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

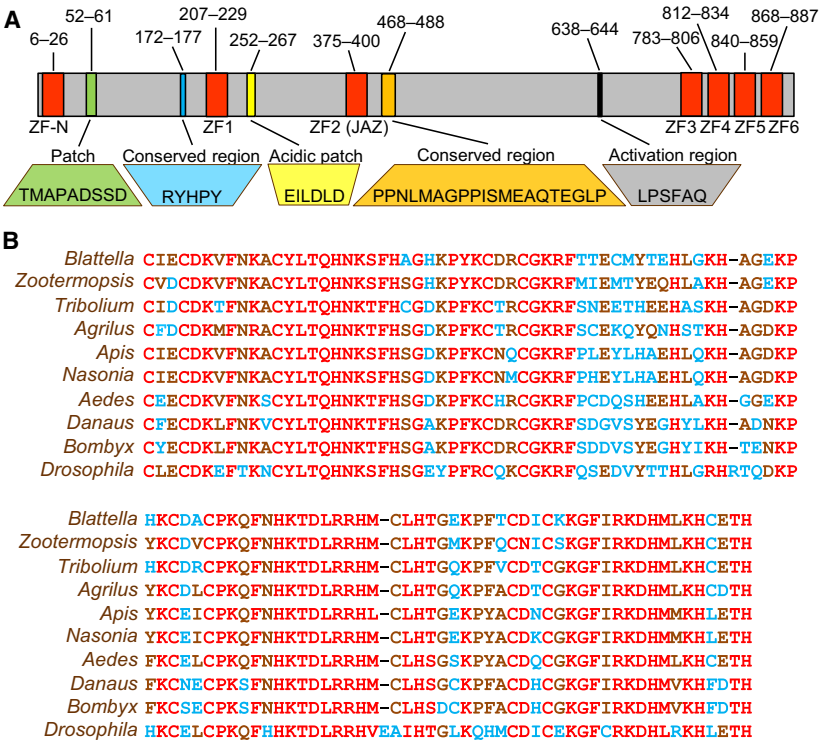

Fig. 1. Structure of Zelda in *Blattella germanica*. (A) Structure of the protein Zelda showing the most characteristic motifs according the nomenclature of Ribeiro *et al.* [15]; the numbers indicate the initial and final starting amino acid of each motif; ZF-N: ZF-Novel. (B) Alignment of the C-terminal 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 MAFFT – 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

