

RESEARCH

Open Access



Mayfly developmental atlas: developmental temporal expression atlas of the mayfly, *Ephemera vulgata*, reveals short germ-specific hox gene activation

Wouter P. D. Makkinje^{1*} , Sabrina Simon¹ , Inge Breukink¹ , Patrick Verbaarschot¹ , Ryuichiro Machida², M. Eric Schranz¹  and Robin van Velzen^{1*} 

Abstract

Background Over the course of evolution, insects have seen drastic changes in their mode of development. While insects with derived modes of development have been studied extensively, information on ancestral modes of development is lacking. To address this, we selected a member of one of the earliest lineages of extant flying insects, serving as an outgroup to the modern winged insects, the short germ, non-model mayfly *Ephemera vulgata* Linnaeus (Insecta: Ephemeroptera, Ephemeridae). We document the embryonic morphology throughout its development and establish a global temporal expression atlas.

Results DAPI staining was used to visualise developmental morphology to provide a frame of reference for the sequenced timepoints. A transcriptome was assembled from 3.2 billion Illumina RNAseq reads divided in 12 timepoints with 3 replicates per timepoint consisting of 35,091 putative genes. We identified 6,091 significantly differentially expressed genes (DEGs) and analysed them for broad expression patterns via gene ontology (GO) as well as for specific genes of interest. This revealed a U-shaped relationship between the sum of DEGs and developmental timepoints, over time, with the lowest number of DEGs at 72 hours after egg laying (hAEL). Based on a principal component analysis of sequenced timepoints, overall development could be divided into four stages, with a transcriptional turning point around katarépsis. Expression patterns of *zld* and *smg* showed a persistent negative correlation and revealed the maternal-to-zygotic transition (MZT), occurring 24 hAEL. The onset of development of some major anatomical structures, including the head, body, respiratory system, limb, muscle, and eye, are reported. Finally, we show that the ancestral short germ sequential mode of segmentation translates to a sequential Hox gene activation and find diverging expression patterns for *lab* and *pb*. We incorporate these patterns and morphological observations to an overview of the developmental timeline.

*Correspondence:

Wouter P. D. Makkinje
wouter.makkinje@wur.nl
Robin van Velzen
robin.vanvelzen@wur.nl

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusions With our comprehensive differential expression study, we demonstrate the versatility of our global temporal expression atlas. It has the capacity to contribute significantly to phylogenetic insights in early-diverging insect developmental biology and can be deployed in both molecular and genomic applications for future research.

Keywords Transcriptomics, Evo-devo, Mayfly, Hox genes, Embryogenesis, Morphology

Background

Understanding the intricacies of development that drive the formation of body plans requires understanding the gene expression patterns that precede it. Insects, in particular the model organism *Drosophila melanogaster*, have served as an effective model system to study these intricacies, and many aspects of its development are generally well understood [1, 2]. Still, *D. melanogaster* embryogenesis has multiple derived characteristics that are far from universal in insects [3, 4]. One of these characteristics is the derived long germ mode of development, where segment patterning occurs at roughly the same time during development [5–7]. Conversely, in the ancestral short germ mode of development, anterior segments are patterned first, while posterior segments are added in sequence afterward. Determining ancestral characteristics of insect development requires a comparative analysis of representatives of early-diverging lineages, such as wingless insects (Apterygota), dragonflies and mayflies [8, 9]. Such early-diverging lineages do not necessarily exhibit more ancestral characteristics than lineages that are considered derived because they represent the same evolutionary time since their common ancestor. Nevertheless, commonalities in developmental characteristics between independent early-diverging lineages suggest these are ancestral. Therefore, studying the developmental transcriptomics of early-diverging, short germ, insect lineages allows us to explore the changes in expression patterns responsible for the evolutionary development of these characteristics and insect body plans in general. However, studies exploring the differential expression of early-diverging insect developmental transcriptomics are lacking.

