establishment of an accurate miRNA toolkit in a given clade has also practical applications such as predictive power, thus facilitating the empirical identification of the conserved miRNA complement in a given species of the clade. Our results on insect miRNA annotations were in accordance with that of miRNA phylogenetic studies (Tarver et al. 2012, 2013) and in contrast with those miRNAs complements available in public repositories such as miRBase (Kozomara et al. 2011), which evidences the need for revised and updated miRNA databases. A recent attempt for gathering high-quality miRNA annotations in a database is the initiative of MirGeneDB (Fromm et al. 2015) which will soon include the miRNAs revised in this thesis.

The annotation of the *B. germanica* miRNAs was also necessary in order to later quantify their expression in different developmental stages (chapter 3.2). The stages in which we quantified the miRNA expression were carefully chosen considering the best ratio between economic cost and potential interest of the obtained information. Thus, in order to prepare the corresponding small RNA libraries, we selected eleven stages: non-fertilized eggs, five embryo stages, four nymphal stages and adult females. For each of the eleven stages, two replicates were obtained and sequenced in independent sequencing "runs" in order to be able to detect possible technical or methodological errors. Interestingly, the analysis of the small RNA-seq data from embryo libraries revealed 67 novel miRNA genes that were not identified during the microRNA toolkit project (chapter 3.1). The reason explaining why these newly found miRNAs were not previously detected is because they were preferentially expressed during embryo stages, for which we did not have RNA-seq data previously. These newly found miRNAs are mostly embryo specific, and in general, expressed in lower amounts in comparison with conserved miRNAs. Previously, the only sequencing data available for miRNA expression on insect embryos was from holometabolan models such as *Drosophila spp.* (Ninova et al. 2014) and *T. castaneum* (Ninova et al. 2016). Therefore, our contribution was relevant in order to observe that the preferential expression of novel miRNAs in early embryo stages it is not an exclusive feature of holometabolan insects.

With all these data, we obtained the expression pattern of the 167 miRNA genes of *B. germanica* along the 11 developmental stages represented in the 11 small RNA libraries. To convert the 167

patterns into more tractable and useful information, we generated coexpression networks. The networks obtained showed three clear coexpression modules, each of them corresponding to three wave of miRNA expression in the embryo, while a wave was observed in postembryonic development. Among other data, these results suggest an important role of miRNAs regulating the maternal to zygotic transition (MZT) and the embryo development of *B. germanica*. The idea of applying coexpression networks for collapsing large datasets of expression data originated with the popularization of micro-arrays, and later adopted for mRNA-seq experiments (Ballouz et al. 2015; Iancu et al. 2012). Inspired by the results using coexpression networks for mRNA-seq datasets, we decided to implement it to our small RNA-seq data, which resulted in a comprehensive way to collapse and interpret the expression of 167 miRNAs at 11 time points of development.

In addition, in order to gain knowledge about miRNA functions along insect evolution, we retrieved similar small RNA datasets from *D. melanogaster*, *Drosophila virilis* and *T. castaneum*.

These models allowed us to compare the embryonic miRNA expression between a short germ-band hemimetabolan (*B. germanica*), a short germ-band holometabolan (*T. castaneum*) and two long germ-band holometabolans (*Drosophila* spp.). Thus, we were able to identify miRNAs involved in both, germ-band and metamorphosis type, such as members of MIR-276, MIR-279, Let-7, and MIR-92 families, which were identified as putatively involved in short-germ band definition. In terms of gains and losses of miRNA families, holometabolan insects gained MIR-1006, MIR-

989 and MIR-1007 during evolution, while *T. castaneum* lost MIR-1006 and MIR-1007. Thus, the only miRNA family specific and common to all studied holometabolan species is MIR-989. On the other hand, MIR-bg5 appeared as the only common miRNA family exclusive of hemimetabolan insects. Therefore, we did not observe dramatic changes in the miRNA complement between hemimetabolans and holometabolans, but differences on their expression patterns. In contrast to the acute peaks of the embryo miRNAs in *B. germanica*, the available data on holometabolan insects (Ninova et al. 2016, 2014) shows a more prolongate miRNA expression patterns. Similar patterns were later observed at mRNA level. These differences may account for the different embryonic developments: the hemimetabolan, which forms a nymph with an adult-like general structure, and the holometabolan that forms a larval structure, very divergent from that of the adult.