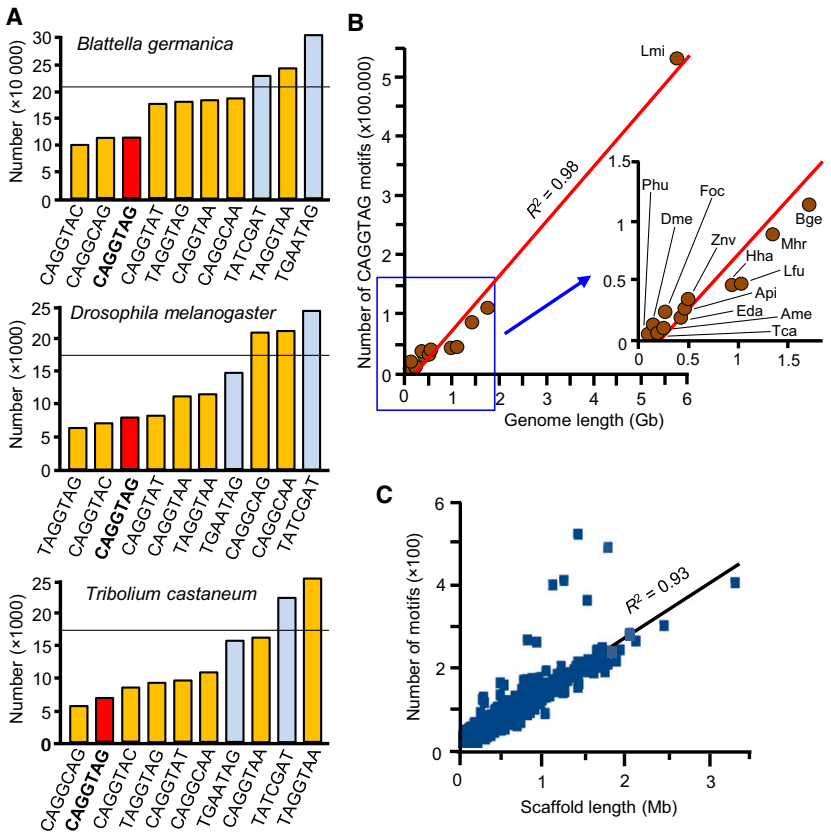

Fig. 2. The TAGteam heptamers in *Blattella germanica*. (A) Occurrence of the eight functional (yellow and red columns, the red column indicates the canonical sequence CAGGTAG) and not functional (blue columns) TAGteam heptamers reported by Nien *et al.* [8] in *Drosophila melanogaster, Tribolium castaneum,* 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

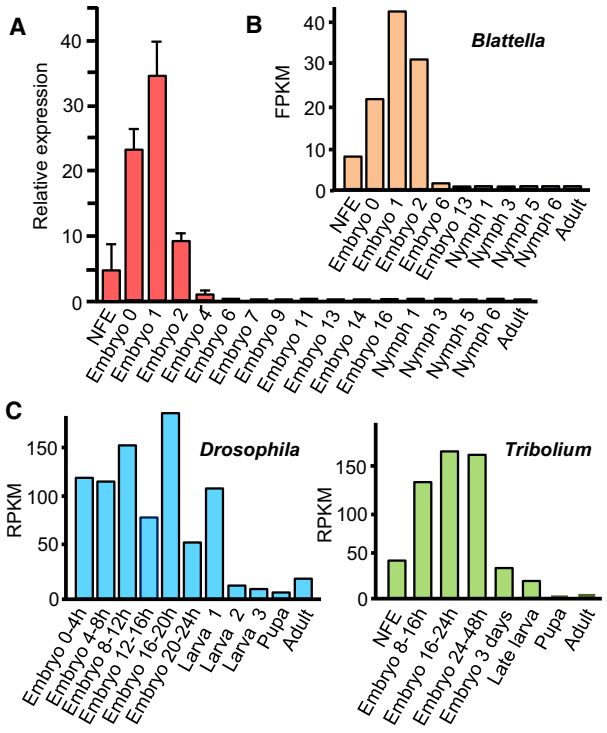

Fig. 3. Expression of *zelda* in *Blattella germanica*. (A) Expression pattern of Bg*zelda* along ontogeny in *B. germanica* obtained by qRT-PCR; the stages examined are the same of the libraries of Ylla *et al.* [19], plus a number of additional embryo stages (ED4, ED7, ED9, ED11, ED14, ED16); each 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