Developmental gene expression studies on model organisms mostly target single genes or gene families. The techniques used in these studies, mostly immunohistochemistry (IHC), in situ hybridisation (ISH), RNA interference (RNAi) or quantitative reverse transcription PCR (RT-qPCR), require sequence data, probe development and/or protocol optimization to achieve. When studying non-model organisms, collecting large-scale transcriptomic data is preferential to establish a foundation on which further in-depth studies can be built. Such datasets allow for studying both broad and single gene expression patterns, co-expression analyses, gene discovery and designing RNA probes for numerous molecular techniques. Fortunately, as RNA-seq technologies have become increasingly accessible, their utilization in

studying non-model organisms has become more and more attractive [10–12]. Some efforts to leverage transcriptomics to expand our scope beyond model organisms have already been made to answer a wide variety of questions for members of several orders, including but not limited to Lepidoptera, Hymenoptera, Hemiptera, Blattodea and Orthoptera [13–19]. However, these studies mainly focus on the transitions between life stages. In-depth analysis of embryonic expression patterns has mostly been reserved for relatively derived, holometabolous model organisms such as *D. melanogaster* and *Bombyx mori* [1, 20–22]. Currently, the only early diverging insect developmental transcriptome described is that of a species within Odonata, *Ischnura elegans*, revealing a transcriptomic turning point occurring around mid-embryogenesis [23]. However, it remains unknown whether these observations are a general characteristic of early-diverging insects or specific for Odonata. In addition, transcriptomics associated with other aspects of early-diverging insect development, like Hox gene activation and morphological development, remain unexplored.

To broaden our scope of embryonic development in early-diverging insect lineages, we use RNAseq to assemble a developmental transcriptome of a member of the mayflies, *Ephemera vulgata* L. (Insecta: Ephemeroptera, Ephemeridae). As an ephemeropteran, *E. vulgata* is member of a lineage that branched early in the onset of insect wing (Pterygote) evolution that can serve as an outgroup to the modern winged insects (Neoptera). Recently, interest in the development of mayflies has grown and efforts have been made to develop techniques to establish model organisms, study wing evolution, investigate the localisation of the expression of several developmental genes and explore the transcriptomics associated with metamorphosis [24–29]. Gonzalez et al. (2022) provided great insights into the genomics and expression localisation of mayfly Hox genes, a developmental gene family that determines body patterning and appendage identity, using *Hexagenia limbata*. With *H. limbata* being a non-model organism, they faced technical limitations with IHC specificity and development of ISH techniques [28]. Consequently, the expression of a subset of only 3 hox genes was studied, which can be expanded on using large-scale developmental transcriptomics. While the transcriptomics gained from RNA-seq doesn't provide the spatial information gained from IHC or ISH, it is impartial to optimisations and target

specificity and covers all transcriptional products. This approach was also implemented on larval and adult stages of the mayfly *Cloeon viridulum* by Si et al. (2017), describing the broad developmental patterns of metamorphosis, leaving those of the embryonic development of mayflies yet to be explored [29].

The embryogenesis of a different species within the Ephemeroptera, *Ephemera japonica*, was determined to follow short germ type segmentation [30]. While long germ development has been clearly established in specific holometabolans insects, some nuances surround the designation of short germ versus intermediate germ development [31]. These distinctions are based on the number of body segments specified prior to gastrulation, with only anterior segments included in short germ development, whereas thoracic and occasionally anterior abdominal segments are involved in intermediate germ development [32]. As *E. japonica* is closely related to *E. vulgata*, we assume they share a similar mode of development; however, we acknowledge the potential variability and assumptions inherent in this classification. The identity of these body structures is determined by Hox gene expression, a conserved set of homeobox genes grouped in two complexes in the long germ *D. melanogaster*, the Antennapedia complex (ANT-C) and Bithorax complex (BX-C), described by Lewis (1978) and Kaufman et al. (1990), respectively [33, 34]. Using *Drosophila* nomenclature, the ANT-C consists of *labial (lab)*, *proboscipedia (pb)*, *Deformed (Dfd)*, *Sex combs reduced (Scr)* and *Antennapedia (Antp)* and control the identity of the regions of the head and anterior thorax [33]. The BX-C consists of *Ultrabithorax (Ubx)*, *abdominal A (abd-A)* and *Abdominal B (Abd-B)*, controlling the identity the posterior thoracic and abdominal regions [34]. In the context of short germ development, it is compelling to determine if the expression of these clusters reflects the sequential mode of segmentation.

