Factors determining the adult body plan in the hemimetabolan embryogenesis

Alba Ventos Alfonso

TESI DOCTORAL UPF / 2021

Thesis supervisor

Dr. Xavier Bellés Ros

INSTITUT DE BIOLOGIA EVOLUTIVA

(CSIC – UNIVERSITAT POMPEU FABRA)

DEPARTMENT OF EXPERIMENTAL AND HEALTH SCIENCES

Universitat Pompeu Fabra Barcelona

A ti. Y a ellos.

Acknowledgements

En primer lugar, quiero agradecer a Xavier no sólo la oportunidad de hacer la tesis en su laboratorio, sino el haberme enseñado tanto, haberme apoyado siempre, y haber confiado en mí, agradecer la paciencia, y el buen ambiente siempre. También quiero agradecer a Dolors, quien siempre ha estado dispuesta a ayudar y me ha enseñado tanto dentro del laboratorio. Y cómo no, a Jose, con quien hemos discutido sobre protocolos, resultados y demás, aprendiendo a tener un ojo crítico y a ser mejor científica, te considero un gran científico, y una gran persona.

I also want to thank my PhD buddy, Tim, we have discussed about science, about life, we have laughed and cried together, we have learnt together, and this PhD would not have been the same without you, so, thanks, for your help, and for your friendship. A Vivi, quien me ha apoyado en los buenos momentos, en los malos, y en los peores, una de las mejores cosas que me ha dado este doctorado es tu amistad, gracias por ayudarme tanto dentro, como fuera del laboratorio. En general, quiero agradecer a todos las personas que han pasado por el lab, haciendo esta tesis más divertida y llevadera, Guillem, Ari, Patricia, Julia, Fleur, Oscar, Jorge, Jimena, Cynthia, Anxo y muchos más. A Natalia y Jose Carlos, mis bioinformáticos favoritos, quienes me han ayudado incluso después de abandonar el laboratorio. A Cristina, por la labor que hace cuidándonos nuestros bichitos, y por las conversaciones que hemos tenido, y el apoyo que

me has dado siempre. Y a Ana, quien me enseñó todo en mi primer año, y a quien he tenido presente durante toda la tesis.

A compañeros de otros laboratorios, Elena, que me enseñó a cultivar células S2 con mucha paciencia, a François que me enseñó a utilizar el iluminómetro, a Alicia, con quien no sólo compartimos la organización de seminarios, sino charlas en las que tu alegría y entusiasmo se me contagiaban.

I would like to also thank the people from Lynch's Lab, Jeremy, Deanna and Alex, who allowed me to visit their lab in Chicago, we shared and discussed our results, they taught me a lot, and made me enjoy my stage there.

También agradecer a Gemma y Núria, aquellas amigas que aún en la distancia he sentido tan cerca, con quienes me puedo desahogar y con quienes compartimos experiencias en nuestros respectivos labs. Un gran apoyo en mi vida, gracias por estar ahí, siempre.

Quiero agradecer a toda mi familia, abuelos, tíos y primos, quienes siempre se han interesado por lo que hago, y se han mostrado orgullosos. A mi hermano, con quien lo he compartido todo, incluso el laboratorio, parece mentira lo mucho que nos parecemos, y a la vez lo diferentes que somos, da gusto poder hablar contigo de todo, y saber que siempre me apoyas, siendo el que más orgulloso está de todo lo que consigo. A mi padre, quien siempre confía en mí, y quien me ha hecho ser como soy, estos últimos años han sido duros,

y me has enseñado que hay que ser fuerte, y seguir pase lo que pase, ¡muchas gracias! A mi madre, quien habría llorado como una magdalena durante toda la defensa de mi tesis, quien siempre se mostró orgullosa de mí, y quien siempre me dijo que, si quiero, puedo, apoyando todas mis decisiones, y haciéndome ser quien soy.

Por último, pero no menos importante, quiero agradecer a Rodol, porque ha sido mi soporte durante todos estos años, sólo tú sabes lo dura que ha sido esta tesis a nivel personal, porque no hace falta que hable para que sepas cómo estoy, ni qué necesito, porque me has apoyado siempre, y me has ayudado todo lo que has podido y más. Me has soportado en mis peores momentos, haciéndolos mucho más llevaderos, gracias por entenderme, y quererme así.

Sin vosotros, nunca habría empezado esta tesis, ni la habría acabado. Gracias.

Abstract

The idea of this work has been to identify factors that determine the adult body plan in hemimetabolan embryogenesis, using the cockroach *Blattella germanica* as model. Comparative genomics studies pointed to three factors in this regard: DNMT1, Zelda and E93. We found that DNMT1 (DNA methyltransferase 1) methylates DNA, which is crucial for early embryo development. Moreover, methylated genes are expressed at high levels and less variably. Zelda activates the expression of zygotic genes in the maternal-to-zygotic transition (MZT), and is required for proper embryogenesis, especially for abdomen morphogenesis. E93 is key for activating the zygotic genome during the MZT, which is crucial for germ band formation and for the adultoid nymph morphogenesis. *E93* is expressed during embryogenesis of *B. germanica*, but not in the holometabolan *Drosophila melanogaster*. Data suggest that this could be general in hemimetabolans and holometabolans, respectively. If E93 determines the nymphal genetic program in the hemimetabolan embryo, we speculate that the loss of expression of this gene in the holometabolan embryo would have facilitated the implementation of a larval genetic program, and the evolution of holometaboly.

Resum

La idea d'aquest treball ha estat identificar factors que determinin el pla corporal de l'adult durant l'embriogènesi hemimetàbola, utilitzant com a model la panerola *Blattella germanica*. Els estudis de genòmica comparativa van indicar tres factors al respecte: DNMT1, Zelda i E93. Hem trobat que la DNMT1 (DNA metiltransferasa 1) metila el DNA, la qual cosa és crucial al principi del desenvolupament embrionari. A més, els gens metilats s'expressen a nivells més alts i de manera menys variable. Zelda activa l'expressió dels gens zigòtics durant la transició maternal a zigòtica (MZT), i és necessària per a la correcta embriogènesi, especialment per a la morfogènesi de l'abdomen. E93 és clau per activar el genoma zigòtic durant la MZT, la qual cosa és crucial per a la formació de la banda germinal i per a la morfogènesi de la nimfa adultoide. *E93* s'expressa durant l'embriogènesi de *B. germanica*, però no a l'holometàbol *Drosophila melanogaster*. Les dades suggereixen que això podria ser general en hemimetàbols i holometàbols, respectivament. Si E93 determina el programa genètic nimfal a l'embrió hemimetàbol, pensem que la pèrdua d'expressió d'aquest gen a embrions holometàbols hauria facilitat la implementació d'un programa genètic larvari i l'evolució de l'holometabolia.

Preface

The evolutionary innovation of wings and metamorphosis made insects the largest and the most diverse group of animals known nowadays. Thus, it is not surprising that entomologists have been long attracted to the study of metamorphosis. During the last 20 years, the evolution of Insect Metamorphosis Group, headed by Xavier Belles at the Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra) has been working on the genetic mechanisms underlying the origin and evolution of metamorphosis in insects. The group mainly focused on the action of juvenile hormone and its molecular signaling pathway in postembryonic development of the German cockroach *Blattella germanica*. This species shows the ancestral hemimetabolan mode of metamorphosis, from which derives the modified holometabolan mode. Therefore, *B. germanica* serves as a baseline model for evolutionary studies, allowing comparative analyses with holometabolan insects (such as *Drosophila melanogaster* which has been more thoroughly studied). Moreover, *B. germanica* is highly sensitive to gene transcript depletion by RNA interference (RNAi), which makes it especially suitable for functional studies. All this gives the possibility to infer molecular mechanisms implied in the evolutionary transition from hemimetaboly to holometaboly.

Strikingly, comparative studies of hemimetabolan and holometabolan metamorphosis have traditionally focused on post-embryonic development. However, the main difference between

both types of metamorphosis occurs in embryogenesis, as in hemimetabolan insects it produces a nymph with the basic characteristics of the adult body structures, whereas in holometabolans, embryogenesis produces a larva, morphologically divergent with respect to the adult.

To help fill this gap, in this work we studied the mechanisms regulating embryogenesis in *B.* germanica in the context of metamorphosis. Thus, we functionally studied three factors (DNA methyltransferase 1, Zelda and E93) that could contribute to the formation of adultoid nymph in *B. germanica*, and we compared our results with those of holometabolan models described in the literature. We believe that the results obtained open new points of view to understand the evolution from hemimetaboly to holometaboly.

Abbreviations

ADD	ATRX, DNMT3, DNMT3L domain
AdD0-AdD8	Adult day 0 – Adult day 8
BAH	Bromo-adjacent homology domain
Bg	*Blattella germanica*
BR-C	Broad Complex
br	Broad Complex
cad	caudal
cDNA	Complementary DNA
CV	Coefficient of variation
DAPI	4’,6’-diamino-2-phenylindole
dl	dorsal
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DNMT1	DNA methyltransferase - 1
DNMT2	DNA methyltransferase - 2
DNMT3	DNA methyltransferase - 3
dpp	decapentaplegic
dsDNMT1	dsRNA targeting DNMT1
dsDNMT3	dsRNA targeting DNMT3
dsE93	dsRNA targeting E93
dsMock	dsRNA targeting Polyhedrin
dsRNA	Double stranded RNA
dsZelda	dsRNA targeting Zelda
EC1, EC2, EC3	Embryo cuticle 1, 2 and 3
ED0-ED21	Embryo on day 0 – Embryo on day 21

eve	even-skipped
E93	Ecdysone-induced protein 93F
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
ftz	Fushi tarazu
hb	hunchback
JH	Juvenile hormone
Kr	Krüppel
Kr-h1	Krüppel homolog 1
L1-L5	First instar larva – fifth instar larva
mCG	Methylated CG
miRNA	MicroRNA
mRNA	Messenger RNA
MZT	Maternal to zygotic transition
Mya	Million years ago
ncRNA	Noncoding RNA
N	Notch
NFE	Non fertilized eggs
nos	nanos
N1-N6	First instar nymph – Sixth instar nymph
odd	odd-skipped
otd	orthodenticle
PCR	Polymerase Chain Reaction
qRT-PCR	quantitative reverse transcription PCR
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-Seq	RNA sequencing

RRBS	Reduced Representation Bisulfite Sequencing
sog	short gastrulation
sna	snail
Tl	Toll
tRNA	Transfer RNA
UTR	Untranslated region
wg	wingless
zld	Zelda
ZF	Zinc Finger
20E	20-Hydroxyecdysone

Table of contents

1. INTRODUCTION

With about 1 million species formally described, insects represent over a half of all extant animal species on Earth. They are the most diverse animal group in terms of number of species and in morphological, physiological and life cycle diversity. This allowed them to colonize almost all major terrestrial and freshwater habitats, and even a few species can live in the open ocean. The extraordinary success of insects is due in great part to their long evolutionary history (Grimaldi and Engel 2005).

The earliest fossil attributed to an insect is preserved in Lower Devonian materials. However, phylogenetic studies (Misof et al. 2014, Wang et al. 2016) suggest that the first wingless and ectognathous (mouthparts not enclosed in a cavity) true insects diverged from entognathous hexapodan ancestors in the Lower Ordovician (Figure 1.1). During evolution, insects have achieved different developmental innovations. Two of these are the emergence of functional wings (and with that, powered flight) and the appearance of metamorphosis in winged insects. Metamorphosis is a developmental process by which the insect continues developing after hatching and changes its morphology until reaching the reproductively competent adult stage. The adult is the final stage, with no further molts. These two innovations, wings and metamorphosis, appeared almost simultaneously (about 410 Mya) (Figure 1.1) and allowed insects to radiate in an extraordinary way (Belles 2020).

There are different types of development and metamorphosis. In the most primitive wingless (or apterygote) insect groups (Archaeognatha and Zygentoma), postembryonic development is ametamorphic or ametabolan. That is, the insect reaches the reproductively capable adult stage through gradual growth by molting, and it continues molting after attaining adulthood. In contrast, winged (or pterygote) insects undertake metamorphosis, and the postembryonic morphological changes can be more or less dramatic depending on the group. Exopterygotes (Paleoptera, Polyneoptera, and Paraneoptera), follow a hemimetabolan metamorphosis, that is, the nymphs are like miniature adults and develop gradually until adulthood. Their developmental cycle comprises three stages: embryo, juvenile instars (called nymphs) and adult. The main differences of nymphs with the adult are the fully developed wings and functional genitalia. In contrast, Endopterygotes follow a holometabolan metamorphosis, with a developmental cycle comprising four stages: embryo, juvenile instars (called larvae), the pupa and the adult. The larvae are morphologically divergent with respect to the adult, and the pupal stage, usually nonfeeding and immobile stage, bridges the morphological gap between the larva and the adult. In both groups (hemimetabolans and holometabolans, but especially in the latter) there are dramatic diversity of forms and lifestyles in the postembryonic development. The differences between the juvenile stage and the adult allowed to exploit different habitats and food sources and acquire new adaptations. This explains, in great part, the evolutionary success of insects (Belles 2020).

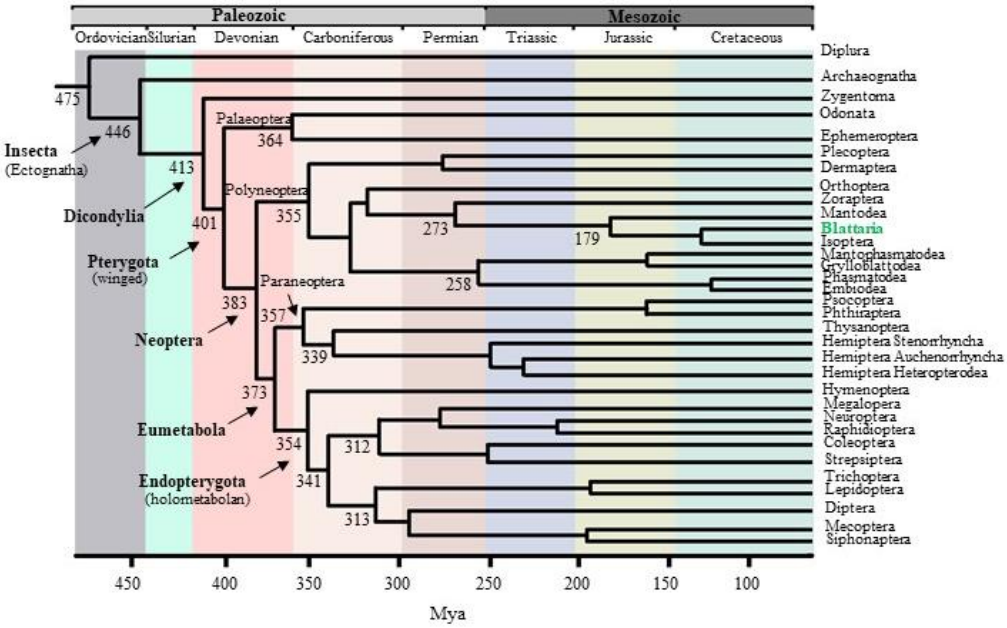


Figure 1.1. Cladogenesis of the main insect groups in a chronological context, as proposed by Misof et al. (2014) and Wang et al. (2016). Divergence times here are based on the average values reported by Wang et al. (2016). The Blattaria branch, with the species *Blattella germanica*, the model used in this thesis, is highlighted in green. Modified from Belles (2020).

Most of studies regarding the evolution from hemimetaboly to holometaboly focus on postembryonic development. However, the differences between both modes of metamorphosis start at embryogenesis. Hemimetabolan embryogenesis gives rise to a miniature adult, while holometabolan embryogenesis gives rise to a larva that can be considerably different from the adult, often exhibiting a vermiform shape. Therefore, a comparative study of embryogenesis in hemimetabolan and holometabolan insects should illuminate how the evolutionary transition from hemimetaboly to holometaboly was produced. The general aim of this thesis is to

study embryonic factors that may have contributed to this transition. The main question would be how holometabolan embryogenesis diverged from the hemimetabolan giving rise to a kind of a worm instead of a miniature adult.

1.1. The metamorphosis of insects and their evolution

Paleontological studies indicate that the oldest insects (currently wingless) were ametabolan. It is at the beginning of Permian (300 Mya) when practically all pterygotes exhibit a postembryonic development with metamorphic transitions between nymphal and adult stages. This represents the origin of hemimetaboly. The first fossil insects that are considered holometabolan are found in Permian strata (approximately 280 Mya). Nowadays, the most stablished hypothesis is that holometabolans originated from hemimetabolan ancestors (Belles 2011). Indeed, from a nymph similar to the adult of a hemimetabolan ancestor, it is easy to imagine how evolution favored morphological changes in those juvenile stages, improving their capabilities to exploit new resources (vermiform larvae are more efficient making galleries into a fruit, for example). This also allows the adult to improve the reproductive capabilities, when forming the external genitalia and the functional wings. Indeed, the evolvability of the larva may confer an advantage to holometabolan insects, who represent between 45% and 60% of all living species (Belles 2011).

1.1.1. Different modes of insect development

a) The ametabolan development

Ametabolan insects (Archaeognatha and Zygentoma) undergo a direct gradual development from the first nymphal instar to the adult (Figure 1.2). The first nymphal instar has a morphology and lifestyle similar to those of the adult, although adult morphological characteristics are gradually added to the nymphs. They molt during their whole life, with more than one sexually mature instar. They do not only increase in size, but the development also includes few metamorphic changes, like the loss of hatching structures or acquisition of scales and fully functional genitalia.

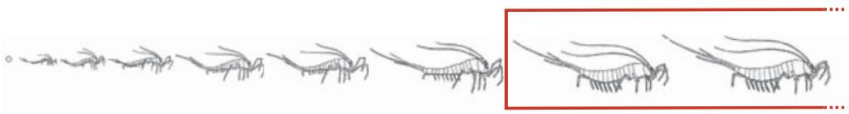


Figure 1.2. Ametabolan development. The adult, reproductively competent stages are showed in a red square, which is open because the adult continues molting. Modified from Belles (2011).

b) The hemimetabolan development

Hemimetaboly appeared with the emergence of wings and the pterygote insects. Palaeoptera, Polyneoptera and Paraneoptera insect superorders follow the hemimetabolan mode of development. When the first nymphal instar emerges, it has a morphology and lifestyle similar to those of the adult. Thus, the nymphal development basically consists in growing (Figure 1.3). While growing, the growth ratio of the exoskeleton linear dimensions

remains approximately constant throughout all nymph-to-nymph molts (Dyar's rule) (Sehnal 1985; Kivelä et al. 2016). When the nymph reaches a specific critical size, it molts to the adult, which is the final stage, as it does not molt anymore. In this transition from the last nymphal instar to the adult, many changes occur, the most important of which are the complete formation of the wings and functional genitalia and other reproductive organs.

Figure 1.3. Hemimetabolan development. The adult, reproductively competent stage, is showed in a red square, which is closed because the adult does not molt anymore. Modified from Belles (2011).

c) The holometabolan development

Endopterygota is a monophyletic group that develops wings inside the body and follows a complex metamorphosis, including a pupal stage. This type of metamorphosis is called holometabolan. Moreover, the larvae of Endopterygota are morphologically divergent from the adult, and usually follow a different lifestyle.

The holometabolan embryo gives rise to a larva that is different from the adult. Then, the postembryonic development consists of several larval instars where the insect grows (Figure 1.4). At the end of the larval period, the pupa is formed, which bridges the gap between larval-adult morphology, and finally the adult emerges.

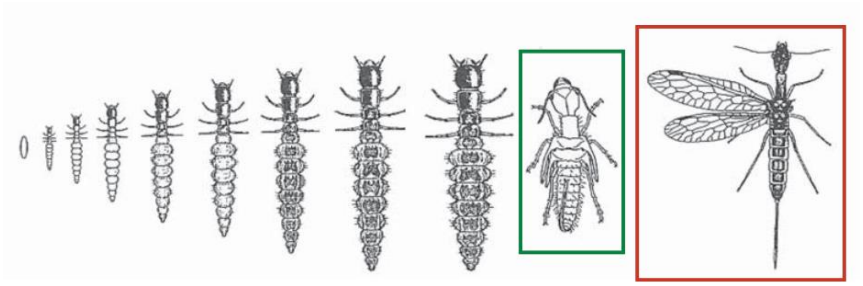


Figure 1.4. Holometabolan development. The adult, reproductively competent stage, is showed in a red square, which is closed because the adult does not molt anymore. The pupal quiescent stage is showed in a green square. Modified from Belles (2011).

1.1.2. Theories explaining the evolution of metamorphosis

There are two main theories explaining the evolution of metamorphosis: the pronymph theory and the theory of direct homology between stages.

a) The pronymph theory

The pronymph theory basically proposes that the embryo of holometabolan species hatches in earlier stages than the hemimetabolan embryo (Figure 1.5). The basis of this theory was first advanced by Lubbock (1873) and later revived by Truman and Riddiford (1999; 2002). The latter authors argued that during hemimetabolan embryogenesis, three embryonic cuticles are deposited (EC1, EC2 and EC3) and after EC2 deposition, there is a stage called pronymph that would derive into the first nymphal instar. In contrast, EC3 would not be deposited in holometabolan embryos, and instead of pronymph formation, the embryo hatches

prematurely as a larva; later, the insect resumes the arrested development in the pupa, where it would form the adult structures. Thus, the corollary is that the hemimetabolan pronymph would be homologous to all the holometabolan larval instars, whereas all the hemimetabolan nymphal instars would be homologous to the pupa (Truman and Riddiford 1999; Truman and Riddiford 2002). Although Konopová and Zrzavý (2005) refuted the concept that holometabolan embryos deposit only two cuticles (the embryo of all species deposit EC1, EC2 and EC3, except some very derived species that have secondarily lost EC3), Truman and Riddiford adapted the theory to these evidences. They maintained that holometabolan embryos arrest development at specific time, and they resume it at the pupal stage, maintaining the homologies stated before (Truman 2019; Truman and Riddiford 2019).

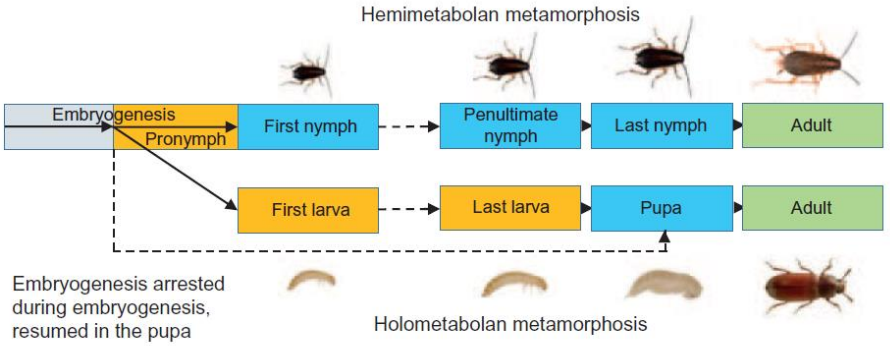


Figure 1.5. Pronymph theory. This theory considers that holometabolan embryos hatch in earlier stages than hemimetabolan embryos. The pronymph would be homologous to all the holometabolan larval instars, whereas all the hemimetabolan nymphal instars would be homologous to the pupa. Modified from Belles (2020).

b) The theory of direct homology between stages

The theory of direct homology between stages proposes that the stages of both modes of metamorphosis are respectively homologous (Figure 1.6). This theory was built on the first concepts of Éraste Poyarkoff, who proposed that the holometabolan pupa originated because the insect needed a "mold" to develop adult skeletal muscles (see Belles, 2020). He considered the holometabolan pupa homologous to the hemimetabolan adult, dividing the holometabolan adult stage in two phases. Poyarkoff also considered that the juvenile stages of both modes of metamorphosis hatch at a comparable stage of development. Later, Hinton (1963) refuted the idea of the pupa being a mold of the holometabolan. Importantly, Hinton concluded that the pupa would be homologous to the last hemimetabolan nymphal instar, thus, the holometabolan juvenile stages would be homologous to those of the hemimetabolans.

Subsequently, comprehensive phylogenetic, morphological, physiological, and paleontological arguments were added to Hinton contribution (Kukalová-Peck 1991; Švácha 1992; Sehnal et al. 1996). Among these arguments, those based on an endocrine point of view were especially relevant. Accordingly, hemimetabolan metamorphosis starts when the production of Juvenile hormone (JH) decreases in the last nymphal instar, whereas, in holometabolans the JH decrease occurs twice, firstly in the last larval instar, stimulating the pupa formation, and secondly in the

pupa, triggering adult formation. At the end, the decrease of JH lets to adult morphogenesis in the last nymphal instar and in the holometabolan pupa. Thus, the JH disappears in the last nymphal instar and in the pupa, which supports the homology of these two stages.

Regarding embryo development, Poyarkoff in 1914, and then Sehnal, Švácha and Zrzavý (1996) reached the conclusion that embryogenesis is equivalent in both modes of metamorphosis, hemimetabolan and holometabolan.

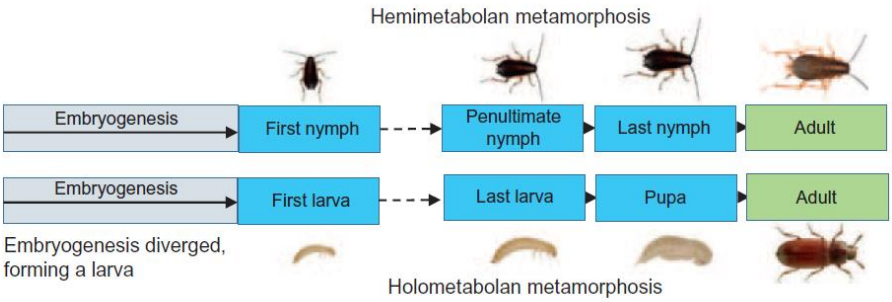


Figure 1.6. Direct homology between stages theory. This theory proposes that embryogenesis is equivalent in both modes of metamorphosis, and the juvenile stages of both modes of metamorphosis are homologous. Thus, the pupa would be a modified nymph. Modified from Belles (2020).

1.2. Insect embryonic development

Insect embryogenesis is the process by which a juvenile (either a larva or a nymph) is formed from an egg. The oocytes developed by the adult female contain a single copy of the haploid genome, plus nutrients and other molecules important for proper embryo

development. Then, the spermatozoid from the male fuses with the oocyte resulting in a diploid cell. This process is known as egg fertilization. Initially, the fertilized egg divides and produces hundreds of cells, which will grow, move, and differentiate into organs and tissues. Although embryogenesis is very diverse in different insects, even between closely related species, there are some common traits that will be explained in the following sections. The following data are borrowed from the comprehensive reviews of Jura (1972) and Anderson (1972a; 1972b).

1.2.1. Egg layers

Eggs have three layers protecting the internal content: the egg membrane, the vitelline membrane and a hard shell named chorion. The internal content is formed by the zygote nucleus, and nutritive materials for nourishment and growth of the embryo (such as yolk proteins, lipids, and carbohydrates). Also patterning molecules (like specialized proteins and mRNAs) that direct key events during embryogenesis, like the establishment of axis polarity, segmentation, and gastrulation.

1.2.2. Egg cleavage

In general, egg cleavage involves the subdivision of cytoplasm (cytokinesis) and subdivision of nuclear material (karyokinesis). The result are individual cells called blastomeres. During egg cleavage, maternal mRNAs and proteins form gradients from anterior to posterior poles of the egg. These gradients provide the

coordinates that position the front and the back of the embryo. However, egg cleavage in insects is very unusual: only karyokinesis takes place, not cytokinesis. This type of cleavage is called syncytial cleavage and results in a compartment occupied by a number of nuclei (syncytium). In most insects, after syncytial cleavage, syncytial nuclei are surrounded by islands of cytoplasm that separate the nuclei from each other. These early cleavages are synchronized, so, as the nuclei divide, they separate, generating a regular spacing between them. Nuclei and associated cytoplasm are referred to as energids (Figure 1.7).

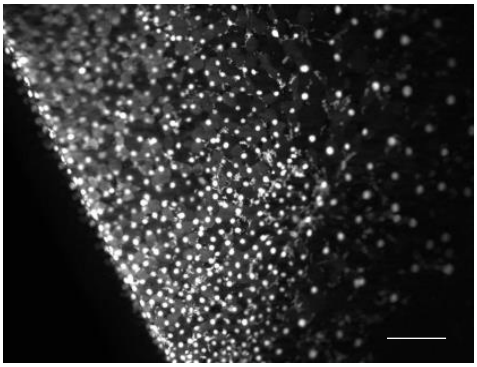


Figure 1.7 Energids in *Blattella germanica* embryo. DAPI staining in day 2 embryos, when energids are formed and migrate to the periphery.

1.2.3. Formation of the blastoderm

Once a critical density of energids is reached, they migrate to the periphery of the egg, where they undergo several rounds of division. Then, the egg membrane invaginates from the egg surface to surround each of the individual syncytial nuclei. The single sheet of cells formed in the periphery of the egg is the cellular blastoderm. Some of the energids do not migrate to the periphery and remain in the yolk mass. These energids become vitellophages, that break the yolk, which will be used for embryo nutrition.

During blastoderm formation, maternal mRNAs and proteins activate the expression of the embryonic Gap genes, which specify broad domains that will correspond to several contiguous segments. This process where the developmental control passes from the maternal components to the zygote is known as the maternal-to-zygotic transition (MZT). The MZT comprises different molecular events that mediate the degradation of the maternally deposited mRNAs and the transcriptional activation of the zygotic genome.

1.2.4. Germ anlage formation

A number of blastoderm cells are destined to form the embryo. These coalesce to form the germ anlage, that later will develop into the germ band. Most of the more modified holometabolan species undergo long germ band development, which means that the germ anlage forms nearly the entire blastoderm surface. Thus, the complete body is established at the blastoderm stage, and all segments (head, gnathal, thoracic and abdominal segments) are formed at once. In contrast, less modified groups (ametabolan and hemimetabolan insects) undergo short germ band or intermediate germ band development, which means that after formation of blastoderm, a relatively small portion of blastoderm cells migrate and aggregate near the posterior pole, where the germ anlage forms. In this case, the head lobes, the most anterior trunk segments, and the posterior terminus are established first, and then, new segments are added progressively through proliferative growth. This process

of posterior growth is similar to the mechanism followed by vertebrates to form their anterior-posterior axis. Although holometabolans typically undergo long germ band development there are exceptions, like the coleopteran *Tribolium casteneum*, which follows an intermediate germ band development.

1.2.5. Serosal and amnion membranes formation

The cells that do not contribute to the germ anlage, form an extraembryonic membrane called serosa. Moreover, the cells that are immediately adjacent to the germ anlage form a second membrane, called amnion. These cells proliferate, flatten and elongate. As they extend over the germ band, they seem to be a sleeping bag that is pulled up from posterior pole and pulled down from the presumptive head lobes of the germ band. Then, the amnion cells meet in the middle of the embryo and form a single cell layer. This layer lies between the germ band and the now separated serosa and covers the germ band leaving a space called amniotic cavity.

1.2.6. Gastrulation

After blastoderm formation, there is a process of cell invagination called gastrulation. Those cells remaining at the blastoderm periphery constitute the ectoderm, and cells that invaginate below the ectoderm constitute the mesoderm. Gastrulation can develop in different ways: mesodermal cells invaginating simultaneously or mesodermal cells invaginating sequentially, starting at the anterior

pole and progressing towards the posterior pole. Regardless of the gastrulation mechanism, the result is a bi-layered embryo with mesodermal precursors underlying the ectoderm. This two-layered structure is called germ band.

1.2.7. Segmentation

Segmentation proceeds nearly simultaneously with gastrulation. It consists in a division of the embryo into a series of similar or identical groups of cells recurring along the length of the embryo. These units are known as segments or metameres. Segmentation proceeds by a progressive refinement of positional information that specify the groups of cells that will form each metameric unit. Refinement starts with maternal mRNAs and protein gradients that activate the expression of Gap genes during the blastoderm formation. In turn, the Gap genes regulate the expression of the Pair-rule genes, which are the responsible for setting the para-segmental boundaries, thus, initiating the first metameric patterns.

In the case of long germ band insects, specific maternal genes are found in both poles (anterior-posterior), Gap genes are distributed all along the embryo, and Pair-rule genes specify anterior and posterior segments by the end of blastoderm stage (Figure 1.8). In contrast, in short and intermediate germ band insects, maternal transcripts are only found in the posterior pole of the egg, Gap genes are also expressed in the posterior pole of the egg, and only anterior segments are specified by Pair-rule gene products by the

end of blastoderm stage (Figure 1.8). Thus, posterior segments are sequentially specified in an antero-posterior progression, while the germ band elongates (Liu and Kaufman 2005).

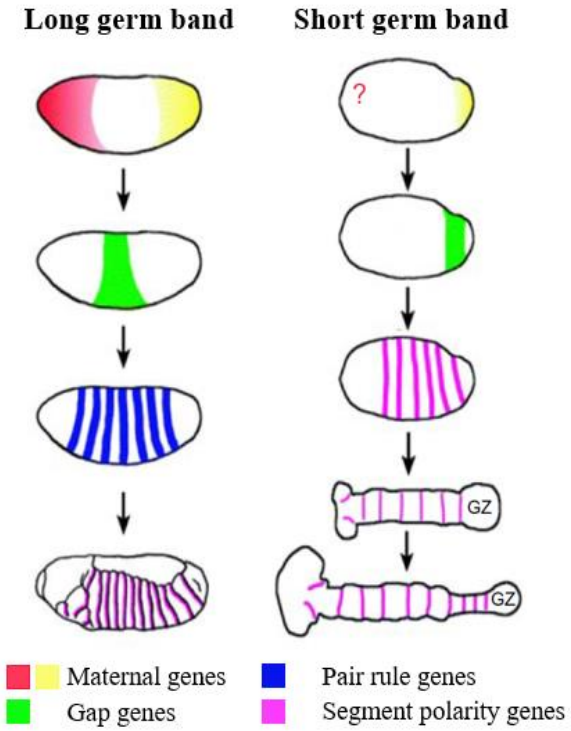


Figure 1.8. Long and short/intermediate germ band segmentation. Long germ band is exemplified by *Drosophila melanogaster* (left), whereas short germ band is exemplified by *Oncopeltus fasciatus* (right). Adapted from Liu and Kaufman (2005).

1.2.8. Blastokinesis

Around mid-embryogenesis, especially in short and intermediate germ band modes, the embryo moves into the yolk mass. This process is called blastokinesis, and normally results in partial revolution of the embryo. Blastokinesis develops in two steps:

anatrepsis and katatrepsis. During anatrepsis, the embryo reverses its axes with respect to the egg, by being pulled into the yolk from the posterior end. Subsequently, during katatrepsis, there are reversal movements that return the embryo to its original position on the ventral side of the egg (Panfilio 2008).

1.2.9. Organogenesis, appendage development and dorsal closure

During the next steps, the germ band transforms into a three-dimensional nymph or larva by differentiating the ectoderm and mesoderm into organ systems. The first differentiated segments appear in the anterior end: the eyes and the labrum and antennae, being formed from the ectoderm. From anterior to posterior end, the three first segments behind the mouth form paired appendages that later become the mandibles, the maxillae, and the labium. The next three segments form the prothorax, mesothorax and metathorax, and the respective appendages that become legs. The ectoderm forms the skin and most of the nymphal or larval external morphology, with bristles and sensory devices. Moreover, ventral ectoderm differentiates into nervous system, whereas the lateral ectoderm invaginates and forms the tracheal system. Other ectodermal invaginations form the ocelli, prothoracic gland, salivary glands, corpora allata-corpora cardiaca, oenocytes and silk glands. The two final invaginations of the ectoderm form the stomodeum and proctodeum. The stomodeum is an invagination at the central-anterior position of the germ band, when it invaginates the cells proliferate in the posterior pole forming the foregut. The

proctodeum is an invagination at the terminal segment of the germ band, and when it invaginates, cells proliferate to anterior pole forming the hindgut. A region of the proctodeum evaginates forming pocket-like cavities, which further develop and form the Malpighian tubules. At the edge of the foregut and hindgut invaginations, a third germ layer arises, which is known as endoderm. The endoderm finally fuses with the foregut and the hindgut, forming the midgut, thus, completing the digestive system. The mesoderm also invaginates, forming different segments. In each segment a pair of transient coelomic sacs are formed. These coelomic sacs develop the dorsal vessel (or "heart"), the internal reproductive organs, the muscles, the fat body, the subesophageal gland and the hemocytes. As the embryo develops, the flanks of the germ band, both ectoderm and mesoderm, extend around the yolk. Finally, the two edges of the germ band meet and fuse along the dorsal mid-line. Thus, the mesodermal and ectodermal layers enclose the yolk. This process is known as dorsal closure. In hemimetabolan insects, legs and wings develop as direct evaginations from the lateral ectoderm. Leg buds appear early, just after gastrulation finishes, whereas wing buds appear later, after lateral ectoderm has grown dorsally. In contrast, in holometabolan insects, instead of evagination of the lateral ectoderm, a cluster of cells that will form the adult leg and wing invaginates below the ectoderm. These cells become the leg and wing imaginal discs or precursor cells, and do not differentiate any further until mature larval instars (Heming 2003).

1.2.10. Hatching

Once the organ formation is complete, the embryo begins to stretch and contract its newly formed muscles, while gas is released into the tracheae. Hatching can be achieved by different means, but it is usually a mechanical process, in which the juvenile grows by imbibing air until the chorion cracks, and often breaking it also with the jaws or with special ovirruptor structures. Sometimes there is an enzymatic digestion of the eggshell, following complex hydrostatic mechanisms. Finally, the hatchling emerges as a first nymphal or larval instar.

1.3. Epigenetic mechanisms during embryogenesis

Nowadays, the term “epigenetics” generally refers to changes on gene expression due to molecular mechanisms that directly affect, alter, or interact with chromatin. Epigenetic information can be transmitted through mitotic cell division within individuals, and through meiotic cell division as intergenerational inheritance (Glastad et al. 2019).

The most prominent epigenetic mechanisms are DNA methylation, modifications on histones, and regulation by noncoding RNAs (ncRNAs). There are few studies about how these epigenetic mechanisms could affect insect embryogenesis. Indeed, most of them focus on DNA methylation and on the action of microRNAs, the best-known type of ncRNAs.