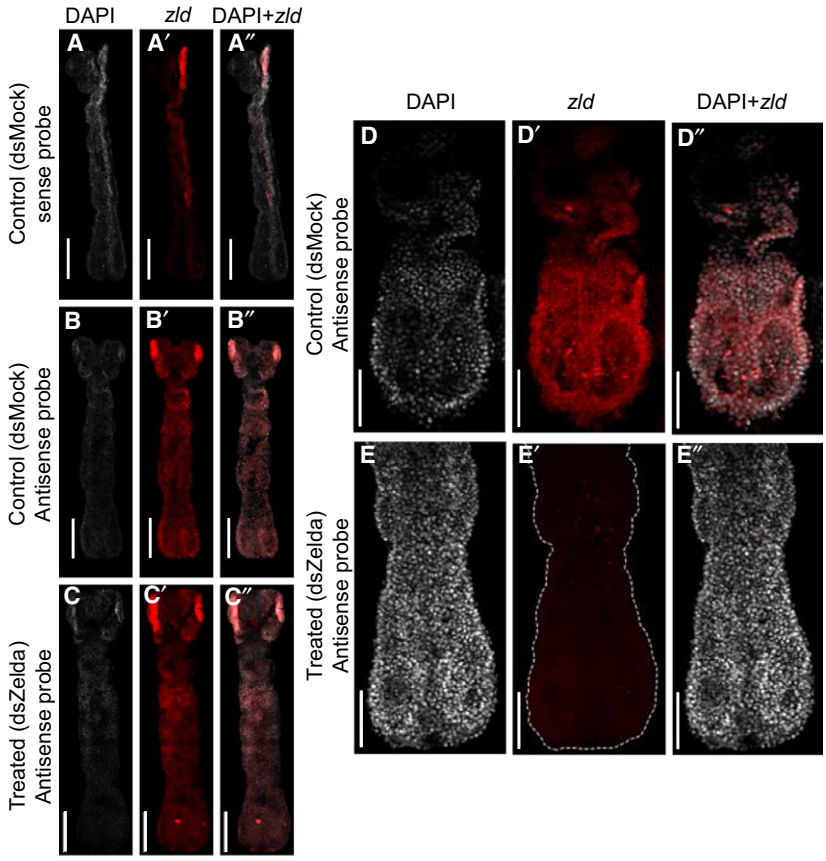

Fig. 4. Expression of Bg*zelda* in 3-day-old embryos of *Blattella germanica*, detected by *in situ* hybridization. (A–C). Entire embryo views; the upper panels, A–A″, correspond to negative controls, showing unspecific background staining observed with a Bg*zelda* (Zld) sense probe; the central panels, 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):