The analysis of miRNA expression gave us information about which are the most expressed miRNAs at each stage, and which of them could be related to germ-band or metamorphosis definition. Nevertheless, miRNA action is indirect, by targeting mRNAs, thus blocking the translation into proteins, which are the ultimate functional molecules. Therefore, to better understand how miRNAs contribute to regulate a given biological process, we must also consider the mRNAs as potential miRNA targets. For this reason, we studied the expression of mRNAs at the same stage-libraries in which we studied the miRNAs (chapter 3.4). Therefore, we prepared and sequenced the mRNA fraction of samples obtained at the 11 developmental stages used in the study of miRNAs.

The 11 mRNA libraries allowed us to track the expression of the coding genes at each development stages and identify key regulators at each stage. In addition, we used similar public datasets from the holometabolan *D. melanogaster* which allowed us to compare the gene expression between the two insect models. The general analysis shows that the most dynamic changes in *B. germanica* are found during embryogenesis, whereas *D. melanogaster* maintains a high dynamics of changes all along ontogeny, including the larval and pupal stages. A similar study analyzed the expression patterns in different developmental stages of the holometabolan *D. melanogaster* (Graveley et al. 2011). In our case, we went a step forward, and we not only analyzed the transcriptomic changes along the development of our insect model, but we additionally compared the developmental transcriptome of our model with that of other insects with different developmental modes. This allowed us to infer evolutionary conclusions.

Detailed analysis focused on the regulatory genes in key development stages revealed interesting differences between *B. germanica* and *D. melanogaster*. For example, in the MZT, the two main players Zelda and Smaug show a clear expression peak at ED0 and ED1 in *B. germanica*, whereas in *D. melanogaster*, they keep high levels of expression during all embryogenesis and even the first larval instar. In mid-late embryo, we compared the expression profiles of gap, pair-rule, and segment polarity genes, and some of them showed clear differences between both insect models, like the occurrence of a peak of hairy in *D. melanoagaster* larvae and the conspicuous embryonic expression of this gene in *B. germanica*. In this context, the expression of other genes families such the Hox

genes, JH and Ecd synthesis and signaling genes, transcription factors and apoptotic genes have been compared between the two insect models. These comparisons revealed strong conservation of genes between the two insect models, but with some radically different expression patterns of some of them. These conserved genes with different expression patterns are important since they could be responsible for the different developmental modes of *B. germanica* and *D. melanogaster*. Consequently, functional studies focused on these genes are presently in progress in our laboratory.

An interesting finding regarding chapter 3.4 was the discovery that DNA methylation may have an important role in early embryo development of *B. germanica*. The role of DNA methylation in insects has been largely discussed (Glastad et al. 2015; Park et al. 2011; Marhold et al. 2004; Glastad et al. 2013), and here, we reported a new role on hemimetabolan embryo development. Most of the articles published regarding insect DNA methylation investigate its role on cast differentiation and social behavior, although there is no clear agreement concerning the respective results (Wang et al. 2006; Bewick et al. 2016; Honeybee Genome Sequencing Consortium 2006; Elango et al. 2009). Our findings open a door to a new possible role of this epigenomic marker in insect embryo development, although they will require further research, especially from a functional point of view.

5. CONCLUSIONS

1) The elaboration of the present thesis has involved the generation of the richest source of transcriptomic data along the development of a hemimetabolan insect. As transcriptomic information from hemimetabolan species is very scarce (although it is essential for understanding the evolution of hemimetaboly), we can conclude that the transcriptomic data generated is in itself an important contribution to the study of the evolution of insect metamorphosis.

2) The study of the miRNA complement of *Blattella germanica*, using RNA-seq data, miRNA biogenesis information, and phylogenetic analyses, revealed that it is composed of 61 conserved and 58 specific miRNA families. This suggests that the 61 conserved miRNA families regulate more ancestral functions, whereas the 58 specific contribute to regulate more derived functions.