1.3.1. DNA methylation

DNA methylation has been associated with several biological processes in animals, including embryo development, genomic imprinting, X-chromosome inactivation and silencing of retrotransposons (He et al. 2011; Jones 2012). Mechanistically, DNA methylation consists of a covalent addition of a methyl group to DNA. In most animal species studied, this addition of a methyl group only happens in cytosine bases, especially those that are confined to CpG dinucleotides (Glastad et al. 2019).

DNA methylation is catalyzed by enzymes named DNA-methyltransferases (DNMTs). Studies in mammals have classified these DNMTs into "*de novo"* methyltransferases and "maintenance" methyltransferases. *De novo* methyltransferases are those of the DNMT3 family, while maintenance methyltransferases belong to the DNMT1 family. In general, DNMT3 enzymes establish new DNA methylation patterns in the genome, while DNMT1 enzymes preferentially act on hemimethylated DNA, maintaining methylation during successive cell generations (He et al. 2011; Jones 2012).

The sequencing of different insect genomes revealed the occurrence of the two DNMTs families (DNMT1 and DNMT3) in different insect orders (Bewick et al. 2017). A number of functional studies showed that those DNMTs promote DNA methylation and are key during insect embryo development. One example is the coleopteran

Tribolium castaneum (Schulz et al. 2018), where depletion of *DNMT1* caused developmental arrest in embryos. Another example is the hymenopteran *Nasonia vitripennis* (Zwier et al. 2012), where *DNMT1* depletion resulted in embryo lethality at the onset of gastrulation. One more example is the hemipteran *Oncopeltus fasciatus* (Bewick et al. 2019), where *DNMT1* depletion reduced egg production, and the oviposited eggs became inviable. Although the importance of DNA methylation for embryo development appears to be general in insects, depletion of *DNMT1* affected different embryo stages in the different species. Thus, the action specific of DNA methylation seems to be typical of each lineage.

1.3.2. MicroRNAs

miRNAs are single stranded RNAs of 22-26 nucleotides that are loaded into the RISC complex and subsequently guided to a target mRNA. Then, the complex miRNA-RISC blocks the translation of the target mRNA. A recent study comparing miRNA expression during the development of *Blattella germanica* (hemimetabolan, short germ band), *T. casteneum* (holometabolan, intermediate germ band) and *Drosophila melanogaster* (holometabolan, long germ band), suggested that miRNAs may play important roles in the developmental transitions, specifically during embryo development, and they might influence the germ band type and metamorphosis mode (Ylla et al. 2017).

1.4. Embryogenesis in different metamorphosis modes

1.4.1. Ametabolan embryogenesis

Ametabolan embryogenesis results in a nymph that possesses the basic shape and features of the adult, and it follows the developmental steps described above: formation of the blastoderm and germ band (short germ band segmentation), formation of the serosa and amnion membranes, then formation of amniotic cavity, gastrulation, conspicuous blastokinesis, dorsal closure, and organogenesis (Larink 1983). Ametabolan embryos secrete and deposit three cuticles, named EC1, EC2 and EC3, as observed in the Zygentoma *Thermobia domestica* (Konopová and Zrzavý 2005).

1.4.2. Hemimetabolan embryogenesis

Hemimetabolan embryogenesis gives rise to a nymph that has the basic adult features. Hemimetabolan insects also follow the developmental steps generally observed in insect embryogenesis, from formation of blastoderm to organogenesis. They follow the short germ band segmentation, with some variations depending on the species. The most common short germ band type forms a blastoderm comprising the head lobes, the most apical trunk segments, and the terminus (Davis and Patel 2002). However, there is another type, named intermediate germ band, where they already have abdominal segments when they reach gastrulation stage. This is the case, for example, of the cricket *Gryllus bimaculatus* (Mito et al. 2005).

Blastokinesis is especially complex in hemimetabolan embryogenesis (Panfilio 2008), but there are also some species where it implies only a modest displacement of the embryo in the same egg side, like for example, in the cockroach *B. germanica* (Tanaka 1976).

During organogenesis, the different organs differentiate from the mesoderm, endoderm and ectoderm. The mesoderm differentiates into muscular and fat body tissues, the endoderm forms the midgut and associated tissues, and the ectoderm differentiates into nervous system and cuticle structures (including the appendages, wing primordia and eyes). Indeed, the limb buds that give rise to the different appendages (antennae, mouth pieces and legs) are formed at early embryogenesis, but it is not until organogenesis that they complete segmentation. The same happens with wing primordia, a cluster of cells that start growing underneath ectodermal germ layer at early embryogenesis (Heming 2003). Most species externalize the wing primordia in later nymphal instars. However, some species externalize them already in the first nymphal instar, thus, forming wing pads, like in the desert locust, *Locusta migratoria*. In other species, like the cockroach *B. germanica*, the wing primordia are encapsulated into a pteroteca, a cuticle pocket placed in the laterobasal part of the prothorax and metathorax (see Belles, 2020). Also, in early embryogenesis, eyes start being developed. The eyes of hemimetabolan insects are compound, formed with a number of ommatidia that contain a dioptric apparatus, pigment cells, and

rhabdomeric-type photoreceptive cells. During embryogenesis, ommatidial precursors differentiate within an optic placode on both sides of the embryonic head (Heming 2003). During nymphal stages, new ommatidia differentiate from the anterior growth zone, thus, the eyes will continue expanding through successive molts (Belles 2020).

Hemimetabolan embryos secrete and deposit three cuticles, named EC1, EC2 and EC3. This is observed in paleopteran, polyneopteran and paraneopteran species (Konopová and Zrzavý 2005).

1.4.3. Holometabolan embryogenesis

Holometabolan insects also follow the basic developmental transitions of a typical insect embryogenesis (Anderson 1972a), but with some differences depending on the species. Most holometabolans follow long germ band segmentation (Peel 2008), like for example, the dipteran *D. melanogaster*, the coleopteran *Callosobruchus maculatus*, the lepidopteran *Manduca sexta*, and the hymenopterans *N. vitripennis* and *Bracon hebetor* (Grbic and Strand 1998; Davis and Patel 2002; Peel 2008). However, there are also species that follow a short or intermediate germ band segmentation, like the coleopteran *Tribolium casteneum*, the lepidopteran *Bombyx mori* (Davis and Patel 2002; Peel 2008), and the hymenopterans *Aphidius ervi* and *Macrocentrus cingulum*. The available data suggests that long germ band segmentation evolved from the short germ band mode, but this transition may have

happened more than once during the endopterygote radiation (Davis and Patel 2002; Peel 2008), and might be reversible, as presumed in some hymenopterans (Grbic and Strand 1998; Sucena et al. 2014).

Blastokinesis in holometabolan embryogenesis is less complex than in hemimetabolans. An extreme case is *D. melanogaster*, in which there is practically no blastokinesis due to the extremely reduced extraembryonic component (Panfilio 2008).

Regarding organogenesis, appendages are formed from limb buds that appear at early embryogenesis, as in hemimetabolan embryogenesis. However, the appendages formed during holometabolan embryogenesis have less segments than in the adult (Heming 2003).

Holometabolan embryos also deposit three cuticles, EC1, EC2 and EC3, although EC2 may be simplified, with only a reduced epicuticle in embryos or even absent in the most evolutionarily modified flies (Konopová and Zrzavý 2005).

1.4.4. Transcriptomic differences between different modes of embryogenesis

In a previous study, Ylla et al. (2018) compared the transcriptomic data during the ontogeny of the cockroach *B. germanica* (hemimetabolan insect) and the fly *D. melanogaster* (holometabolan). In this study, enrichment analyses revealed that biological functions related with DNA methylation are enriched in

early *B. germanica* embryo, but not in *D. melanogaster* (Figure 1.9). In the same sense, Bewick et al. (2017) reported that hemimetabolan insects show higher levels of methylated CGs than holometabolans, relating DNA methylation in insects with DNA-methyltransferase 1 (DNMT1) rather than with DNA-methyltransferase 3 (DNMT3). This suggests that DNMT1 might be a good candidate as embryonic factor that would have contributed to the evolutionary transition from hemimetaboly to holometaboly.

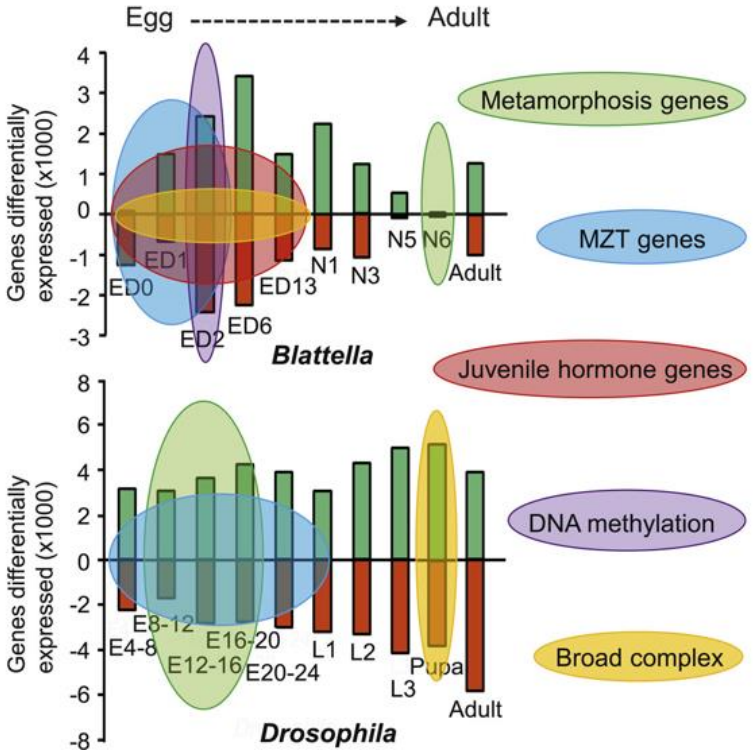


Figure 1.9. Comparative transcriptomic data of *Blattella germanica* and *Drosophila melanogaster*. Number of genes significantly ($p < 0.05$) upregulated (green) and downregulated (red) according to a differential expression analysis between consecutive stage-libraries of *B. germanica* (hemimetabolan) and *D. melanogaster* (holometabolan). Functions of gene groups determined according to GO Terms from Enrichment Analysis. From Ylla et al. (2018).

The Ylla et al. (2018) study also revealed that gene expression related with the maternal to zygotic transition (MZT) is concentrated in early stages of embryogenesis in *B. germanica*, while in *D. melanogaster* this expression extends during the whole embryogenesis. A gene expressed during the MZT that showed to have very different pattern in both species is *zelda*, a key activator of the zygotic genome during the MZT (Liang et al. 2008; Harrison et al. 2011). Zelda binds *cis*-regulatory elements (TAGteam heptamers) and makes the chromatin accessible for gene transcription (Foo et al. 2014; Li et al. 2014; Schulz et al. 2015; Hug et al. 2017; Dufourt et al. 2018). In *B. germanica*, *zelda* is expressed only during early embryogenesis, during the MZT, while in *D. melanogaster* the expression is maintained beyond the MZT (Ylla et al. 2018). Since *B. germanica* develops an adult-shaped nymph during embryogenesis, while *D. melanogaster* develops a vermiform larva, the authors speculated that *zelda* expression beyond MZT in *D. melanogaster* may be required to regulate the development of the very derived vermiform larval morphology. Thus, Zelda appeared as an interesting candidate of embryonic factor that would have a role in the evolution of holometaboly from hemimetaboly.

Another interesting gene that came out from Ylla et al. (2018) study as a differentially expressed factor between *B. germanica* and *D. melanogaster* is *E93*. In postembryonic development, E93 is known as an adult specifier, as it triggers metamorphosis (Mou et al. 2012; Ureña et al. 2014). The most important hormones regulating

metamorphosis are 20-hydroxyecdysone (20E), and Juvenile hormone (JH). 20E promotes every molt, in the embryo, as in the postembryonic development until the adult. JH inhibits metamorphosis, thus maintaining the juvenile growth. A decrease on JH in the premetamorphic stage (last juvenile instar) results in a decrease of Krüppel homolog 1 (Kr-h1). The decrease of Kr-h1 causes a disinhibition of E93, and that, triggers metamorphosis (Belles and Santos 2014). Thus, *E93* becomes highly expressed during the last juvenile instar (last nymphal instar in hemimetabolans or pupa in holometabolans). Ylla et al. (2018) analyses showed this high expression of *E93* in the last nymphal instar of *B. germanica* and in the pupa of *D. melanogaster*. Interestingly, however, the analyses also showed that *E93* is expressed at early stages of development in *B. germanica*, showing a peak of expression during the MZT. This is not the case in *D. melanogaster* embryos, where *E93* is not expressed during embryogenesis. This makes E93 a good candidate as a determining factor in the hemimetabolan embryogenesis, and to have a role in the evolution of holometaboly.

1.5. The German cockroach, *Blattella germanica*, as experimental model

The German cockroach *B. germanica* is one of the best-studied hemimetabolan species, especially at molecular level. This is in great measure due to the work of our laboratory since the 1980's. Nowadays, we have access to *B. germanica* genome, which is annotated (Harrison et al. 2018), and also to a series of

representative RNA sequencing datasets (Cristino et al. 2011; Belles and Ylla 2015; Ylla et al. 2017; Ylla et al. 2018).

One interesting characteristic of *B. germanica* is that it belongs to a basal insect lineage that exhibits ancestral features, like hemimetabolan metamorphosis and short germ band segmentation. Moreover, cockroaches are very sensitive to RNAi, and easily and inexpensively reared and maintained in the laboratory. All this made them a good model for insect evolutionary studies, particularly to study the transition from hemimetaboly to holometaboly. In our laboratory, it has been used as a model insect to study embryonic and postembryonic development, as well as to investigate reproduction.

B. germanica has three morphologically differentiate stages: embryo, nymph, and adult. Females oviposit around 30-40 eggs in a hard capsule or ootheca. The eggs are packed in two rows, and are about 3 mm long, 1 mm wide and 0.3 mm thick (Tanaka 1976). Females carry the oothecae attached to the genital atrium for 18 days (at 29ºC), until embryos hatch. Tanaka (1976) described in detail the embryo development of *B. germanica*, defining 18 different stages. Briefly, stages 0 to 2 correspond to the egg cleavage, when first energids appear and start migrating to the periphery and begins blastoderm formation. Stages 3 to 5 correspond to gastrulation and segmentation processes. From stage 5 on, the stages correspond to elongation and different organ formation, and dorsal closure. Finally, at stage 18 the first nymphal

instar is formed and able to emerge. Postembryonic development takes between 22 and 24 days and comprises six nymphal instars (at 29ºC) (Ewing 1970).

2. OBJECTIVES

2.1. Hypotheses proposed in this thesis

The general hypothesis of the present thesis is based on the fact that embryogenesis determines the type of metamorphosis, and that there will be differential factors in the embryogenesis of hemimetabolan and holometabolan species. Consequently, the study of these factors can provide valuable information to explain the evolution of holometaboly from hemimetaboly.

From this general notion, we have focused on three specific hypotheses that correspond to three relevant differences between the transcriptional activity in the embryo of the hemimetabolan *B. germanica* and the holometabolan *D. melanogaster*, which were revealed in the study of Ylla et al. (2018).

2.1.1. DNA-methylation has a specific role in hemimetabolan embryogenesis

DNA methylation has been detected in different insect groups. However, DNA methylation levels are much higher in hemimetabolans than in holometabolans (Falckenhayn et al. 2013; Bewick et al. 2017; Provataris et al. 2018). Ylla et al. (2018) found a negative correlation between gene expression and possible candidate genes to be methylated in early embryo stages of *B. germanica*, suggesting that DNA methylation operates during the maternal-to-zygotic transition (MZT). Moreover, it was found that

DNMT1, a gene coding for a DNA methyltransferase, was expressed in the same stages, whereas this gene is not present in the holometabolan *D. melanogaster*. Therefore, we presumed that DNMT1 promotes DNA methylation in the embryo of *B. germanica*, and that DNA methylation is essential in early embryo development in this species. If this is true, it could represent a typical feature of hemimetabolans, and a possibility to explain, at least in part, the evolutionary transition from hemimetabolan to holometabolan embryogenesis.

2.1.2. Zelda may have common but also differential functions in hemimetabolan and holometabolan embryogenesis

In *D. melanogaster*, Zelda activates the zygotic genome during the MZT. Zelda binds specific regulatory elements in the genome, called TAGTeam heptamers, and makes the chromatin accessible, thus promoting transcriptional activation of early-gene networks in the MZT (Foo et al. 2014; Li et al. 2014; Schulz et al. 2015; Hug et al. 2017; Dufourt et al. 2018). Outside *D. melanogaster*, Zelda has been studied in the holometabolan, long germ band insects, *Apis mellifera* and *Nasonia vitripennis* (Pires et al. 2016; Arsala and Lynch 2017). In these studies, the role of activator of the zygotic genome during the MZT is reported. However, it is also reported that Zelda is important after the MZT in these holometabolan species, in posterior segmentation and patterning of imaginal discs.

Studying transcriptomic patterns of *zelda* along the ontogeny of *D. melanogaster* and *B. germanica*, we found that *zelda* expression concentrates in early embryo in *B. germanica* while in *D. melanogaster zelda* expression practically covers the entire embryogenesis. We hypothesize that the concentrated pattern in *B. germanica* reflects functions restricted to early embryogenesis and is related with the mode of metamorphosis.

2.1.3. E93 has unique functions in hemimetabolan embryos

E93 is a metamorphosis master regulator, driving adult specification forward (Belles and Santos 2014; Ureña et al. 2014). In a great measure, E93 plays this role of adult specifier by acting as a chromatin modifier, enabling or preventing expression in given genomic regions (Uyehara et al. 2017). In the transcriptomic comparative analysis of Ylla et al. (2018), *E93* appears to be expressed not only in pre-adult stages, but also in the early embryo of *B. germanica*. In contrast, no expression of *E93* is observed in the embryo of *D. melanogaster*. This leads to wonder about the role of E93 in *B. germanica* embryo. Also, whether the expression of *E93* in the embryo is a feature exclusive of hemimetabolan species, and if so, then whether changes in embryonic *E93* expression may have contributed to the evolution of holometaboly.

2.2. Objectives

Based on the above three hypotheses, we have stablished the following three corresponding objectives:

2.2.1. To study whether DNMT1 methylates the DNA in *Blattella germanica* during embryogenesis, and to establish which are the biological consequences of such methylation.

2.2.2. To study the role of Zelda during *Blattella germanica* embryogenesis, and to assess whether it shows commonalities and differences with respect to the role in holometabolans, as described in the literature.

2.2.3. To study the role of E93 during *Blattella germanica* embryogenesis, and to find out whether significant expression of *E93* in the embryo is a typical feature only of hemimetabolan species.

3. RESULTS

3.1. DNMT1 promotes genome methylation and early embryo development in cockroaches

Alba Ventos-Alfonso, Guillem Ylla, Jose Carlos Montañés and Xavier Bellés

Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain

Alba Ventos-Alfonso, Guillem Ylla, Jose Carlos Montañés and Xavier Bellés

DNMT1 Promotes Genome Methylation and Early Embryo Development in Cockroaches

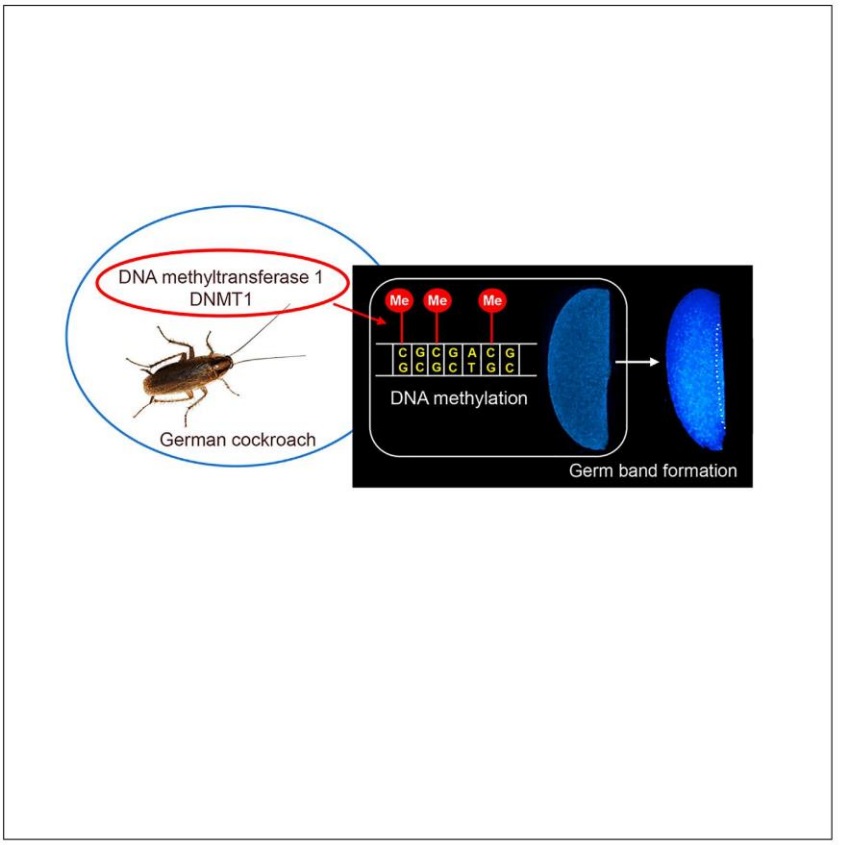
iScience23, 101778

iScience

CellPress
OPEN ACCESS

Article

DNMT1 Promotes Genome Methylation and Early Embryo Development in Cockroaches



Alba Ventós-Alfonso, Guillem Ylla, Jose-Carlos Montañes, Xavier Belles

xavier.belles@ibe.upf-csic.es

HIGHLIGHTS

Blattella germanica has *DNMT1* and *DNMT3* genes, which are expressed in early embryo

DNMT1 depletion reduces DNA methylation levels and impairs early embryo development

Methylated genes are highly expressed and are involved in metabolic processes

High DNA methylation is associated with low expression variability

Ventós-Alfonso et al., iScience *23*, 101778
December 18, 2020
https://doi.org/10.1016/j.isci.2020.101778

iScience

CellPress
OPEN ACCESS

Article

DNMT1 Promotes Genome Methylation and Early Embryo Development in Cockroaches

Alba Ventós-Alfonso,[1] Guillem Ylla,[1,2] Jose-Carlos Montañes,[1] and Xavier Belles[1,3,*]

SUMMARY

The influence of DNA methylation on gene behavior and its consequent phenotypic effects appear to be very important, but the details are not well understood. Insects offer a diversity of DNA methylation modes, making them an excellent lineage for comparative analyses. However, functional studies have tended to focus on quite specialized holometabolan species, such as wasps, bees, beetles, and flies. Here, we have studied DNA methylation in the hemimetabolan insect *Blattella germanica*. In this cockroach, a gene involved in DNA methylation, *DNA methyltransferase 1* (*DNMT1*), is expressed in early embryogenesis. In our experiments, RNAi of *DNMT1* reduces DNA methylation and impairs blastoderm formation. Using reduced representation bisulfite sequencing and transcriptome analyses, we observed that methylated genes are associated with metabolism and are highly expressed, whereas unmethylated genes are related to signaling and show low expression. Moreover, methylated genes show greater expression change and less expression variability than unmethylated genes.

INTRODUCTION

DNA methylation is the covalent addition of a methyl group to a DNA nucleotide. In most of the animals studied, this only occurs in cytosines, particularly at CpG dinucleotide sites (He et al., 2011; Hunt et al., 2013; Bewick et al., 2017). It is a widespread epigenetic mechanism that contributes to gene expression regulation in eukaryotes (He et al., 2011; Jones, 2012; Sarda et al., 2012; Anastasiadi et al., 2018). In mammals, it has been associated with a number of biological processes, including embryo development, genomic imprinting, X-chromosome inactivation, and silencing of retrotransposons (He et al., 2011; Jones, 2012).

DNA methylation patterns differ between vertebrates and invertebrates. While in vertebrates, DNA methylation is typically localized in the 5′ regulatory regions and appears associated with gene inactivation (Anastasiadi et al., 2018), in invertebrates it is mainly localized in intragenic regions and seems associated with gene activation (Bonasio et al., 2012; Falckenhayn et al., 2013; Hunt et al., 2013; Wang et al., 2013; Glastad et al., 2014, 2016; Bewick et al., 2019). Moreover, DNA methylation levels in arthropods, particularly insects, are generally lower than in vertebrates (Bewick et al., 2017). Indeed, given the possibilities of comparing different orders that have distinct DNA methylation patterns (Bewick et al., 2017; Lewis et al., 2020), insects have become the model of choice for studying the functional significance of this DNA modification (Hunt et al., 2013).

Most insects develop through metamorphosis, which can be classified into two modes: hemimetabolan, or direct development through the embryo, nymph, and adult stages; and holometabolan, or diverging development through the embryo, larva, pupa, and adult stages (Belles, 2020). In this respect, although DNA methylation has been detected in the different insect groups, higher levels have been observed in hemimetabolan than holometabolan models (Falckenhayn et al., 2013; Bewick et al., 2017; Provataris et al., 2018). This, along with a comparative analysis between hemimetabolan and holometabolan insects, led us (Ylla et al., 2018) to hypothesize that DNA methylation could be instrumental in the type of embryo development, and the mode of metamorphosis. Many roles have been associated with DNA methylation in insects, including phenotypic plasticity and caste determination (Bonasio et al., 2012; Glastad et al., 2016; Robinson et al., 2016; Cardoso-Júnior et al., 2017; Li et al., 2018), alternative splicing (Bonasio et al., 2012; Glastad et al., 2014, 2016), and reproduction (Schulz et al., 2018; Bewick et al., 2019). However, there are few

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

[2]Present address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

[3]Lead Contact

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

https://doi.org/10.1016/j.isci.2020.101778

functional studies on the role of DNA methylation during early embryo development (Schulz et al., 2018; Bewick et al., 2019), despite this being the period when de novo DNA methylation is expected to occur (He et al., 2011).

DNA methylation is catalyzed by DNA-methyltransferases (DNMTs). In mammals, DNMTs are classified into DNMT3, which establishes new methylation (methylation de novo), and *DNA methyltransferase 1* (DNMT1) which preferentially methylates hemimethylated DNA, maintaining methylation during successive cell generations (maintenance methylation) (He et al., 2011; Jones, 2012). Although a third DNMT was initially reported, further studies demonstrated that the so-called DNMT2 actually methylates tRNA, rather than DNA (Goll et al., 2006; Jurkowski et al., 2008; Lyko, 2018). Insects can possess either just DNMT1 (like the lepidopteran *Bombyx mori* and the coleopteran *Tribolium castaneum*), both DNMT1 and DNMT3 (like the hymenopteran *Apis mellifera*), or neither (like the dipteran, *Drosophila melanogaster*) due to secondary loss of both DNMT1 and DNMT3 (Bewick et al., 2017).

In a previous work, we reported the gene expression patterns in the German cockroach, *Blattella germanica*, on the basis of 11 transcriptomes representing key stages of embryonic and post-embryonic development (Ylla et al., 2018). One of the genes with the most characteristic profile was *DNMT1*, whose expression was concentrated in the first days of embryogenesis. At that time, we hypothesized that DNMT1 would catalyze DNA methylation and that it may play an important role in early embryo development. The present work was planned to study the *DNMT1* expression pattern in detail, the relationships between DNMT1 and DNA methylation, and the possible role of DNMT1 in embryogenesis. Results obtained with quantitative PCR have confirmed that *DNMT1*expression concentrates in the first days of embryo development, and RNA interference (RNAi) and reduced representation bisulfite sequencing (RRBS) studies have revealed that DNMT1 promotes genome methylation and early embryo development in *B. germanica*.

Beyond these findings, by analyzing the genome-wide methylation profiles in regions with a high CpG content, and looking at the relationships between methylation levels and gene expression, we discovered certain regularities that may be of more general interest. Comparing methylated and unmethylated genes, we found that the former are related to metabolism and are highly expressed throughout development, while the latter are more associated with signaling pathways and generally have low expression levels. Moreover, methylated genes present a relatively high degree of expression change after the DNMT1 peak, but with little expression variability, whereas unmethylated genes display the opposite properties.

RESULTS

Blattella germanica Has *DNMT1* and *DNMT3* Genes that Express in the Early Embryo

By combining a BLAST search in *B. germanica* transcriptomes (Ylla et al., 2018), mapping of the resulting sequences in the *B. germanica* genome (Harrison et al., 2018), and PCR strategies, we obtained a cDNA of 4,662 nucleotides comprising the complete ORF (GenBank: MT881788), whose conceptual translation gave a 1,554 amino acid sequence that was highly similar to insect DNMT1 proteins. We also obtained a cDNA of 1,803 nucleotides, comprising the complete ORF (GenBank: MT881790), whose conceptual translation gave a 601 amino acid sequence that was highly similar to insect DNMT3 proteins. A phylogenetic analysis using DNMT1 and DNMT3 sequences from representative species showed that the DNMT1 and DNMT3 identified in *B. germanica* clustered at the DNMT1 and DNMT3 nodes, respectively (Figure S1), strongly suggesting that these were DNMT1 and DNMT3 orthologs.

With regard to protein organization, *B. germanica* DNMT1 contains all the characteristic DNMT1 domains described by Lyko (2018) (Figure 1A): a DNMT1-associated protein 1 (DMAP1) binding domain, which allows interaction with the transcriptional repressor DNMAP1 and the histone diacetylase HDAC2 (HD2); a replication foci targeting sequence (RFTS), which allows DNMT1 to target replication foci; a CXXC domain, which allows DNMT1 to bind unmethylated DNA; two bromo-adjacent homology domains, whose function is still unknown; and a catalytic domain at the C-terminal. *B. germanica* DNMT3 also contains all the characteristic DNMT3 domains described by Lyko (2018) (Figure 1A): a PWWP domain, which allows binding to histone H3 molecules that are trimethylated at lysine 36; an ATRX-DNMT3-DNMT3L (ADD) domain, which mediates targeting to histone H3 molecules with unmethylated lysine 4; and a catalytic domain, the C5-Cytosine-specific DNA methylase domain. In previous analyses, we also found a *bona fide* DNMT2 ortholog (Ylla et al., 2018). However, as DNMT2 methylates tRNA (Goll et al., 2006; Jurkowski et al., 2008; Lyko, 2018), the *B. germanica* ortholog has not been considered in this work, which focuses on DNA methylation.

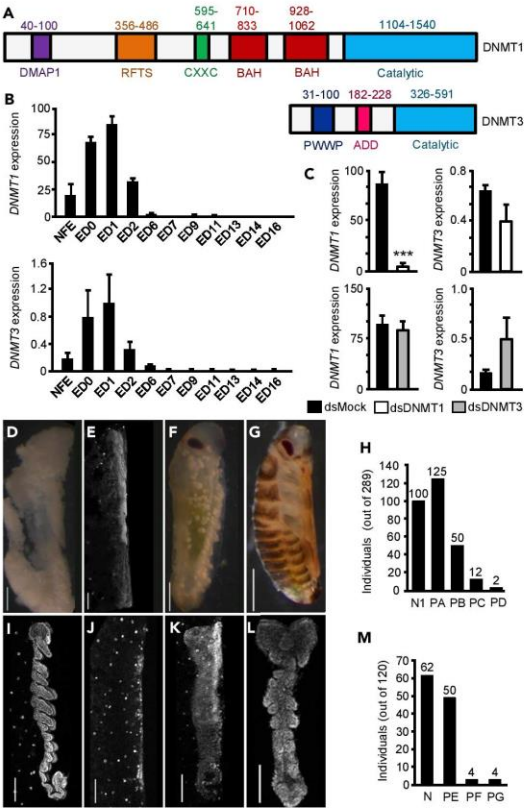


Figure 1. *Blattella germanica* DNMT1 and DNMT3, and Effects of Maternal RNAi

(A) Protein organization of DNMT1 and DNMT3; the domains follow the nomenclature established by Lyko (2018); numbers indicate the start and end amino acids of the different domains.

(B and C) (B) qRT-PCR mRNA levels of *DNMT1* and *DNMT3* during embryogenesis; NFE: non-fertilized egg; ED0 to ED16: embryo day 0 to embryo day 16; (C) Effects of DNMT1 (upper panels) and DNMT3 (lower panels) maternal RNAi on *DNMT1* and *DNMT3* transcript levels; dsDNMT1, dsDNMT3 or dsMock were injected into 5-day-old adult females, and measurements were taken on ED1.

(D–G) Phenotypes observed in unhatched oothecae from DNMT1-depleted embryos; D: Phenotype PA, embryos with development interrupted at the pre-blastoderm stage; E: Phenotype PB, embryos with malformed head and appendage-like structures; F: Phenotype PC, embryos around Tanaka stage 13, with no appendages and a narrower abdomen than normal; G: Phenotype PD, embryos at Tanaka stage 18, ready to hatch, but with darker coloration than normal.

(H) Number of individuals showing the phenotypes PA to PD; the total number of individuals studied was 289, and the number of individuals in each category is indicated at the top of each bar; the sample also includes the 100 nymphs hatched from 3 viable oothecae (N1).

(I–L) Phenotypes observed in ED4 from DNMT1-depleted embryos; I: Normal ED4 embryo (N); J: Phenotype PE, embryos with development interrupted at Tanaka stage 2; K: Phenotype PF, embryos with development interrupted at Tanaka stage 3; L: Phenotype PG, embryo with a general morphology similar to Tanaka stage 4, with the cephalic and thoracic segments delimited but incompletely developed, and the abdominal region amorphous and unsegmented.

(M) Number of embryos showing the phenotypes PN and PE to PG; the total number of embryos studied was 120, and the number of embryos in each category is indicated at the top of each bar. In D-G, and I-L, the upper part of each picture corresponds to the cephalic part of the embryos; the scale bars are equivalent to 500 μm in panels D-G, and 100 μm in panels I-L. In Figures B and C, each qRT-PCR value represents three biological replicates and is expressed as copies of mRNA per 1000 copies of *Actin-5c* mRNA (mean ± SEM); the triple asterisk indicates statistically significant differences with respect to controls ($p < 0.001$), calculated on the basis of a Pairwise Fixed Reallocation Randomization Test implemented in REST (Pfaffl, 2002).

By using real-time quantitative reverse transcription PCR (qRT-PCR), we studied the expression of *DNMT1* and *DNMT3* during *B. germanica* embryogenesis (embryo days 0, 1, 2, 4, 6, 7, 9, 11, 13, 14, and 16). The results show that both *DNMT1* and *DNMT3* are expressed between days 0 and 2 of embryogenesis (0–12% of embryo development), both showing an expression peak at day 1 (Figure 1B). However, the expression levels of *DNMT3* are very low (maximum expression of 0.98 ± 0.46 copies per 1000 *Actin-5c* copies at day 1), not only when compared with that of *DNMT1* (maximum expression of 86.06 ± 8.83 copies per 1000 *Actin-5c* copies at day 1) but also taking into account the very low quantity of absolute mRNA at this early embryo stage, including the absolute levels of *Actin-5c* mRNA.

Maternal RNAi of *DNMT1* and *DNMT3*

To study the possible functions of DNMT1 and DNMT3 in the early embryo, we used maternal RNAi. Five-day-old adult females (AdD5) of *B. germanica* were injected with 3 μg of a dsRNA targeting DNMT1 (dsDNMT1) or DNMT3 (dsDNMT3). These females were then allowed to mate (fertilization was checked at the end of the experiment by examining the presence of spermatozoids in the spermatheca) and to produce the first ootheca. Control females were treated equivalently but with a non-specific dsRNA (dsMock). To estimate the efficiency of the RNAi, we measured the levels of the respective transcripts in 1-day-old embryos, the day of peak expression. In the dsDNMT1-treated females, the mRNA levels of *DNMT1* were 87.5% lower than in the controls (Figure 1C), indicating that the maternal RNAi was remarkably efficient. In addition, the *DNMT3* mRNA levels were similar in both groups, indicating that dsDNMT1 is specific and does not affect *DNMT3* transcripts. In contrast, dsDNMT3 treatment did not significantly affect the mRNA levels of *DNMT3*, despite the high dose of dsRNA and the replication using three experimental batches containing 10 females each. Figure 1C illustrates representative results demonstrating that our dsDNMT3 treatments did not reduce *DNMT3* mRNA levels. As we were unable to deplete *DNMT3* transcript levels, we continued the functional studies with DNMT1.

Depletion of DNMT1 Impairs Embryo Development

A total of 10 control (dsMock-treated) females formed the first ootheca on day 8 of the adult stage, which hatched 19 days later, giving a total of 373 first instar nymphs (35–40 nymphs per ootheca, on average). The dsDNMT1-treated females (n = 10) also produced the first ootheca on day 8 but only 3 of 10 oothecae (30%) hatched 19 days later, giving a total of 100 first instar nymphs (30–35 nymphs per ootheca). No nymphs hatched from the remaining 7 oothecae (70%) produced by the dsDNMT1-treated females. The examination of the embryos in the 7 unviable oothecae, 20 days after the formation of the ootheca (n = 189 embryos), showed various phenotypes, which were classified into the following four categories. Phenotype PA (Figure 1D): embryos with development interrupted in a pre-blastoderm stage, thus, only white yolk was observed. Phenotype PB (Figure 1E): embryos that were completely transparent under the stereomicroscope, but for which 4′,6-diamidino-2-phenylindole (DAPI) staining revealed malformations of the head and appendages. Phenotype PC (Figure 1F): embryos at Tanaka stage 13 (Tanaka, 1976) (58% embryo development), but with no appendages, and narrower abdomens than the controls. Phenotype PD (Figure 1G): embryos at Tanaka stage 18, thus, just prior to hatching, but presenting a darker coloration than the controls. A total of 125 embryos of the 189 studied showed phenotype PA (66% of the abnormal embryos and 43% of all DNMT1-depleted embryos). Phenotype PB was represented by 50 embryos (26% of the abnormal embryos and 17% of all DNMT1-depleted embryos). Phenotypes PC and PD were the least frequent; 12 embryos presented phenotype PC (6% of the abnormal embryos and 4% of all DNMT1-depleted embryos), and only 2 embryos presented Phenotype PD (1% of the abnormal embryos and 0.7% of all DNMT1-depleted embryos) (Figure 1H).