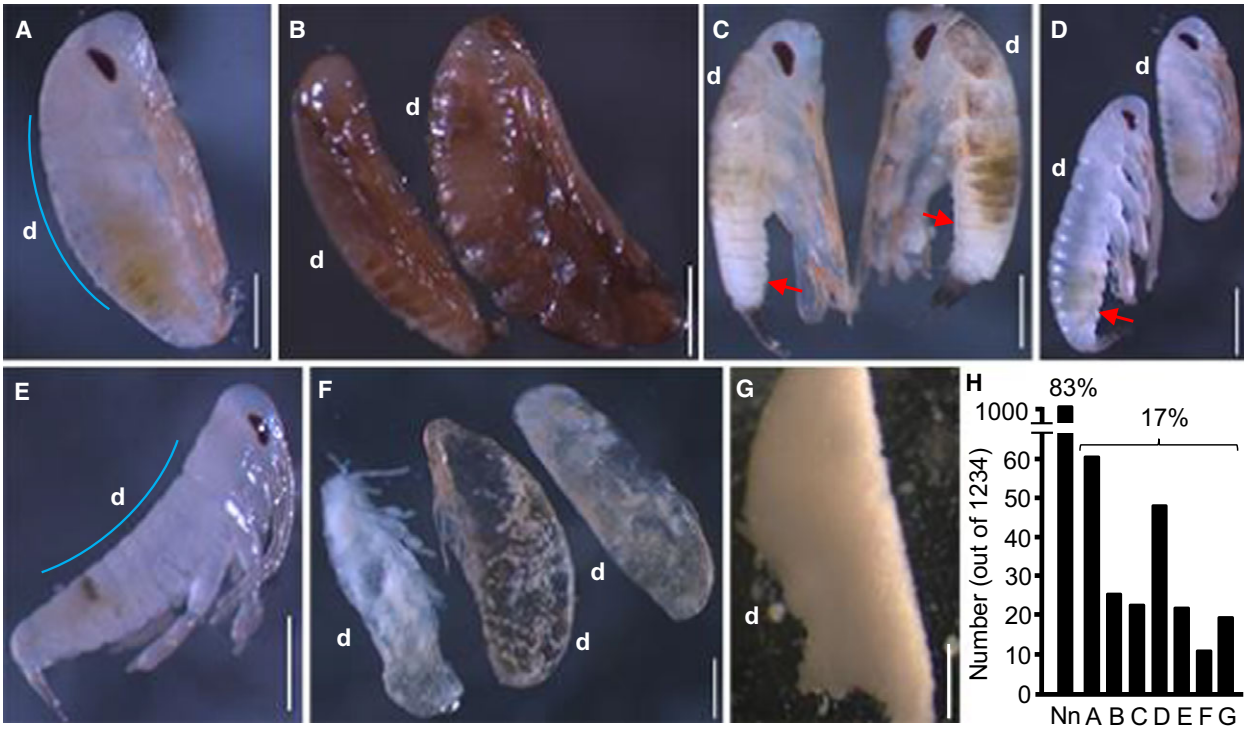

Fig. 5. Phenotypes observed in unhatched oothecae from BgZelda-depleted *Blattella germanica*. (A) Phenotype A, normal embryos in stage 19, thus ready to hatch as normal nymphs (used as control of phenotypes B to G), but which did not hatch. (B) Phenotype B, embryos in stage 16–18, similar to controls 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

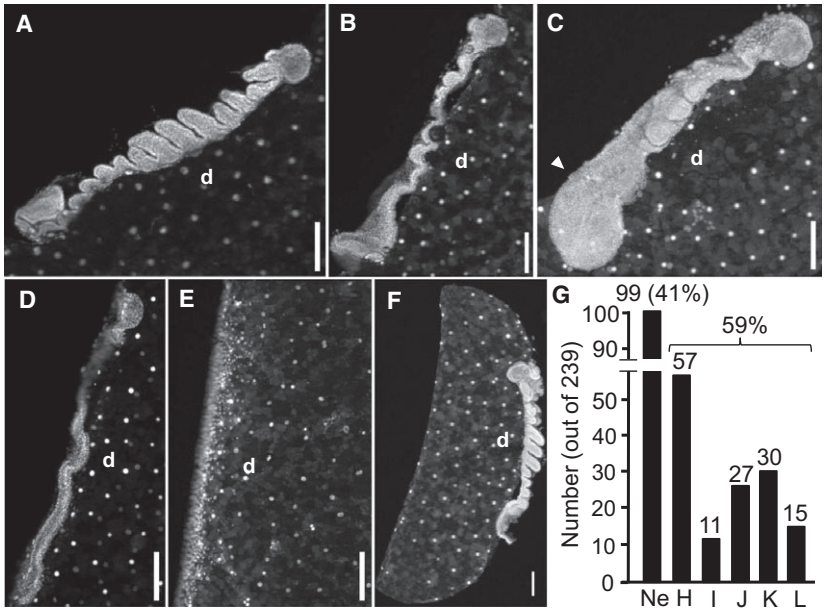

Fig. 6. Phenotypes observed in 4-day-old oothecae (embryos in stage ED4) from BgZelda-depleted *Blattella germanica*. (A) Normal ED4 embryo. (B) Phenotype H, embryo with seriously impaired segmentation, with the three thoracic segments delimited by undulated furrows, and the abdominal region amorphous. (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*