3) An equivalent study of the miRNA complement in two other hemimetabolan insects (*Locusta migratoria* and *Acyrtosiphon pisum*) and in four holometabolans (*Apis mellifera*, *Tribolium castaneum*, *Bombyx mori* and *Drosophila melanogaster*) has allowed the identification of 62 miRNA

families highly conserved across these species. This, together with the wide and generally high expression levels of these miRNAs, suggests that they regulate basic and more ancestral functions that are conserved in hemimetabolan and holometabolan species, and probably in all insects.

4) The 62 highly conserved miRNA families have been defined as the insect minimal miRNA toolkit, which led to predict that any insect species may possess these miRNAs. This predictive power can facilitate the empirical identification of the conserved miRNA complement in other, unstudied insect species.

5) Species-specific miRNAs are characteristically highly expressed during embryogenesis. This suggests that these miRNAs are especially important in embryo development, and become less relevant in nymphal and adult stages.

6) In *B. germanica*, MIR-309 miRNAs are dramatically expressed in a narrow temporal window during the maternal to zygotic transition (MZT). In *D. melanogaster*, MIR-309 miRNAs have a key role in eliminating maternal mRNAs. These two facts suggest that MIR-309 might play the same role in *B. germanica* during the MZT.

7) The hemimetabolan embryo shows a high and very diverse transcriptomic activity that contrasts with the conservative activity in postembryonic stages. Conversely, transcriptomic activity is relatively constant along holometabolan ontogeny. This can be related to the fact that the hemimetabolan embryo constructs the essential adult body structure during the embryogenesis, whereas holometabolans delay the completion of the adult body structure building until the last juvenile stage (the pupa) of postembryonic development.

8) The expression patterns of a number of early embryonic, gap, pair-rule, segment polarity, and Hox genes in *B. germanica* differ from what has been observed in *D. melanogaster*. These differences can account for the divergent body structure patterning between these two species, which represent hemimetabolan and holometabolan embryogenesis, respectively.

9) A correlation between gene expression and DNA methylation marks has been observed in early embryogenesis in *B. germanica* but not in *D. melanogaster*. This difference suggests that DNA methylation can be an important process for hemimetabolan embryogenesis that might have been lost in holometabolans, at least in *D. melanogaster*, during evolution.

6. APPENDIX

This appendix consists of two articles authored by the PhD candidate, which do not perfectly fit into the main story line of the thesis. However, they are included here because they have conceptual connexions with the thesis. Moreover, they can be relevant for the comprehension of the work carried out during the preparation of the thesis.

The Appendix 1 is a paper published in 2015 with the results of my final master project of the candidate undertaken in Belles' Lab. In this work, we explored the role of transcription factors on the morphogenesis of the *Blattella germanica* tergal glands during metamorphosis. Although this article was published before starting the PhD project, it clearly had an impact on the present thesis. Most of the technical and conceptual knowledge and skills obtained on that project were later applied on the PhD project.

The Appendix 2 is a co-authored manuscript that resulted from a collaboration in a large consortium. The consortium was established for sequencing and annotateing the genomes of two Blattodean insects, the cockroach *B. germanica* and the termite *Cryptotermes secundus*. These genomes, plus other insect genomes available and several transcriptomic datasets, were used for studying the evolution of insect sociality. My participation in this project focused on the *B. germanica* genome. I collaborated by manually annotating genes, providing and analyzing the RNA-seq datasets, interpreting results and revising the successive versions of the manuscript.

6.1. Appendix 1

This Appendix is the published article about the *B. germanica* tergal glands. The results of this work were obtained during my master project in 2015, which was based on the analysis of RNA-seq data.

For that project, we obtained 4 transcriptomes of tergites 7 and 8 at different stages and conditions. At that time, the *B. germanica* genome assembly was not yet released, and thus we built a reference transcriptome by assembling a total of 11 RNA-seq datasets from Roche-454 technology. The 4 RNA-seq libraries of tergites 7 and 8 were mapped against the transcriptome assembly to quantify the expression of each transcript in each condition. The assembled transcripts were later annotated by Blast and their function extracted from the GO terms of orthologous genes

Towards understanding the molecular basis of cockroach tergal gland morphogenesis. A transcriptomic approach.