Since most of the embryos from dsDNMT1-treated females died early in their development, coinciding with the temporal expression of *DNMT1*, we repeated the maternal RNAi experiment, but this time we examined the embryos 4 days after oviposition (ED4). We studied 120 embryos from 5 oothecae produced by control (dsMock-treated) females and 120 embryos from 5 oothecae produced by dsDNMT1-treated females. All the embryos from control females (100%) presented the normal aspect of an ED4 embryo, in other words, 20–25% embryo development and Tanaka stage 5–6 (Tanaka, 1976) (Figure 1I). A total of 62 of 120 embryos (51.7%) examined in oothecae from dsDNMT1-treated females, were normal embryos, similar to the controls. The remaining 58 embryos (48.3%) showed several different phenotypes that were classified into three categories, as follows. Phenotype PE (Figure 1J): embryos with development interrupted at Tanaka stage 2, when the germ band is delimited and slightly expanded on both sides (12% embryogenesis). Phenotype PF (Figure 1K): embryos with development interrupted at Tanaka stage 3 (16%

Gene Region	Control	DNMT1-Depleted	Decrease (%)
All regions	17.8	8.8	50.5
Intergenic regions	10.0	5.2	47.5
Promoter region	26.4	10.9	58.7
Gene	41.8	19.1	54.2
5′ UTR	27.4	11.0	60.0
All exons	51.3	21.5	58.0
First exon	20.5	7.4	63.8
Other exons	62.7	27.2	56.7
Last exon	59.2	28.9	51.1
Exon of monoexonic	30.9	8.1	73.7
All introns	40.6	18.9	53.5
First intron	30.3	13.1	56.9
Other introns	45.8	21.4	53.2
Last intron	41.0	19.5	52.5
Intron of monointronic	23.3	11.3	51.3
3′UTR	75.5	36.2	52.1

Table 1. Effect of DNMT1 Depletion in *Blattella germanica* Embryos on CG Methylation Levels in Different Gene Regions

Measurements were taken on 4-day-old embryos, in controls and in DNMT1-depleted Insects. Results are expressed as a percentage of methylated CG.

development), when the germ band has started to expand on both sides. Phenotype PG (Figure 1L): embryos with a general morphology similar to Tanaka stage 4, at the start of abdominal segmentation and tail folding (17% embryogenesis) but presenting various malformations: cephalic and thoracic segments delimited but incompletely developed; and amorphous and unsegmented abdominal regions. Phenotype PE was represented by 50 embryos (86% of the abnormal embryos and 42% of all the embryos), while phenotypes PF and PG had 4 embryos each (7% of the abnormal embryos and 3% of all embryos, in both cases) (Figure 1M).

Depletion of DNMT1 Reduces DNA Methylation

To assess whether DNMT1 is required for DNA methylation in *B. germanica*, we studied DNA methylation levels in DNMT1-depleted and control embryos, following the RRBS method. For this purpose, we performed RRBS in two different conditions: 4-day-old control embryos (ED4C) and 4-day-old DNMT1-depleted embryos (ED4T), using four biological replicates per condition. We analyzed the levels of methylated cytosines within CG dinucleotides in these two conditions in different genomic features. Firstly, we considered all the regions available from the RRBS, then the sequences corresponding to intergenic regions, the genes (the region that is transcribed, including the UTRs), the promoter region (considering an arbitrary length of 2 Kb upstream of the transcription start site), the 5′ UTR, and the 3′ UTR. Moreover, we examined the levels of methylated cytosines in exonic and intronic regions. We considered all exons as a whole (including the 3′ UTR), the first exon, the last exon, all exons except the first and the last, and the exon of monoexonic genes. We performed an equivalent analysis for introns.

Considering the whole gene and different gene features, we observed that the 3′ UTR regions have the higher average levels of methylation (Table 1). Moreover, CG methylation levels are higher in genic regions than in intergenic regions; and within genes these are higher in exonic regions, particularly 3′ UTR regions. In intronic regions, there is a tendency to show higher levels of CG methylation toward the 3′ region (Table 1). Characteristically, the methylation density (Figure 2A) of the different genetic features shows two clear

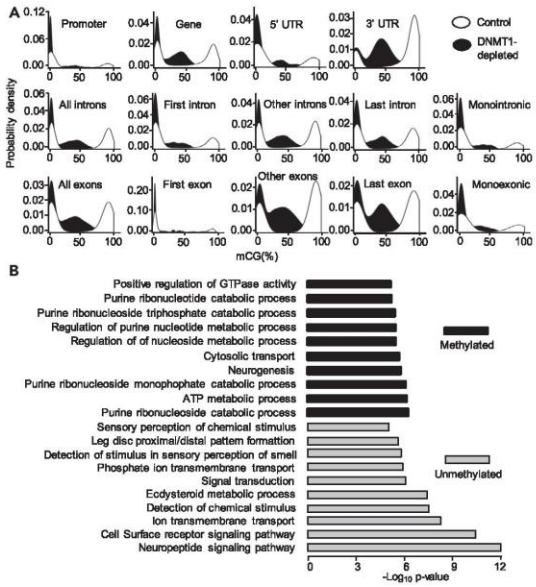


Figure 2. DNA Methylation in *Blattella germanica* and Effects of DNMT1 Depletion

(A) Kernel density plot of CG methylation in control and DNMT1-depleted 4-day-old embryos. The genomic features examined are the same as described in Table 1. In controls, the levels of CG methylation are generally either very high (80–100% methylation), or very low (0–4%), thus presenting a bimodal distribution; in DNMT1-depleted insects the bimodal distribution is modified as the peak of high values is reduced.

(B) Selection of GO terms of biological functions resulting from enrichment analyses carried out on methylated genes in 4-day-old control embryos. The 10 enriched biological functions with the lowest p values are shown for methylated and unmethylated genes. The p values were calculated according to Fisher's exact test.

peaks, indicating that they are either very methylated (80–100% methylation), or have very low levels of methylation (0–4%), practically without any intermediate values. RRBS sequencing also revealed that the levels of CG methylation are lower in DNMT1-depleted embryos than in controls, irrespective of the genomic feature examined. In most cases, the reduction was between 50% and 60%. Greater reductions were observed in the first exon (63.8% reduction) and the single exon of monoexonic genes (73.7% reduction) (Table 1). Consequently, DNMT1 depletion modified the bimodal distribution of CG methylation since a significant proportion of intermediate values appeared, and the peak of high values (80–100% methylation) was reduced (Figure 2A).

Methylated Genes Are Associated with Metabolism and Are Highly Expressed, Whereas Unmethylated Genes Are Associated with Signaling and Show Low Expression Levels

To obtain information on the functions of the highly methylated genes (80–100% methylation, hereinafter referred to as "methylated" for simplicity) and practically unmethylated genes (0–4% methylation, hereinafter referred to as "unmethylated" for simplicity), we carried out a gene ontology (GO) enrichment analysis. The results (Figure 2B) indicate that methylated genes are enriched in biological functions related to metabolic processes, neurogenesis, and cytosolic transport. On the other hand, the potential biological functions of unmethylated genes appear to be related to signaling pathways, including neuropeptide signaling, cell surface receptor signaling, ion transport, signal transduction, detection of chemical stimulus, ecdysteroid metabolic processes, and leg patterning (Figure 2B).

Next, we examined the expression levels of the genes that had been designated as methylated or unmethylated in the 4-day-old embryonic stage in later stages of *B. germanica*. These stages were those associated to transcriptomes previously described (Ylla et al., 2018). The results show that the expression levels

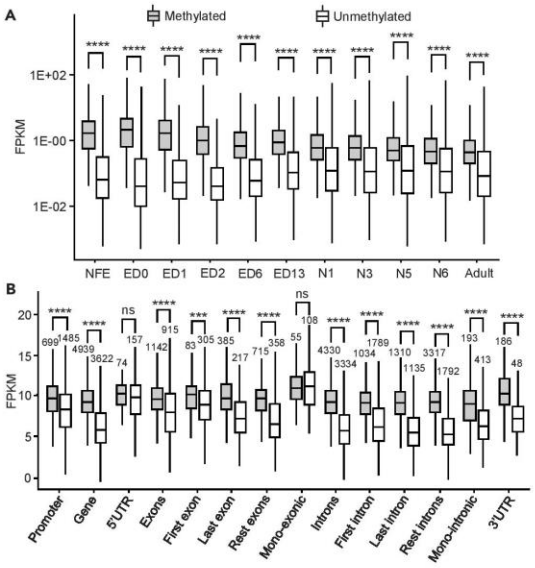


Figure 3. DNA Methylation and the Amount of Gene Expression in *Blattella germanica*

(A) Expression levels of methylated and unmethylated genes in the 11 embryo stages studied (NFE: non-fertilized egg, and ED0 to ED16: embryo day 0 to embryo day 16), four nymphal instars (N1, N3, N5, and N6), and the adult.

(B) Expression levels of methylated and unmethylated genes in 6-day-old embryos (ED6), considering gene expression levels in 6-day-old embryos (ED6), grouped by methylations status (unmethylated vs methylated) of various gene features. In all cases, expression is expressed as FPKM; the asterisks indicate statistically significant differences using the Mann-Whitney U test, adjusting p values by False Discovery Rate using the Benjamini-Hochberg method (* FDR <0.05; ** FDR <0.01; *** FDR <0.001; **** FDR <0.0001), non-significant differences (ns; FDR >0.05), are also indicated.

of the methylated genes are higher than those of the unmethylated genes in all the ontogenetic stages studied, the difference being more evident in early embryo stages, from ED0 to ED6 (Figure 3A).

We then analyzed the differences in expression between methylated and unmethylated genes, considering the gene region where the methylation is located. For this analysis, we used the transcriptomic data corresponding to ED6 since this is the stage following the pulse of *DNMT1* expression (Figure 1B). ED6 corresponds to Tanaka stage 8 (Tanaka, 1976), which precedes major developmental processes, like dorsal closure and organogenesis. The results show that the expression levels of methylated genes are significantly higher than those of unmethylated genes when methylation occurs in all the studied regions, except in the 5′ UTR or in the exon of monoexonic genes (Figure 3B). It is worth noting, however, that the number of annotated 5′ UTRs and monoexonic genes are relatively low, which could explain the lack of significant differences between methylated and unmethylated gene expression.

Methylated Genes Show Greater Expression Change Than Unmethylated Genes

Once again using the set of transcriptomes of Ylla et al. (2018), we examined the gene expression change between ED2 (when the peak expression of DNMT1 is already declining) and ED6 (4 days later) (Figure 1B). We determined the fold change ($\log_2$FC) of differentially expressed genes between these two stages, considering those having a $|\log_2 FC| \geq 2$ and false discovery rate of <0.05. In this way, we identified 1,599 genes, 553 of which were methylated and 1,046 of which were unmethylated. As shown in Figure 4A, both methylated and unmethylated genes increased or decreased their expression levels in similar proportions. Intriguingly, the change was less in methylated genes, regardless of whether this change was incremental or decremental, and this was more significant when the methylation was in the introns (Figure 4B). This notion can be condensed by expressing the coefficient of variation (CV) of gene expression between ED2 and ED6, which is significantly lower in methylated than in unmethylated genes (Figure 4C). The data

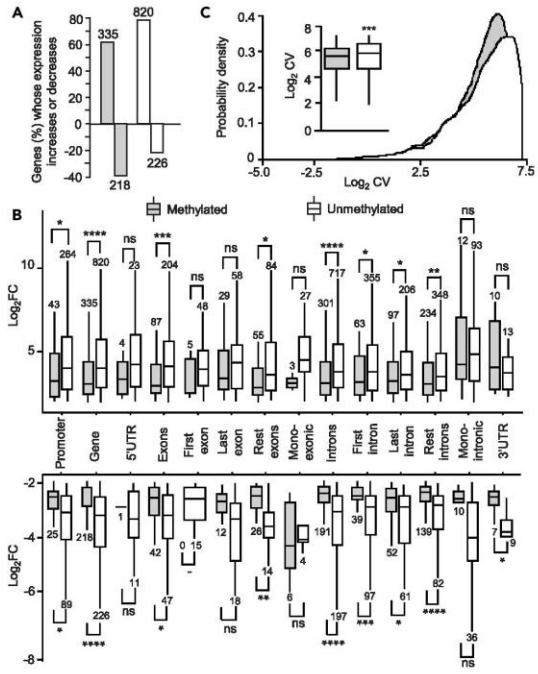


Figure 4. DNA Methylation and Gene Expression Dynamics in *Blattella germanica*

(A) Expression increase or decrease between ED2 and ED6 in methylated and unmethylated genes; a minimum of Log2FC > 2 with FDR <0.05, was considered an increase or decrease.

(B and C) (B) Expression change (log_2FC) between ED2 and ED6 of differentially upregulated or downregulated genes ($Log_2FC > 2$ and FDR <0.05); in all cases, genetic features were classified as methylated or unmethylated; data outliers have been omitted for clarity; the gene features considered were those in Table 1 (C) Density plot and boxplot (inset) of the coefficient of variation (CV) of the gene expression between ED2 and ED6, considering methylated and unmethylated genes; the inset describes the mean CV between ED2 and ED6 in methylated and unmethylated genes; data outliers have been omitted for clarity. In all cases, the gray bars indicate methylated genes, and the white bars are unmethylated genes; in B and C, asterisks indicate statistically significant differences using the Mann-Whitney U test, adjusting p values by False Discovery Rate using the Benjamini-Hochberg method (* FDR <0.05; ** FDR <0.01; *** FDR <0.001; **** FDR <0.0001), non-significant differences (ns; FDR >0.05), are also indicated.

suggest that the expression change of methylated genes has lower dispersion than that in unmethylated genes. This led us to analyze the CV of the expression levels of each gene between biological replicates at the same stage.

Methylated Genes Have Less Expression Variance Than Unmethylated Genes

Using two biological replicates, each comprising a pool of specimens from the transcriptome set of Ylla et al. (2018), we first compared the gene expression levels between replicates at the different stages. The results showed that there are no differences between replicates (Figure S2). We then calculated the gene expression CV between the replicates in the same stage, comparing methylated with unmethylated genes at each stage. The results show that methylated genes present less expression variability, measured as covariance, between biological replicates than unmethylated genes do, a property that is more evident in earlier embryo stages (from ED0 to ED6) (Figure 5).

DISCUSSION

The cockroach *B. germanica* has two *DNMT* genes, one coding for DNMT1 and one coding for DNMT3, which possess the functional motifs characteristic of these kinds of proteins, according to Lyko (2018).

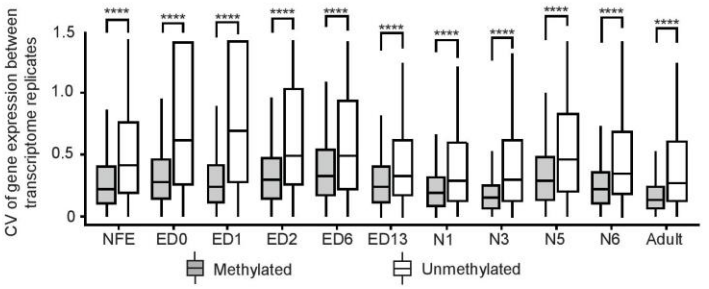


Figure 5. DNA Methylation and Expression Variability in *Blattella germanica*

The coefficient of variation (CV) of gene expression for methylated and unmethylated genes between the two biological replicates generated from a pool of specimens for each of the developmental transcriptomes studied. Data outliers have been omitted for clarity; the developmental stages studied were: NFE: non-fertilized egg; ED0 to ED13: embryo day 0 to embryo day 13; N1-N6: first to sixth nymphal instar; and the adult (Ylla et al., 2018). The four asterisks indicate statistically significant differences using the Mann-Whitney U test, adjusting p values by False Discovery Rate using the Benjamini-Hochberg method (FDR <0.0001).

Quantitative determinations showed that *DNMT1* and *DNMT3* are expressed during the early embryo development (between 0% and 12% embryogenesis) of *B. germanica*. This suggests that both genes play roles in early embryogenesis, although *DNMT3* expression levels are about 100 times lower than those of *DNMT1*. To study these roles, we used maternal RNAi, which efficiently knocked down the *DNMT1*, but not the *DNMT3*, whose mRNA levels were not reduced despite the relatively high doses of dsRNA used, and the three independent experimental batches employed. Although *B. germanica* is highly sensitive to RNAi (Belles, 2010), there are situations where this technique has been ineffective. For example, in the case of the lipophorin receptor, RNAi has proven highly efficient in the fatty body, where the gene is highly expressed, and much less efficient in the ovary, where it is expressed at low levels (Ciudad et al., 2007). In other cases, such as that of the yellow-g gene, the transience of its expression makes its depletion by RNAi impossible (Irles et al., 2009). We believe that the very low expression levels of *DNMT3* (about 1 copy of mRNA per 1000 copies of *Actin-5c* mRNA at most) are very difficult to significantly lower any further using RNAi.

In the case of *DNMT1*, although the penetrance of the effects was not 100%, as is usual in maternal RNAi (Belles, 2010), the transcript decrease obtained and the phenotypes observed were clear. Indeed, the RNAi experiments and subsequent RRBS analyses showed that DNMT1 promotes DNA methylation in *B. germanica*, as observed in other insects, such as the milkweed bug *Oncopeltus fasciatus* (Bewick et al., 2019) and the beetle *T. castaneum* (Schulz et al., 2018), when implementing an equivalent approach. The RNAi experiments also revealed that DNMT1, and thus DNA methylation, promotes the formation of the germband in early embryogenesis, at 12% development. In the hymenopteran *Nasonia vitripennis* DNMT1-depleted embryos die at the onset of gastrulation (Zwier et al., 2012), at around 40% embryo development. In the beetle *T. castaneum*, although DNA methylation does not preferentially occur at CpG sites (Zemach et al., 2010; Feliciello et al., 2013; Song et al., 2017), DNMT1 is required in very early embryo development to progress beyond the first few cleavage cycles, in other words around 4% embryogenesis (Schulz et al., 2018). In the bug *O. fasciatus*, eggs laid by DNMT1-depleted females are inviable, although the stage at which development is interrupted has not been determined (Bewick et al., 2019). The phenotype of embryos that complete development but are incapable of hatching is reminiscent of what is observed in a percentage of embryos with depleted JH signaling (Fernandez-Nicolas and Belles, 2017). In the present case, this phenotype might correspond to embryos with alterations in the expression of hatching regulatory genes, alterations that could derive from a deficient state of methylation. The fact that DNMT1 is required for embryo development in vertebrates, including mice (Li et al., 1992; Jackson-Grusby et al., 2001), frogs (Stancheva et al., 2001), and zebrafish (Rai et al., 2006), may suggest that their functions in embryogenesis are conserved from insects to vertebrates (Zwier et al., 2012). However, DNMT1 depletion in the insects *B. germanica*, *N. vitripennis*, and *T. castaneum* affects different embryo stages, and DNMT1 depletion in mice, frogs, and zebrafish also elicits different phenotypes, resulting in the misexpression of genes that specify embryonic cell identity but with limited effects on early

developmental mitosis (He et al., 2011). Thus, although DNA methylation is instrumental for embryogenesis in cockroaches up to mammals, the current evidence indicates that its specific action varies in different lineages, even within insects.

The matching expression patterns of *DNMT1* and *DNMT3* in *B. germanica* suggest that both act on the same early embryo stage, while the presence of a methyltransferase catalytic domain in DNMT1 and DNMT3 suggests that they both have the capacity to promote DNA methylation. However, when DNMT1 is depleted, a clear phenotype is observed in the embryo. This indicates that *DNMT3* expression, which is not affected by DNMT1 RNAi, does not compensate for the DNMT1 deficiency. These lines of evidence point to the possibility that both proteins are functionally redundant, and if so, *DNMT3*, which has very low expression levels, could be dispensable.

Most mammals have two DNMT3, DNMT3A, and DNMT3B, which establish DNA methylation patterns. Even many rodent species have a third enzyme, DNMT3C that selectively methylate the promoters of young retrotransposon insertions in their germline (Molaro et al., 2020). In contrast, DNMT3 has been evolutionarily lost in a number of insect orders, including Odonata, Ephemeroptera, Orthoptera, Thysanoptera, Phthiraptera, Lepidoptera, Trichoptera, and Diptera (Bewick et al., 2017; Lewis et al., 2020). Moreover, Bewick et al. (2017) showed that the presence of DNMT1 correlates positively with DNA methylation, whereas that is not seen for DNMT3. These authors suggest that either DNMT3 is unnecessary for DNA methylation or that DNMT1 compensates for DNMT3. Finally, studies *in vitro* have demonstrated that DNMT1 can also act as a de novo methyltransferase (Fatemi et al., 2002). Taken together, the data suggest that *B. germanica* DNMT1 plays both the de novo and maintenance roles in DNA methylation, while DNMT3 has a minor role, and is possibly redundant with respect to DNMT1. It is worth noting that the DNMT3 sequence of *B. germanica*, especially the catalytic domain (Figure S3), is remarkably conserved with respect to other proven functional DNMT3, such as that of the honeybee *A. mellifera* (Wang et al., 2006), suggesting that *B. germanica* DNMT3 is functional, and thus natural selection maintains the conserved sequence.

As in other species, the CG methylation levels in *B. germanica* present a bimodal distribution, being either very high or very low. In insects, this has also been reported in the locust *Schistocerca gregaria* (Falckenhayn et al., 2013) and the wasp *N. vitripennis* (Wang et al., 2013). Moreover, CG methylation in *B. germanica* tends to concentrate toward the 3′ region of the gene, in line with general DNA methylation trends in insects (Bewick et al., 2017; Lewis et al., 2020). In holometabolan species, DNA methylation appears to be biased toward the exons close to the 5′ region of the gene (Bonasio et al., 2012; Hunt et al., 2013; Wang et al., 2013; Glastad et al., 2016), while in hemimetabolans it presents higher levels toward the 3′ region of the gene coding part (Glastad et al., 2016; Bewick et al., 2019). In this sense, the DNA methylation pattern of hemimetabolans is similar to that of vertebrates, where the first intron and first exon are less methylated than the remaining regions in different tissues, species, and developmental stages (Anastasiadi et al., 2018). Furthermore, DNA methylation is biased toward exons rather than introns in some hemimetabolan insects, like the locust *S. gregaria* (Falckenhayn et al., 2013), the termite *Zootermosis nevadensis* (Glastad et al., 2016), and the bug *O. fasciatus* (Bewick et al., 2019), as is also the case in *B. germanica*. In general, our findings in embryos are similar to those observed by Bewick et al. (2019) in *B. germanica* adults, using whole-genome bisulfite sequencing data. These authors found similar general levels of CG methylation, with the highest being observed in intragenic regions rather than intergenic regions, and which tended to concentrate toward the 3′ UTR (Bewick et al., 2019; Lewis et al., 2020).

With respect to DNA methylation and gene functions, GO enrichment analyses revealed that methylated genes are mainly involved in metabolic processes, and are more highly expressed than unmethylated genes, which are instead related to signaling pathways. In other insects, like the ant *Camponotus floridanus* (Bonasio et al., 2012) and the wasp *N. vitripennis* (Wang et al., 2013), both holometabolan insects, GO terms analyses performed on methylated genes revealed that they were enriched for housekeeping functions. Furthermore, it has recently been found that putatively methylated genes are under stronger purifying selection in both hemimetabolan and holometabolan insects (Ylla et al., 2020), highlighting the evolutionary importance of those genes undergoing DNA methylation.

A controversial aspect of DNA methylation is whether it can stimulate or repress gene expression. In vertebrates, a negative correlation between DNA methylation and gene expression has been reported,

especially when methylation is located in the promoters, first intron, and first exon (Anastasiadi et al., 2018). In insects, a number of studies report a positive correlation between DNA methylation in intragenic regions and gene expression, such as in the termite *Z. nevadensis* (hemimetabolan) (Glastad et al., 2016), the ants *C. floridanus* and *Harpegnathos saltator* (Bonasio et al., 2012), and the wasp *N. vitripennis* (Wang et al., 2013) (holometabolans). However, in other insects like the desert locust *S. gregaria* (Falckenhayn et al., 2013), and the bug *O. fasciatus* (Bewick et al., 2019) (both hemimetabolans), no relationships have been found between DNA methylation and gene expression, although in the migratory locust, *Locusta migratoria*, alternative solitary or gregarious phases are associated with the methylation status of genes that are differentially expressed in these two phases, and in the expression of genes involved in DNA methylation (Robinson et al., 2016). Our observations indicate that methylated genes are significantly more expressed than unmethylated genes, especially in early embryogenesis (from ED0 to ED6), regardless of the methylation location in the gene.

The transcriptomic analysis of the methylated and unmethylated genes between ED2 and ED6 (i.e., after the *DNMT1* expression pulse) showed that the percentage of methylated genes with increased expression was higher than the percentage of those with decreased expression (61% vs. 39%). Nevertheless, the unmethylated genes behaved similarly (78% vs. 22%). Comparing the magnitude of change, in other words, how much the gene expression increased or decreased between ED2 and ED6, and the amount of expression variability (in terms of CV), revealed more that high DNA methylation levels are associated with high expression levels. At the same time, the amount of change between ED2 and ED6 is lower in methylated genes, regardless of whether the change involved increasing or decreasing expression. This fits with the results of our GO enrichment analyses, as signaling factors are rarely expressed at high levels, but suffer higher expression variations, whereas high expression levels and low expression variation of housekeeping genes, and increased metabolism, is typical in early embryo development (Miyazawa and Aulehla, 2018). Our results are reminiscent of those obtained for *N. vitripennis*, where methylated genes were shown to have higher median expression levels and lower expression variation across developmental stages than unmethylated genes (Wang et al., 2013).

Finally, comparing the gene expression between biological replicates at the different stages of *B. germanica* development revealed that methylated genes show lower expression variability than unmethylated genes. This indicates that the expression of the methylated genes is tightly regulated, a feature that fits with the essential roles identified for these genes. The amount of intrinsic expression variability between individuals has been considered an inherent property of genes (Alemu et al., 2014; De Jong et al., 2019), representing a layer of gene regulation information that is just as important as changes in the mean expression levels (Wang and Zhang, 2011). Expression variability has been shown to be low in genes involved in growth, general metabolism, and universal functions, whereas it is high in genes involved in environmental responses and non-housekeeping functions, in general, thus affecting gene network functioning by lowering noise (Alemu et al., 2014; De Jong et al., 2019).

Several features pertaining to the genomic, epigenomic, regulatory, polymorphic, functional, structural, and network characteristics of the gene, have been correlated with expression variability (Alemu et al., 2014). In the epigenomic context, a long-standing hypothesis posits that DNA methylation in gene regions reduces transcriptional noise, although the mechanisms involved are unclear (Bird, 1995; Suzuki et al., 2007). However, the information supporting this hypothesis is scarce and it focuses on human tissues. Using nucleotide-resolution data on genomic DNA methylation and microarray data for human brain and blood tissues, Huh et al. (2013) showed that gene body methylation appears to lower expression variability. Further studies of human brain tissues, using Illumina sequencing, have indicated that genes with low and high expression variability are likely to have low and medium gene methylation, respectively, whereas non-variable genes are likely to be highly methylated (Bashkeel et al., 2019). Also in plants, recent studies of *Arabidopsis thaliana* indicate that genes with high expression variability are depleted in DNA methylation (Cortijo et al., 2019). Our data on *B. germanica*, based on RRBS sequencing and transcriptomic data during embryo development, when DNA methylases are expressed and DNA methylation occurs, afford the first association between high DNA methylation and low expression variability in an insect.

Limitations of the Study

Despite our attempts, we have not been able to reduce the levels of *DNMT3* transcripts. This leaves open the question of the possible functions of DNMT3 in relation to DNMT1. Furthermore, the association

between high DNA methylation and low expression variability seems very relevant, but our evidence is based on the two replicates of gene expression values from our 11 stage-specific transcriptomes. This association would deserve further extensive research, not only collecting more data from a single model but also covering other animal and plant models, which may lead to the conclusion that reducing transcriptional noise is a universal property of DNA methylation.

Resource Availability

Lead Contact

Any questions or requests should be addressed to the Lead Contact (xavier.belles@ibe.upf-csic.es).

Materials Availability

We used the transcriptomes data as mentioned below. This study did not generate new reagents.

Data and Code Availability

The accession numbers for the two sequences, *DNMT1* and *DNMT3*, reported in this paper are GenBank: MT881788, and GenBank: MT881790, respectively. The transcriptomic analyses were carried out on the RNA-seq libraries produced in our laboratory (Ylla et al., 2018) and available at Gene Expression Omnibus with accession number GSE99785.

METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101778.

ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Economy and Competitiveness (grants CGL2012-36251, CGL2015-64727-P and PID2019-104483GB-I00 to XB, including FEDER funds), the CSIC (grant 2019AEP029), and the Catalan Government (grants 2014 SGR 619 and 2017 SGR 1030).

AUTHOR CONTRIBUTIONS

X.B. designed the research; A.V.-A. X.B., J.C.M and G.Y. performed the research; GY and J.C.M performed the bioinformatics analyses; X.B., A.V.-A., G.Y. and J.C.M discussed and interpreted the results; A.V.-A. and X.B. wrote the paper.

DECLARATIONS OF INTERESTS

The authors declare no competing interests.

Received: September 3, 2020
Revised: October 8, 2020
Accepted: November 3, 2020
Published: December 18, 2020

REFERENCES

Alemu, E.Y., Carl, J.W., Jr., Corrada Bravo, H., and Hannenhalli, S. (2014). Determinants of expression variability. Nucleic Acids Res. *42*, 3503–3514.

Anastasiadi, D., Esteve-Codina, A., and Piferrer, F. (2018). Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. Epigenetics Chromatin *11*, 37.

Bashkeel, N., Perkins, T.J., Kærn, M., and Lee, J.M. (2019). Human gene expression variability and its dependence on methylation and aging. BMC Genomics *20*, 941.

Belles, X. (2010). Beyond *Drosophila*: RNAi in vivo and functional genomics in insects. Annu. Rev. Entomol. *55*, 111–128.

Belles, X. (2020). Insect Metamorphosis. From Natural History to Regulation of Development and Evolution (Academic Press).

Bewick, A.J., Sanchez, Z., McKinney, E.C., Moore, A.J., Moore, P.J., and Schmitz, R.J. (2019). Dnmt1 is essential for egg production and embryo viability in the large milkweed bug, *Oncopeltus fasciatus*. Epigenetics Chromatin *12*, 6.

Bewick, A.J., Vogel, K.J., Moore, A.J., and Schmitz, R.J. (2017). Evolution of DNA methylation across insects. Mol. Biol. Evol. *34*, 654–665.

Bird, A.P. (1995). Gene number, noise reduction and biological complexity. Trends Genet. *11*, 94–100.

Bonasio, R., Li, Q., Lian, J., Mutti, N.S., Jin, L., Zhao, H., Zhang, P., Wen, P., Xiang, H., Ding, Y., et al. (2012). Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Curr. Biol. *22*, 1755–1764.

Cardoso-Júnior, C.A., Fujimura, P.T., Santos-Júnior, C.D., Borges, N.A., Ueira-Vieira, C., Hartfelder, K., Goulart, L.R., and Bonetti, A.M. (2017). Epigenetic modifications and their relation to caste and sex determination and adult division of labor in the stingless bee *Melipona scutellaris*. Genet. Mol. Biol. *40*, 61–68.

Ciudad, L., Bellés, X., and Piulachs, M.D. (2007). Structural and RNAi characterization of the German cockroach lipophorin receptor, and the evolutionary relationships of lipoprotein receptors. BMC Mol. Biol. *8*, 53.

Cortijo, S., Aydin, Z., Ahnert, S., and Locke, J.C. (2019). Widespread inter-individual gene expression variability in *Arabidopsis thaliana*. Mol. Syst. Biol. *15*, e8591.

Falckenhayn, C., Boerjan, B., Raddatz, G., Frohme, M., Schoofs, L., and Lyko, F. (2013). Characterization of genome methylation patterns in the desert locust *Schistocerca gregaria*. J. Exp. Biol. *216*, 1423–1429.

Fatemi, M., Hermann, A., Gowher, H., and Jeltsch, A. (2002). Dnmt3a and Dnmt1 functionally cooperate during *de novo* methylation of DNA. Eur. J. Biochem. *269*, 4981–4984.

Feliciello, I., Parazajder, J., Akrap, I., and Ugarković, D. (2013). First evidence of DNA methylation in insect *Tribolium castaneum*: environmental regulation of DNA methylation within heterochromatin. Epigenetics *8*, 534–541.

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

Glastad, K.M., Gokhale, K., Liebig, J., and Goodisman, M.A.D. (2016). The caste- and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*. Sci. Rep. *6*, 37110.

Glastad, K.M., Hunt, B.G., and Goodisman, M.A.D. (2014). Evolutionary insights into DNA methylation in insects. Curr. Opin. Insect Sci. *1*, 25–30.

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

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

He, X.J., Chen, T., and Zhu, J.K. (2011). Regulation and function of DNA methylation in plants and animals. Cell Res. *21*, 442–465.

Huh, I., Zeng, J., Park, T., and Yi, S.V. (2013). DNA methylation and transcriptional noise. Epigenetics Chromatin *6*, 9.

Hunt, B.G., Glastad, K.M., Yi, S.V., and Goodisman, M.A.D. (2013). The function of intragenic DNA methylation: insights from insect epigenomes. Integr. Comp. Biol. *53*, 319–328.

Irles, P., Bellés, X., and Piulachs, M.D. (2009). Identifying genes related to choriogenesis in insect panoistic ovaries by Suppression Subtractive Hybridization. BMC Genomics *10*, 206.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., and Jaenisch, R. (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat. Genet. *27*, 31–39.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

De Jong, T.V., Moshkin, Y.M., and Guryev, V. (2019). Gene expression variability: the other dimension in transcriptome analysis. Physiol. Genomics *51*, 145–158.

Jurkowski, T.P., Meusburger, M., Phalke, S., Helm, M., Nellen, W., Reuter, G., and Jeltsch, A. (2008). Human DNMT2 methylates tRNAAsp molecules using a DNA methyltransferase-like catalytic mechanism. RNA *14*, 1663–1670.

Lewis, S.H., Ross, L., Bain, S.A., Pahita, E., Smith, S.A., Cordaux, R., Miska, E.A., Lenhard, B., Jiggins, F.M., and Sarkies, P. (2020). Widespread conservation and lineage-specific diversification of genome-wide DNA methylation patterns across arthropods. PLoS Genet. *16*, e1008864.

Li, B., Hou, L., Zhu, D., Xu, X., An, S., and Wang, X. (2018). Identification and caste-dependent expression patterns of DNA methylation associated genes in *Bombus terrestris*. Sci. Rep. *8*, 2332.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

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

Miyazawa, H., and Aulehla, A. (2018). Revisiting the role of metabolism during development. Development *145*, dev131110.

Molaro, A., Malik, H.S., and Bourc'his, D. (2020). Dynamic evolution of de novo DNA methyltransferases in rodent and primate genomes. Mol. Biol. Evol. *37*, 1882–1892.

Provataris, P., Meusemann, K., Niehuis, O., Grath, S., and Misof, B. (2018). Signatures of DNA methylation across insects suggest reduced DNA methylation levels in Holometabola. Genome Biol. Evol. *10*, 1185–1197.

Pfaffl, M.W. (2002). Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. *30*, e36.

Rai, K., Nadauld, L.D., Chideste, r. S., Manos, E.J., James, S.R., Karpf, A.R., Cairns, B.R., and Jones, D.A. (2006). Zebra fish Dnmt1 and Suv39h1 regulate organ-specific terminal differentiation during development. Mol. Cell. Biol. *26*, 7077–7085.

Robinson, K.L., Tohidi-Esfahani, D., Ponton, F., Simpson, S.J., Sword, G.A., and Lo, N. (2016). Alternative migratory locust phenotypes are associated with differences in the expression of genes encoding the methylation machinery. Insect Mol. Biol. *25*, 105–115.

Sarda, S., Zeng, J., Hunt, B.G., and Yi, S.V. (2012). The evolution of invertebrate gene body methylation. Mol. Biol. Evol. *29*, 1907–1916.

Schulz, N.K.E., Wagner, C.I., Ebeling, J., Raddatz, G., Diddens-de Buhr, M.F., Lyko, F., and Kurtz, J. (2018). Dnmt1 has an essential function despite the absence of CpG DNA methylation in the red flour beetle *Tribolium castaneum*. Sci. Rep. *8*, 16462.

Song, X., Huang, F., Liu, J., Li, C., Gao, S., Wu, W., Zhai, M., Yu, X., Xiong, W., Xie, J., and Li, B. (2017). Genome-wide DNA methylomes from discrete developmental stages reveal the predominance of non-CpG methylation in *Tribolium castaneum*. DNA Res. *24*, 445–458.

Stancheva, I., Hensey, C., and Meehan, R.R. (2001). Loss of the maintenance methyltransferase, xDnmt1, induces apoptosis in *Xenopus* embryos. EMBO J. *20*, 1963–1973.

Suzuki, M.M., Kerr, A.R.W., De Sousa, D., and Bird, A. (2007). CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res. *17*, 625–631.

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

Wang, X., Wheeler, D., Avery, A., Rago, A., Choi, J.H., Colbourne, J.K., Clark, A.G., and Werren, J.H. (2013). Function and evolution of DNA methylation in *Nasonia vitripennis*. PLoS Genet. *9*, 1003872.

Wang, Y., Jorda, M., Jones, P.L., Maleszka, R., Ling, X., Robertson, H.M., Mizzen, C.A., Peinado, M.A., and Robinson, G.E. (2006). Functional CpG methylation system in a social insect. Science *314*, 645–647.

Wang, Z., and Zhang, J. (2011). Impact of gene expression noise on organismal fitness and the efficacy of natural selection. Proc. Natl. Acad. Sci. U S A *108*, E67–E76.

Ylla, G., Piulachs, M.D., and Belles, X. (2018). Comparative transcriptomics in two extreme neopterans reveals general trends in the evolution of modern insects. iScience *4*, 164–179.