Another aspect of embryogenesis that varies between insects, is blastokinesis [4]. This refers to movements of the embryo inside the egg and generally involves two processes: anatrepsis, in which the embryo assumes an inverted position along the egg's anteroposterior axis by moving into the yolk during early embryogenesis; and katrepsis, the reversal of the embryo to the same anteroposterior axis as the egg by moving outside the yolk during mid-embryogenesis. Many nuances to this process exist as reviewed by Panfilio (2008). While members of the Ephemeroptera obtain an inverted position during early embryogenesis, this is not considered anatrepsis as no movement is involved in acquiring this position.

An integral transcriptional event that occurs during early insect embryogenesis is the maternal to zygotic transition. Throughout the cleavage stage, rapid division

of the zygotic nuclei, or syncytial cleavage, is coordinated by maternally deposited mRNA [35, 36]. A transition from dependency on maternal to zygotic mRNA occurs, called the maternal-to-zygotic transition (MZT), which lasts until blastoderm cellularization in *D. melanogaster* [37–39]. Two genes necessary for this event whose expression can be used as markers for the MZT are *smaug (smg)* and *zelda (zld)*, which are associated with maternal transcript destruction and zygotic genome activation, respectively [17, 40, 41]. Consistent with their functions, the expression levels of *smg* and *zld* in *Blattella germanica* were found to be negatively correlated, with *smg* showing an early peak, followed by *zld* [17].

The research questions this study aims to address include: What differential expression patterns underlie the development of ancestral versus derived characteristics, like short versus long germ development, in insects; are previously identified developmental transcriptomic events and expression patterns, including those from other early-diverging species, observable in *E. vulgata*; and can large-scale transcriptomic analyses reflect morphological developmental events? To this end, we present a global temporal expression atlas of the embryonic development of the mayfly *E. vulgata*. In addition, we describe the embryonic morphology through the course of development at each of the 12 time intervals we investigated. Differential gene expression (DGE) analyses shed light on the timing of developmental processes like the MZT, anatomical structure development, and short germ specific Hox gene expression.

Results

Temporal analysis of *E. vulgata* embryonic development

DAPI fluorescence staining of *E. vulgata* eggs revealed the embryonic morphology at the twelve sequenced timepoints (Fig. 1 and Fig. S1–S10 for additional angles). We use descriptions of the development of a closely related species, *E. japonica*, from Tojo and Machida (1997) to assess the timeline of the embryonic development of *E. vulgata*.

Twelve hours after egg laying (hAEL), nuclei have divided mitotically during the cleavage stage and migrated to the egg surface to form the blastoderm (12hAEL, Fig. 1A). After cellularisation (not shown), cells concentrated at the posterior pole of the egg become the embryo, forming the germ disc (gd), whilst remaining cells form an extraembryonic membrane, the serosa (ser; 24hAEL, Fig. 1B) [42]. As the germ disc develops into the germband and the anterior protocephalon (pce) and posterior protocorm (pco) differentiate, it invaginates into the yolk with its posterior end ahead while another extraembryonic membrane, the amnion, is produced from the embryonic margin (not shown; 36hAEL, Fig. 1C) [42]. The germband elongates as abdominal segments