Guillem Ylla & Xavier Belles

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

Ylla G, Belles X. Towards understanding the molecular basis of cockroach tergal gland morphogenesis. A transcriptomic approach. Insect Biochemistry and Molecular Biology (2015), 63: 104–12.

6.2. Appendix 2

This Appendix is a manuscript resulting from a large international consortium in which we actively participated. The consortium was established in order to study and annotate the *B. germanica* genome. In addition to it, the consortium also studied and annotated that of the termite *Cryptotermes secundus*.

The comparison of these two genomes -plus that of two other termites and other non-Blattodean insects- gave clues about the emergence of eusociality in hemimetabolans. This work provided us with an essential tool for our studies, which is the annotated genome of *B. germanica*.

I actively participated in this project and thus, I'm one of the co-authors of the manuscript (note that the central authors are arranged alphabetically). The contributions of the Barcelona group (X. Belles, M.D. Piulachs and G. Ylla) focused on gene annotation, mRNA sequencing and interpretation of results.

Hemimetabolous genomes reveal molecular basis of termite eusociality

Mark Harrison**, Evelien Jongepier**, Hugh Robertson**, Nicolas Arning, Tristan Bitard-Feildel, Hsu Chao, Christopher Childers, Ms. Huyen Dinh, Harshavardhan Doddapaneni, Ms. Shannon Dugan, Johannes Gowin, Ms. Carolin Greiner, Yi Han, Haofu Hu, Daniel Hugher, Ann Kathrin Huylmans, Carsten Kemena, Lukas Kremer, Mrs. Sandra Lee, Alberto Lopez-Ezquerra, Ludovic Mallet, Jose Monroy-Kuhn, Ms. Annabell Moser, Ms. Shwetha Murali, Ms. Donna Muzny, Saria Otani, **Maria-Dolors Piulachs**, Monica Poelchau, Jiaxin Qu, Ms. Florentine Schaub, Ayako Wada-Katsumata, Kim Worley, Qiaolin Xi, **Guillem Ylla**, Michael Poulsen, Richard Gibbs, Coby Schal, Stephen Richards, **Xavier Belles***, Judith Korb*, Erich Bornberg-Bauer*

Hemimetabolous genomes reveal molecular basis of termite eusociality. bioRxiv (2017). Publication in progress in Nature Ecology and Evolution (2017).

Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, et al. Hemimetabolous genomes reveal molecular basis of termite eusociality. Nat Ecol Evol. 2018 Mar 5;2(3):557–66. DOI: 10.1038/s41559-017-0459-1

7. REFERENCES

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, et al. 2000. The Genome Sequence of Drosophila melanogaster. *Science* **287**: 2185–2195.

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, Green ED, Hurles ME, Knoppers BM, Korbel JO, Lander ES, Lee C, et al. 2015. A global reference for human genetic variation. *Nature* **526**: 68–74.

Averof M, Akam M. 1995. Hox genes and the diversification of insect and crustacean body plans. *Earth Atmos Aquat Sci Database pg* **376**.

Ballouz S, Verleyen W, Gillis J. 2015. Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics* btv118-.

Beccaloni G, Eggleton P. 2013. Order Blattodea. In : Zhang, Z.-Q. (Ed.) Animal Biodiversity: An Outline of Higher-level Classification and Survey of Taxonomic Richness (Addenda 2013). *Zootaxa* **3703**: 46–48.

Belles X. 2011. Origin and Evolution of Insect Metamorphosis. *Encycl Life Sci* 1–11.

Belles X, Cristino AS, Tanaka ED, Rubio M, Piulachs MD. 2012. Insect MicroRNAs: From Molecular Mechanisms to Biological Roles. In *Insect Molecular Biology and Biochemistry*, pp. 30–56.

Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. 2016. Evolution of DNA Methylation across Insects. *Mol Biol Evol* 654–665.