Ylla, G., Nakamura, T., Itoh, T., Kajitani, R., Toyoda, A., Tomonari, S., Bando, T., Ishimaru, Y., Watanabe, T., Fuketa, M., et al. (2020). Cricket genomes: the genomes of future food. bioRxiv. https://doi.org/10.1101/2020.07.07.191841.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916–919.

Zwier, M.V., Verhulst, E.C., Zwahlen, R.D., Beukeboom, L.W., and Van De Zande, L. (2012). DNA methylation plays a crucial role during early *Nasonia* development. Insect Mol. Biol. *21*, 129–138.

iScience, Volume *23*

Supplemental Information

DNMT1 Promotes Genome Methylation and Early Embryo Development in Cockroaches

Alba Ventós-Alfonso, Guillem Ylla, Jose-Carlos Montañes, and Xavier Belles

SUPPLEMENTAL INFORMATION

DNMT1 promotes genome methylation and early embryo development in cockroaches

Alba Ventós-Alfonso, Guillem Ylla, Jose-Carlos Montañes, and Xavier Belles

Figure S1. Phylogenetic relationships of the DNMT1 and DNMT3 proteins of *Blattella germanica* with those of other insect species.

Figure S2. Violin plots displaying the FPKM distribution in log scale between the two transcriptome replicates of each studied developmental stage of *Blattella germanica*.

Figure S3. Alignment of the catalytic region of DNMT3 of *Blattella germanica* compared with that of *Apis mellifera*, *Nasonia vitripennis* and *Zootermopsis nevadensis*.

Figure S4. Percentage of CG methylation in the four RRBS libraries obtained in *Blattella germanica*.

Table S1. Primers used to measure the expression levels of *DNMT1* and *DNMT3* by qRT-PCR, and to prepare the corresponding dsRNAs for RNAi experiments.

Table S2. Statistics of the RRBS libraries, and bisulfite conversion ratio.

Materials and methods.

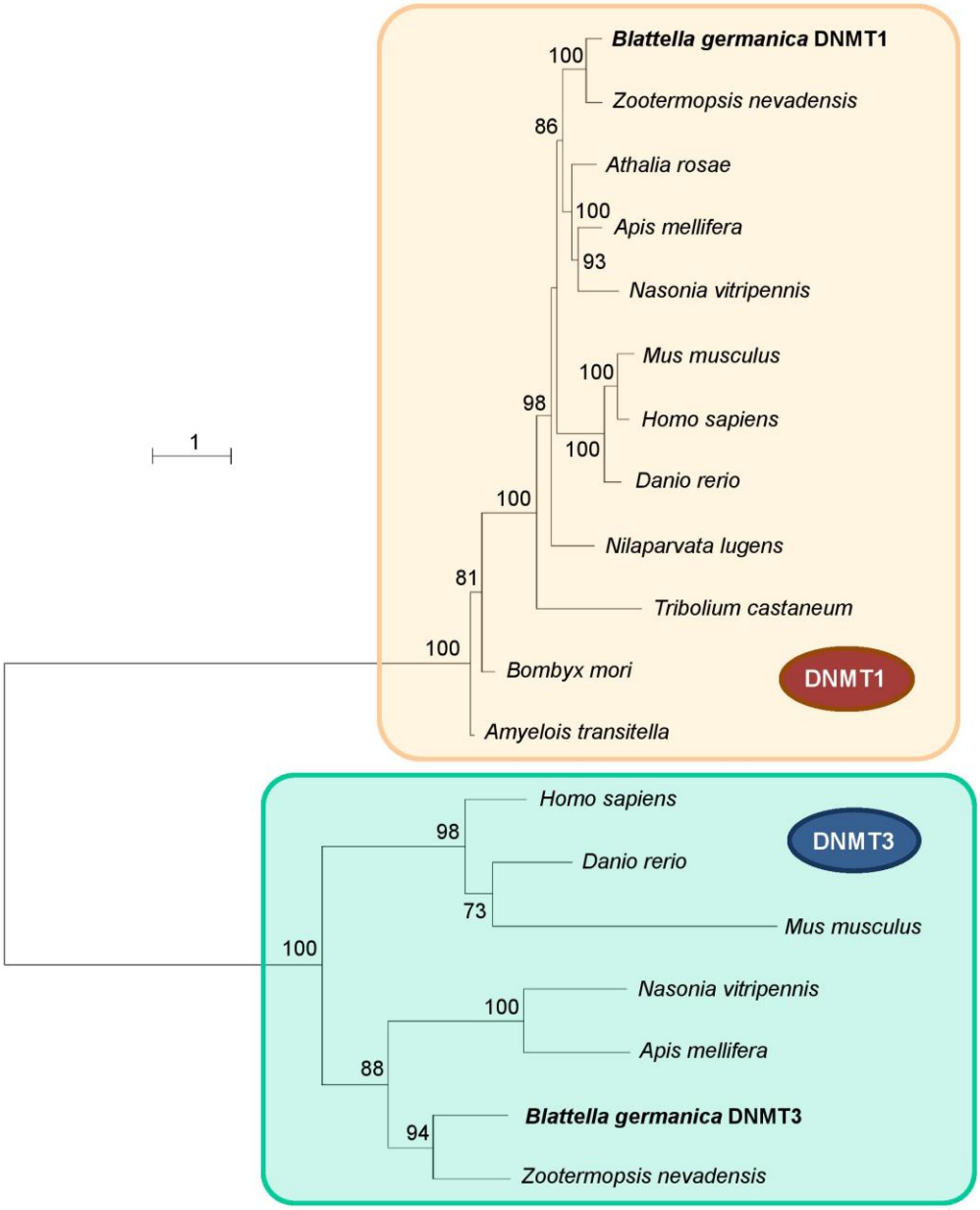


Figure S1. Phylogenetic relationships of the DNMT1 and DNMT3 proteins of *Blattella germanica* with those of other insect species. Sequences used were obtained by Blast from GenBank. Alignments were carried out with ClustalX (http://www.clustal.org/clustal2) and phylogenetic reconstruction with PhyML 3.1 (http://www.atgc-montpellier.fr/phyml), based on the maximum-likelihood principle, a JTT matrix, a gamma model of heterogeneity rate, and using empirical base frequencies and estimating proportions. The data was bootstrapped for 100 replicates. The sequences used for comparison with those of *B. germanica* were the following. For DNMT1, *Amyelois transitella* (XP_013186878.1), *Apis mellifera* (XP_026298868.1), *Athalia rosae* (XP_012254091.1), *Bombyx mori* (XP_012550860.1), *Danio rerio* (NP_571264.2), *Homo sapiens* (EAW84079.1), *Mus musculus* (EDL25141.1), *Nasonia vitripennis* (XP_008212391.1), *Nilaparvata lugens* (AHZ08393.1), *Tribolium castaneum* (XP_008193458.1) and *Zootermopsis nevadensis* (XP_021941799.1). For DNMT3 we used *A. mellifera* (XP_026302146.1), *D. rerio* (AAI62467.1), *H. sapiens* (3A1B_A), *M. musculus* (NP_001075164.1), *N. vitripennis* (XP_008204446.1), *Z. nevadensis* (XP_021915977.1). Bootstrap values >50 are indicated in the corresponding nodes. Scale bar: number of substitutions per site. Related to Figure 1.

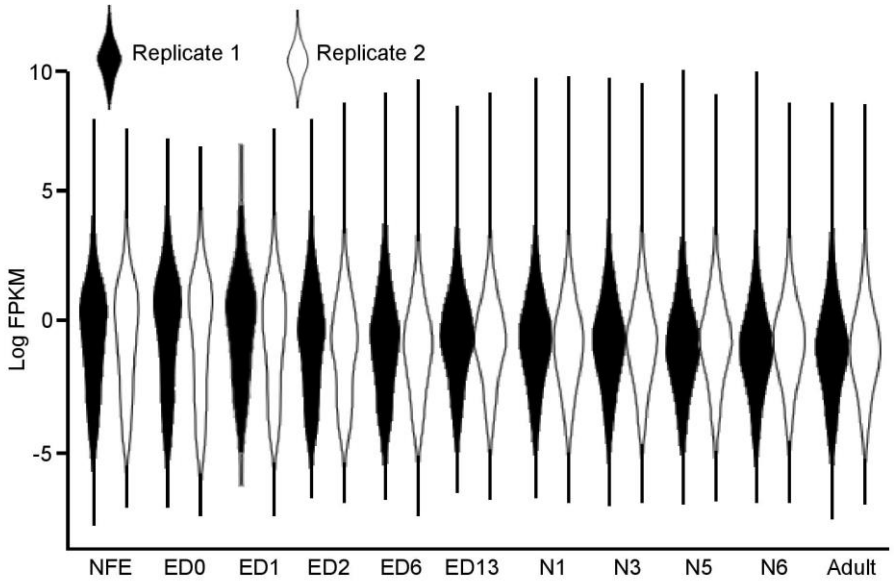


Figure S2. Violin plots displaying the FPKM distribution in log scale between the two transcriptome replicates of each studied developmental stage of *Blattella germanica*. The developmental stages are: NFE: non-fertilized egg, ED0 to ED13: embryo day 0 to embryo day 13, N1-N6: first to sixth nymphal instar, and the adult (Ylla et al., 2018). Related to Figure 5.

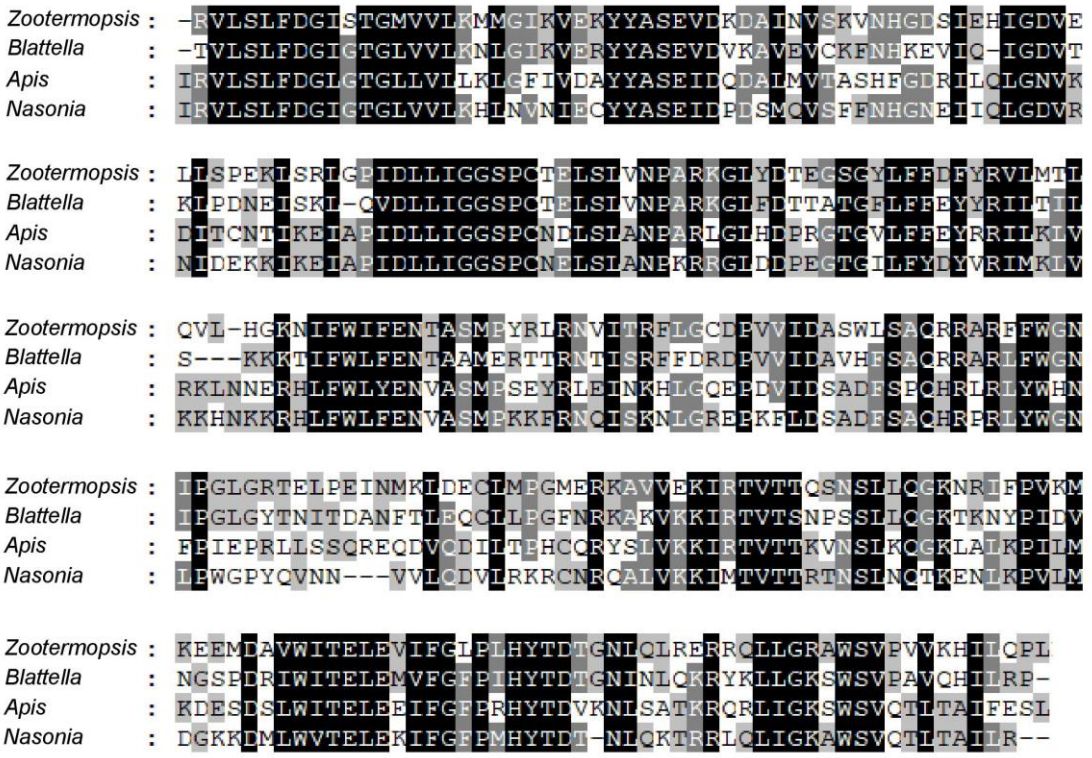


Figure S3. Alignment (Clustal X, Larkin et al., 2007) of the catalytic region of DNMT3 of *Blattella germanica* compared with that of *Apis mellifera* (XP_026302146.1), *Nasonia vitripennis* (XP_008204446.1) and *Zootermopsis nevadensis* (XP_021915977.1). Related to Figure 1.

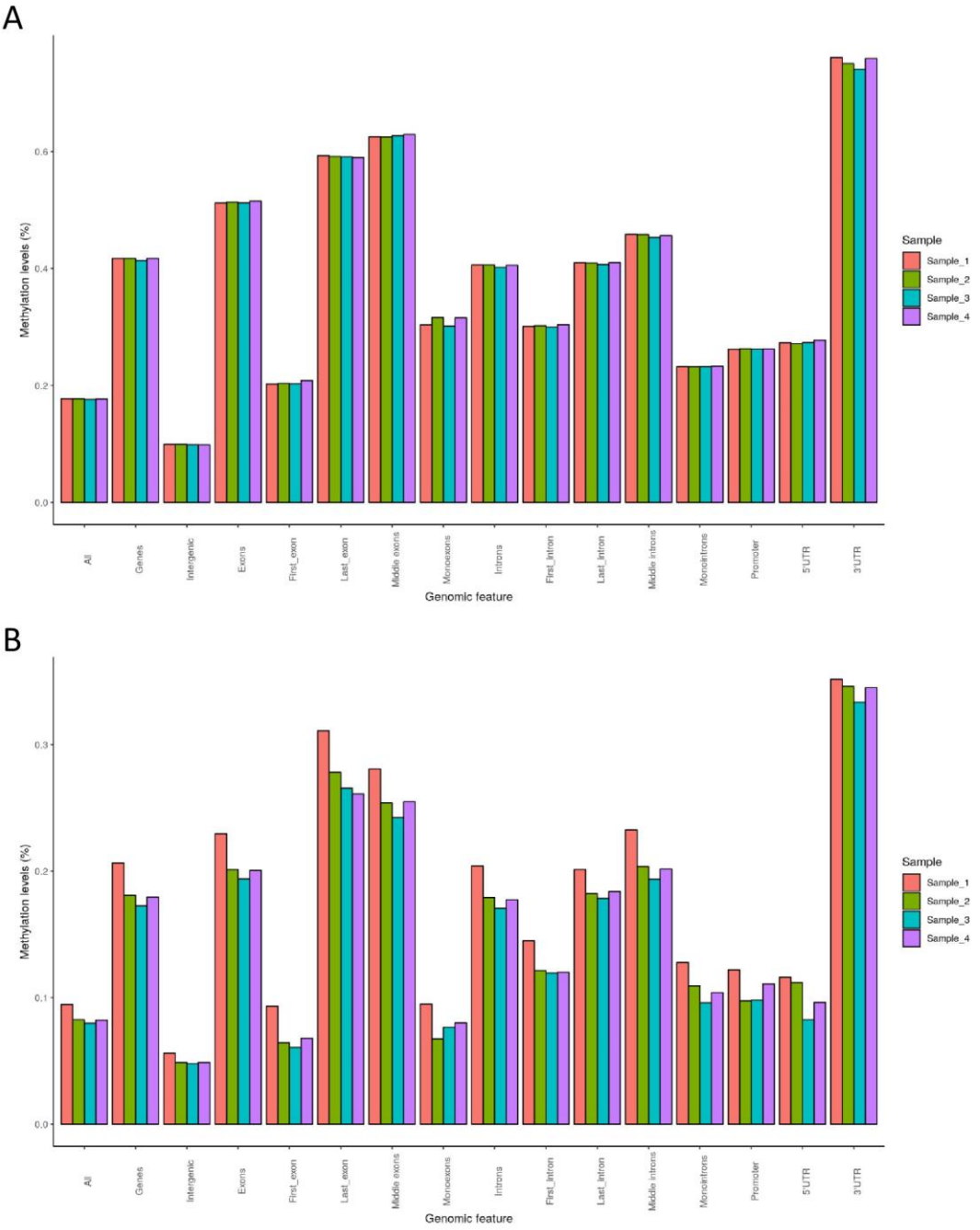


Figure S4. Percentage of CG methylation in the four RRBS libraries obtained in *Blattella germanica*. (A) RRBS of control insects. (B) RRBS of DNMT1-depleted insects. (%) levels in each biological replicate per every genomic feature studied. The genomic features examined are the same described in Table 1. The methylKit package v1.4.1 for R with methylation calling, was used for comparisons; called bases with less than 10 reads or more than the 99th percentile of coverage were discarded. Related to Table 1.

Table S1. Primers used to measure the expression levels of *DNMT1* and *DNMT3* by qRT-PCR, and to prepare the dsRNAs for RNAi experiments. Related to Figure 1.

Gene	Forward primer	Reverse primer	Accession number
Actin-5c	AGCTTCCTGATGGTCAGGTGA	TGTCGGCAATTCCAGGGTACATGGT	AJ862721.1
DNMT1 (qRT-PCR)	ATGAAAAAGGCGGTTGTGAC	AAAAACGTCTTGCCGTCATC	MT881788
DNMT1 (dsRNA)	GTCGGAGAAAGTGGCAAGAG	GTCACAACCGCCTTTTTCAT	MT881788
DNMT3 (qRT-PCR)	TTGAAAACACGGCTGCTATG	TGTGCCGAAAAATGTACAGC	MT881790
DNMT3 (dsRNA)	CGCCGAGCTAGGTTATTCTG	CAATCTTGGTGCTCTGAGTCC	MT881790

Table S2. Statistics of the RRBS libraries, and bisulfite conversion ratio. ED4C1-ED4C4 are the four libraries of DNMT1-depleted 4-day-old embryos, and ED4T1-ED4T4 the four control libraries. Related to Table 1.

Library	ED4C1	ED4C2	ED4C3	ED4C4	ED4T1	ED4T2	ED4T3	ED4T4
Raw reads	18,519,760	24,355,944	40,711,653	41,651,622	28,035,791	24,170,456	20,865,347	20,626,862
Trimmed reads	6,909,794	93,27,442	15,334,651	15,793,729	12,434,698	9,532,532	8,947,228	9,258,945
Trimmed ratio (%)	37.31	38.30	37.67	37.92	44.35	39.44	42.88	44.89
Unique aligned reads	9,061,483	12,677,679	20,339,272	21,053,385	10,820,311	11,012,600	8,698,668	7,621,916
Unique alignment ratio (%)	48.93	52.05	49.96	50.55	38.59	45.56	41.69	36.95
Covered cytosines	80,059,545	111,178,068	177,077,633	183,312,338	93,668,773	95,184,764	75,468,956	66,398,833
Covered CpG	19,942,568	27,626,895	44,141,130	45,914,155	23,151,959	23,545,622	18,702,854	16,430,827
CpGs > 10 reads	563,909	722,180	818,622	816,939	531,916	530,875	451,577	412,469
Fold coverage	28.94	32.7	47.9	50.34	38.35	38.86	36.45	34.63
Bisulfite conversion ratio (%)	98.4	98.4	98.7	98.7	98.4	98.4	98.2	97.6

MATERIALS AND METHODS

Insects and dissections

Insects were obtained from a *B. germanica* colony fed with Panlab dog chow (Panlab S.L.U, Barcelona, Spain) and water *ad libitum*, and reared in the dark at 29±1ºC and 60-70% relative humidity. Freshly ecdysed adult females were selected and used at appropriate ages. Mated females were used for all the experiments, and mating was assessed by checking the presence of sperm in the spermatheca. Prior to injection treatments, dissections, and tissue sampling, insects were anesthetized with carbon dioxide.

Alignments and phylogenetic analysis of DNMT1 and DNMT3

Sequences used for the analyses were obtained by Blast from GenBank. Alignments were carried out with ClustalX (Larkin et al. 2007) and the phylogenetic tree reconstruction with 100 bootstraps of PhyML 3.1 (Guindon et al., 2010), based on the maximum-likelihood principle, a JTT matrix, a gamma model of heterogeneity rate, and using empirical base frequencies and estimating proportions. The sequences used for comparison with those of *B. germanica* were the following. For DNMT1, *Amyelois transitella* (XP_013186878.1), *Apis mellifera* (XP_026298868.1), *Athalia rosae* (XP_012254091.1), *Bombyx mori* (XP_012550860.1), *Danio rerio* (NP_571264.2), *Homo sapiens* (EAW84079.1), *Mus musculus* (EDL25141.1), *Nasonia vitripennis* (XP_008212391.1), *Nilaparvata lugens* (AHZ08393.1), *Tribolium castaneum* (XP_008193458.1) and *Zootermopsis nevadensis* (XP_021941799.1). For DNMT3 we used *A. mellifera* (XP_026302146.1), *D. rerio* (AAI62467.1), *H. sapiens* (3A1B_A), *M. musculus* (NP_001075164.1), *N. vitripennis* (XP_008204446.1) and *Z. nevadensis* (XP_021915977.1).

RNA extraction and reverse transcription

Total RNA extraction was performed using RNeasy Plant minikit (QIAGEN, Hilden, Germany) in the case of young oothecae (from NFE to 4-day-old) and GenElute Mammalian Total RNA Miniprep kit (Sigma-Aldrich, Madrid, Spain) in the case of

older oothecae (from 6- to 16-day-old). The volume extracted was lyophilized in a freeze-dryer FISHER-ALPHA 1–2 LDplus. Then, it was resuspended in 8 µL of miliQ H_2O, treated with DNase I (Promega, Madison, WI, USA), and reverse transcribed with first Strand cDNA Synthesis Kit (Roche, Barcelona, Spain).

Quantification of mRNA levels by quantitative real time PCR

Quantitative real-time PCR (qRT-PCR) reactions were carried out in triplicate in an iQ5 Real-Time PCR Detection System (Bio-Lab Laboratories, Madrid, Spain), using SYBR®Green (iTaq™ Universal SYBR®Green Supermix; Applied Biosystems, Foster City, CA, USA). A template-free control was included in all batches. Primers used to detect the transcripts are detailed in Table S1. The efficiency of each set of primers was validated by constructing a standard curve through three serial dilutions. mRNA levels were calculated relative to Actin-5c mRNA (accession number AJ862721), using the Bio-Rad iQ5 Standard Edition Optical System Software (version 2.0). Results are given as copies of mRNA of interest per 1000 copies of Actin-5c mRNA. To test the statistical significance between treated and control samples we used the Relative Expression Software Tool (REST), which evaluates the significance of the derived results by Pairwise Fixed Reallocation Randomization Test (Pfaffl, 2002).

RNA interference

Maternal RNAi assays have been described previously (Fernandez-Nicolas and Belles, 2017). *DNMT1* and *DNMT3* sequences were amplified by PCR and then cloned into pSTBlue-1 vector. Primers used to prepare dsRNA are described in Table S1. A 307 bp sequence from *Autographa californica* nucleopolyhedrosis virus (Accession number K01149.1) was used as control dsRNA (dsMock). dsDNMT1 and dsDNMT3 were respectively injected at a dose of 3 µg in 1µl volume into the abdomen of 5-day-old adult females with a 5 µl Hamilton microsyringe. Control females were treated at the same age with the same dose and volume of dsMock.

Microscopy

Twenty-day-old oothecae were detached from female abdomen, and artificially opened. Embryos were dechorionated and individualized. The embryos were directly observed under the stereo microscope Carl Zeiss–AXIO IMAGER.Z1 (Oberkochen, Germany). For 4',6-di-amidino-2-phenylindole (DAPI) staining, 4-day-old oothecae were detached from female abdomen and incubated in PBT (Triton-X 0.1% in PBS 0.2M) in a water bath at 95ºC. Then, each ootheca was artificially opened and embryos were dechorionated and individualized. Between 12 and 24 embryos per ootheca, chosen from the central part of it, were dissected for staining. The embryos were fixed in 4% paraformaldehyde in PBS 0.2M for 1 h, washed with PBT, and then incubated for 10 min in 1 mg/mL DAPI. They were mounted in Mowiol (Calbiochem, Madison, WL, USA) and observed with a fluorescence microscope Carl Zeiss–AXIO IMAGER.Z1.

Reduced Represented Bisulfite Sequencing (RRBS) and methylation analyses

Four-day-old oothecae from dsDNMT1-treated females, and from controls (dsMock-treated) were detached from female abdomen and kept in a water bath at 95ºC. Then, each ootheca was artificially opened and embryos were dechorionated and individualized. In the dsMock group, all embryos were collected. In the dsDNMT1 group, only embryos showing abnormal phenotypes (between 40 and 50%) were collected. Genomic DNA was extracted using GenElute™ Mammalian Genomic DNA Miniprep Kit (Merck), following manufacturer's instructions. Then, samples were sent to the Genomics Unit of the Centre for Genomic Regulation (PRBB, Barcelona) where RRBS libraries were prepared using the Premium Reduced Representation Bisulfite Sequencing (RRBS) Kit (Diagenode), and sequenced ton an Illumina HiSeq 2000 platform in 50-bp single-end mode. We reserved a part of the sample to extract RNA, reverse transcribe, and check *DNMT1* transcript decrease by qRT-PCR. Sequences from RRBS libraries were first quality trimmed using Cutadapt (Marcel Martin, 2011), and then aligned to the reference *B. germanica* genome (Accession code: PRJNA427252) using Bismark v.0.20.0. (Krueger and Andrews, 2011). Once aligned, the bisulfite conversion ratio was calculated using spike-in controls provided by Premium Reduced Representation Bisulfite Sequencing (RRBS) Kit (Diagenode) (Table S2). Then, using the methylKit package v1.4.1 for R we performed methylation calling;

called bases with less than 10 reads or more than the 99th percentile of coverage were discarded. We checked that the four replicates have similar %mCG (Figure S4). Using the mixtools R package, and assuming that DNA methylation follow a mixture of three different Gaussian distributions, we identified the Gaussian distribution of hypomethylated genes (mean methylation = 0.84 ± 0.99 %) and the Gaussian distribution of hypermethylated genes (mean methylation = 93.77 ± 4.45%). Then, we selected as hypermethylated regions those with methylation values between 80.43 - 100% and as hypomethylated regions those with methylation values between 0 – 3.82% (that is, using mean of each of the two peaks of the bimodal distribution of mCG ± 3 standard deviations to categorize the genes in hypermethylated and hypomethylated).

Gene ontology (GO) analyses

Using gene annotations available at NCBI bioproject with accession number PRJNA427252, we assigned GO terms to *B. germanica* proteins using eggNOG mapper (Huerta-Cepas et al., 2017). Then, using topGO package (Alexa and Rahnenfuhrer, 2010), we tested enriched GO terms in hypermethylated and hypomethylated genes. We used Fisher's exact text with weighted algorithm to test statistical significance, which was stablished at $p < 0.05$.

Combined analyses of genomic methylation and gene expression

For the transcriptomic analyses, we used RNA-seq libraries produced in our laboratory (Ylla et al., 2018), available at Gene Expression Omnibus with accession number GSE99785. To examine the changes in the gene expression between ED2 and ED6, we normalized raw counts using trimmed mean of M values (TMM), and performed a differential expression analysis between the embryo day 2 (ED2) and embryo day 6 (ED6) using the edgeR package (Robinson et al., 2009). Then, we selected the genes with an absolute $\log_2FC \geq 2$ and FDR < 0.05 as differentially expressed genes. We separated the differentially expressed genes between upregulated and downregulated in ED6, and within each of these two groups we compared the expression levels of the hypomethylated vs hypermethylated genes with Mann–Whitney U test, adjusting p-values by false discovery rate (FDR) using Benjamini-Hochberg method. For the rest of

the analyses we considered those genes having a $|\log_2FC| \geq 2$ and FDR < 0.05. To test statistical differences between hypermethylated and hypomethylated genes FPKMs, $|\log_2FC|$ and Coefficient of Variation (CV), we also used Mann–Whitney U test, adjusting p-values by FDR using Benjamini-Hochberg method. In the graphs, data outliers have been omitted for clarity. We defined as outliers those values below Q1-1.5*Inter Quartil Range (IQR) or higher than Q3+1.5*IQR, being IQR the distance between Q1 and Q3.

References

Alexa, A., and Rahnenfuhrer, J. (2010). Bioconductor - topGO. R Packag. version 28.

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

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, .LJ., Von Mering, C., and Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. *34*, 2115–2122.

Krueger, F., and Andrews, S.R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompso, J.D., Gibson, T.J., and Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet *17*, 1–3.

Pfaffl, M.W. (2002). Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. *30*, e36.

Robinson, M.D., McCarthy, D.J., and Smyth, G,K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Ylla, G., Piulachs, M.D., and Belles, X. (2018). Comparative transcriptomics in two extreme neopterans reveals general trends in the evolution of modern insects. iScience *4*, 164–179.

3.2. Zelda and the maternal-to-zygotic transition in cockroaches

Alba Ventos-Alfonso, Guillem Ylla and Xavier Bellés

Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain

Alba Ventos-Alfonso, Guillem Ylla and Xavier Bellés

Zelda and the maternal-to-zygotic transition in cockroaches.

The FEBS Journal 286 (2019) 3206–3221

The FEBS Journal

FEBS PRESS science publishing by scientists

Zelda and the maternal-to-zygotic transition in cockroaches

Alba Ventos-Alfonso, Guillem Ylla* and Xavier Belles

Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain

Keywords
embryogenesis; maternal-to-zygotic transition; metamorphosis; *vielfältig*; *zelda*

Correspondence
X. Belles, Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Passeig Marítim 37, Barcelona 08003, Spain
Tel: +34 9323096011
E-mail: xavier.belles@ibe.upf-csic.es
Website: http://www.biologiaevolutiva.org/xbelles/

***Present address**
Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

Alba Ventos-Alfonso and Guillem Ylla contributed equally to this work

(Received 19 July 2018, revised 22 March 2019, accepted 15 April 2019)

doi:10.1111/febs.14856

In the endopterygote *Drosophila melanogaster*, Zelda is an activator of the zygotic genome during the maternal-to-zygotic transition (MZT). Zelda binds *cis*-regulatory elements (TAGteam heptamers), making chromatin accessible for gene transcription. Zelda has been studied in other endopterygotes: *Apis mellifera* and *Tribolium castaneum*, and the paraneopteran *Rhodnius prolixus*. We studied Zelda in the cockroach *Blattella germanica*, a hemimetabolan, short germ-band, and polyneopteran species. *B. germanica* Zelda has the complete set of functional domains, which is typical of species displaying ancestral features concerning embryogenesis. Interestingly, we found *D. melanogaster* TAGteam heptamers in the *B. germanica* genome. The canonical one, CAGGTAG, is present at a similar proportion in the genome of these two species and in the genome of other insects, suggesting that the genome admits as many CAGGTAG motifs as its length allows. Zelda-depleted embryos of *B. germanica* show defects involving blastoderm formation and abdomen development, and genes contributing to these processes are down-regulated. We conclude that in *B. germanica*, Zelda strictly activates the zygotic genome, within the MZT, a role conserved in more derived endopterygote insects. In *B. germanica*, *zelda* is expressed during MZT, whereas in *D. melanogaster* and *T. castaneum* it is expressed beyond this transition. In these species and *A. mellifera*, Zelda has functions even in postembryonic development. The expansion of *zelda* expression beyond the MZT in endopterygotes might be related with the evolutionary innovation of holometabolan metamorphosis.

Databases
The RNA-seq datasets of *B. germanica*, *D. melanogaster*, and *T. castaneum* are accessible at the GEO databases GSE99785, GSE18068, GSE63770, and GSE84253. In addition, the RNA-seq library from *T. castaneum* adult females is available at SRA: SRX021963. The *B. germanica* reference genome is available as BioProject PRJNA203136.

Introduction

An important event in early embryo development in metazoans is the maternal-to-zygotic transition (MZT), by which maternal mRNAs are eliminated and the zygotic genome becomes activated and governs

Abbreviations
AdD1, adult on day 1; AOF, after ootheca formation; *BR-C*, *Broad complex*; DAPI, 4′,6-diamidino-2-phenylindole; *dl*, *dorsal*; *Dnmt1*, *DNA methyltransferase-1*; *Dnmt2*, *DNA methyltransferase-2*; *dpp*, *decapentaplegic*; dsRNA, double-stranded RNA; DV, dorso-ventral; ED1, embryos on day 1; *eve*, *even*-skipped; FPKM, fragments per kilobase million; *ftz*, *fushi tarazu*; *Kr*, *Krüppel*; miRNA, microRNA; MZT, maternal-to-zygotic transition; N5D0, fifth nymphal instar on day 0; N5, fifth nymphal instar; N6, sixth (last) nymphal instar; NFE, nonfertilized eggs; Nn, normally hatched nymphs; *N*, *Notch*; *otd*, *orthodenticle*; qRT-PCR, real-time quantitative reverse transcription PCR; RNAi, RNA interference; *sna*, *snail*; *sog*, *short gastrulation*; *Tl*, *Toll*; *wg*, *wingless*.

further development [1]. In the MZT context, the protein Zelda was found to be a key activator of the early zygotic genome in the fruit fly, *Drosophila melanogaster* [2], an endopterygote species that shows a long germ-band embryo development and an holometabolan mode of metamorphosis. In vertebrates, no clear orthologs of Zelda have been found. Nevertheless, studies in the zebrafish, *Danio rerio*, have shown that the transcription factors Nanog, SoxB1, and Pou5f3 (Oct4) are required to initiate approximately 75% of the first major phase of zygotic gene activation (reviewed in Ref. [3,4]), thus they could play the role that Zelda plays in invertebrates.

The gene *zelda* was discovered in *D. melanogaster* in 2006 [5] as a X-chromosomal gene encoding a nuclear zinc finger protein, which was initially called *vielfältig*. Gene transcripts were maternally loaded, and loss of gene activity resulted in mitosis problems, and elicited asynchronous DNA replication [5]. The same year, the heptamer CAGGTAG (a consensus sequence of a larger series of similar heptamers collectively referred to as TAGteam) was found to be overrepresented in regulatory regions of many of the early transcribed genes in the zygote of *D. melanogaster* [6]. Subsequently, Liang *et al.* [2] discovered that the protein that binds specifically to the TAGteam sites and activate transcription was Vielfältig, although the authors renamed it as Zelda. *D. melanogaster* embryos lacking Zelda showed defects in the formation of the blastoderm and could not activate genes essential for further development. It was concluded that Zelda plays a key role in the activation of the early zygotic genome of *D. melanogaster* [2]. Subsequently, it was reported that during early embryogenesis of *D. melanogaster*, Zelda marks regions that are activated during the MZT, and promotes transcriptional activation of early-gene networks by binding to more than 1000 DNA regulatory regions [7,8]. It became soon clear that the most important role of Zelda in *D. melanogaster* is to increase the accessibility of chromatin, allowing the transcription of multiple genes [9–13]. Regarding functional organization, *D. melanogaster* Zelda presents a low complexity region, corresponding to the transactivation domain, and four C-terminal Zinc fingers, which are required for DNA binding and transcriptional activation [14]. Zelda was found to be conserved throughout insects and crustaceans, and examination of a number of species revealed that the protein has additional conserved regions, like an acidic patch (which could be involved in the recruitment of chromatin remodeling proteins) and two N-terminal Zinc fingers [15]. The functional study of Zelda domains was approached using a Cas9 engineering system to obtain *D. melanogaster* insects with point mutations in the Zinc finger 1, in the acidic patch, or in the Zinc finger 2 (JAZ) [16]. Mutations in the Zinc finger 1 or in the acidic patch did not affect the viability of the insects, whereas mutations in the Zinc finger 2 (JAZ) led to a maternal-effect lethal phenotype [16]. A transcriptomic analysis of mutated flies showed that the Zinc finger 2 (JAZ) acts as a maternal repressive domain during embryo development, being associated with the transcriptional regulation of the zygotic Zelda targets and with the clearance of maternal transcripts [16].

Outside *D. melanogaster*, expression of *zelda* has been studied in the honeybee *Apis mellifera*, another endopterygote, long germ-band, holometabolan species. Expression was observed in early embryogenesis in relation to blastoderm formation and gastrulation [17]. These expression features and the occurrence of TAGteam heptamers in *A. mellifera*, suggest that in this species Zelda plays the same role of activator of the early zygotic genome described in *D. melanogaster*. The gene *zelda* is also expressed in late embryogenesis of *A. mellifera*, associated with the development of central nervous system precursor cells [17]. Also, outside *D. melanogaster*, the expression of *zelda* has been studied in *Nasonia vitripennis* [18], which, as in all Hymenoptera, follow haplodiploid sex determination, thus, unfertilized eggs develop into haploid males and fertilized eggs develop into diploid females. The study showed that ploidy has little effect on timing early embryonic events in this wasp. Furthermore, RNA interference (RNAi) experiments demonstrated that Zelda is key not only during the MZT but also in posterior segmentation and patterning of imaginal disc structures of the beetle *Tribolium castaneum* [15], a holometabolan species with short germ-band embryo development. It was also shown that Zelda is important for posterior segmentation in the kissing bug *Rhodnius prolixus* [15], a hemimetabolan and short germ-band species.

Studying transcriptomic differences along the ontogeny of *D. melanogaster* and the cockroach, *Blattella germanica* (a short germ-band and hemimetabolan species), we found that *zelda* showed a very different expression pattern in the two species [19]. In *B. germanica*, the expression was concentrated in early embryo, whereas in *D. melanogaster zelda* expression covered the entire embryogenesis [19]. We related this difference with the modes of metamorphosis of the studied species, hemimetabolan in *B. germanica* and holometabolan in *D. melanogaster*. Hemimetaboly is characterized by an embryogenesis that produces a nymph displaying the essential adult body structure. In holometabolan species, embryogenesis gives rise to a larva with a body

structure considerably divergent from that of the adult, often more or less vermiform [20]. The present work aims at illuminating these conjectures by examining in *B. germanica* the organization of the protein, the distribution of TAGteams in the genome, the expression of *zelda* along the ontogeny, and, importantly, their functions explained in morphological and molecular terms.

Results

Zelda organization is conserved in *B. germanica*

By combining BLAST (National Center for Biotechnology Information, Bethesda, MD, USA) search in *B. germanica* transcriptomes, mapping the resulting sequences in the *B. germanica* genome, and PCR strategies, we obtained a cDNA of 2754 nucleotides comprising the complete open reading frame (GenBank accession number LT71728.1), which once translated gave a 918-amino acid sequence with high similarity to Zelda proteins, which we called BgZelda. In *B. germanica*, the gene coding for BgZelda has two exons, one containing part of 5′UTR, and the other containing part of 5′ UTR, the complete coding region, and the 3′UTR. As the coding region is contained in a single exon, protein spliced isoforms cannot be generated. In addition, RNA-seq data [19], robustly supports that there is a single Zelda protein in *B. germanica*. This contrasts with the different protein variants generated by the *zelda* gene of *D. melanogaster* [14,21].