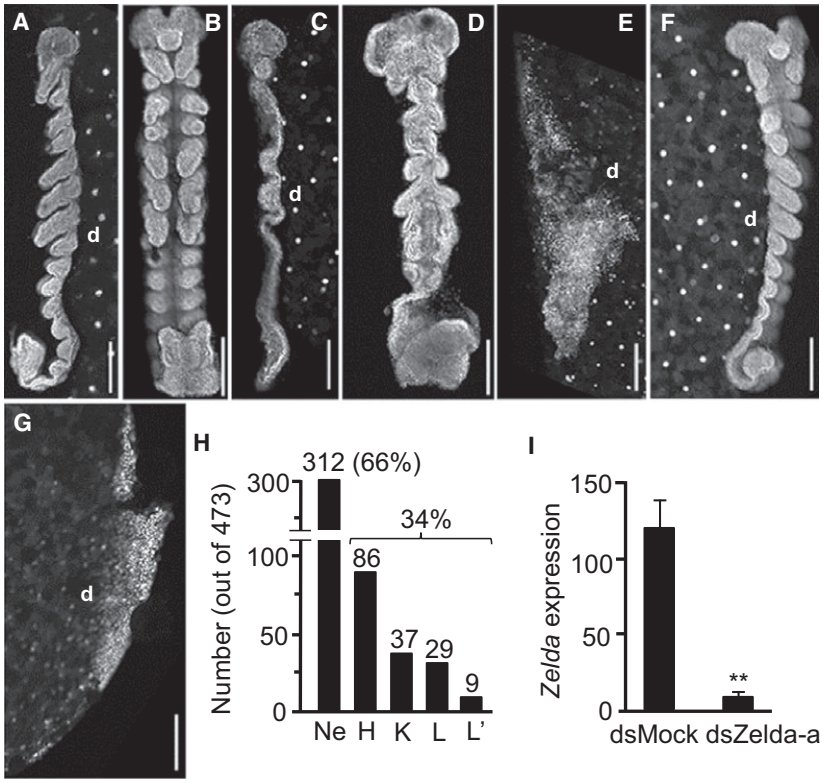

Fig. 7. Phenotypes observed in 4-day-old oothecae (embryos in stage ED4) from females of *Blattella germanica* treated with dsZelda-a. (A) Normal ED4 embryo in lateral view. (B) Normal ED4 embryo in ventral view. (C) Phenotype H in lateral view. (D) Phenotype H in ventral view. (E) Phenotype K in lateral 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.

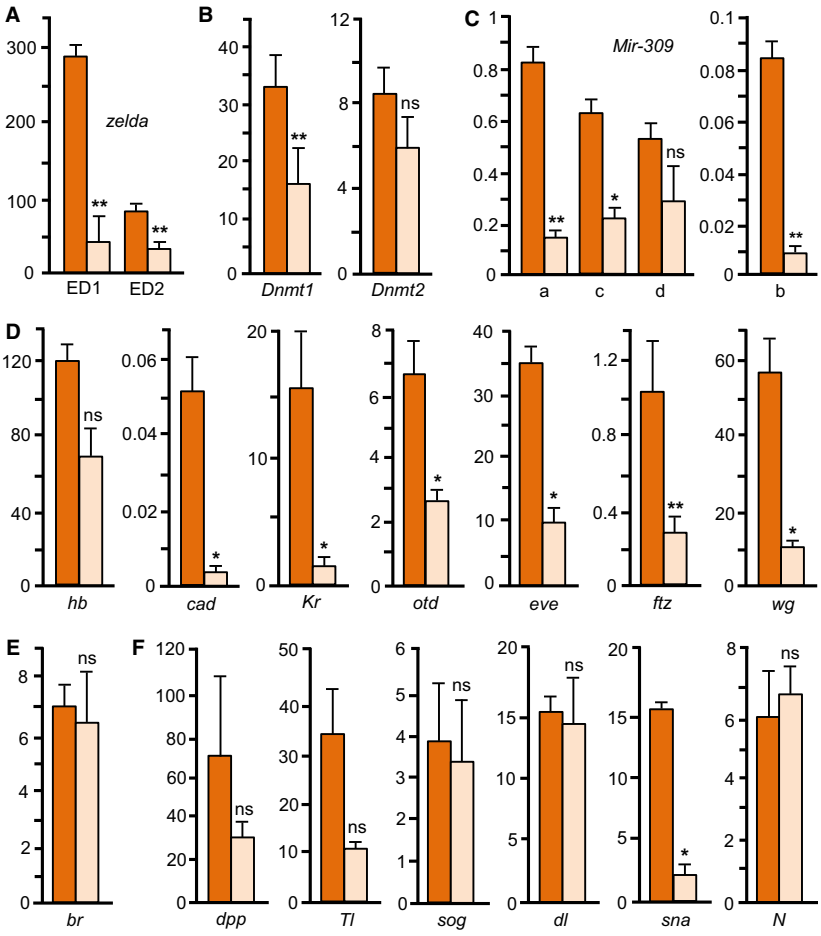

Fig. 8. Effects of BgZelda depletion on the expression of early embryogenesis genes in *Blattella germanica*. (A) Transcript decrease resulting from maternal RNAi of BgZelda, measured on ED1 and ED2. (B) Expression of *DNA methyltransferase-1* (*DNMT1*) and *DNA methyltransferase-2* (*DNMT2*). (C) Expression 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.