Bourguignon T, Lo N, Cameron SL, Šobotník J, Hayashi Y, Shigenobu S, Watanabe D, Roisin Y, Miura T, Evans TA. 2015. The Evolutionary History of Termites as Inferred from 66 Mitochondrial Genomes. *Mol Biol Evol* **32**: 406–421.

Bushati N, Stark A, Brennecke J, Cohen SM. 2008. Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in *Drosophila*. *Curr Biol* **18**: 501–6.

Cedar H, Bergman Y. 2012. Programming of DNA Methylation Patterns. *Annu Rev Biochem* **81**: 97–117.

Celniker SE, L Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD. 2009. Unlocking the secrets of the genome. *Nature* **18**: 927–930.

Chen L, Dumelie JG, Li X, Cheng MH, Yang Z, Laver JD, Siddiqui NU, Westwood JT, Morris Q, Lipshitz HD, Smibert CA. 2014. Global regulation of mRNA translation and stability in the early *Drosophila* embryo by the Smaug RNA-binding protein. *Genome Biol* **15**: R4.

Chipman AD. 2015. Hexapoda: Comparative Aspects of Early Development. In *Evolutionary Developmental Biology of Invertebrates 5*, pp. 93–110, Springer Vienna, Vienna.

Cristino A, Tanaka E, Rubio M, Piulachs M-D, Belles X. 2011. Deep sequencing of organ- and stage-specific microRNAs in the evolutionarily basal insect Blattella germanica (L.) (Dictyoptera, Blattellidae). *PLoS One* **6**: e19350.

de Mendoza A, Sebé-Pedrós A, Sestak MS, Matejcic M, Torruella G, Domazet-Loso T, 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**: E4858-66.

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

Elango N, Hunt BG, Goodisman MAD, 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**: 11206–11.

Engel MS, Grimaldi DA. 2004. New light shed on the oldest insect. *Nature* **427**: 627–30.

Foo SM, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien CY, Kirov N, Shvartsman SY, Rushlow CA. 2014. Zelda potentiates morphogen activity by increasing chromatin accessibility. *Curr Biol* **24**: 1341–1346.

Fox JG, Bennett BT. 2015. Laboratory Animal Medicine. In *Laboratory Animal Medicine*, pp. 1–21, Elsevier.

Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, Newcomb JM, Sempere LF, Flatmark K, Hovig E, Peterson KJ. 2015. A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. *Annu Rev Genet* **49**: 213–42.

Glastad KM, Hunt BG, Goodisman MAD. 2015. DNA methylation and chromatin organization in insects: insights from the Ant *Camponotus floridanus*. *Genome Biol Evol* **7**: 931–42.

Glastad KM, Hunt BG, Goodisman MAD. 2013. Evidence of a conserved functional role for DNA methylation in termites. *Insect Mol Biol* **22**: 143–154.

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, Brown JB, Cherbas L, Davis C a, Dobin A, Li R, Lin W, Malone JH, Mattiuzzo NR, Miller D, et al. 2011. The

developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**: 473–9.

Grimaldi D, Engel MS. 2005. *Evolution of the Insects*. Cambridge University Press.

Grünwald S, Adam I V., Gurmai A-M, Bauer L, Boll M, Wenzel U. 2013. The Red Flour Beetle *Tribolium castaneum* as a Model to Monitor Food Safety and Functionality. In *Advances in biochemical engineering/biotechnology*, Vol. 135, pp. 111–122.

Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, Childers CP, Dinh H, Doddapaneni H, Dugan S, Gowin J, Greiner C, Han Y, Hu H, Hughes DST, Huylmans A-K, Kemena C, Kremer LPM, Lee SL, et al. 2017. Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat Ecol Evol Press*.

Honeybee_Genome_Sequencing_Consortium. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931–949.

Hrycaj SM, Wellik DM. 2016. Hox genes and evolution. *F1000Research* **5**.

I5K-Consortium. 2013. The i5K Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. *J Hered* **104**: 595–600.

Iancu OD, Kawane S, Bottomly D, Searles R, Hitzemann R, McWeeney S. 2012. Utilizing RNA-Seq data for de novo coexpression network inference. *Bioinformatics* **28**: 1592–7.