The canonical Zelda originally reported in *D. melanogaster*, DmZelda, was described as composed by a JAZ-C2H2 zinc finger domain and four C2H2 zinc fingers in the C terminus region [2,14,21]. However, a recent work on insect and crustacean Zelda sequences [15] reported two additional zinc finger domains upstream the Zinc finger 2 (JAZ), although some insects have lost the first of them. Following the nomenclature of Ribeiro *et al.* [15], BgZelda possesses all seven C2H2 zinc fingers: two of them at the N-terminal part of the protein (ZF-Novel and ZF1), then the ZF2 (JAZ domain) follows, and at the C-terminal region there are the other four ZFs (ZF3 to ZF6). Between ZF-Novel and ZF1 there is the 'patch' region, with the sequence TMAPADSSD, and the 'conserved region' with the typical motif RYHPY. Then, between ZF1 and ZF2 we found the 'acidic patch', characterized by the motif [DE]I[LW]DLD, which in the case of BgZelda is EILDLD. Finally, between ZF2 and ZF3, there is the 'conserved region', which in BgZelda has the sequence PPNLMAGPPISMEAQTEGLP, and the 'activation region', with the motif LPSFAQ [15] (Fig. 1A). A high conservation was found in the C-terminal four zinc fingers ZF3 to ZF6 (Fig. 1B), which are responsible for recognizing and binding to the DNA TAGteam heptamers, as demonstrated in *D. melanogaster* [14]. In this ZF3 to ZF6 region, the percentage of similarity between DmZelda and BgZelda is 76.7%.

D. melanogaster TAGteam heptamers are present in *B. germanica* genome

In *D. melanogaster*, the C-terminal cluster of four zinc fingers recognizes particular cis-regulatory heptamer motifs collectively known as TAGteam, CAGGTAG being the one to which Zelda bound with the highest affinity [2,6,7]. Nien *et al.* [8] experimentally assessed that DmZelda binds the sequence CAGGTAG and, to a lesser extent, the seven additional TAGteam heptamers: TAGGTAG, CAGGTAC, CAGGTAT, CAGGTAA, TAGGTAA, CAGGCAG, and CAGGCAA. Conversely, two additional heptamers (TGAATAG and TATCGAT) were not bound by DmZelda. Using chromatin immunoprecipitation and high-throughput sequencing to map regions bound by Zelda around the MZT, Harrison *et al.* [7] found that the CAGGTAG heptamer was highly enriched in the top 1000 regions bound *in vivo*. Conversely, the occurrence of other TAGteam heptamers was not especially significant [7].

In our analyses of the *D. melanogaster* genome, we found 95 495 heptamers that coincided with any of the eight DmZelda-binding TAGteam elements previously identified [8], whereas 1 299 345 were found in the longer genome of *B. germanica*. Significantly, these amounts are clearly lower than those expected if the occurrence of these eight heptamers were by chance (68% and 78% of the expected by chance, respectively). The canonical heptamer CAGGTAG appears 8068 times in the *D. melanogaster* genome and 112 830 in that of *B. germanica* (Fig. 2A) (45% and 50% of the expected by chance, respectively). Interestingly, the number of CAGGTAG heptamers appears roughly proportional to the length of the genome (1 per 15 160 nucleotides in *B. germanica* and 1 per 17 894 nucleotides in *D. melanogaster*). Proportions are similar in the genome of *T. castaneum*, where CAGGTAG appears 6221 times, that is 1 per 24 845 nucleotides (33% of the expected by chance) (Fig. 2A).

To ascertain whether this proportionality is more general in insects, we searched the number of CAGGTAG heptamers in the genome of 13 insect species from different orders, with genome lengths ranging from 110 Mb (*Pediculus humanus*) to 5.7 Gb

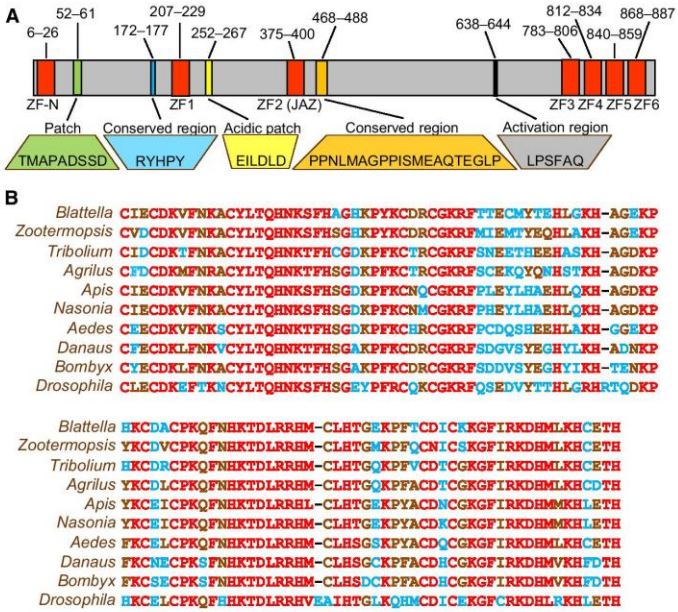


Fig. 1. Structure of Zelda in *Blattella germanica*. (A) Structure of the protein Zelda showing the most characteristic motifs according the nomenclature of Ribeiro *et al.* [15]; the numbers indicate the initial and final starting amino acid of each motif; ZF-N: ZF-Novel. (B) Alignment of the C-terminal four zinc fingers ZF3 to ZF6 of Zelda sequence of two species of Blattodea (*B. germanica*, SIW61620.1 and *Zootermopsis nevadensis*, KDR19000.1), two Hymenoptera (*Apis mellifera*, XP_006566924.1 and *Nasonia vitripennis*, XP_016842278.1), two Coleoptera (*Tribolium castaneum*, EFA04702.1 and *Agrilus planipennis*, XP_018332809.1), two Lepidoptera (*Bombyx mori*, XP_004933203.1 and *Danaus plexippus*, EHJ63786.1) and two Diptera (*Aedes aegypti*, XP_001659452.1 and *Drosophila melanogaster*, NP_001285466.1). Alignments were carried out with MAFT – online version (https://mafft.cbrc.jp/alignment/server/).

(*Locusta migratoria*) (Table S1). The analysis revealed a strong and significant correlation ($R^2 = 0.98$ and Pearson correlation P-value < 0.05) between genome length and the number of CAGGTAG heptamers (Fig. 2B). Moreover, we found that the number of CAGGTAG heptamers in each scaffold of the *B. germanica* genome is proportional to the length of the scaffold (Fig. 2C and Table S2), and that they are distributed more or less regularly within the scaffold (Fig. S1). This suggests that the 112 830 CAGGTAG motifs found in *B. germanica* are distributed more or less regularly throughout the genome, without significant concentrations in given regions.

We do not know whether BgZelda binds to the same heptamers as in *D. melanogaster*. However, the four C-terminal zinc fingers, ZF3 to ZF6, which have been shown to be responsible for recognizing and binding to DNA TAGteam heptamers in *D. melanogaster* [14], are highly conserved in BgZelda (Fig. 1B), as in other species [15]. This suggests the hypothesis that BgZelda binds to the heptamers reported in *D. melanogaster*, especially the canonical CAGGTAG. The fact that the CAGGTAG heptamer occurs in a similar proportion in the genome of *D. melanogaster* and *B. germanica* (Fig. 2A), suggests that it is subjected to similar selective pressures and that it plays similar roles in both species, which gives an additional support to the above hypothesis.

Bg*zelda* is expressed in a narrow temporal window around the MZT

Using real-time quantitative reverse transcription PCR (qRT-PCR), we studied the expression of Bg*zelda* along the ontogeny of *B. germanica* at the 11 stages used in the RNA-seq libraries reported by Ylla *et al.* [19], plus six additional embryo stages (days 4, 7, 9, 11, 14, and 16) in order to cover the embryogenesis more thoroughly. Results show that Bg*zelda* is predominantly expressed in the first stages of embryogenesis, between day 0 and day 2 (0–12.5% of embryogenesis), with a clear peak on day 1 (35 copies

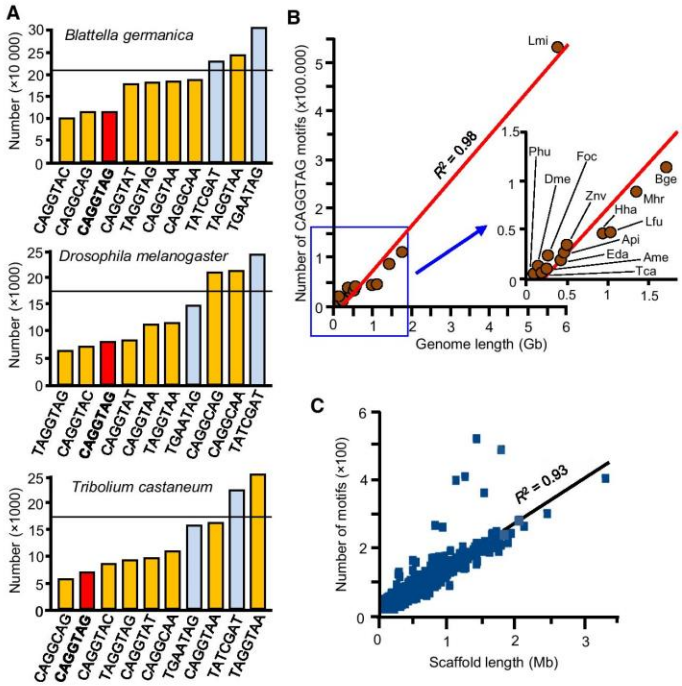


Fig. 2. The TAGteam heptamers in *Blattella germanica*. (A) Occurrence of the eight functional (yellow and red columns, the red column indicates the canonical sequence CAGGTAG) and not functional (blue columns) TAGteam heptamers reported by Nien *et al.* [8] in *Drosophila melanogaster, Tribolium castaneum*, and *B. germanica* genomes; the horizontal line crossing the columns indicates the expected occurrences of a random heptamer. (B) Correlation between genome length and number of CAGGTAG motifs in insects; the genomes included are the Archaeognatha *Machilis hrabei* (Mhr), the Odonata *Ladona fulva* (Lfu), the Ephemeroptera *Ephemera danica* (Eda), the Orthoptera *Locusta migratoria* (Lmi), the Blattodea *B. germanica* (Bge) and *Zootermopsis nevadensis* (Znv), the Hemiptera *Halyomorpha halys* (Hha), *Acyrthosiphon pisum* (Api) and *Frankliniella occidentalis* (Foc), the Phthiraptera *Pediculus humanus* (Phu), the Hymenoptera *Apis mellifera* (Ame), the Coleoptera *T. castaneum* (Tca), and the Diptera *D. melanogaster* (Dme); the shown regression line has a Pearson correlation coefficient of 0.98. (C) Length of the scaffolds of the *B. germanica* genome (masked regions excluded) and number of CAGGTAG motifs in each scaffold; the shown regression line has a Pearson correlation coefficient of 0.93.

of Bg*zelda* mRNA as average per 100 copies of actin mRNA), that is within the MZT. Beyond embryo day 4 (21% of embryogenesis) and until the adult stage, Bg*zelda* is expressed at comparatively very low levels (between 0.006 and 0.08 copies of Bg*zelda* as average per 100 copies of actin) (Fig. 3A).

The pattern obtained using qRT-PCR measurements is very similar to that obtained by computing the reads (FPKM) from the 11 stages of the RNA-seq libraries previously reported [19] (Fig. 3B). Indeed, a remarkable correlation ($R^2 = 0.82$) is obtained when using the same 11 stage points and comparing the FPKM and the qRT-PCR values. The expression of Bg*zelda* concentrated at the beginning of embryogenesis (Fig. 3A, B) contrasts with the expression in *D. melanogaster* obtained from transcriptomic data [19], which spans all along embryogenesis and first larval instar and then declines during the remaining postembryonic stages (Fig. 3C). Northern blot analyses had shown that Dm*zelda* expression appears quite high in the embryo, the first two larval instars, then decreases in the third larval instar and the pupa, and slightly increases in the adult [5], which is fairly coincident with the reads per kilobase million pattern obtained by us (Fig. 3C). The analysis of available transcriptomic data of *T. castaneum* [22,23], shows that the expression of Tc*zelda* progressively increases at least until 48 h after egg laying (30% embryo development), and then decreases in late embryo, larval, pupal and adult stages (Fig. 3C), a pattern that is similar to that of *D. melanogaster*, at

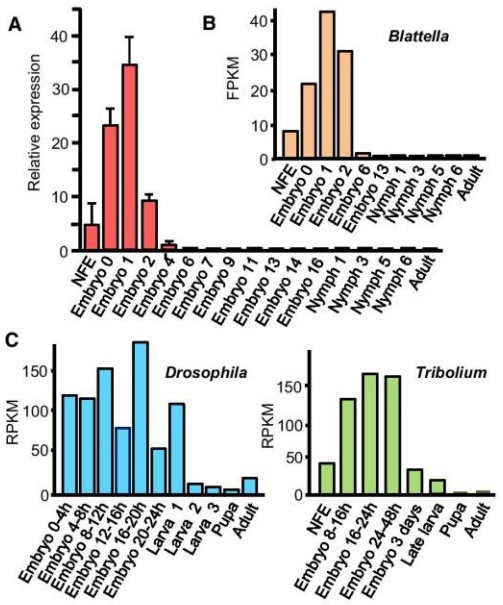


Fig. 3. Expression of *zelda* in *Blattella germanica*. (A) Expression pattern of Bg*zelda* along ontogeny in *B. germanica* obtained by qRT-PCR; the stages examined are the same of the libraries of Ylla *et al.* [19], plus a number of additional embryo stages (ED4, ED7, ED9, ED11, ED14, ED16); each value represents three biological replicates and it is represented as copies of Bg*zelda* mRNA per 100 copies of BgActin-5c mRNA (mean ± SEM). (B) Transcriptomic expression of Bg*zelda* in *B. germanica* according to Ylla *et al.* [19]. (C) Transcriptomic expression of *zelda* in *Drosophila melanogaster* and *Tribolium castaneum*; data of *D. melanogaster* are as in Ref. [19]. In *T. castaneum*, the following five stages are represented: NFE and three sequential embryo stages: 8–16 h, 16–24 h, 24–48 h, which represent 30% development at 25 °C [22], as well as 3-day-old embryo, late larva, pupa, and adult (female) [43].

least in the postembryonic period and up to 30% of the embryogenesis.

Bg*zelda* is mainly expressed in the abdominal growth zone

To study the spatial pattern of Bg*zelda* expression, we used *in situ* hybridization. Specificity of the probe was assessed not only with the use of a sense probe as negative control but also depleting Bg*zelda* mRNA by maternal RNAi. One-day-old adult females (AdD1) were injected with 3 μg of dsZelda (treated females), whereas equivalent females were injected with 3 μg of dsMock (control females), and kept with males in both cases. A first ootheca was formed on day 8 (AdD8) in both groups, and it was dissected 3 days later, obtaining embryos early on day 3 of development. Comparison of hybridizations obtained with a sense probe, showing unspecific background labeling (Fig. 4A), with those obtained using an antisense Bg*zelda* probe (Fig. 4B), show that specific labeling appears in the posterior cephalic region corresponding to the buccal pieces, and especially in the distal part of the abdominal region. This labeling vanishes in the BgZelda-depleted embryos (Fig. 4C), which supports their specificity. However, specific labeling of Bg*zelda* expression conspicuously concentrates in the distal part of the abdomen, the growth zone where the last abdominal segments should be formed. Detailed examination of this region revealed intense Bg*zelda* expression in controls, whereas it vanished in BgZelda-depleted embryos (Fig. 4D,E).

Depletion of BgZelda results in embryo defects

The maternal RNAi approach used in the *in situ* experiments was also used to study the function of BgZelda. Control females (injected with 3 μg of dsMock on AdD1) and treated females (injected with 3 μg of dsZelda on AdD1), were allowed to mate, and formed a first ootheca on AdD8. The oothecae of control females (n = 32) gave normal embryos that hatched 19 days later (AdD27), as 1281 first instar nymphs (35–40 nymphs per ootheca). The dsZelda-treated females (n = 39), also produced a first ootheca on AdD8. Nineteen of the 39 oothecae (49%) gave normally hatched nymphs (35–40 nymphs per ootheca) in AdD27, whereas from the remaining 20 oothecae (51%) no nymphs hatched, or hatched very few (a maximum of 20 hatched nymphs per ootheca, but often much less). In total, from the 39 oothecae formed by dsZelda-treated females, we recorded 1234 individuals, from which 1024 (83%) were nymphs that hatched normally, and 210 (17%) were embryos or nymphs that did not hatch, and that were dissected out from the totally or partially unviable oothecae on AdD28. These 210 embryos or unhatched nymphs showed a diversity of phenotypes that were classified into the following seven categories in different stages defined according to Tanaka [24]. Phenotype A (Fig. 5A): normal embryos in stage 19, thus ready to hatch as normal nymphs, but which did not hatch. Phenotype B (Fig. 5B): embryos in stage 16–18, morphologically similar to controls but showing a general brown coloration. Phenotype C (Fig. 5C): embryos in stage 17 but with the abdomen narrower and somewhat shorter than in controls. Phenotype D (Fig. 5D): embryos in stage 16, but with the abdomen more

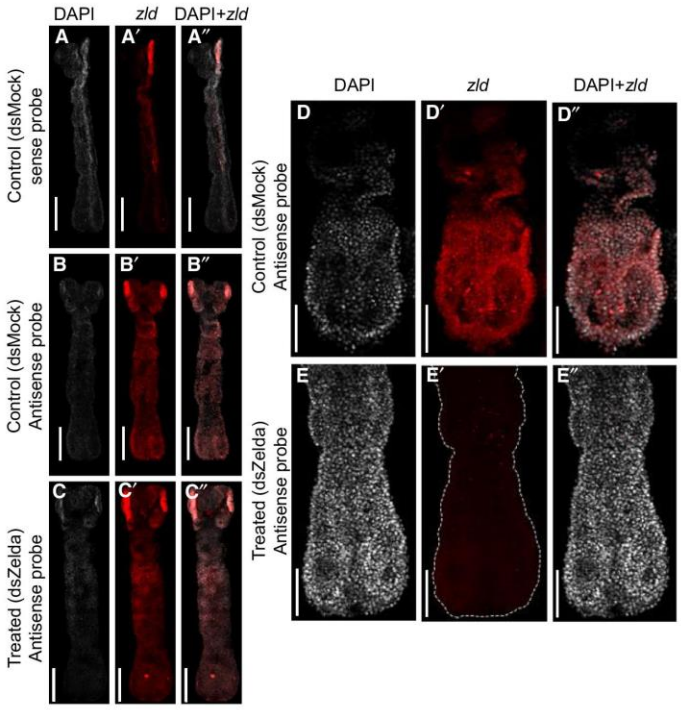


Fig. 4. Expression of Bg*zelda* in 3-day-old embryos of *Blattella germanica*, detected by *in situ* hybridization. (A–C). Entire embryo views; the upper panels, A–A″, correspond to negative controls, showing unspecific background staining observed with a Bg*zelda* (Zld) sense probe; the central panels, B–B″, show Bg*zelda* expression in control embryos (dsMock-treated females); the lower panels, C–C″, show the expression in BgZelda-depleted embryos; scale bars: 200 μm. (D–E). Detail of the distal part of the abdomen; the upper panels, D–D″, show Bg*zelda* expression in control embryos; the lower panels, E–E″, show the expression in BgZelda-depleted embryos; scale bars: 100 μm. DAPI staining (white, A–E), Bg*zelda* pattern (red, A′–E′) and DAPI + Bg*zelda* (pink, A″–E″), are shown from left to right.

elongated and smaller than in controls; this phenotype shows the same essential defect as phenotype C, but being in an early developmental stage. Phenotype E (Fig. 5E): dorsal region of the embryo forming a concave curve (while in controls is convex, like in Fig. 5A); they can reach the stage 15 and even beyond. Phenotype F (Fig. 5F): embryos with the development interrupted between stages 9 and 11, in general almost transparent, with the abdomen shorter than normal, and showing a variety of defects of segmentation and appendage formation. Phenotype G (Fig. 5G): embryos corresponding to very early development with no signs of segmentation. The most common (29% of abnormal embryos) was phenotype A, followed by D (23%), and then B, C, E, and G ca. 10% penetrance each) and F (5%) (Fig. 5H). If we merge categories C and D, as they show the same essential phenotype (small abdomen), then we can conclude that ca. 33% of the embryos from unviable oothecae show this defect.

As the most conspicuous expression of Bg*zelda* extends from ED0 to ED2, and in ED4 already falls to very low levels (Fig. 3A), we considered that BgZelda functions concentrate over that period. Consequently, we decided to study the BgZelda-depleted embryos just after the period of maximal expression, following the same protocol of maternal RNAi but dissecting the oothecae 4 days after its formation (at ED4 stage). Control females (n = 7) were treated with dsMock, and 172 ED4 embryos were studied from their oothecae, whereas from the oothecae of the dsZelda-treated females (n = 15) we examined 239 embryos. The 172 embryos (100%) from control females showed the normal aspect of an ED4 embryo (18–20% embryo development, Tanaka stage 4–5) (Fig. 6A). From the 239 embryos examined in oothecae from dsZelda-treated females, a total of 99 (41%) were normal ED4 embryos, identical to controls (Fig. 6A). The remaining 140 embryos (59%) showed a diversity of developmental delays and malformations, which were classified into five categories, as follows. Phenotype H (Fig. 6B): embryo with seriously impaired segmentation, the cephalic segments are more or less delimited, the three thoracic segments can be distinguished by three undulated furrows in the thorax region, and the abdominal region is amorphous; in terms of general shape, it resembles the Tanaka stages 2–3 (12–15% development). Phenotype I (Fig. 6C):

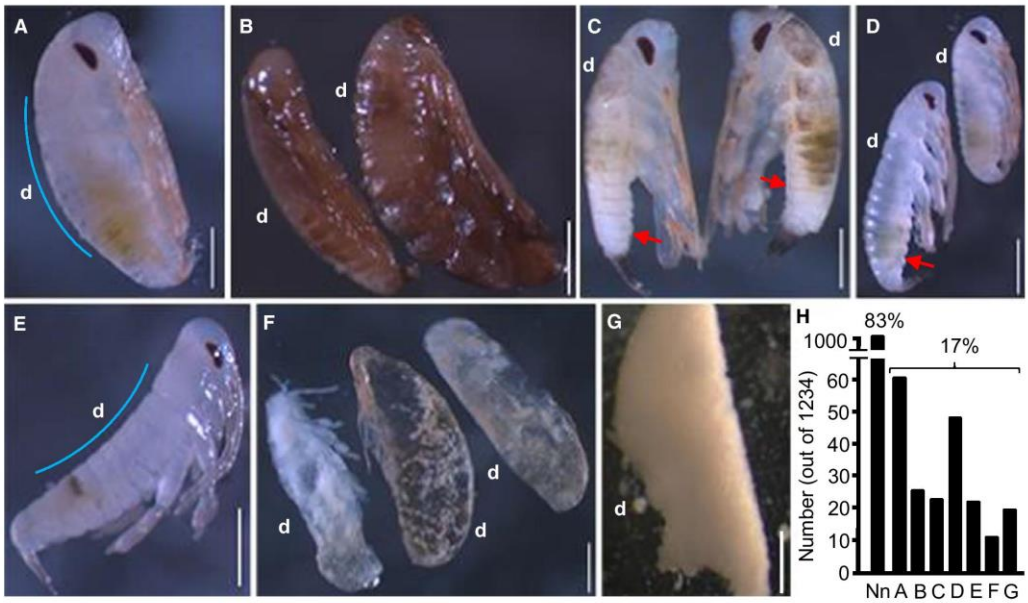


Fig. 5. Phenotypes observed in unhatched oothecae from BgZelda-depleted *Blattella germanica*. (A) Phenotype A, normal embryos in stage 19, thus ready to hatch as normal nymphs (used as control of phenotypes B to G), but which did not hatch. (B) Phenotype B, embryos in stage 16–18, similar to controls but showing a brown coloration. (C) Phenotype C, embryos in stage 17 but with the abdomen smaller than in controls. (D) Phenotype D, embryos in stage 16, but with the abdomen more elongated and smaller than in controls. (E) Phenotype E, embryos with the curvature of the body inverted, with the dorsal part concave instead of convex. (F) Phenotype F, embryos in medium stages of development, between stages 9 and 11, almost transparent and with a short abdomen. (G) Phenotype G, embryos in very early development, with no signs of segmentation. (H) Number of individuals showing the different phenotypes, from a sample of 1234 studied; Nn: normally hatched nymphs; (A–G) phenotypes of the unhatched individuals. In A–G, the upper part of the panel corresponds to the cephalic part of the embryo, and 'd' indicates the dorsal part; the arrowheads indicate the abdomen smaller than normal; the blue line in A and E shows the different curvature of the dorsal part of the embryo; scale bars in panels A–G: 500 μm.

embryo with seriously impaired segmentation, twisted, shorter, and wider, and with the caudal region (which would correspond to the growth zone) amorphous and swollen. Phenotype J (Fig. 6D): embryo with no traces of segmentation, only furrows and undulations are insinuated in the cephalic and thoracic region; in terms of general shape, it resembles the Tanaka stages 1–2 (8–12% development). Phenotype K (Fig. 6E): embryo at the stage of germ-band anlage, resembling the Tanaka stage 1 (4–8% development). Phenotype L (Fig. 6F): embryos developed in dorsal side of the egg instead of the ventral side; they show the shape of a Tanaka stage 5 (20% development), but with only three or four first abdominal segments recognizable, the caudal part being amorphous. The most frequent (41% of abnormal embryos) was phenotype H, followed by J and K (ca. 20% each), and then I and L (ca. between 8% and 11%) (Fig. 6G). The defect of abdomen malformation is common to all phenotypes, even in the less severe. Taken together, these results suggest that BgZelda depletion affects the elongation, segmentation, and the formation of the abdomen, and the curvature of the body, possibly resulting from defects in the dorso-ventral (DV) patterning.

We wondered whether the low percentage of embryos affected by the BgZelda RNAi could be due to the double-stranded RNA (dsRNA) used. Thus, we prepared an alternative dsRNA, dsZelda-a, and followed the same protocol of maternal RNAi, examining the embryos in ED4. From 10 oothecae of control females, we studied 240 embryos, whereas from 19 oothecae of dsZelda-a-treated females, we examined 473 embryos. The 240 embryos (100%) from controls showed the aspect of a normal ED4 embryo (Fig. 7A, B). From the 473 embryos from dsZelda-a-treated females, 312 (66%) were normal ED4 embryos (like those of Fig. 7A,B), whereas the remaining 161 embryos (34%) showed the phenotypes K, H, and L

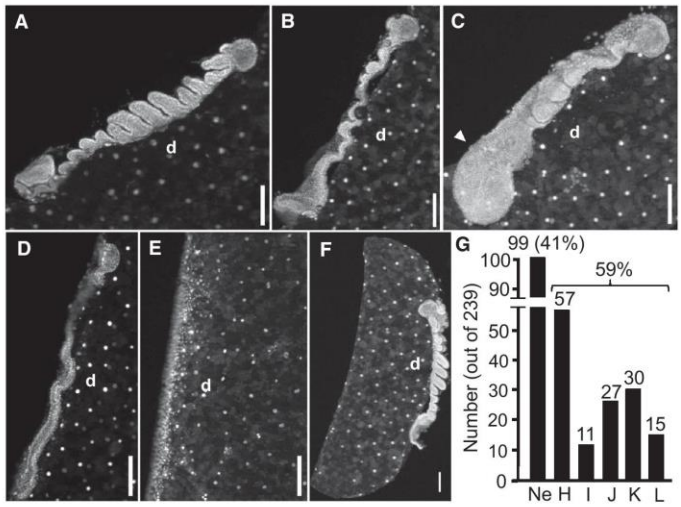


Fig. 6. Phenotypes observed in 4-day-old oothecae (embryos in stage ED4) from BgZelda-depleted *Blattella germanica*. (A) Normal ED4 embryo. (B) Phenotype H, embryo with seriously impaired segmentation, with the three thoracic segments delimited by undulated furrows, and the abdominal region amorphous. (C) Phenotype I, embryo with seriously impaired segmentation, twisted, shorter, and wider, and with the caudal region amorphous and swollen. (D) Phenotype J, embryo with no traces of segmentation. (E) Phenotype K, embryo at the stage of germ-band anlage. (F) Phenotype L, embryo developed in dorsal side of the egg instead of the ventral side. (G) Number of embryos showing the described phenotypes from a sample of 239 embryos studied. In A–F, the upper part of the panel corresponds to the cephalic part of the embryo, and 'd' indicates the dorsal part; the arrowhead indicates the growth zone, amorphous and swollen; scale bars in panels A–F: 200 μm.

(Fig. 7C–H). Regarding phenotype L (embryos developed in dorsal side of the egg instead of the ventral side), we find not only well-formed embryos in the wrong side (Fig. 7F) but also embryos in the same side, but with the development interrupted shortly after formation of the blastoderm (phenotype L′) (Fig. 7G). qRT-PCR measurements showed that the treatment with dsZelda-a reduced the Bg*zelda* mRNA levels by 92% on ED1 (Fig. 7I). These results indicate that even with this efficient transcript reduction, the penetrance was somewhat lower than that obtained with dsZelda.

Finally, although Bg*zelda* expression is very low in postembryonic stages (Fig. 3A), we aimed at assessing if it can play some role in these stages. We focused on the metamorphic transition between the fifth (N5) and last (N6) nymphal instar and from N6 to adult, as the most relevant from a developmental point of view. Thus, freshly emerged N5 (N5D0) was injected with 3 μg of dsZelda (treated nymphs, $n = 15$), whereas equivalent nymphs were injected with 3 μg of dsMock (controls, $n = 12$). All treated and control nymphs molted to morphologically normal N6, and subsequently to morphologically normal adults. The only difference between treated and controls was a slightly longer duration of the stages in the treated group. That of N5 was 6.0 ± 0 days in controls and 6.13 ± 0.35 days in treated, whereas that of N6 was 8.0 ± 0 days in controls and 8.20 ± 0.45 days in treated.

Depletion of BgZelda impairs the expression of genes involved in early embryo development

The window of maximal expression of Bg*zelda* is between ED0 and ED4, with an acute peak on ED1 (Fig. 3A), and the most characteristic phenotypes of BgZelda-depleted specimens were observed in early embryogenesis (Fig. 6). Therefore, we examined the expression of a number of genes involved in early embryo, at the ED2 stage. We first assessed that maternal treatment with dsZelda reduced the Bg*zelda* mRNA levels by 85% on ED1 and kept even lower levels in ED2 (Fig. 8A). Then, in relation to epigenetic mechanisms, we measured the expression of *DNA methyltransferase-1* (*Dnmt1*), involved in DNA methylation, and *DNA methyltransferase-2* (*Dnmt2*), involved in tRNA methylation [25], and results indicated that *dnmt1* was significantly down-regulated, whereas *dnmt2*

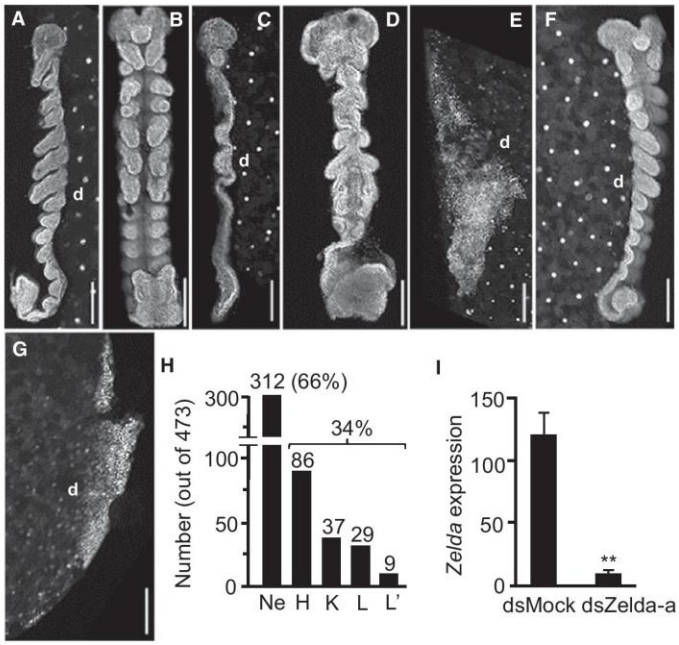


Fig. 7. Phenotypes observed in 4-day-old oothecae (embryos in stage ED4) from females of *Blattella germanica* treated with dsZelda-a. (A) Normal ED4 embryo in lateral view. (B) Normal ED4 embryo in ventral view. (C) Phenotype H in lateral view. (D) Phenotype H in ventral view. (E) Phenotype K in lateral view. (F) Phenotype L in lateral view. (G) Phenotype L′ in lateral view (H) Number of embryos showing the described phenotypes from a sample of 473 embryos studied. (I) Bg*zelda* transcript decrease resulting from maternal treatment with dsZelda-a, as measured on ED1; each value represents three biological replicates and is expressed as copies of mRNA per 1000 copies of BgActin-5c mRNA (mean ± SEM). Asterisks indicate statistically significant differences with respect to controls (** $P < 0.005$) calculated on the basis of Pairwise Fixed Reallocation Randomization Test implemented in REST [46]. In A–G, the upper part of the panel corresponds to the cephalic part of the embryo, and "d" indicates the dorsal part of the embryo; scale bars in panels A–G: 200 μm.

was not (Fig. 8B). Regarding microRNAs (miRNAs), it has been shown in *D. melanogaster* that DmZelda activates the expression of MIR-309 miRNAs in the MTZ context, which in turn eliminate maternal mRNAs [2,26]. We, thus, measured the expression of the precursors of the four variants of MIR-309 reported in *B. germanica* [27]. Results showed that, in all cases, the expression decreased in the BgZelda-depleted embryos at ED2 (Fig. 8C). Then, we studied a number of genes involved in very early embryogenesis and in the formation of the abdomen [28–31], including two maternal genes that are also expressed at the onset of zygotic activation: *hunchback* (*hb*) and *caudal* (*cad*), two gap genes: *Krüppel* (*Kr*) and *orthodenticle* (*otd*), two pair-rule genes: *even-skipped* (*eve*) and *fushi tarazu* (*ftz*), and one segment polarity gene: *wingless* (*wg*). Results indicated that the expression of *cad*, *Kr*, *otd*, *eve*, *ftz*, and *wg* was significantly reduced in BgZelda-depleted embryos (Fig. 8D). Moreover, previous studies have shown that maternal RNAi of *Broad Complex* (*BR-C* or *br*) provoked embryo defects similar to those of BgZelda depletion, in particular deficiently developed abdomens [32]. Therefore, we also included *BR-C* among the genes to be studied. However, results showed that BgZelda depletion did not affect *BR-C* mRNA levels (Fig. 8E). Although less frequent, another defect resulting from BgZelda depletion was the positioning of the embryo in the dorsal part of the egg. Thus, we studied the expression of genes involved in DV patterning, like *decapentaplegic* (*dpp*), *Toll* (*Tl*), *short-gastrulation* (*sog*), *dorsal* (*dl*), *snail* (*sna*), and *Notch* (*N*) [2,33]. Results showed that only the expression of *sna* resulted significantly reduced in BgZelda-depleted embryos (Fig. 8F).

Discussion

BgZelda has all the characteristic domains described by Ribeiro *et al.* [15], including the ZF-Novel that was left in doubt for *B. germanica* in that work (Fig. 1A). The above authors [15] found the same complete set of motifs in the Zelda orthologs of an archaeognathan (*Machilis hrabei*), two palaeopterans, (the odonatan *Ladona fulva* and the ephemeropteran *Ephemera danica*), and a basal neopteran (the isopteran *Zootermopsis nevadensis*). Then, one or more motifs are absent in different species of higher clades, whereas the set is very incomplete in the Zelda orthologs of noninsect hexapods and in crustaceans [15]. This suggests that the complete set of Zelda motifs is an ancestral condition in insects, which is still present in Archaeognatha, Palaeoptera, and basal Neoptera like *B. germanica*, while some of the motifs became secondarily lost in different lineages of Paraneoptera and Endopterygota.

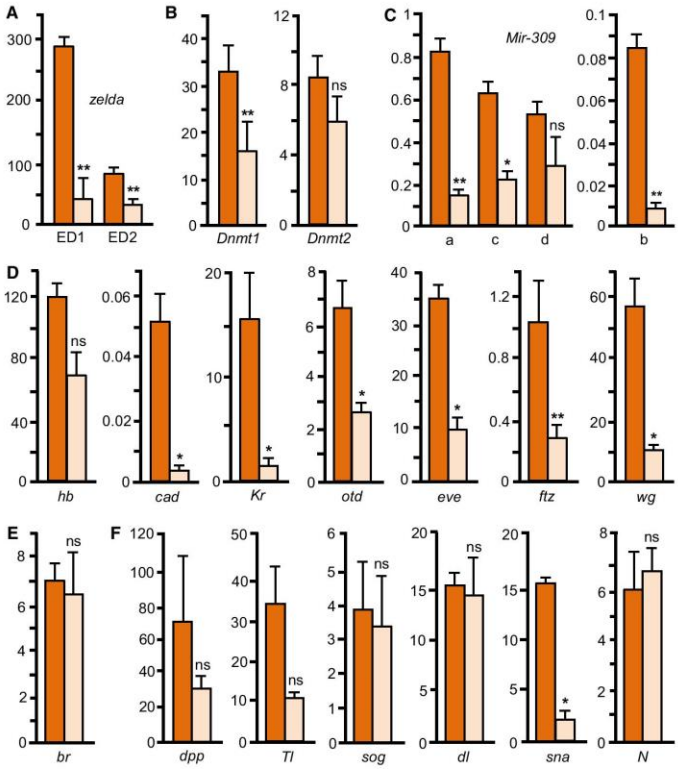


Fig. 8. Effects of BgZelda depletion on the expression of early embryogenesis genes in *Blattella germanica*. (A) Transcript decrease resulting from maternal RNAi of BgZelda, measured on ED1 and ED2. (B) Expression of *DNA methyltransferase-1* (*DNMT1*) and *DNA methyltransferase-2* (*DNMT2*). (C) Expression of the precursors of the three variants of *miR-309*: *miR-309a* (a), *miR-309b* (b), *miR-309c* (c), and *miR-309d* (d). (D) Expression of the early patterning genes *hb, cad, Kr, otd, eve, ftz,* and *wg* (E). Expression of *BR-C*. (F) Expression of the genes related to DV patterning *dpp, Tl, sog, dl, sna,* and *N*. The dark orange columns represent controls and the light orange columns BgZelda-depleted embryos. Each value represents three biological replicates and is expressed as copies of mRNA per 1000 copies of BgActin-5c mRNA (mean ± SEM). Asterisks indicate statistically significant differences with respect to controls (*$P < 0.05$; **$P < 0.0005$) calculated on the basis of Pairwise Fixed Reallocation Randomization Test implemented in REST [46]; ns: statistically not significant.

In the genome of *B. germanica*, we have found all the TAGteam heptamers to which DmZelda binds in *D. melanogaster*. The canonical one, CAGGTAG, is present at a similar relative number in these two and in other insect genomes (Fig. 2B). This suggests that, although within certain evolutionary constraints, the genome admits as many CAGGTAG motifs as its length allows. Moreover, the distribution along the genome, at least in *B. germanica*, is very regular, without accumulations or biases in given regions, in general (Fig. 2C; Fig. S1, Table S2). In *D. melanogaster*, DmZelda has been proposed as a molecule that functions as a transcription factor and as a modulator of chromatin accessibility. The above results support a role of Zelda during chromatin organization and nucleosome positioning [7,12,13,34]. Our data on the relative abundance of the CAGGTAG motif in the genomes of *B. germanica*, *D. melanogaster* and in a variety of other species, suggest that this role of pioneer factor that binds DNA regions through CAGGTAG-like motifs, making the chromatin accessible for transcription, would be general in insects.

Unhatched embryos from oothecae of dsZelda-treated females showed a diversity of malformations. The most frequent were related to abdomen development, followed by morphologically normal first nymphal instar but unable to hatch (Fig. 5). When BgZelda-depleted embryos were studied on ED4, a significant number of them showed the development interrupted around blastoderm stage or around segmentation. A few percentage of embryos was formed in the dorsal part of the egg instead the ventral part. A defect common to all these abnormal embryos was the incompletely formed abdomen (Fig. 6). This is consistent with our observation that expression of Bg*zelda* mainly localizes in the most caudal part of the embryonic abdomen (Fig. 4). The importance of Zelda for the development of the posterior zone and the development of the abdomen has been described in other short germ-band insects, like the beetle *T. castaneum* and the bug *R. prolixus* [15].

The low percentage of malformed phenotypes in the RNAi experiments performed with two different dsRNAs is difficult to explain. It is possible that it is

an intrinsic characteristic of maternal RNAi, at least of *B. germanica*, since experiments with other genes, such as the transcription factor *BR-C* [32] or several genes of the synthesis or the signaling of the juvenile hormone [35], also resulted in a low percentage of malformed phenotypes. In this context, it is plausible that the dsRNA or the siRNAs penetrate well in the peripheral oocytes of each ovary, but not in those that are in the inner part. It could also be that the observed low percentage of malformed phenotypes does not depend on the differential penetrance of the dsRNA-siRNAs, but is due to factors related to resilience properties of *zelda* gene in *B. germanica*. Finally, we must also consider that the maternal zygotic transition is not only dependent on *zelda* activity.

Our qRT-PCR transcript measurements in BgZelda-depleted embryos indicate that BgZelda promotes the expression of most of the early genes that we examined. The stimulation of *Dnmt1* expression (Fig. 8B) is interesting, as this points to a role of BgZelda in DNA methylation in *B. germanica*. *Dnmt1* is the only DNA methyltransferase gene found in *B. germanica*, and its expression pattern is similar to that of Bg*zelda*, with maximal levels between ED0 and ED2 [19]. The expression of MIR-309 miRNAs also depends on BgZelda (Fig. 8C), a function that is conserved in *D. melanogaster* [2,26]. In *B. germanica*, MIR-309 miRNAs peak on ED2 [27], that is, 1 day after the peak of Bg*zelda* (Fig. 3A), and we have postulated a role of these miRNAs in eliminating maternal mRNAs during the MZT, as occurs in *D. melanogaster* [36]. Then, all early embryo genes examined (gap, pair-rule, and segment polarity genes), were or tended to be down-regulated in BgZelda-depleted embryos (Fig. 8D). Previous reports had shown that the expression of most of these genes is impaired in DmZelda-deficient *D. melanogaster* [2,8], which suggests that the role of Zelda as a key activator of early zygotic genes [2,8,14] is present in the phylogenetically more basal *B. germanica*. We considered that the wrong positioning of the embryo in the dorsal part of the egg may have to do with genes regulating DV patterning, but the level of expression of the genes examined was not affected in BgZelda-depleted embryos, except that of *sna*, which was significantly down-regulated (Fig. 8F). Nevertheless, *sna* is not only involved in the DV patterning [2,33] but it is also required for coordinating mesoderm invagination during gastrulation [37,38], at least in *D. melanogaster*.

Bg*zelda* is mostly expressed in a narrow window in early embryogenesis, with an acute expression peak on ED1, within the MZT (Fig. 3A). This contrasts with the pattern observed in *D. melanogaster* and *T. castaneum*, where high expression is maintained beyond the MZT (Fig. 3C). These two latter species are holometabolan, but embryo develops through short germ-band mode in *T. castaneum* and long germ-band in *D. melanogaster*. Thus, the different pattern of expression of Bg*zelda* in *B. germanica* with respect to *D. melanogaster* and *T. castaneum* appear to relate with the type of metamorphosis. Hemimetabolan species, like *B. germanica*, develop the basic adult body structure during embryogenesis, whereas embryogenesis of holometabolan species, like *D. melanogaster* and *T. castaneum*, gives rise to a larva morphologically divergent with respect to the adult [20,39]. Thus, the expression of *zelda* beyond the MZT in *D. melanogaster* and *T. castaneum*, might be needed to activate successive gene sets needed to build, during embryogenesis, the derived vermiform larval morphology. Moreover, *zelda* expression and functions significantly extend beyond the embryo in holometabolan species (Fig. 3C), as shown by its role on patterning of imaginal disc-derived structures in *T. castaneum* [15], and in the genesis of neuroblasts in *D. melanogaster* larvae [40]. The expansion of *zelda* expression to mid and late embryogenesis that we observe in *D. melanogaster* and *T. castaneum*, with respect to *B. germanica*, might have been a factor contributing to the innovation of holometaboly in Endopterygota, from hemimetabolan ancestors.

Materials and methods

Insects

Adult females of *B. germanica* were obtained from a colony fed on Panlab dog chow (Panlab S.L.U, Barcelona, Spain) and water *ad libitum*, and reared in the dark at 29 ± 1 °C and 60–70% relative humidity. Freshly ecdysed adult females were selected and used at appropriate ages. Mated females were used in all experiments, and the presence of spermatozoa in the spermatheca was assessed in all experiments. Prior to injection treatments, dissections, and tissue sampling, the insects were anesthetized with carbon dioxide.

Transcriptomic and genomic data

We obtained the transcriptome-based pattern of expression of *zelda* in *B. germanica*, *D. melanogaster*, and *T. castaneum*. Those of *B. germanica* and *D. melanogaster* were identical to those previously obtained by Ylla *et al.* [19], who precisely describe the different stage-libraries used. The RNA-seq dataset of *B. germanica* and *D. melanogaster* are accessible at GEO: GSE99785 [19] and GEO: GSE18068 [41,42], respectively. The RNA-seq dataset of *T. castaneum* embryogenesis used (GEO: GSE63770)

comprises eight libraries from four developmental stages (two replicates each) covering nonfertilized eggs (NFE), and three sequential embryo stages: 8–16 h, 16–24 h, 24–48 h [22]. In addition, we studied a RNA-seq library from *T. castaneum* adult females [43] available at SRA: SRX021963. The TAGteam heptamers were examined along the genome assemblies and its complementary sequences using custom-made Python scripts. To calculate the correlation between the genome length and the genome size, we considered the number of heptamers found in each genome and the genome length, excluding ambiguous bases 'N', and using the R 'cor' function. Additionally, we calculated the density as the number of canonical heptamer per scaffold length (Table S2), and plotted setting the bin to 10 kb at Fig. S1. The complete list of genomes used and their accession can be found in Table S1.

RNA interference

The detailed procedures for RNAi assays have been described previously [44]. Primers used to prepare BgZelda dsRNA are described in Table S3. The sequence was amplified by PCR and then cloned into a pST-Blue-1 vector. A 307-bp sequence from *Autographa californica* nucleopolyhedrosis virus (Accession number K01149.1) was used as control dsRNA (dsMock); primers used to synthetize dsMock are also described in Table S3. The dsRNAs were prepared as reported elsewhere [44]. A volume of 1 μL of the dsRNA solution (3 $\mu g \cdot \mu L^{-1}$) was injected into the abdomen of 1-day-old adult females, with a 5-μL Hamilton microsyringe. Control specimens were treated at the same age with the same dose and volume of dsMock.

RNA extraction and reverse transcription to cDNA

We performed a total RNA extraction from oothecae using the RNeasy Plant minikit (QIAGEN, Hilden, Germany) in the case of early oothecae (since NFE until 4 days after ootheca formation, AOF) and GenElute Mammalian Total RNA Miniprep kit (Sigma-Aldrich, Madrid, Spain) in the case of later oothecae (since 6 days AOF to 16 days AOF). In both cases, all the volume extracted was lyophilized in the freeze-dryer FISHER-ALPHA 1–2 LDplus and then resuspended in 8 μL of miliQ H_2O. For mRNA and miRNA precursor quantification, these 8 μL were treated with DNase I (Promega, Madison, WI, USA) and reverse transcribed with first Strand cDNA Synthesis Kit (Roche, Barcelona, Spain) and then random hexamer primers (Roche).

Quantification of mRNA levels by qRT-PCR

Quantitative RT-PCR was carried out in an iQ5 Real-Time PCR Detection System (Bio-Lab Laboratories, Madrid, Spain), using SYBR®Green (iTaq ™Universal SYBR® Green Supermix; Applied Biosystems, Foster City, CA, USA). Reactions were triplicate, and a template-free control was included in all batches. Primers used to detect mRNA levels studied here are detailed in Table S3. We have validated the efficiency of each set of primers by constructing a standard curve through three serial dilutions. In all cases, levels of mRNA were calculated relative to BgActin-5c mRNA (accession number AJ862721). Results are given as copies of mRNA of interest per 1000 or per 100 copies of BgActin-5c mRNA.

Microscopy

Oothecae were detached from the female abdomen by gentle pressure or obtained directly because the animal left them. Each ootheca was opened and the embryos were dechorionated and individualized. Then these embryos were directly observed under the stereo microscope Carl Zeiss – AXIO IMAGER.Z1 (Oberkochen, Germany). For 4′,6-diamidino-2-phenylindole (DAPI) staining, after 10 min in a water bath at 95 °C, each ootheca was opened and the embryos dechorionated and individualized. Between 12 and 24 embryos per ootheca, chosen from the central part, were dissected for staining. These embryos were fixed in 4% paraformaldehyde in PBS for 1 h, then washed with PBS 0.3% Triton (PBT) and incubated for 10 min in 1 $mg \cdot mL^{-1}$ DAPI in PBT. They were mounted in Mowiol (Calbiochem, Madison, WL, USA) and observed with the fluorescence microscope Carl Zeiss – AXIO IMAGER.Z1.

In situ hybridization studies

Fluorescent in situ hybridization protocols were carried out according to Shippy *et al.* [45] with some modifications. Briefly, a 310-bp Bg*zelda* cDNA fragment (indicated in Table S3) was used as a template for transcription. Sense and antisense probes were labeled with DIG RNA Labeling Mix (Roche) and detected with Anti-Digoxigenin-AP, Fab fragments (Roche) using SIGMAFAST™ Fast Red as a substrate (Sigma). Before hybridization, entire embryos were fixed with Solution A (8% Formaldehyde, 0.1% TritonX, 0.1% EGTA in PBS 0.2 M), dehydrated with methanol and stored at −20 °C. Embryos were mounted in Mowiol (Calbiochem) and observed with the fluorescence microscope Carl Zeiss-AXIO IMAGER.Z1.

Statistical analyses of qRT-PCR

In all experiments, to test the statistical significance between treated and control samples it has been used the Relative Expression Software Tool (REST), which evaluates the significance of the derived results by Pairwise Fixed Reallocation Randomization Test [46].

Acknowledgements

This work was supported by the Spanish Ministry of Economy and Competitiveness (grants CGL2012–36251 and CGL2015–64727-P to XB and predoctoral fellowship to AV-A) and the Catalan Government (grant 2017 SGR 1030 to XB). It also received financial assistance from the European Fund for Economic and Regional Development (FEDER funds). Natalia Llonga helped to find TAGteam motifs associated with the genes examined in the expression studies. The suggestions of four anonymous reviewers contributed to considerably improve the manuscript.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

XB conceived the study and wrote the final version of the manuscript. GY performed the bioinformatics analyses. AV-A performed the RNAi experiments and the qRT-PCR measurements as well as drafted a first version of the manuscript. XB, AV-A, and GY discussed and interpreted the results, and commented on the manuscript.

References

1 Tadros W & Lipshitz HD (2009) The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–3042.

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

3 Lee MT, Bonneau AR & Giraldez AJ (2014) Zygotic genome activation during the maternal-to-zygotic transition. *Annu Rev Cell Dev Biol* **30**, 581–613.

4 Jukam D, Shariati SAM & Skotheim JM (2017) Zygotic genome activation in vertebrates. *Dev Cell* **42**, 316–332.

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

6 ten Bosch JR, Benavides JA & Cline TW (2006) The TAGteam DNA motif controls the timing of *Drosophila* pre-blastoderm transcription. *Development* **133**, 1967–1977.

7 Harrison MM, Li X-Y, Kaplan T, Botchan MR & Eisen MB (2011) Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet* **7**, e1002266.

8 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**, e1002339.

9 Li X-Y, Harrison MM, Villalta JE, Kaplan T & Eisen MB (2014) Establishment of regions of genomic activity during the *Drosophila* maternal to zygotic transition. *Elife* **3**, e03737.

10 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* **25**, 1715–1726.

11 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.

12 Hug CB, Grimaldi AG, Kruse K & Vaquerizas JM (2017) Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169**, 216–228.e19.

13 Dufourt J, Trullo A, Hunter J, Fernandez C, Lazaro J, Dejean M, Morales L, Nait-Amer S, Schulz KN, Harrison MM *et al.* (2018) Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. *Nat Commun* **9**, 5194.

14 Hamm DC, Bondra ER & Harrison MM (2015) Transcriptional activation is a conserved feature of the early embryonic factor Zelda that requires a cluster of four zinc fingers for DNA binding and a low-complexity activation domain. *J Biol Chem* **290**, 3508–3518.

15 Ribeiro L, Tobias-Santos V, Santos D, Antunes F, Feltran G, de Souza MJ, Aravind L, Venancio TM & Nunes da Fonseca R (2017) Evolution and multiple roles of the Pancrustacea specific transcription factor zelda in insects. *PLoS Genet* **13**, e1006868.

16 Hamm DC, Larson ED, Nevil M, Marshall KE, Bondra ER & Harrison MM (2017) A conserved maternal-specific repressive domain in Zelda revealed by Cas9-mediated mutagenesis in *Drosophila melanogaster*. *PLoS Genet* **13**, e1007120.

17 Pires CV, Freitas FC, Cristino AS, Dearden PK & Simões ZLP (2016) Transcriptome analysis of honeybee (*Apis mellifera*) haploid and diploid embryos reveals early zygotic transcription during cleavage. *PLoS One* **11**, e0146447.

18 Arsala D & Lynch JA (2017) Ploidy has little effect on timing early embryonic events in the haplo-diploid wasp *Nasonia*. *Genesis* **55**, e23029.

19 Ylla G, Piulachs MD & Belles X (2018) Comparative transcriptomics in two extreme neopterans reveals general trends in the evolution of modern insects. *iScience* **4**, 164–179.
20 Belles X (2011) Origin and evolution of insect metamorphosis. In *Encyclopedia of Life Sciences (ELS)*, pp. 1–11. John Wiley & Sons, Ltd, Chichester.
21 Giannios P & Tsitilou SG (2013) The embryonic transcription factor Zelda of *Drosophila melanogaster* is also expressed in larvae and may regulate developmentally important genes. *Biochem Biophys Res Commun* **438**, 329–333.
22 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.
23 Song X, Huang F, Liu J, Li C, Gao S, Wu W, Zhai M, Yu X, Xiong W, Xie J *et al.* (2017) Genome-wide DNA methylomes from discrete developmental stages reveal the predominance of non-CpG methylation in *Tribolium castaneum*. *DNA Res* **24**, 445–457.
24 Tanaka A (1976) Stages in the embryonic development of the German cockroach, *Blattella germanica* Linné (Blattaria, Blattellidae). *Kontyû, Tokyo* **44**, 1703–1714.
25 Lyko F (2018) The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet* **19**, 81–92.
26 Fu S, Nien C-Y, Liang H-L & Rushlow C (2014) Co-activation of microRNAs by Zelda is essential for early *Drosophila* development. *Development* **141**, 2108–2118.
27 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 Genom* **18**, 774.
28 Chesebro JE, Pueyo JI & Couso JP (2013) Interplay between a Wnt-dependent organiser and the Notch segmentation clock regulates posterior development in *Periplaneta americana*. *Biol Open* **2**, 227–237.
29 St Johnston D & Nüsslein-Volhard C (1992) The origin of pattern and polarity in the *Drosophila* embryo. *Cell* **68**, 201–219.
30 Mito T, Kobayashi C, Sarashina I, Zhang H, Shinahara W, Miyawaki K, Shinmyo Y, Ohuchi H & Noji S (2007) Even-skipped has gap-like, pair-rule-like, and segmental functions in the cricket *Gryllus bimaculatus*, a basal, intermediate germ insect (Orthoptera). *Dev Biol* **303**, 202–213.
31 Jaeger J & Reinitz J (2012) *Drosophila* blastoderm patterning. *Curr Opin Genet Dev* **22**, 533–541.
32 Piulachs M-D, Pagone V & Belles X (2010) Key roles of the Broad-Complex gene in insect embryogenesis. *Insect Biochem Mol Biol* **40**, 468–475.
33 Lynch JA & Roth S (2011) The evolution of dorsal–ventral patterning mechanisms in insects. *Genes Dev* **25**, 107–118.
34 Zaret KS & Carroll JS (2011) Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**, 2227–2241.
35 Fernandez-Nicolas A & Belles X (2017) Juvenile hormone signaling in short germ-band hemimetabolan embryos. *Development* **144**, 4637–4644.
36 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**, 501–506.
37 Leptin M (1991) Twist and snail as positive and negative regulators during *Drosophila* mesoderm development. *Genes Dev* **5**, 1568–1576.
38 Kosman D, Ip YT, Levine M & Arora K (1991) Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. *Science* **254**, 118–122.
39 Truman JW & Riddiford LM (1999) The origins of insect metamorphosis. *Nature* **401**, 447–452.
40 Reichardt I, Bonnay F, Steinmann V, Loedige I, Burkard TR, Meister G & Knoblich JA (2018) The tumor suppressor Brat controls neuronal stem cell lineages by inhibiting Deadpan and Zelda. *EMBO Rep* **19**, 102–117.
41 Celniker SE, Dillon LAL, Gerstein MB, Gunsalus KC, Henikoff S, Karpen GH, Kellis M, Lai EC, Lieb JD, MacAlpine DM *et al.* (2009) Unlocking the secrets of the genome. *Nature* **459**, 927–930.
42 modENCODE Consortium S, Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, *et al.* (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330**, 1787–1797.
43 Altincicek B, Elashry A, Guz N, Grundler FMW, Vilcinskas A & Dehne H-W (2013) Next generation sequencing based transcriptome analysis of septic-injury responsive genes in the beetle *Tribolium castaneum*. *PLoS One* **8**, e52004.
44 Ciudad L, Piulachs M-D & Belles X (2006) Systemic RNAi of the cockroach vitellogenin receptor results in a phenotype similar to that of the *Drosophila* yolkless mutant. *FEBS J* **273**, 325–335.
45 Shippy TD, Coleman CM, Tomoyasu Y & Brown SJ (2009) Concurrent in situ hybridization and antibody staining in red flour beetle (Tribolium) embryos. *Cold Spring Harb Protoc* **2009**, pdb.prot5257.
46 Pfaffl MW, Horgan GW & Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative

expression results in real-time PCR. *Nucleic Acids Res* **30**, e36.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Number of CAGGTAG motifs within 10 kb windows along each scaffold of the genome sequence of *Blattella germanica*.

Table S1. Insect genomes used for correlating the number of canonical Zelda-binding heptamer CAGGTAG and the genome size.

Table S2. Scaffolds of the *Blattella germanica* genome and CAGGTAG motifs in them.

Table S3. Primers used to measure expression levels by qRT-PCR, to prepare the dsRNAs for RNAi experiments, and to prepare the riboprobe for *in situ* hybridization studies.

Due to the length of some datasets, the additional supporting information is not included here. It can be downloaded from:

https://febs.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Ffebs.14856&file=febs14856-sup-0001-SupInfo.zip

3.3. Maternal E93 activates gene expression in the zygote and is crucial in early hemimetabolan embryogenesis

Alba Ventos-Alfonso [a], Ana Fernandez-Nicolas [a,b], Guillem Ylla[a,c], Xavier Belles[1,a]

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

[b] Present address: Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, USA

[c] Present address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

[1] Corresponding author. Email: xavier.belles@ibe.upf-csic.es

INTRODUCTION

Metamorphosis has been one of the most successful innovations in insect evolution. It originated in early-middle Devonian, some 400 Mya (Misof et al., 2014; Wang et al., 2016), associated with the practically contemporaneous emergence of wings (Belles, 2019). Therefore, from apterous and ametabolan ancestors, emerged a new type of insect with direct, or hemimetabolan, metamorphosis, whose basic adult body structure is formed during embryogenesis. Thus, the hatching nymphs resemble miniature adults, and only at the end

of the nymphal period, the insect metamorphose into a functional adult, fully winged and sexually competent, and stops molting. Subsequently, during the early Carboniferous, some 350 Mya (Misof et al., 2014; Wang et al., 2016), took place the evolutionary transition from hemimetaboly to holometaboly, or indirect metamorphosis. In holometabolans, embryo development gives rise to a larva that is morphologically divergent to the adult. At the end of larval period, the insect molt to the pupal stage, a kind of intermediate between the larva and the adult, and then to the fully winged and sexually competent adult, which does not molt anymore (see Belles, 2020 for a comprehensive review).

The above definition of hemimetaboly and holometaboly already tells us that embryonic development will determine what type of life cycle and post-embryonic metamorphosis will follow. However, metamorphosis researchers, perhaps fascinated by the beauty of the post-embryonic changes, rarely concentrate their attention on embryogenesis. Most modern studies with observations on a molecular scale, have focused on the Juvenile hormone (JH), which is a repressor of metamorphosis (Jindra et al., 2013; Jindra et al., 2015; Truman and Riddiford, 2007), and the transcription factor Broad complex (Br-C), which has a fundamental role in the growth of wings in hemimetabolan nymphs (Erezyilmaz et al., 2006; Huang et al., 2013; Konopová and Jindra, 2008), and in the formation of the pupa in holometabolans (Karim et al., 1993; Kiss et al., 1988; Konopová and Jindra, 2008; Parthasarathy et al., 2008; Zhou et al., 1998). The studies on JH and Br-C in embryogenesis were triggered

by previous hypotheses considering that they would have been instrumental in the emergence of the larva and holometabolan metamorphosis, mediated by changes in the embryogenesis (Erezyilmaz et al., 2006; Erezyilmaz et al., 2009; Truman and Riddiford, 1999).

Molecular-scale studies in hemimetabolan embryogenesis have been carried out with maternal RNAi. In the case of JH, studies on the cockroach *Blattella germanica* have shown that the interference of JH biogenesis and signaling pathways trigger a variety of defects, mainly associated with cuticle maturation and nymph hatchability (Fernandez-Nicolas and Belles, 2017). Analogous maternal RNAi experiments in the milkweed bug *Oncopeltus fasciatus* (Erezyilmaz et al., 2009) and in *B. germanica* (Piulachs et al., 2010), have shown that Br-C depletion in the embryo causes a series of defects, many of them associated with the development of the abdomen as well as with nymph hatching. In holometabolan species, clearer (and more surprising) results have been obtained using genome editing approaches. Thus, in the silkworm *Bombyx mori*, the suppression of JH signaling, as well as that of Br-C, did not cause any serious defect during embryogenesis, at most a reduction in larval hatchability (Daimon et al., 2015).

In *B. germanica*, we have also studied the functions of Zelda in embryogenesis (Ventos-Alfonso et al., 2019). Zelda plays a major role in the activation of the zygotic genome, during the maternal-to-zygotic transition (MZT) (Harrison et al., 2011; Liang et al., 2008;

Nien et al., 2011) by increasing the accessibility of chromatin, allowing the transcription of multiple genes, as shown in *Drosophila melanogaster* (Dufourt et al., 2018; Foo et al., 2014; Schulz et al., 2015). We focused on Zelda after examining its expression in a transcriptomic series of *B. germanica* that covers the entire ontogenesis (Ylla et al., 2018). We observed that the expression of *zelda* concentrates in very early embryogenesis, as maternal mRNA. On the other hand, studying an equivalent transcriptomic series of the holometabolan *D. melanogaster*, we observed that *zelda* is expressed practically during the entire embryogenesis ((Ylla et al., 2018), see also (Staudt et al., 2006)). Quantitative RT-PCR measurements confirmed the transcriptomic pattern in *B. germanica*, and maternal RNAi reveled that Zelda has key functions in the embryo development of the cockroach, especially during the elongation of the abdomen (Ventos-Alfonso et al., 2019).

The comparative analysis of the RNA-seq data of Ylla et al. (2018) afforded another interesting gene: *Eip93F*, more familiarly termed *E93*. It was shown to be expressed only in pupae and adults in *D. melanogaster*, while in *B. germanica* it was expressed in the last nymphal instar, but also at the early stages of embryogenesis. Thus, we studied the role of E93 in embryo development of *B. germanica*, and the results revealed that maternal E93 activates gene expression in the zygote and is crucial in the early hemimetabolan embryogenesis. Moreover, expression of *E93* in the embryo appears to be characteristic of hemimetabolan species, as it is practically

absent in holometabolans. This suggests that the loss of embryonic expression of *E93* may have been instrumental in the evolution of holometaboly.

RESULTS

Maternal E93 is required for early embryogenesis

Real-time quantitative reverse transcription PCR (qRT-PCR) measurements confirmed our previous observations based on RNA-seq data that maximal levels of *E93* transcripts are observed at the very beginning of embryogenesis, with the peak centered on day 1 of embryo development (ED1) (Figure 3.3.1A). Then, the levels dramatically decrease and keep at low values during all embryogenesis. This pattern suggests that E93 plays a specific role in early embryogenesis. To address this hypothesis, we used maternal RNAi. Five-day-old adult females (AdD5) were injected with 3 µg of dsRNA targeting *E93* mRNA (dsE93). Control females received the same dose of unspecific dsRNA (dsMock). Then, all females were allowed to mate. *E93* expression in embryos from dsE93-treated insects was significantly lower (71%) than in controls, which indicates that maternal RNAi was efficient (Figure 3.3.1B). All control females (n=40) formed an ootheca on day 8 of the adult stage (AdD8), which contained viable embryos that became first instar nymphs (between 35-40 emerged nymphs per oothecal) 18 days later. Females treated with dsE93 (n=36) also formed an ootheca on AdD8, and 18 of them (50%) gave first instar

nymphs 18 days later (between 35-40 emerged nymphs per oothecal). However, 10 of them (27.8%) dropped the oothecae between ED2-ED3 (12-17% embryogenesis), whereas 8 of them (22.2%) transported the ootheca beyond day 18, dropping it between ED19-ED21. No nymphs hatched from these 18 oothecae.

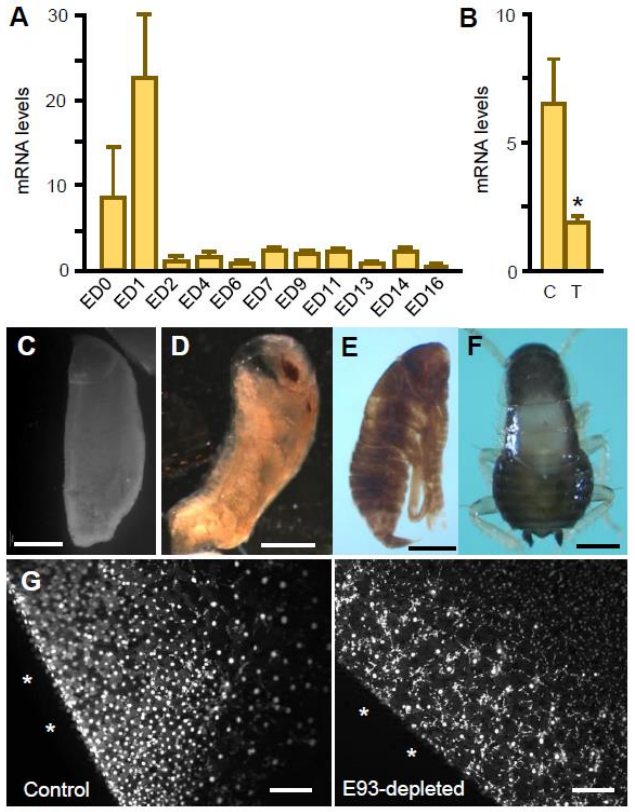


Figure 3.1.1. Effects of *E93* depletion on embryo development in *Blattella germanica*. *(A)* Expression of *E93* in embryos of different ages from day 0 (ED0) to day 16 (ED16). *(B)* Expression of *E93* after maternal RNAi at ED0 in control (C: dsMock) and treated (T: dsE93). *(C)* Phenotype PA resulting from *E93* depletion: embryos that interrupted development around the formation of germ-band anlage. *(D)* Phenotype PB: embryos that interrupted development between Tanaka stages 10 and 15. *(E)* Phenotype PC: apparently well-formed nymphs, with dark cuticle. *(F)* Phenotype PD: embryos that although they are similar to

On embryo day 19 (ED19) we examined 278 embryos from the 8 oothecae formed by dsE93-treated females that did not hatch on day 18 (ED18). Seventy-one of these embryos (25.5%) had interrupted development in a pre-blastoderm stage, and only white yolk was observed, which corresponds to stage 1-2 of embryogenesis defined by Tanaka (1976); we named this phenotype PA (Figure 3.3.1C). Forty embryos (14.4%) had interrupted development between Tanaka stages 10 and 15, showing a vermiform aspect, with poorly defined segments and appendages, and with diverse malformations such as imperfectly sealed dorsal closure, a reduced abdomen or imperfect eyes; we named this phenotype PB (Figure 3.3.1D). Fifty-four embryos (19.4%) were apparently well-formed nymphs, but with a featured and intensely sclerotized cuticle; we named this phenotype PC (Figure 3.3.1E). Finally, we named phenotype D to 113 embryos (40.6%) that were nymphs practically indistinguishable from control pre-hatched nymphs. If the chorion was artificially broken, then most phenotype PD specimens showed the shape of naturally hatched first instar nymphs (Figure 3.3.1F).

Figure 3.1.1 (continued) pre-hatched nymphs, they could not hatch, but if the chorion was artificially broken, they looked like naturally eclosed first instar nymphs, as shown in the picture. *(G)* Accumulation of energids in the ventral side of the embryo (white asterisks) in control (dsMock-treated) and E93-depleted specimens, in ED2. In *(A)* and *(B)*, each qRT-PCR expression measurement represents three biological replicates, and results are expressed as copies of the examined transcript per 1000 copies of BgActin-5c mRNA; data are represented as the mean ± SEM. In *(B)*, the asterisk indicates statistically significant differences with respect to controls (p-value < 0.05), calculated on the basis of the REST (Pfaffl et al, 2002). Scale bars in *(C)* to *(F)*: 500 µm; in *(G):* 200 µm.

The dropping of the ootheca at early embryogenesis (ED2-ED3) in dsE93-treated females is interesting as it happens just after the maximum levels of *E93* mRNA, which occurs at ED1. Thus, we studied embryos from oothecae that were dropped on ED2, compared with ED2 embryos from control females. We analyzed 350 embryos from control females (control embryos) and 235 embryos from dsE93-treated females (E93-depleted embryos). Control embryos were between Tanaka stages 1-2, showing a high density of energids accumulated on the ventral side of the egg, where they will contribute to the formation of the germ-band anlage (Figure 3.3.1G, left panel). Although E93-depleted embryos were also around Tanaka stages 1-2, in all embryos examined, the accumulation of energids in the ventral side of the egg was lower than in controls, and their distribution was more heterogeneous, forming discrete groups separated from each other (Figure 3.3.1G, right panel).

Depletion of *E93* modifies gene expression at the MZT and prevents the decline of *zelda* expression

We presumed that the critical problems observed in the development of E93-depleted embryos in ED2 might be associated to the MZT, which takes place around this stage (Ventos-Alfonso et al., 2019) during which early zygotic genes become upregulated, while maternal mRNAs vanish. Examining the differential expression of genes around the MZT in the transcriptomes of *B.*

germanica (Ylla et al., 2018), we found examples of both categories of genes. We obtained the profiles of *Krüppel* (*Kr*), *caudal* (*cad*) and *even-skipped* (*eve*), genes whose expression increases at the MZT (Figure 3.3.2A). Then, we measured the expression of *Kr*, *cad* and *eve* in ED2 in control and in E93-depleted embryos, observing that the corresponding mRNA levels were significantly lower in E93-depleted embryos than in controls (Figure 3.3.2B). This suggests that the activation of the zygotic genes in E93-depleted embryos was impaired. On the other hand, the same differential expression examination in the *B. germanica* transcriptomes (Ylla et al., 2018), afforded the profiles of *orthodenticle* (*otd*), *nanos* (*nos*) and *odd skipped* (*odd*), as examples of genes whose mRNA levels decline at the MZT (Figure 3.3.2C). The mRNA levels of *otd*, *nos* and *odd* were much higher in E93-depleted embryos than in controls (Figure 3.3.2D), which suggests that the clearance of maternal mRNAs in E93-depleted embryos was impaired.

Given the important functions of Zelda as an activating factor of the zygotic genome, we measured the mRNA levels of *zelda* in ED2, finding that they were higher in E93-depleted embryos than in controls (Figure 3.3.2E). In *B. germanica* embryos, *zelda* mRNA is of maternal origin, and their levels sharply decline in ED2 (Ventos-Alfonso et al., 2019). Thus, the observations suggest that this decline is regulated by the same mechanism that operates in *otd*, *nos* and *odd*, a mechanism that would have been impaired in E93-depleted embryos.

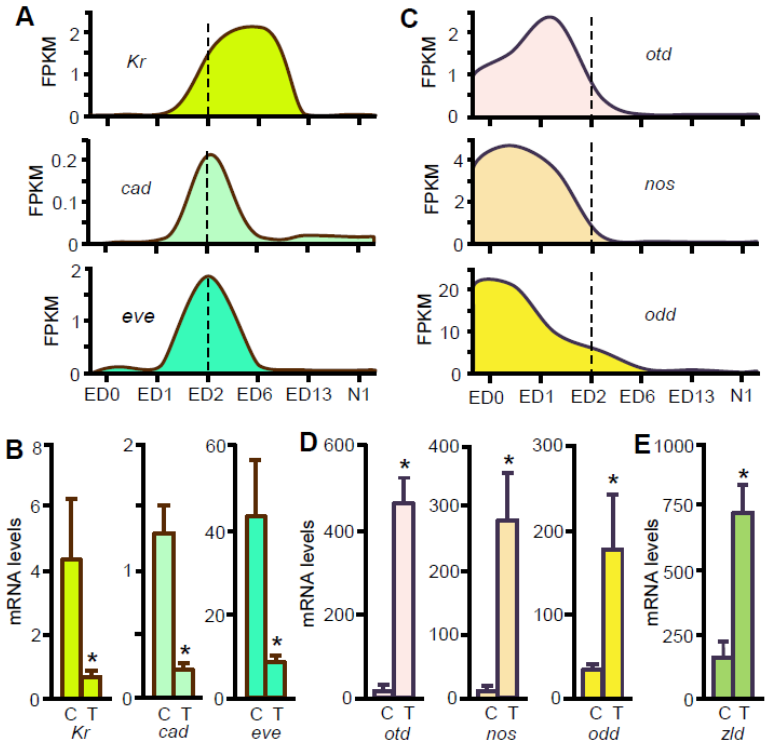


Figure 3.3.2. Effects of *E93* depletion on gene expression in *Blattella germanica* embryo. (*A*) Transcriptomic pattern of genes that become upregulated at the MZT: *Kruppel* (*Kr*), *caudal* (*cad*), and *even skipped* (*eve*). (*B*) Expression of *Kr*, *cad* and *eve* after maternal RNAi of *E93*. (*C*) Transcriptomic pattern of genes that become downregulated at the MZT: *orthodenticle* (*otd*), *nanos* (*nos*), and *odd skipped* (*odd*). (*D*) Expression of *otd*, *nos* and *odd* after maternal RNAi of *E93*. In (*A*) and (*C*), expression is expressed in FPKM and, in addition to embryo stages ED0 to ED13, it was also measured in freshly emerged first instar nymph (N1); the dashed line indicates the stage ED2, around which takes place the MZT. (*E*) Expression of *zelda* (*zld*) after maternal RNAi of *E93*. In (*B*) (*D*) and (*E*), qRT-PCR measurements were carried out in ED2, in control (C, dsMock-treated) and treated (T, dsE93-treated). Each qRT-PCR expression measurement represents three biological replicates, and results are expressed as copies of the examined transcript per 1000 copies of BgActin-5c mRNA. Data are represented as the mean ± SEM. The asterisk indicates statistically significant differences with respect to controls (p-value < 0.05), calculated on the basis of the REST (Pfaaffl et al, 2002).

mir-309 miRNAs would contribute to remove *zelda* transcripts in the MZT

Obvious candidates to regulate the reduction of maternal mRNAs are the *miR-309* family of miRNAs, given their role as scavengers of these mRNAs in the MZT (Bushati et al., 2008). In the *B. germanica* genome, the *miR-309* cluster includes four miRNA precursors: *miR-309a*, *miR-309b*, *miR-309c* and *miR-309d* (Ylla et al., 2017). The *miR-309a* precursor gives the mature miR-309A: TCACTGGAAAGGCAATATCATT. Both, the *miR-309b* and *miR-309c* precursors give the same mature miR-309B: TCACTGGGAAGACATTATCGT. Finally, the *miR-309d* precursor gives the mature miR-309D: TCACCGGAAGGAATCCACAGT, which is significantly divergent with respect to the other two, where even the seed region (CACCGG) differs from the canonical seed region of *miR-309* (CACTGG) that is conserved in all insects (Ylla et al., 2017). Therefore, miR-309D was not further studied in the present work.