Kim HS, Murphy T, Xia J, Caragea D, Park Y, Beeman RW, Lorenzen MD, Butcher S, Manak JR, Brown SJ. 2010. BeetleBase in 2010: revisions to provide comprehensive genomic information for *Tribolium castaneum*. *Nucleic Acids Res* **38**: D437–D442.

Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**: D152-7.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Laver JD, Marsolais AJ, Smibert CA, Lipshitz HD. 2015. Regulation and Function of Maternal Gene Products During the Maternal-to-Zygotic Transition in *Drosophila*. In *Current Topics in Developmental Biology*, Vol. 113, pp. 43–84.

Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843–854.

Legendre F, Nel A, Svenson GJ, Robillard T, Pellens R, Grandcolas P. 2015. Phylogeny of Dictyoptera: Dating the Origin of Cockroaches, Praying Mantises and Termites with Molecular Data and Controlled Fossil Evidence. *PLoS One* **10**: e0130127.

Liu PZ, Kaufman TC. 2005. Short and long germ segmentation: Unanswered questions in the evolution of a developmental mode. In *Evolution and Development*, Vol. 7, pp. 629–646.

Lynch JA, El-Sherif E, Brown SJ. 2012. Comparisons of the embryonic development of *Drosophila*, *Nasonia*, and *Tribolium*. *Wiley Interdiscip Rev Dev Biol* **1**: 16–39.

Macedo LMF, Nunes FMF, Freitas FCP, Pires C V., Tanaka ED, Martins JR, Piulachs M-D, Cristino AS, Pinheiro DG, Simões ZLP. 2016. MicroRNA signatures characterizing caste-independent ovarian activity in queen and worker honeybees (*Apis mellifera L.*). *Insect Mol Biol* **25**: 216–226.

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

Metzker ML. 2010. Sequencing technologies - the next generation. *Nat Rev Genet* **11**: 31–46.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, Niehuis O, Petersen M, Izquierdo-Carrasco F, Wappler T, Rust J, Aberer AJ, Aspock U, Aspock H, Bartel D, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**: 763–767.

Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, Kadono-Okuda K, Yamamoto K, Ajimura M, Ravikumar G, Shimomura M, Nagamura Y, Shin-I T, Abe H, Shimada T, et al. 2004. The genome sequence of silkworm, Bombyx mori. *DNA Res* **11**: 27–35.

Nien CY, Liang HL, Butcher S, Sun Y, Fu S, Gocha T, Kirov N, Manak JR, Rushlow C. 2011. Temporal coordination of gene networks by Zelda in the early *Drosophila embryo*. *PLoS Genet* **7**.

Ninova M, Ronshaugen M, Griffiths-Jones S. 2014. Conserved Temporal Patterns of MicroRNA Expression in *Drosophila* Support a Developmental Hourglass Model. *Genome Biol Evol* **6**: 2459–2467.

Ninova M, Ronshaugen M, Griffiths-Jones S. 2016. MicroRNA evolution, expression, and function during short germband development in *Tribolium castaneum*. *Genome Res* **26**: 85–96.

Pace RM, Grbić M, Nagy LM. 2016. Composition and genomic organization of arthropod Hox clusters. *Evodevo* **7**: 11.

Pang D, Thompson DNP. 2011. Embryology and bony malformations of the craniovertebral junction. *Childs Nerv Syst* **27**: 523–64.

Park J, Peng Z, Zeng J, Elango N, Park T, Wheeler D, Werren JH, Yi S V. 2011. Comparative Analyses of DNA Methylation and Sequence Evolution Using *Nasonia* Genomes. *Mol Biol Evol* **28**: 3345–3354.

Peel AD, Chipman AD, Akam M. 2005. Arthropod Segmentation: beyond the *Drosophila* paradigm. *Nat Rev Genet* **6**: 905–916.

Poulsen M, Hu H, Li C, Chen Z, Xu L, Otani S, Nygaard S, Nobre T, Klaubauf S, Schindler PM, Hauser F, Pan H, Yang Z, Sonnenberg ASM, de Beer ZW, Zhang Y, Wingfield MJ, Grimmelikhuijzen CJP, de Vries RP, et al. 2014. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc Natl Acad Sci U S A* **111**: 14500–5.

Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, Gibbs R, Beeman RW, Brown SJ, Bucher G, Friedrich M, Grimmelikhuijzen CJP, Klingler M, Lorenzen M, Richards S, Roth S, Schroder R, Tautz D, Zdobnov EM, et al. 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**: 949–955.

Rubio M, de Horna A, Belles X. 2012. MicroRNAs in metamorphic and non-metamorphic transitions in hemimetabolan insect metamorphosis. *BMC Genomics* **13**: 386.

Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. 2011. The real cost of sequencing: higher than you think! *Genome Biol* **12**: 125.

Schier AF. 2007. The Maternal-Zygotic Transition: Death and Birth of RNAs. *Science* **316**: 406–407.

Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, McKay DJ, Harrison MM. 2015. Zelda is differentially required for chromatin accessibility, transcription-factor binding and gene expression in the early *Drosophila* embryo. *Genome Res* 1715–1726.

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

Tadros W, Lipshitz HD. 2009. The maternal-to-zygotic transition: a play in two acts. *Development* **136**: 3033–42.

Tarver JE, Donoghue PCJ, Peterson KJ. 2012. Do miRNAs have a deep evolutionary history? *Bioessays* **34**: 857–66.

Tarver JE, Sperling E a., Nailor A, Heimberg AM, Robinson JM, King BL, Pisani D, Donoghue PCJ, Peterson KJ. 2013. Mirnas: Small genes with big potential in metazoan phylogenetics. *Mol Biol Evol* **30**: 2369–2382.

Terrapon N, Li C, Robertson HM, Ji L, Meng X, Booth W, Chen Z, Childers CP, Glastad KM, Gokhale K, Gowin J, Gronenberg W, Hermansen RA, Hu H, Hunt BG, Huylmans AK, Khalil SM, Mitchell RD, Munoz-Torres MC, et al. 2014. Molecular traces of alternative social organization in a termite genome. *Nat Commun* **5**: 3636.

The International Aphid Genomics Consortium. 2010. Genome Sequence of the Pea Aphid Acyrthosiphon pisum. *PLoS Biol* **8**: e1000313.

The International Silkworm Genome. 2008. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem Mol Biol* **38**: 1036–1045.

Venter J, Adams M, Myers E, Li P, Mural R, Sutton G, Smith H, Yandell M, Evans C, Holt R, Gocayne J, Amanatides P, Ballew

R, Huson D, Wortman J, Zhang Q, Kodira C, Zheng X, Chen L, et al. 2001. The sequence of the human genome. *Science* **291**: 1304–1351.

Wang L, Wang S, Li Y, Paradesi MSR, Brown SJ. 2007. BeetleBase: the model organism database for *Tribolium castaneum*. *Nucleic Acids Res* **35**: D476-9.

Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, Li B, Cui F, Wei J, Ma C, Wang Y, He J, Luo Y, Wang Z, Guo X, Guo W, Wang X, Zhang Y, Yang M, et al. 2014. The locust genome provides insight into swarm formation and long-distance flight. *Nat Commun* **5**: 2957.

Wang Y, Jorda M, Jones PL, Maleszka R, Ling X, Robertson HM, Mizzen CA, Peinado MA, Robinson GE. 2006. Functional CpG Methylation System in a Social Insect. *Science* **314**: 645–647.

Wightman B, Ha I, Ruvkun G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* **75**: 855–862.

Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, Cheng T, Chai C, Pan G, Xu J, Liu C, Lin Y, Qian J, Hou Y, Wu Z, Li G, Pan M, et al. 2004. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* **306**: 1937–1940.

Yin C, Shen G, Guo D, Wang S, Ma X, Xiao H, Liu J, Zhang Z, Liu Y, Zhang Y, Yu K, Huang S, Li F. 2016. InsectBase: a resource for insect genomes and transcriptomes. *Nucleic Acids Res* **44**: D801.

Ylla G, Belles X. 2015. Towards understanding the molecular basis of cockroach tergal gland morphogenesis. A transcriptomic approach. *Insect Biochem Mol Biol* **63**: 104–12.