From the small RNA libraries prepared and studied by Ylla et al. (2017), we obtained the expression profiles of *miR-309a*, *miR-309b* and *miR-309c* precursors during embryogenesis. The three profiles were practically coincident, showing a peak on ED2 (Figure 3.3.3A). Then we measured the levels of mature miR-309A and miR-309B in E93-depleted embryos in ED2, observing that the levels of both were reduced in comparison with the controls (Figure 3.3.3B). This suggests that E93 promotes the production of *miR-309*

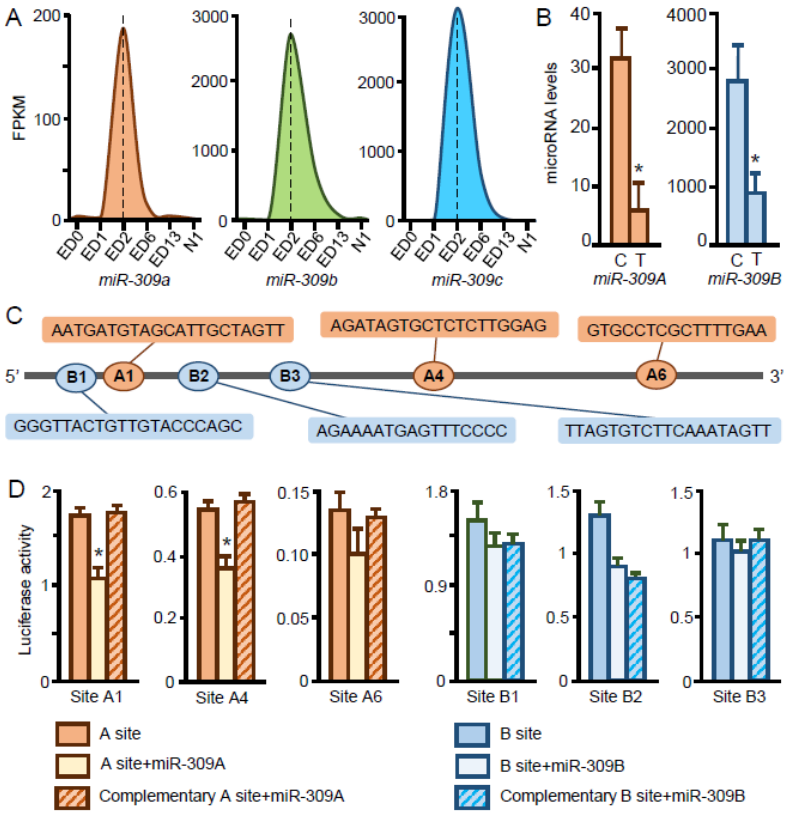


Figure 3.3.3. Expression of *miR-309* microRNAs during the embryogenesis of *Blattella germanica*, effects of E93 depletion, and miR-309 binding sites in the 3'UTR of *zelda* mRNA. (*A*) Transcriptomic pattern of *miR-309a*, *miR-309b* and *miR-309c* and corresponding mature miRNAs during embryogenesis; expression is expressed in FPKM and, in addition to embryo stages ED0 to ED13, it was also measured in freshly emerged first instar nymph (N1); the dashed line indicates the stage ED2, around which takes place the MZT. (*B*) Levels of miR-309A and miR-309B after maternal RNAi of *E93*; measurements were carried out in ED2, in control (C, dsMock-treated) and treated (T, dsE93-treated), and results are expressed as copies of the examined microRNA per 1000 copies of U6 miRNA. (*C*) Scheme of the *zelda* mRNA 3'UTR showing the binding sites for miR-309A (A1, A4 and A6 sites) and miR-309B (B1, B2 and B3 sites) predicted with $\Delta G < -15$ Kj mol^{-1} (see Table 3.3.S1 and Supplementary Figure 2). (*D*) Luciferase assays testing the sites A (A1, A4 and A6) and B (B1, B2 and B3), with the respective

miRNAs, and that these miRNAs would regulate the decline of *zelda* transcripts (as well as those of *otd*, *nos*, *odd*). To further support this conjecture, we searched binding sites for miR-309A and miR-309B in the 3'UTR region of *zelda* mRNA. Using the algorithms StarMir, miRanda, PITA, RNA22 and RNAhybrid, we predicted seven binding sites for miR-309A and four for miR-309B (SI Appendix, Figure 3.3.S1 and Table 3.3.S1). Six of these putative sites having a $\Delta G < -15$ Kj mol^{-1} (Figure 3.3.3C) were selected to test their functionality. Accordingly, we examined these six sites A and B with a conventional luciferase reporter assay conducted in *Drosophila* S2 cells and using synthetic miR-309A and B mimics, respectively. As a negative control, we used the complementary sequences of the A and B sites. Luciferase activity significantly decreased only for the sites A1 and A4 under the presence of miR-309A. This indicates that these sites are functional for miR-309A (Figure 3.3.3D).

The whole data suggest that E93 contributes to the expression decline of *zelda* at ED2, as well as to the expression of miR-309 miRNAs at the same stage. Moreover, the finding of functional

Figure 3.3.3 (continued). miR-309A and miR-309B mimics; as negative controls, the complementary sequences of miR-309A and miR-309B sites were used; each measurement represents 3 biological replicates and results are given as *Renilla* luciferase activity relative to that of *Firefly* luciferase. In (*B*) and (*D*), each measurement represents three biological replicates, and data are represented as the mean ± SEM; the asterisk indicates statistically significant differences with respect to controls (p-value < 0.05), calculated on the basis of the REST (Pfaffl, 2002).

miR-309A binding sites in the 3'UTR of *zelda* mRNA, and the respective patterns, suggest that miR-309A removes *zelda* transcripts at ED2. The model emerging (Figure 3.3.4A) is that the contribution of E93 to the decline of *zelda* mRNA levels is mediated by its stimulatory action on miR-309A production. Intriguingly, Zelda also stimulates the production of *miR-309* miRNAs in *B. germanica* during embryogenesis (Ventos-Alfonso et al., 2019), thus producing a short, negative feedback loop. The data suggest that regulation of mir-309 and Zelda are very fine-tuned thanks to a relatively simple system based on three players (Figure 3.3.4A, inset). It is possible that the decline of *otd*, *nos* and *odd* mRNA levels that was impaired in E93-depleted embryos (Figure 3.3.2D) was also due to the reduced stimulation of miR-309A production found in these embryos (Figure 3.3.3B).

E93 expression is higher in hemimetabolan than in holometabolan embryos

On the other hand, the expression of genes that should become stimulated in the MZT, like *Kr*, *cad* and *eve*, did not increase in the E93-depleted embryos (Figure 3.3.2B). This suggests that E93 may have an activation effect on the zygotic genome of *B. germanica*, eventually additional to that of Zelda, which would drive the development of the hemimetabolan nymph. Intriguingly, *E93* is practically not expressed during embryogenesis of *D. melanogaster*, according to Northern blot data (Baehrecke and Thummel, 1995) and transcriptomic analyses (Ylla et al., 2018). This led us to

wonder whether significant levels of *E93* expression in the embryo are typical of hemimetabolan species. To address this conjecture, we look for sets of publicly available transcriptomes covering the embryonic development, finding those of the hemimetabolan species *Cloeon dipterum* (Palaeoptera, Ephemeroptera), and *Nilaparvata lugens* (Paraneoptera, Hemiptera), in addition to *B. germanica*. Of holometabolan species (Endopterygota), we find that of *Tribolium castaneum* (Coleoptera) and the aforementioned *D. melanogaster* (Diptera). We retrieved the expression profiles of *E93* and *zelda*, observing that the latter is consistently expressed in the embryo of all species, hemimetabolan and holometabolan (Figure 3.3.4B). In contrast, we observed a tendency to show very low or null *E93* expression in holometabolan embryos. In those species where the transcriptomes set covers the entire life cycle, like *B. germanica*, *T. castaneum* and *D. melanogaster*, we compared the expression of *E93* in the preadult stage, when E93 reach the highest values which trigger metamorphosis (Belles, 2020), with embryo stages. In *B. germanica* the mRNA levels of *E93* are within the same order of magnitude in the last nymphal instar and in early embryo (Figure 3.3.4B). The transcriptomic profiles of *T. castaneum* and *D. melanogaster* indicate that *E93* values in the embryo are one to two orders of magnitude lower than those in the pupa. To further validate the transcriptomic observations, in *B. germanica* we have comparable qRT-PCR data (obtained in the same laboratory) for *E93* expression in the embryo, where values reach an average of 23 copies of E93 mRNA per 1000 copies of Actins (this report, Figure 3.3.1A), and in the last nymphal instar,

when the values are similar, reaching as average 21 copies of E93 mRNA per 1000 copies of Actins (Belles and Santos, 2014).

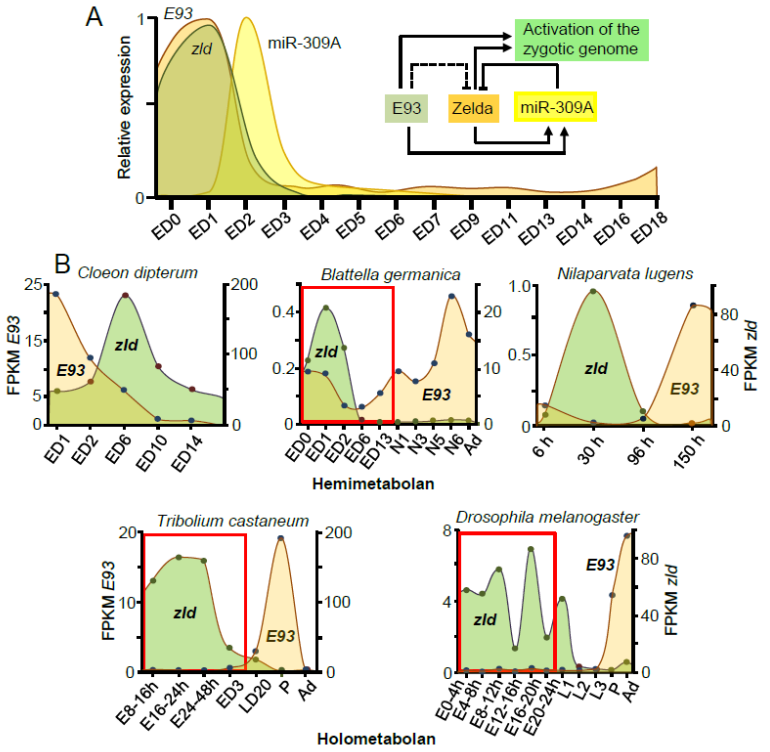


Figure 3.3.4. *E93*, *zelda* and *miR-309* during *Blattella germanica* embryogenesis, and comparison with *E93* and *zelda* in other insects. (*A*) Schematic profiles of *E93*, *zelda* and *miR-309* in embryos of different ages from day 0 (ED0) to day 18 (ED18, the last day of embryo development); the inset shows the inferred interactions between the three factors. (*B*) Transcriptomic pattern of *E93* and *zelda* in hemimetabolan and holometabolan insects; in *Cloeon dipterum* and *Nilaparvata lugens*, the transcriptome set covers practically the entire embryogenesis, each stage is defined in days or in hours; in *B. germanica*, *Tribolium castaneum* and *Drosophila melanogaster*, the transcriptome set covers the entire life cycle, including embryo stages (E), nymphal stages (N), larval stages (L), pupal stages (P), and the adult (Ad); in these three species, the embryonic period is indicated with a red square; expression is given in FPKM, and in all diagrams the left ordinate refers to *E93* and the right ordinate to *zelda* (*zld*).

DISCUSSION

E93 is required for early cockroach embryogenesis

Maternal RNAi showed that E93 is required for proper early embryogenesis, as the most frequent abnormal phenotype of E93-depleted embryos was associated with an interruption of development before germ band formation, around ED2. In particular, all observed embryos from oothecae of dsE93-treated females that were dropped on ED2 did not form a proper germ band due to a deficit in the accumulation of energids. In contrast, the germ band was completed in most Zelda-depleted embryos, where a large part of the defects were related with the elongation of the abdomen, as observed by Ventos-Alfonso et al. 2019. Indeed, there are practically no studies regarding the role of E93 during insect embryogenesis. The only antecedent is the report of Lee et al. (2002) showing that overexpression of *E93* in embryos of the holometabolan *D. melanogaster* -where *E93* is practically not expressed (Baehrecke and Thummel, 1995)- induces cell death, and impaired development, including problems to retract the germ band, and to pattern the cuticle. This obviously indicates that E93 is detrimental to the development of the embryo in *D. melanogaster*.

E93 promotes the production *miR-309* miRNAs, which contribute to clear maternal transcripts, including those of *zelda*

We observed that maternal transcripts that should have disappeared in the MZT, including those of *zelda*, kept their levels high in E93-depleted embryos, which led us to examine the *miR-309* miRNAs. In *D. melanogaster*, *miR-309* miRNAs are required for the degradation of at least 14% of maternal transcripts in the MZT (Bushati et al., 2008). In *B. germanica*, miR-309 miRNAs have an expression peak at ED2 ((Ylla et al., 2017) and this report), thus in the MZT. Moreover, the 3'UTR of *zelda* mRNA contains functional binding sites for miR-309A, which suggests that *zelda* transcripts are eliminated by miR-309 in the context of *B. germanica* MZT. The steep drop in *zelda* transcripts beyond ED1 may suggest that the clearing action of miR-309A is very efficient. Previous studies had reported that Zelda promotes miR-309 production, as shown in *D. melanogaster* (Fu et al., 2014; Liang et al., 2008), and in *B. germanica* (Ventos-Alfonso et al., 2019). Intriguingly, we have shown here that E93 also promotes the production of *miR-309* miRNAs in *B. germanica*, and, in turn, miR-309A contributes to eliminate *zelda* transcripts. This short Zelda inactivation loop can explain the transience of the *zelda* mRNA peak (Ventos-Alfonso et al., 2019).

E93 is crucial for the activation of the zygotic genome

However, and most interestingly, the activating action of embryonic E93 goes beyond *miR-309* miRNAs, extending to other genes. The pioneer factor par excellence in the activation of the zygotic genome in the MZT is Zelda, which facilitates an increase of chromatin accessibility, thus allowing the transcription of multiple genes, as elegantly shown in *D. melanogaster* (Dufourt et al., 2018; Foo et al., 2014; Schulz et al., 2015). In *B. germanica*, *zelda* maternal transcripts are abundant in early embryo, and its depletion results in a diversity of developmental defects, most of them related to the elongation of the abdomen (Ventos-Alfonso et al., 2019). Our present results suggest that, in *B. germanica*, E93 plays an activating role of the zygotic genome similar to that of Zelda, which affects earlier in development.

A pioneer role for E93 in the activation of the zygotic genome is not surprising, given its role as modulator of chromatin accessibility in postembryonic development. In *D. melanogaster*, chromatin accessibility is highly dynamic during wing development, and correlates with the activity of cis-regulatory elements. Importantly, E93 plays a critical role in regulating both opening and closing chromatin, as well as activating and deactivating the activity of cis-regulatory elements (Uyehara et al., 2017). Even more, precociously expressed *E93* in larval wings promotes accessibility changes at many of its target sites, including functional of cis-regulatory elements with dynamic activity, which were precociously activated

or deactivated (Nystrom et al., 2020). Indeed, the observed changes mimic those that occur later in normal development of the adult wing, thus indicating that E93 triggers the metamorphic program through modulating chromatin accessibility.

Embryonic E93 and the evolution of metamorphosis

The data obtained in *B. germanica* suggest that E93 plays a role of master regulator of genome activation in the MZT, similar to that observed for Zelda in *D. melanogaster*. A relevant question that emerges is how general is this role in insects. The available data indicates that *E93* expression is high in the embryo of the hemimetabolan species, and low in that of holometabolans. This leads to presume that the expression of *E93* in the hemimetabolan embryo is associated with the formation of the nymph, a juvenile stage already exhibiting the adult body plan. In contrast, the practical absence of *E93* in the holometabolan embryo could lead to the activation of an alternative genetic program in the zygote, leading to the development of the larva, a juvenile stage morphologically divergent from that of the adult. We realize that the differences observed might be due to the different germ band type, as holometabolan species generally follow the long germ band segmentation, whereas the hemimetabolans develop through the short germ band mode (Davis and Patel, 2002; Liu and Kaufman, 2005). However, *E93* expression is low in the embryo of the beetle *T. castaneum*, a holometabolan that follows a short germ band development ((Davis and Patel, 2002), but see also (Benton, 2018)).

The above considerations led to speculate that a loss of expression of *E93* in the embryo may have facilitated the transition from hemimetaboly to holometaboly. Two main theories are competing to explain the evolution of the holometabolan metamorphosis (see (Belles, 2020)). One is the theory of the pronymph, by which, the holometabolan embryo would arrest development at the late stage of pronymph, giving rise to the larva; development to the adult body plan would resume at the pupal stage. The corollary is that the hemimetabolan pronymph would be homologous to the set of holometabolan larval instars, while the set of hemimetabolan nymphal instars would be homologous to the holometabolan pupa (Truman, 2019; Truman and Riddiford, 1999; Truman and Riddiford, 2019). The other is the theory of direct homology between stages, which maintains that there is no arrest in the development of the embryo, but that ancestral hemimetabolan embryogenesis was modified to give rise to the larva. Accordingly, the nymphal instars would homologous to the larvae, and the last nymphal instar would be homologous to the pupa (Belles, 2020; Belles and Santos, 2014; Hinton, 1963; Jindra, 2019; Sehnal et al., 1996).

The mechanistic explanation of the pronymph theory is based on a shift in Juvenile hormone (JH) production during the development of the embryo, which would have been instrumental in the transition to holometaboly (Truman and Riddiford, 1999). In the theory of direct homology between stages, a mechanism proposed to contribute to the emergence of holometaboly is a shift in the action

of the JH on the *Br-C* gene in juvenile stages, from stimulatory in hemimetabolan nymphs to inhibitory in holometabolan larvae (Belles, 2020; Huang et al., 2013). This would mainly lead to a postponement of wing formation until the pupal stage in holometabolans. With the data revealed in the present work, we can hypothesize an additional mechanism to explain the origin of holometabolan metamorphosis, which considers that E93 and Zelda modulate chromatin in the hemimetabolan embryo and establish the genetic program that gives rise to the formation of the adult body plan exhibited by the nymph. In this context, a loss of expression of E93 in the embryo (or in the preceding oocytes) in the last holometabolan common ancestor, would lead to a change in the modulation of chromatin without the effect of E93 and based on Zelda, which would lead to the formation of a larva instead of a nymph. This mechanism is compatible with the one proposed for Br-C, which would specifically delay wing formation in holometabolans. In the end, the adult-building properties of E93 manifested at the end of the juvenile period in direct developing insects, appear intriguingly mirrored at the onset of embryogenesis. Or vice versa.

MATERIALS AND METHODS

Insects and dissections

B. germanica specimens used in the experiments were from a colony reared in the dark at 30 ± 1ºC and 60-70% r.h. Freshly

emerged females were maintained with males during the first gonadotrophic cycle; mating was confirmed at the end of experiments by assessing the occurrence of spermatozoids in the spermathecae. For dissections and tissue sampling, specimens were anesthetized with carbon dioxide. For RNA extractions we used non-fertilized eggs (NFE, see below), entire oothecae (to establish the expression patterns along embryogenesis and in oothecae from treated females dropped between days 2 and 3) or individual embryos dissected out from artificially opened oothecae.

RNA extraction and reverse transcription to cDNA

RNA extractions were performed with Gen Elute Mammalian Total RNA kit (Sigma- Aldrich, Madrid, Spain). An amount of 100 ng from each RNA extraction was treated with DNase (Promega, Madison, WI, USA) and reverse transcribed with Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) and random hexamers (Promega). RNA quantity and quality were estimated by spectrophotometric absorption using a Nanodrop Spectrophotometer ND-1000® (452 NanoDrop Technologies, Wilmington, DE, USA).

To measure mature miRNA levels, total RNA extractions were performed with miRNeasy Mini Kit (Qiagen). Then, all the volume extracted was lyophilized in the freeze dryer FISHER ALPHA 1 LDplus and resuspended in 8 µl of miliQ H2O. These 8 µl were treated with DNase I (Promega, Madison, WI, USA) and reverse

transcribed with miRNA First-Strand cDNA Synthesis and qRT-PCR Kit (Agilent).

Quantitative reverse transcription PCR

Quantitative reverse transcription PCR (qRT-PCR) reactions were carried out in triplicate in an iQ5 Real-Time PCR Detection System (Bio-Rad Laboratories, Madrid, Spain), using SYBR®Green Supermix (iTaq™ Universal Supermix; Applied Biosystems, Madrid, Spain). A control without a template was included in all batches. The primers used to measure the transcripts studied are indicated in SI Appendix, Table 3.3.S2. The efficiency of the primers was first validated by constructing a standard curve through four serial dilutions. mRNA levels were calculated relative to BgActin-5c expression, and miRNA levels were calculated relative to U6 miRNA expression using the Bio-Rad iQ5 Standard Edition Optical System Software (version 2.0). In the case of mRNA, results are given as copies of mRNA per 1000 copies of BgActin-5c mRNA, and in the case of miRNA measurements results are given as copies of miRNA per 1000 copies of U6 miRNA.

RNA interference

Detailed procedures for dsRNA preparation were as described previously (Ciudad et al., 2006). A dsRNA from *Autographa californica* nucleopoydrovirus was used for control treatments (dsMock). The primers used to prepare the dsRNAs are detailed in

SI Appendix, Table 3.3.S2. Maternal RNAi treatments were carried out essentially as previously reported (Fernandez-Nicolas and Belles, 2017; Ventos-Alfonso et al., 2019). A volume of 1 μL of dsRNA solution (3 μg/μL) was injected into the abdomen of 5-day-old adult females. Then the effects of the treatment were examined in the first oothecae formed by treated and control females.

Examination of embryos

Expression studies were carried out in embryos on days 0, 1, 2, 4, 6, 7, 9, 11, 13, 14 and 16 (ED0 to ED16). ED0 to ED2 cover the maternal to zygotic transition, ED3, ED6 and ED13 coincide with respective pulses of 20-hydroxyecdysone (20E) (Piulachs et al., 2010), and from ED6 to ED16 there is a cycle of JH production (Maestro et al., 2010). To examine the embryos microscopically, the oothecae were opened after 5 min in a water bath at 95ºC and the embryos were dechorionated and individualized. Then, they were fixed in 4% paraformaldehyde, permeabilized in PBS-0.2% tween (PBT) and incubated for 10 min in 1 μg/ml DAPI in PBT. They were then mounted in Mowiol (Calbiochem, Madison, WI, USA) and examined and photographed using epifluorescence with an AxioImager Z1 microscope (ApoTome System, Zeiss).

Cell culture and Luciferase Report Assay

We follow similar protocol than (Zhang et al., 2016). Briefly, *Drosophila* S2 cells (Invitrogen) were grown in *Schneider's*

Drosophila medium (Gibco, Life Technologies) containing 10% (vol/vol) heat-inactivated FBS (Gibco) and 1×Antibiotic-Antimycotic (Gibco) at 23°C in a humidified incubator. The psiCheck-2 reporters were constructed by inserting the *miR-309* putative site into 3′-UTR from the *renilla* luciferase of psiCheck-2 vector (Promega). As negative controls, complementary sequences of the *miR309* putative sites were inserted into 3′-UTR from the *renilla* luciferase of psiCheck-2 vector (Promega). Thereafter, 1000 ng of psiCheck-2 reporters with 1000 nM of synthetic bge-miR-309 miRNA Mimic (GenePharma) were cotransfected into *Drosophila* S2 cells by using Effectene Transfection Reagent (Qiagen). Cells were collected and lysed at 48 h after transfection, and luciferase activities were measured by using the dual luciferase reporter assay system (Promega). Each sample was performed in duplicate, and transfections were repeated three times.

Statistical analyses of qRT-PCR and luciferase assays

To test the statistical significance differences between control and treated samples in qRT-PCR experiments, we used the Relative Expression Software Tool (REST), which evaluates the significance of the derived results by Pair-wise Fixed Reallocation Randomization Test (Pfaffl et al., 2002). In the luciferase assays, to test the statistical significance of the differences between the three groups we used Statistical Package for the Social Sciences (SPSS) performing ANOVA test, and using a p-value<0.005.

Transcriptome analyses

The transcriptome sets used to estimate the expression of *E93* and *zelda* in the embryo were as follows: *Blattella germanica* (GSE99785), *Cloeon dipterum* (PRJEB34721), *Nilaparvata lugens* (SRP227154), *Tribolium castaneum* (GSE63770), *Drosophila melanogaster* (GSE18068). The *E93* and *zelda* respective homologues were annotated in each species as follows: *B. germanica* (*E93* ID: HF536494.1; *zelda* ID: LT717628.1), *C. dipterum* (*E93* ID: gene-CLODIP_2_CD15741; *zelda* ID: gene-CLODIP_2_CD14567), *N. lugens* (*E93* ID: XM_022347627; *zelda* ID: XM_022345079), *T. castaneum* (*E93* ID: gene12027 (LOC655012); *zelda* ID: gene7124 (LOC655012)), *D. melanogaster* (*E93* ID: FBgn0264490; *zelda* ID: FBgn0259789). In all cases, the amount of expression was estimated as FPKM in the context of the whole transcriptomic set.

ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Economy and Competitiveness (grants CGL2012-36251, CGL2015-64727-P and PID2019-104483GB-I00 to XB, including FEDER funds), the CSIC (grant 2019AEP029) and the Catalan Government (grants 2014 SGR 619 and 2017 SGR 1030). AF-N and AV-A had respective PhD fellowships from the Spanish Ministry of Economy and Competitiveness, associated to grants CGL2012-36251 and CGL2015-64727-P, respectively.

REFERENCES

Baehrecke, E. H. and Thummel, C. S. (1995). The *Drosophila* E93 gene from the 93F early puff displays stage- and tissue-specific regulation by 20-hydroxyecdysone. Dev. Biol. 171, 85–97.

Belles, X. (2019). The innovation of the final moult and the origin of insect metamorphosis. Philos. Trans. R. Soc. B Biol. Sci. 374, 20180415.

Belles, X. (2020). Insect metamorphosis. From natural history to regulation of development and evolution. London: Academic Press.

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

Benton, M. A. (2018). A revised understanding of *Tribolium* morphogenesis further reconciles short and long germ development. PLoS Biol. 16, e2005093.

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

Ciudad, L., Piulachs, M.-D. and Belles, X. (2006). Systemic RNAi of the cockroach vitellogenin receptor results in a phenotype similar to that of the *Drosophila* yolkless mutant. FEBS J. 273, 325–335.

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

Davis, G. K. and Patel, N. H. (2002). Short, long, and beyod: molecular and embryological approaches to insect segmentation. Annu. Rev. Entomol. 47, 669–699.

Dufourt, J., Trullo, A., Hunter, J., Fernandez, C., Lazaro, J., Dejean, M., Morales, L., Nait-Amer, S., Schulz, K. N., Harrison, M. M., et al. (2018). Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. Nat. Commun. 9, 5194.

Erezyilmaz, D. F., Riddiford, L. M. and Truman, J. W. (2006). The pupal specifier broad directs progressive morphogenesis in a direct-developing insect. Proc. Natl. Acad. Sci. U. S. A. 103, 6925–6930.

Erezyilmaz, D. F., Rynerson, M. R., Truman, J. W. and Riddiford, L. M. (2009). The role of the pupal determinant broad during embryonic development of a direct-developing insect. Dev. Genes Evol. 219, 535–544.

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

Foo, S. M., Sun, Y., Lim, B., Ziukaite, R., O'Brien, K., Nien, C. Y., Kirov, N., Shvartsman, S. Y. and Rushlow, C. A. (2014). Zelda potentiates morphogen activity by increasing chromatin accessibility. Curr. Biol. 24, 1341–1346.

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

Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. and Eisen, M. B. (2011). Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. PLoS Genet. 7, e1002266.

Hinton, H. E. (1963). The origin and function of the pupal stage. Proc. R. Entomol. Soc. London 38, 77–85.

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

Jindra, M. (2019). Where did the pupa come from? The timing of juvenile hormone signalling supports homology between stages of hemimetabolous and holometabolous insects. Philos. Trans. R. Soc. B 374, 20190064.

Jindra, M., Palli, S. R. and Riddiford, L. M. (2013). The juvenile hormone signaling pathway in insect development. Annu. Rev. Entomol. 58, 181–204.

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

Karim, F. D., Guild, G. M. and Thummel, C. S. (1993). The Drosophila Broad-Complex plays a key role in controlling

ecdysone-regulated gene expression at the onset of metamorphosis. Development 118, 977–988.

Kiss, I., Beaton, A. H., Tardiff, J., Fristrom, D. and Fristrom, J. W. (1988). Interactions and developmental effects of mutations in the Broad-Complex of *Drosophila melanogaster*. Genetics 118, 247–259.

Konopová, B. and Jindra, M. (2008). Broad-Complex acts downstream of Met in juvenile hormone signaling to coordinate primitive holometabolan metamorphosis. Development 135, 559–568.

Lee, C.-Y., Cooksey, B. A. K. and Baehrecke, E. H. (2002). Steroid regulation of midgut cell death during *Drosophila* development. Dev. Biol. 250, 101–111.

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

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

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

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

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

Nystrom, S. L., Niederhuber, M. J. and McKay, D. J. (2020). Expression of E93 provides an instructive cue to control dynamic enhancer activity and chromatin accessibility during development. Development 147, dev181909.

Parthasarathy, R., Tan, A., Bai, H. and Palli, S. R. (2008). Transcription factor broad suppresses precocious development of adult structures during larval-pupal metamorphosis in the red flour beetle, Tribolium castaneum. Mech. Dev. 125, 299–313.

Pfaffl, M. W., Horgan, G. W. and Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. 30, e36.

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

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

Sehnal, F., Švácha, P. and Zrzavý, J. (1996). Evolution of insect metamorphosis. In Metamorphosis. Postembryonic

reprogramming of gene expression in amphibian and insect cells (ed. Gilbert, L. I.), Tata, J. R.), and Atkinson, B. G.), pp. 3–58. San Diego: Academic Press.

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

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

Truman, J. W. (2019). The evolution of insect metamorphosis. Curr. Biol. 29, 1252–1268.

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

Truman, J. W. and Riddiford, L. M. (2007). The morphostatic actions of juvenile hormone. Insect Biochem. Mol. Biol. 37, 761–770.

Truman, J. W. and Riddiford, L. M. (2019). The evolution of insect metamorphosis: a developmental and endocrine view. Philos. Trans. R. Soc. B Biol. Sci. 374, 20190070.

Uyehara, C. M., Nystrom, S. L., Niederhuber, M. J., Leatham-Jensen, M., Ma, Y., Buttitta, L. A. and McKay, D. J. (2017). Hormone-dependent control of developmental timing through regulation of chromatin accessibility. Genes Dev. 31, 862–875.

Ventos-Alfonso, A., Ylla, G. and Belles, X. (2019). Zelda and the maternal-to-zygotic transition in cockroaches. FEBS J. 286, 3206–3221.

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

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

Ylla, G., Piulachs, M. D. and Belles, X. (2018). Comparative transcriptomics in two extreme neopterans reveals general trends in the evolution of modern insects. iScience 4, 164–179.

Zhang, Y., Zhao, B., Roy, S., Saha, T. T., Kokoza, V. A., Li, M. and Raikhel, A. S. (2016). MicroRNA-309 targets the Homeobox gene SIX4 and controls ovarian development in the mosquito *Aedes aegypti*. Proc. Natl. Acad. Sci. U S A 113, E4828–E4836.

Zhou, B., Hiruma, K., Shinoda, T. and Riddiford, L. M. (1998). Juvenile hormone prevents ecdysteroid-induced expression of broad complex RNAs in the epidermis of the tobacco hornworm, *Manduca sexta*. Dev. Biol. 203, 233–244.

Maternal E93 activates gene expression in the zygote and is crucial in early hemimetabolan embryogenesis

Alba Ventos-Alfonso, Ana Fernandez-Nicolas, Guillem Ylla, Xavier Belles Ros

Supporting Information Appendix

Figure 3.3.S1. miR-309A and miR-309B putative binding sites predicted in the 3'UTR of Zelda.

Table 3.3.S1. Features of the miR-309 predicted sites on the 3'UTR of Zelda mRNA.

Table 3.3.S2. Primer sequences used for qRT-PCR, RNAi experiments and luciferase construct preparation.

Figure 3.3.S1. miR-309A and miR-309B putative binding sites predicted in the 3'UTR of *zelda*. The seven miR-309A and the four miR-309B sites (highlighted in blue and brown, respectively) were predicted with the StarMir, miRanda, PITA, RNA 22 and RNAhybrid algorithms (see Table 3.3.S1).

Table 3.3.S1. Features of the miR-309 predicted sites on the 3 'UTR of *zelda* mRNA. The information includes the binding alignment (the seed region of the miRNA is highlighted in blue), the predicted **Δ**G (KJ mol^{-1}) and the algorithm that predicted each site.

miR-309	Site name	Binding alignment	ΔG (KJ mol^{-1})	Algorithm
A	A1	3' **UUACUAUAACG**GAAAGGUCACU 5' 5' **AAUGAUGUAGC**AUUGCUAGUU 3'	-20.30	StarMir, RNAHybrid
A	A2	3' U**UACUAUA - ACG**- GAAAGGUCACU 5' 5' U**AUGAUAUGUGC**ACAAGU**CACUGG** 3'	-10.70	RNA22
A	A3	3' UUAC**UAUAA**CGGAA**AGGUCACU** 5' 5' AUAAA**AAUU** - - **AUUUUUAGUGG** 3'	-10.10	miRanda
A	A4	3' UUA**CUAUAACG**GAA**AGGUCACU** 5' 5' **AGAUAGUGCUC**- UCUUG- **GAG** 3'	-15.50	StarMir
A	A5	3' UUA**CUAUAA**CGGAA**AGGUCACU** 5' 5' AUG**GAUAUU** - -**UAUUCCAUUGC** 3'	-10.54	miRanda, PITA
A	A6	3' UUACUAUA**ACGGAAAGGUCACU** 5' 5' **GUGCCUCGCUUUUGAA** 3'	-16.80	StarMir
A	A7	3' UU**ACU**AUAA**CGGAAAGGUCACU** 5' 5' AC**UGAAACUCCGUGUCCAGGGU** 3'	-12.91	miRanda
B	B1	3' UGC**UA**-UU-**ACAG**-AA-**GGGUCACU** 5' 5' **GGGUUAC**U**GU**UGUA**CCCAGC** 3'	-17.40	StarMir, miRanda, RNAHybrid
B	B2	3' UG**CUAUUACA**- **GAAGGG**UCACU 5' 5' **AGAAAAUGAGUUUCCCC** 3'	-16.20	StarMir
B	B3	3' UGCU**AUUACAGAAG** -**G** -**GUCACU** 5' 5' **UUAGUGUCUUCAAAUAGUU** 3'	-21.10	StarMir, miRanda, RNAHybrid
B	B4	3' UGC**UAU** -**UACAGA** -**AGGGUCACU** 5' 5' UUA**AUAGAUAUAUAUCGCAGUGU** 3'	-11.98	miRanda, PITA

Table 3.3.S2. Primer sequences used for qRT-PCR, RNAi experiments and luciferase construct preparation

Gene Name	Accession number	Forward primer sequence	Reverse primer sequence
qRT-PCR			
Actin 5c	AJ862721.1	AGCTTCCTGATGGTCAGGTGA	TGTCGGCAATTCCAGGGTACATGGT
cad	Bger_M0002*	GCAAAGTCATCAACCGTCCT	ATGCTGAGGGGGTGTTACTG
eve	Bger_18070*	AGGTGCGGGATTGTTAACTG	AGGGTTGGAAAAGCTTTGGT
E93	HF536494	TCCAATGTTTGATCCTGCAA	TTTGGGATGCAAAGAAATCC
hb	LT717629.1	TCTAAATTGCCCACCAGGTC	CCATGAGTTGGAGCCTGAAT
Kr	LT717630	CGTACACACACGGGAGAAAA	AACAAATTGCCGGTCACAAT
nos	Bger_23144*	CGCATTTGACTGTAGTAACGC	CTCATCTCCGCTAGCATTGC
Odd-skipped (odd)	Submitted to GenBank	CATCCACTCGAAGGAGAAGC	CTCCTCCATGTGCAGGATCT
zelda	LT717628.1	TGTCCCAAACAGTTCAACCA	AAAGGGTTTCTCTCCCGTGT
mir-309A	MF574891	TCACTGGAAAGGCAATATCATT	Universal from kit
mir-309B	MF574892	TCACTGGGAAGACATTATCGT	Universal from kit
dsRNA			
E93 dsRNA	HF536494	AAAGAGTTGTCGGGAGCAGA	CCACTGCTAGAAGCCACTCC
polyhedrin	K01149	ATCCTTTCCTGGGACCCGGCA	ATGAAGGCTCGACGATCCTA
Luciferase constructs			
Site A1 wt	LT717628.1	ATCGCTCGAGTTGGTTCAAGTTCCCTGACA	TAGAGCGGCCGCAACATACAGAGGAATCAAGA
Site A4 wt	LT717628.1	ATCGCTCGAGTGATTGATCCAATATGCCACA	TAGAGCGGCCGCAGCATTCCGAGGCTACACAC
Site A6 wt	LT717628.1	ATCGCTCGAGGAGCTTTTTGGCAAGTTGTCT	TAGAGCGGCCGCCCTGGACACGGAGTTTCAGT

Table 3.3.S2 (continued). Primer sequences used for qRT-PCR, RNAi experiments and luciferase construct preparation

Site B1 wt	LT717628.1	ATCGCTCGAGTTGGTTCAAGTTCCCTGACA	TAGAGCGGCCGCAAGAGAATGCCCAGCAACAT
Site B2 wt	LT717628.1	ATCGCTCGAGGCCAGTCAGATATTGGTTATGG	TAGAGCGGCCGCTGTGTGAATCCCTTGCAGTG
Site B3 wt	LT717628.1	ATCGCTCGAGTATGCACTGCAAGGGATTCA	TAGAGCGGCCGCTGACTGTCCCAGTGACTTGTG
Site A1 complementary	LT717628.1	ATCGCTCGAGCATAACCAATATCTGACTGG	TAGAGCGGCCGCTGCATCAGATGATATACT
Site A4 complementary	LT717628.1	ATCGCTCGAGAGCATTCCGAGGCTACACAC	TAGAGCGGCCGCCCCCAATAGTGCTCTCTT
Site A6 complementary	LT717628.1	ATCGCTCGAGTGTGAAAAATAAAAATGTCA	TAGAGCGGCCGCCTTGCAGTCTTGCATCGT
Site B1 complementary	LT717628.1	TAGAGCGGCCGCTTGGTTCAAGTTCCCTGA	ATCGCTCGAGAAGAGAATGCCCAGCAACAT
Site B2 complementary	LT717628.1	TAGAGCGGCCGCGCCAGTCAGATATTGGTT	ATCGCTCGAGTGTGTGAATCCCTTGCAGTG
Site B3 complementary	LT717628.1	TAGAGCGGCCGCTATGCACTGCAAGGGATT	ATCGCTCGAGTGACTGTCCCAGTGACTTGT

4. GENERAL DISCUSSION

Insect metamorphosis is a developmental process by which the insect continues developing after hatching, changing its morphology until reaching the adult stage. There are two basic modes of metamorphosis: hemimetabolan metamorphosis, where the juvenile stages are like miniature adults and develop gradually until adulthood; and holometabolan metamorphosis where the juvenile stages are morphologically divergent with respect to the adult. Metamorphosis studies mainly focus on post-embryonic development, but crucial differences between the two modes of metamorphosis appear during embryogenesis. Moreover, most of the available data on insect studies rely in holometabolan models (especially *Drosophila melanogaster* and *Tribolium castaneum*). In this work, we have focused on the embryogenesis of a hemimetabolan insect, *Blattella germanica*, and then, we have compared the data generated with available data in holometabolan models. With this, we aim at stablishing differences between the respective modes of embryogenesis, which could help to understand the evolution from hemimetaboly to holometaboly.

In previous studies of our laboratory, Ylla et al. (2018) analyzed a series of transcriptomes at different stages of *B. germanica* embryogenesis. The comparative analysis of these transcriptomes with an equivalent transcriptome of the holometabolan *D. melanogaster*, led to the identification of a number of genes that show differential expression patterns between the two species.

Some of them have general and important functions like DNA methylation, juvenile hormone and determination of metamorphosis, and regulatory aspects in the maternal to zygotic transition (MZT). DNA methylation genes and metamorphosis genes were found to be expressed in the embryo of *B. germanica*, but not in the embryo of *D. melanogaster*. Moreover, MZT genes were found to be expressed in a narrow window in early embryogenesis in *B. germanica*, while they are expressed during the whole embryogenesis in *D. melanogaster*. Therefore, three representative genes were considered good candidates to play significant roles determining the adult structures of a hemimetabolan nymph, thus, stablishing the mode of metamorphosis: *DNMT1* (a DNA methylation gene), *Zelda* (a MZT gene) and *E93* (a metamorphosis gene, related to Juvenile hormone). Figure 4.1 summarizes the main findings of this work, specifically related with the profiles of expression and the interactions of these three factors.

4.1. DNMT1 promotes gene methylation, early embryo development and regulates transcriptional noise in *B. germanica*

DNA methylation is a covalent union of a methyl group to a DNA nucleotide; typically it occurs in cytosines in a CpG context (He et al. 2011; Hunt et al. 2013; Bewick et al. 2017). In mammals, it has been associated with several biological processes, including embryo development, genomic imprinting, X-chromosome inactivation, and silencing of retrotransposons (He et al. 2011; Jones 2012).

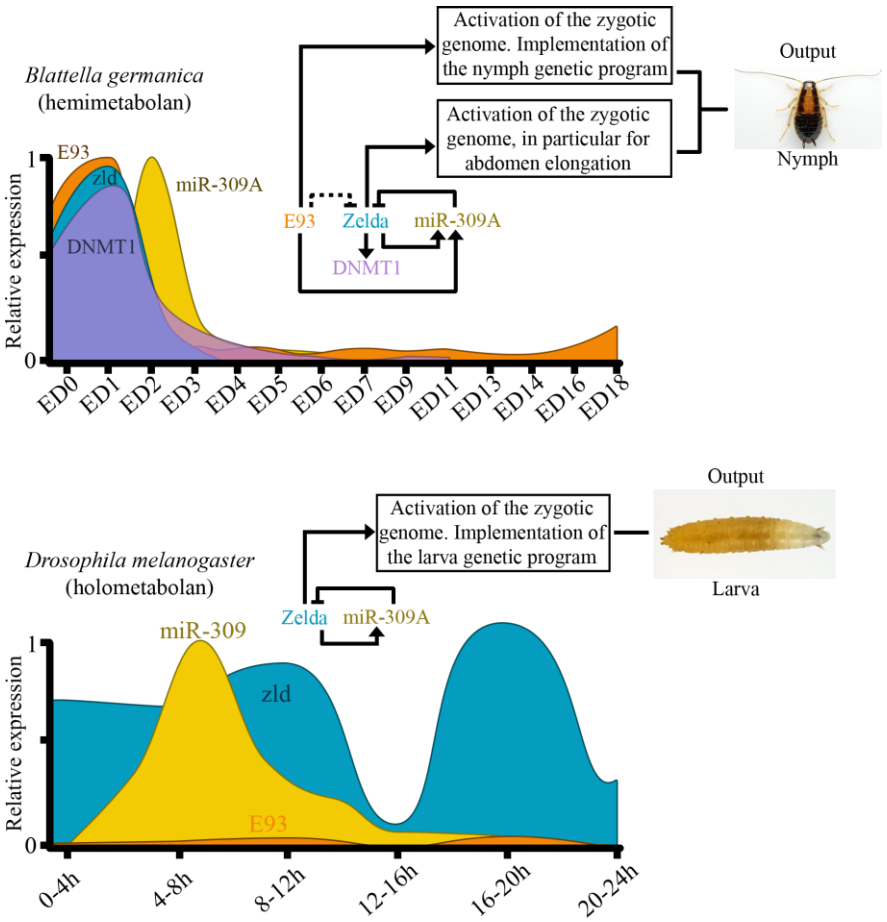


Figure 4.1. *E93, zelda, mir-309* and *DNMT1* during *Blattella germanica* embryogenesis and comparison with *Drosophila melanogaster*. (A) Schematic profiles of *E93*, *zelda*, *mir-309* and *DNMT1* in *B. germanica* embryos from different ages from day 0 (ED0) to day 18 (ED18, last day of embryo development); the inset shows the inferred functions and interactions between the four factors. (B) Schematic profiles of *E93*, *zelda*, *mir-309* and *DNMT1* in *D. melanogaster* embryos from different ages from 0-4 hours to 20-24h (last hour of embryo development); the inset shows the functions and interactions described in the bibliography (Liang et al. 2008; Harrison et al. 2011; Foo et al. 2014; Schulz et al. 2015).

DNA methylation is catalyzed by DNA methyltransferases (DNMTs). In mammals, DNMTs are classified into DNMT3 that establishes new DNA methylation patterns, and DNMT1 that maintains methylation patterns during different cell generations (He et al. 2011; Jones 2012). In *B. germanica*, we have reported both DNMTs, that are well conserved and show a specific expression pattern at the beginning of embryogenesis. However, the low levels of expression of *DNMT3*, the impossibility to deplete them, and previous reports showing that insect's DNA methylation is more related with DNMT1 than with DNMT3 (Bewick et al. 2017), led us to focus our study on DNMT1. Maternal RNAi of DNMT1 and RRBS analyses showed that DNMT1 promotes DNA methylation during early *B. germanica* embryogenesis. Moreover, RNAi experiments also showed that DNMT1 (thus, DNA methylation), promote the formation of germ band in early embryogenesis (around 12% of embryo development). DNMT1 is required for embryo development in other species, including mice (Li et al. 1992; Jackson-Grusby et al. 2001), frogs (Stancheva et al. 2001), zebrafish (Rai et al. 2006), or other insects like the hymenopteran *Nasonia vitripennis* and the beetle *Tribolium castaneum*. However, *DNMT1* depletion in insects affected different embryo stages, in *N. vitripennis*, *DNMT1* depletion caused embryos death at the onset of gastrulation (around 40% of embryo development) (Zwier et al. 2012), and in *T. castaneum*, it affected the first cleavage cycles, around 4% of embryo development (Schulz et al. 2018). This

suggests that although DNA methylation is necessary for embryogenesis from insects to mammals, its specific action varies in different lineages.

Another relevant finding of this study is the relationship between expression variability and DNA methylation. Our results showed that methylated genes show lower expression variability than unmethylated genes. There is a hypothesis positing that DNA methylation in gene regions reduces transcriptional noise (Bird 1995), but information supporting this focused on human tissues (Huh et al. 2013; Bashkeel et al. 2019). Thus, our data in *B. germanica* afforded the first association between high DNA methylation and low expression variability in an insect. Bewick et al. (2017) showed that hemimetabolans present higher levels of DNA methylation than holometabolans. Indeed, some of the most derived holometabolan species, such as *D. melanogaster*, have lost both DNMT1 and DNMT3 (Bewick et al. 2017). Thus, holometabolan insects might have compensated the lack regulation of expression variability by DNA methylation with other features pertaining to genomic, epigenomic, regulatory, polymorphic, functional, structural and gene network characteristics (Alemu et al. 2014).

4.2. Zelda contributes to activate the zygotic genome and is key for abdomen elongation in *B. germanica*

The MZT is an important event during early embryo development in all metazoans. During this transition, maternal mRNAs are

eliminated, while zygotic genome is activated. In this context, *zelda* was found to be a key activator of the zygotic genome in *D. melanogaster* (Liang et al. 2008; Harrison et al. 2011). In *D. melanogaster*, Zelda increases the accessibility of chromatin to those regions that need to be activated during the MZT, promoting the transcriptional activation of early gene networks (Foo et al. 2014; Li et al. 2014; Schulz et al. 2015; Hug et al. 2017; Dufourt et al. 2018). Maternal RNAi experiments showed that Zelda has an important role on the development of the abdomen and the growth of posterior zone during embryo development of *B. germanica*. Moreover, qRT-PCR transcript measurements in the same cockroach showed that Zelda promotes the expression of early zygotic genes (gap, pair-rule and segment polarity genes). This functionality has been described in other insects, such as the beetle *T. castaneum* and the bug *Rhodnius prolixus* (Liang et al. 2008; Ribeiro et al. 2017).

Our comparative analyses showed that the expression pattern of *zelda* between species is different: in *B. germanica* it is concentrated in early embryogenesis, specifically during the MZT, while in *D. melanogaster* and *T. castaneum* z*elda* expression is maintained beyond MZT, covering the entire embryogenesis (Figure 4.1). Moreover, the patterns not only differ in timing but also in localization. *In situ* hybridization showed that *zelda* is localized in the growth zone of *B. germanica* embryos, as occurs in *T. castaneum* (Ribeiro et al. 2017), while in *D. melanogaster* it is generally expressed in the whole embryo (Koromila and

Stathopoulos 2017). The differences observed in the expression pattern can be due to the different type of germ band or different mode of metamorphosis. *T. castaneum* and *B. germanica* embryos develop through short germ band, while *D. melanogaster* embryos develop through long germ band. However, *T. castaneum* and *D. melanogaster* develop through holometabolan metamorphosis, while *B. germanica* follows a hemimetabolan metamorphosis. The whole data suggest that Zelda localization may be important to determine the type of germ band, while Zelda timing may be important to determine the mode of metamorphosis. Thus, hemimetabolan insects may have specific factors regulating z*elda* temporal expression pattern that may be different from holometabolan insects. Knowing those factors may be important to understand the mechanisms that drove the transition from hemimetabolan to holometabolan embryogenesis.

In holometabolan species, embryogenesis gives rise to a larva, with a body structure divergent from that of the adult, while in hemimetabolan species, embryogenesis gives rise to a nymph displaying the essential adult body shape. It could be speculated that the ancestral role of Zelda might be to activate the expression of zygotic genes related with posterior growth and abdomen elongation during early embryogenesis, as suggested by its role in the hemimetabolan *B. germanica*. Then, Zelda could have acquired new functions that extend beyond the early embryo development, as shown by its role on patterning of imaginal disc-derived structures (wings, legs, elytra and antenna) in *T. castaneum* (Ribeiro et al.

2017), and in the genesis of neuroblasts in *D. melanogaster* larvae (Reichardt et al. 2018). Thus, the presence of *Zelda* during the entire embryogenesis in holometabolan species could be needed to acquire these new functionalities and help to maintain chromatin accessible and to activate different transcription factors needed to build the derived larva during holometabolan embryogenesis.

Interestingly, *zelda* depleted embryos showed downregulation of DNMT1, another modulator of chromatin accessibility. Holometabolan species do not show high levels of DNA methylation, and *D. melanogaster* even lost DNMT1. Maybe, Zelda acquired more functional roles in holometabolan species regarding the regulation of chromatin accessibility, which would replace in part the role of DNA methylation in this regard.

4.3. E93 has an essential role in the MZT of the hemimetabolan *B. germanica*

The molecular mechanisms that regulate insect metamorphosis is condensed in the MEKRE93 pathway (Belles and Santos 2014), by which, Juvenile hormone (JH) induces the expression of the transcription factor Krüppel homolog 1 (Kr-h1), and this, in turn, represses the expression of *E93*. In holometabolan insects, such as *D. melanogaster* and *T. casteneum*, E93 depletion causes death at the end of the pupal stage (Mou et al. 2012; Ureña et al. 2014). In the hemimetabolan *B. germanica*, E93 depletion in the nymph prevents the nymph-adult transition (Belles and Santos 2014; Ureña et al. 2014). Thus, E93 is known as the master trigger of

metamorphosis or adult specifier. More recently, Uyehara et al. (2017), showed that E93 controls temporal identity by regulating chromatin accessibility across the genome in *D. melanogaster*.

In *B. germanica* we found that *E93* is highly expressed during the embryogenesis and during the last nymphal instar. In contrast, *E93* is expressed only in pupae and adult stages, but not during embryogenesis in *D. melanogaster* (Figure 4.1). Maternal RNAi experiments showed that *E93* activates the expression of zygotic genes (gap and pair-rule genes) and is crucial for a proper germ band in the *B. germanica* embryo. Although both Zelda and E93 show similar roles in activating the zygotic genome in *B. germanica*, Zelda-depleted embryos show developmental problems later (during abdomen elongation) than E93-depleted embryos (during blastoderm formation). Therefore, it seems that E93 is acting earlier (and has a broader impact on the zygotic genome activation) than Zelda during *B. germanica* embryogenesis. Comparing the expression patterns of *E93* and *zelda* in different insects, we have found that in hemimetabolan insects *E93* has a peak of expression in early embryogenesis (not only in *B. germanica* but also in *Cloeon dipterum* and *Nilaparvata lugens*) whereas in the holometabolan insects that we examined (*T. castaneum* and *D. melanogaster*), *E93* is practically not expressed during embryogenesis. Moreover, in hemimetabolan insects *zelda* shows a specific expression pattern, concentrated in the early embryogenesis, while in holometabolan insects, *zelda* is expressed during the whole embryogenesis. This suggests that two chromatin

accessibility regulators operate at the beginning of embryogenesis in hemimetabolans: E93 and Zelda. Both would act as pioneer factors, regulating chromatin accessibility in the whole embryo, but whereas E93 might have a broader action, implementing the genetic program that drives the formation of an adultiform nymph, Zelda would have a more specialized function in the formation and elongation of the abdomen.

Interestingly, depletion of *E93* in *B. germanica* prevented the expression decline of *zelda* that normally occurs at ED2. Moreover, E93 stimulates the expression of miR-309 miRNAs, which contribute to remove maternal transcripts during the MZT, as demonstrated in *D. melanogaster* (Bushati et al. 2008). These data and the observation that *zelda* mRNA contains functional binding sites for miR-309 miRNAs, strongly suggest that E93 produced in early oogenesis contributes to shape the peak of *zelda* expression. Thus, E93 becomes the key factor that regulates *zelda* temporal expression, and this appears to be a feature characteristic of the hemimetabolan embryo.

4.4. DNA methylation, Zelda and E93, and the evolution of holometaboly

The evolutionary transition from hemimetaboly to holometaboly involves the transformation of a type of embryogenesis that gives rise to a nymph, which has adult features, to an embryogenesis that gives rise to a larva, with features very different from those of an adult, which in very derived species takes a vermiform shape. In

other words, during hemimetabolan embryogenesis a genetic program is installed that leads to the development of a nymph, while in holometabolan a larval genetic program is implemented.

Therefore, the emergence of holometaboly would entail a kind of replacement (or modification) of the nymphal genetic program by a larval genetic one. The comparative genomic data in both types of embryogenesis (for example, those of Ylla et al., 2018) suggest at least three genes that change significantly from the hemimetabolan type of embryogenesis to the holometabolan: *DNMT1*, *zelda* and *E93*. *DNMT1* expression is reduced in holometabolan embryogenesis, even the gene can disappear in the most derived species. The expression of *zelda* widens, from only the beginning of embryogenesis in the hemimetabolan to all embryogenesis in the holometabolan species. Finally, the expression of *E93* is dramatically reduced in the holometabolan embryo, until being practically undetectable.

Certainly, other genes must involve in the hemimetabolan to holometabolan embryogenesis transformation. This is shown by the large number of genes that are differentially expressed in both types of metamorphosis (see, for example, Ylla et al., 2018). But the three genes mentioned may have an especially relevant role, especially due to their modulatory properties of chromatin, which can have effects on a broad genomic scale. Of these, *E93* seems especially interesting, since it is an adult specifying factor in the post-embryonic development of hemimetabolan and holometabolan

species (Ureña et al., 2014). It is plausible, then, that this adult specifying role is beginning to be exercised in the embryogenesis of the hemimetabolan.

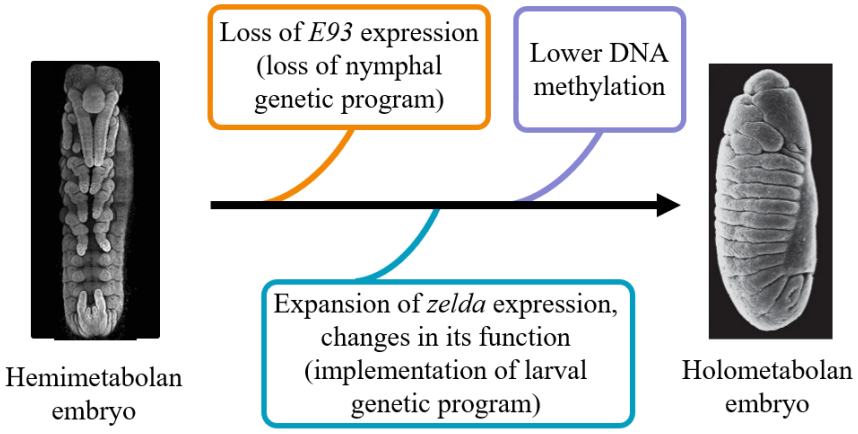


Figure 4.2. Factors studied in the present work, and changes in the ancestrally hemimetabolan embryo that would have facilitated the transition towards the formation of a holometabolan embryo. Hemimetabolan embryo is represented with *Blattella germanica* late embryo, and holometabolan embryo is represented with *Drosophila melanogaster* late embryo.

Figure 4.2 summarizes the changes in the three genes studied, which would be involved in the evolutionary transition from hemimetabolan to holometabolan embryogenesis. The loss of DNA methylation in holometabolan embryos, or even the loss of *DNMT1* in the derived holometabolan *D. melanogaster*, suggest that holometabolan embryos found alternative mechanisms to enhance the expression of metabolic genes and to reduce transcriptional noise. The expression of *E93* would have been lost in the transition to holometaboly, with the consequent total or partial loss of the

genetic program of the adult-shaped nymph. In return, *zelda* would have expanded the time frame of its expression, whereas their chromatin modulating functions would have contributed to implement the larval genetic program.

It is clear that this is a simplistic and speculative scenario, in which many other players and interactions are missing. However, it can be the beginning of new reflections on the never-ending theme of the evolution of metamorphosis.

5. CONCLUSIONS

From the results obtained, the following ten conclusions can be inferred. All of them refer to the German cockroach, *Blattella germanica* (Insecta, Polyneoptera, Blattodea), unless specifically indicated.

1. Depletion of *DNMT1* resulted in a reduction of DNA methylation levels in the genome and impaired embryo development. This indicates that DNMT1 promotes DNA methylation in *B. germanica*, as observed in other insect species, and that this DNA methylation is necessary for proper embryo development.

2. Methylated genes showed higher levels of expression than unmethylated genes and GO enrichment analyses revealed that methylated genes are involved in metabolic processes, while unmethylated genes are more related with signaling pathways. This indicates that DNA methylation influences the level of expression of the genes, maintaining highly expressed those genes with housekeeping functions, as shown in other insect species.

3. High levels of DNA methylation correlated with low expression variability, both between biological replicates and between different developmental stages. This supports the hypothesis positing that DNA methylation in gene regions reduces transcriptional noise as previously suggested by authors working on human tissues.

4. *zelda* depletion impaired embryo development mainly involving different degrees of abdomen malformation. This indicates that Zelda is required for proper embryogenesis, in particular regarding the formation and elongation of the abdomen.

5. Expression of early zygotic genes and miR-309 miRNAs was downregulated in Zelda depleted embryos. These observations indicate that Zelda induces the activation of the zygotic genome, including the production of miR-309 miRNAs in the maternal-to-zygotic transition (MZT) context.

6. In *B. germanica zelda* expression concentrates in early embryogenesis, during the MZT, whereas in *D. melanogaster* and *T. casteneum* it continues beyond MZT. This suggests that the mechanisms regulating *zelda* temporal expression is different in hemimetabolan and holometabolan species, and that the narrow peak of *zelda* at the beginning of embryogenesis of *B. germanica* may be associated with the formation of the adultoid nymph, whereas its extension in holometabolan species may be associated with the formation of a derived larva.

7. Depletion of *E93* impaired embryo development before germ band formation and triggered a number of malformations in embryos that passed that stage, including the formation of individuals with a vermiform shape. The whole data suggests that E93 is crucial for the germ band formation of the adult-shaped nymph.

8. *E93* depleted embryos showed early zygotic genes and miR-309 miRNAs downregulated. This indicates that E93 activates the zygotic genome, including the production of *miR-309* miRNAs, thus having a key role in the MZT.

9. The RNAi experiments showed that E93 contributes to the increase of miR-309 miRNA levels and to the decline of *zelda* expression in early embryogenesis. Moreover, luciferase assays demonstrated that *zelda* mRNA contains functional miR-309a target sites. These evidences suggest that E93 indirectly downregulates *zelda* expression by enhancing the production of miR-309a.

10. *E93* is consistently expressed in early embryogenesis of *B. germanica*, which is not the case in *Drosophila melanogaster*. The data available in other species suggest that this could be general in hemimetabolans and holometabolans, respectively. If E93 determines the nymphal genetic program in the hemimetabolan embryo, it could be that the loss of expression of this gene in the holometabolan embryo has facilitated the implementation of a larval genetic program, and the evolution of holometaboly.

6. REFERENCES

Alemu EY, Carl JW, Bravo HC, Hannenhalli S. 2014. Determinants of expression variability. Nucleic Acids Res. 42:3503–3514. https://pubmed.ncbi.nlm.nih.gov/24435799/

Anderson DT. 1972a. The Development of Holometabolous Insects. In: Counce SJ, Waddington CH, editors. Developmental Systems: Insects. London: Academic Press. p. 165–242.

Anderson DT. 1972b. The development of Hemimetabolous insects. In: Counce S., Waddington CH, editors. Developmental Systems: Insects. London: Academic Press. p. 95–163.

Arsala D, Lynch JA. 2017. Ploidy has little effect on timing early embryonic events in the haplo-diploid wasp Nasonia. Genesis 55. https://pubmed.ncbi.nlm.nih.gov/28432826/

Bashkeel N, Perkins TJ, Kærn M, Lee JM. 2019. Human gene expression variability and its dependence on methylation and aging. BMC Genomics 20:941. https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-6308-7

Belles X. 2011. Origin and Evolution of Insect Metamorphosis. In: eLS. www.els.net

Belles X. 2020. Insect metamorphosis: from natural history to regulation of development and evolution. London, Academic Press

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. https://pubmed.ncbi.nlm.nih.gov/25008785/

Belles X, Ylla G. 2015. Towards understanding the molecular basis of cockroach tergal gland morphogenesis. A transcriptomic approach. Insect Biochem. Mol. Biol. 63:104–112. https://pubmed.ncbi.nlm.nih.gov/26086932/

Bewick AJ, Sanchez Z, McKinney EC, Moore AJ, Moore PJ, Schmitz RJ. 2019. Dnmt1 is essential for egg production and embryo viability in the large milkweed bug, *Oncopeltus fasciatus*. Biological Sciences 0604 Genetics. Epigenetics and Chromatin, 12, 6. https://doi.org/10.1186/s13072-018-0246-5

Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. 2017. Evolution of DNA methylation across insects. Mol. Biol. Evol. 34:654–665. https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msw264

Bird AP. 1995. Gene number, noise reduction and biological complexity. Trends Genet. 11:94–100. https://pubmed.ncbi.nlm.nih.gov/7732579/

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. 18:501–506. https://pubmed.ncbi.nlm.nih.gov/18394895/

Cristino AS, Tanaka ED, Rubio M, Piulachs MD, Belles X. 2011. Deep sequencing of organ- and stage-specific micrornas in the evolutionarily basal insect *Blattella germanica* (L.)

(Dictyoptera, Blattellidae). Zhang B, editor. PLoS One 6:e19350. https://dx.plos.org/10.1371/journal.pone.0019350

Davis GK, Patel NH. 2002. Short, long, and beyond: Molecular and embryological approaches to insect segmentation. Annu. Rev. Entomol. 47:669–699. https://www.annualreviews.org/doi/10.1146/annurev.ento.47.091201.145251

Dufourt J, Trullo A, Hunter J, Fernandez C, Lazaro J, Dejean M, Morales L, Nait-Amer S, Schulz KN, Harrison MM, et al. 2018. Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. Nat. Commun. 9:5194.https://doi.org/10.1038/s41467-018-07613-z

Ewing LS. 1970. The cockroach. Vol. I. A laboratory insect and an industrial pest. Anim. Behav. 18:192.

Falckenhayn C, Boerjan B, Raddatz Gü, Frohme M, Schoofs L, Lyko F. 2013. Characterization of genome methylation patterns in the desert locust *Schistocerca gregaria*. J. Exp. Biol. 216:1423–1429. http://titan.biotec.uiuc.edu/locust

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. https://pubmed.ncbi.nlm.nih.gov/24909324/

Glastad KM, Hunt BG, Goodisman MAD. 2019. Epigenetics in insects: Genome regulation and the generation of phenotypic diversity. Annu. Rev. Entomol. 64:185–203. https://doi.org/10.1146/annurev-ento-011118-

Grbic M, Strand MR. 1998. Shifts in the life history of parasitic wasps correlate with pronounced alterations in early development. Proc. Natl. Acad. Sci. 95:1097 LP – 1101. http://www.pnas.org/content/95/3/1097.abstract

Grimaldi D, Engel MS. 2005. Evolution of the Insects. doi: 10.1653/0015-4040(2007)90[588:eoti]2.0.co;2

Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, Childers CP, Dinh H, Doddapaneni H, Dugan S, et al. 2018. Hemimetabolous genomes reveal molecular basis of termite eusociality. Nat. Ecol. Evol. 2:557–566.

Harrison MM, Li X-Y, Kaplan T, Botchan MR, Eisen MB. 2011. Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. PLoS Genet. 7:e1002266. http://www.ncbi.nlm.nih.gov/pubmed/22028662

He XJ, Chen T, Zhu JK. 2011. Regulation and function of DNA methylation in plants and animals. Cell Res. 21:442–465.

Heming BS. 2003. Insect Development and Evolution. Comstock books. Ithaca, New York.

Hinton HE. 1963. The origin and function of the pupal stage. Proc. R. Entomol. Soc. London. Ser. A, Gen. Entomol. 38:77–85. http://doi.wiley.com/10.1111/j.1365-3032.1963.tb00759.x

Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. 2017. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell 169:216-228.e19. https://pubmed.ncbi.nlm.nih.gov/28388407/

Huh I, Zeng J, Park T, Yi S V. 2013. DNA methylation and transcriptional noise. Epigenetics and Chromatin 6:9. http://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/1756-8935-6-9

Hunt BG, Glastad KM, Yi S V., Goodisman MAD. 2013. The function of intragenic DNA methylation: Insights from insect epigenomes. Integr. Comp. Biol. 53:319–328.

Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G, Dausman J, Lee P, Wilson C, Lander E, et al. 2001. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat. Genet. 27:31–39.

Jones PA. 2012. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13:484–492.

Jura C. 1972. Development of Apterygote insects. In: Counce S., Waddington C., editors. Developmental Systems: Insects. London: Academic Press. p. 49–94.

Kivelä SM, Friberg M, Wiklund C, Leimar O, Gotthard K. 2016. Towards a mechanistic understanding of insect life history evolution: Oxygen-dependent induction of moulting explains moulting sizes. Biol. J. Linn. Soc. 117:586–600. https://academic.oup.com/biolinnean/article-lookup/doi/10.1111/bij.12689

Konopová B, Zrzavý J. 2005. Ultrastructure, development, and homology of insect embryonic cuticles. J. Morphol. 264:339–362. http://doi.wiley.com/10.1002/jmor.10338

Koromila T, Stathopoulos A. 2017. Broadly expressed repressors integrate patterning across orthogonal axes in embryos. Proc. Natl. Acad. Sci. U. S. A. 114:8295–8300.

Kukalová-Peck J. 1991. Fossil history and the evolution of Hexapod structures. In: CSIRO, Division of Entomology. Insects of Australia. 2nd edition, 2 volumes. Cornell University Press. pp: 141-179. https://natlib.govt.nz/records/30747665

Larink O. 1983. Embryonic and postembryonic development of *Machilidae* and *Lepismatidae* (Insecta: Archeognatha et Zygentoma). Entomol. Genet. 8:119–133.

Li E, Bestor TH, Jaenisch R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915–926.

Li XY, Harrison MM, Villalta JE, Kaplan T, Eisen MB. 2014. Establishment of regions of genomic activity during the *Drosophila* maternal to zygotic transition. Elife 3:e03737

Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. Nature 456:400–403. https://pubmed.ncbi.nlm.nih.gov/18931655/

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. Evol Dev. p. 629–646. https://pubmed.ncbi.nlm.nih.gov/16336416/

Lubbock J. 1873. On the Origin and Metamorphoses of Insects. (Classics F, editor.). Macmillan, London

Mito T, Sarashina I, Zhang H, Iwahashi A, Okamoto H, Miyawaki K, Shinmyo Y, Ohuchi H, Noji S. 2005. Non-canonical functions of hunchback in segment patterning of the intermediate germ cricket *Gryllus bimaculatus*. Development 132:2069–2079. https://dev.biologists.org/content/132/9/2069

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

Mou X, Duncan DM, Baehrecke EH, Duncan I. 2012. Control of target gene specificity during metamorphosis by the steroid response gene *E93*. Proc. Natl. Acad. Sci. U. S. A. 109:2949–2954.

Panfilio KA. 2008. Extraembryonic development in insects and the acrobatics of blastokinesis. Dev. Biol. 313:471–491.

Peel AD. 2008. The evolution of developmental gene networks: lessons from comparative studies on holometabolous insects. Philos. Trans. R. Soc. B Biol. Sci.:1539–1547.

Pires CV, De Paula Freitas FC, Cristino AS, Dearden PK, Simões ZLP. 2016. Transcriptome analysis of honeybee (*Apis Mellifera*) haploid and diploid embryos reveals early zygotic transcription during cleavage.Hudson ME, editor. PLoS One 11:e0146447.
https://dx.plos.org/10.1371/journal.pone.0146447

Provataris P, Meusemann K, Niehuis O, Grath S, Misof B. 2018. Signatures of DNA methylation across insects suggest

reduced DNA methylation levels in holometabola. Genome Biol. Evol. 10:1185–1197. https://academic.oup.com/gbe/article/10/4/1185/4943971

Rai K, Nadauld LD, Chidester S, Manos EJ, James SR, Karpf AR, Cairns BR, Jones DA. 2006. Zebra Fish Dnmt1 and Suv39h1 Regulate Organ-Specific Terminal Differentiation during Development. Mol. Cell. Biol. 26:7077–7085. Reichardt I, Bonnay F, Steinmann V, Loedige I, Burkard TR, Meister G, Knoblich JA. 2018. The tumor suppressor Brat controls neuronal stem cell lineages by inhibiting Deadpan and Zelda. EMBO Rep. 19:102–117.

Ribeiro L, Tobias-Santos V, Santos D, Antunes F, Feltran G, de Souza Menezes J, Aravind L, Venancio TM, Nunes da Fonseca R. 2017. Evolution and multiple roles of the Pancrustacea specific transcription factor zelda in insects. PLoS Genet. 13:e1006868. https://doi.org/10.1371/journal.pgen.1006868

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. 25:1715–1726.

Schulz NKE, Wagner CI, Ebeling J, Raddatz G, Diddens-de Buhr MF, Lyko F, Kurtz J. 2018. Dnmt1 has an essential function despite the absence of CpG DNA methylation in the red flour beetle *Tribolium castaneum*. Sci. Rep. 8:16462. https://doi.org/10.1038/s41598-018-34701-3

Sehnal F. 1985. Growth and life cycles. Compr. Insect Physiol. Biochem. Pharmacol. Pergamon Press. Oxford.:1–86.

Sehnal F, Švácha P, Zrzavý J. 1996. Metamorphosis. Postembryonic Reprogramming of Gene Expression in Amphibian and Insect Cells. In: L.I. G, Tata JR, B.G. A, editors. Evolution of insect metamorphosis. San Diego: Academic Press. p. 3–58.

Stancheva I, Hensey C, Meehan RR. 2001. Loss of the maintenance methyltransferase, xDnmt1, induces apoptosis in *Xenopus* embryos. EMBO J. 20:1963–1973.

Sucena É, Vanderberghe K, Zhurov V, Grbić M. 2014. Reversion of developmental mode in insects: evolution from long germband to short germband in the polyembrionic wasp *Macrocentrus cingulum* Brischke. Evol. Dev. 16:233–246.

Švácha P. 1992. What are and what are not imaginal discs: Reevaluation of some basic concepts (insecta, holometabola). Dev. Biol. 154:101–117.

Tanaka A. 1976. Stages in the Embryonic Development of the German Cockroach *Blattella germanica* (Blattaria, Blattellidae). Entomol. Soc. Japan, Kontyu,Tokyo, 44(4):512–525.

Truman JW. 2019. The Evolution of Insect Metamorphosis. Curr. Biol. 29:R1252–R1268.

Truman JW, Riddiford LM. 1999. The origins of insect metamorphosis. Nature 401:447–452.

Truman JW, Riddiford LM. 2002. Endocrine insights into the evolution of metamorphosis in insects. Annu. Rev. Entomol. 47:467–500.

Truman JW, Riddiford LM. 2019. The evolution of insect metamorphosis: A developmental and endocrine view. Philos. Trans. R. Soc. B Biol. Sci. 374:20190070. https://royalsocietypublishing.org/doi/10.1098/rstb.2019.0070

Ureña E, Manjón C, Franch-Marro X, Martín D. 2014. Transcription factor E93 specifies adult metamorphosis in hemimetabolous and holometabolous insects. Proc. Natl. Acad. Sci. U. S. A. 111:7024–7029.

Uyehara CM, Nystrom SL, Niederhuber MJ, Leatham-Jensen M, Ma Y, Buttitta LA, McKay DJ. 2017. Hormone-dependent control of developmental timing through regulation of chromatin accessibility. Genes Dev. 31:862–875. https://pubmed.ncbi.nlm.nih.gov/28536147/

Wang YH, Engel MS, Rafael JA, Wu HY, Rédei D, Xie Q, Wang G, Liu XG, Bu WJ. 2016. Fossil record of stem groups employed in evaluating the chronogram of insects (Arthropoda: Hexapoda). Sci. Rep. 6:1–12.

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. https://doi.org/10.1186/s12864-017-4177-5

Ylla G, Piulachs MD, Belles X. 2018. Comparative Transcriptomics in Two Extreme Neopterans Reveals General Trends in the Evolution of Modern Insects. iScience 4:164–179.

Zwier M V., Verhulst EC, Zwahlen RD, Beukeboom LW, Van De Zande L. 2012. DNA methylation plays a crucial role during early *Nasonia* development. Insect Mol. Biol. 21:129–138. https://pubmed.ncbi.nlm.nih.gov/22122805/