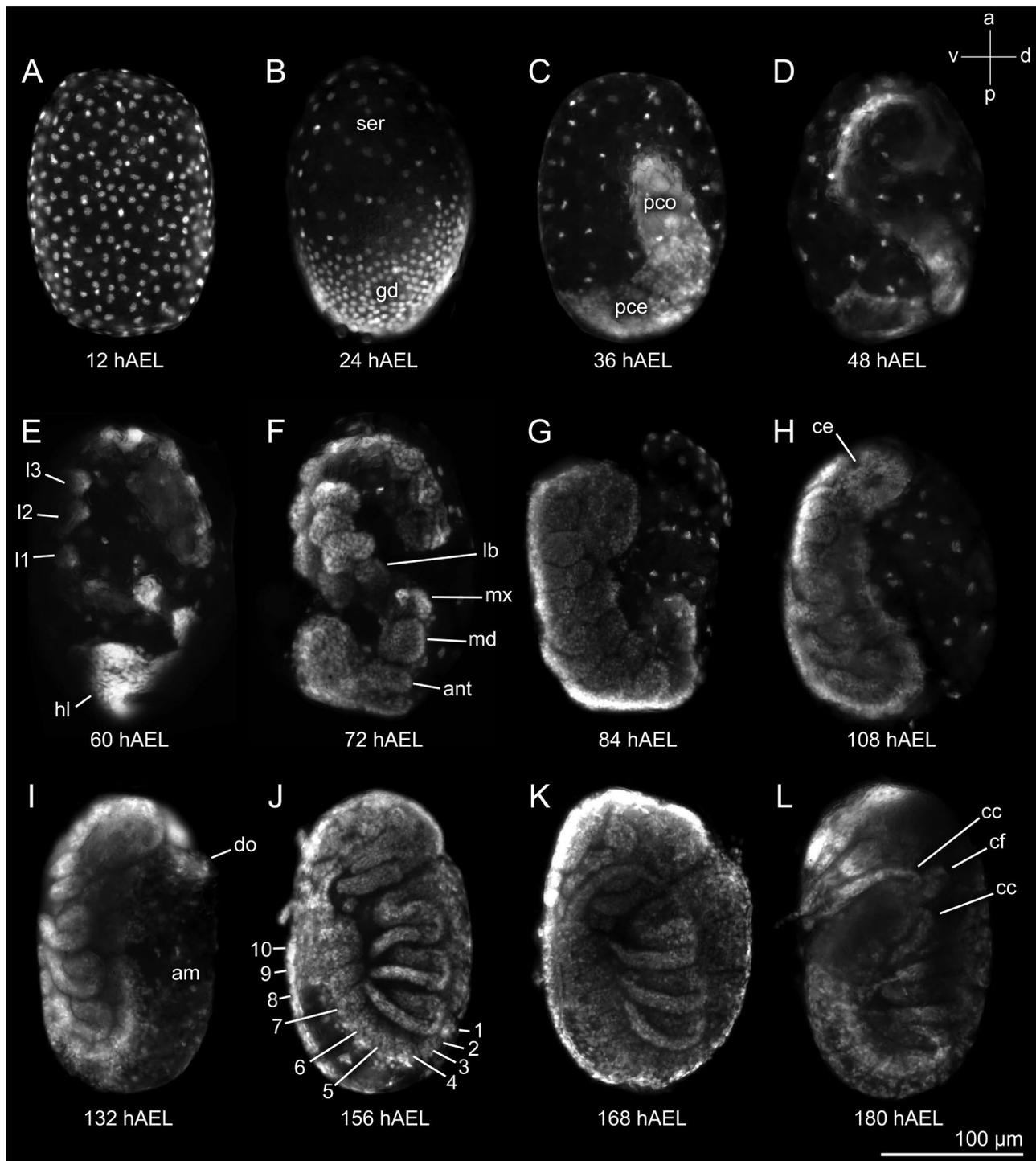


Fig. 1 The embryogenesis of *Ephemera vulgata* visualized with DAPI staining. Lateral views of embryonic morphology at successive sampled timepoints. **(A)** 12 hours after egg laying (hAEL). **(B)** 24 hAEL. **(C)** 36 hAEL. **(D)** 48 hAEL. **(E)** 60 hAEL. **(F)** 72 hAEL. **(G)** 84 hAEL. **(H)** 108 hAEL. **(I)** 132 hAEL. **(J)** Numbers 1–10 indicate the positions of the abdominal segments at 156 hAEL. **(K)** 168 hAEL. **(L)** 180 hAEL. Anatomical directions (top right) are based on final orientation of embryo where: a, anterior; d, dorsal; p, posterior; v, ventral. Scale bar (bottom right) = 100 μ m. ant, antenna; am, amnion; cc, cercus; ce, compound eye; cf, caudal filament; gd, germ disc; hl, head lobe; l1-3, prothoracic, mesothoracic, and metathoracic legs, respectively; lb, labium; md, mandible; mx, maxilla; pce, protocephalon; pco, protocorm; do, dorsal organ; ser, serosa

are added in sequence, assuming an S-shape characteristic for ephemeropteran development (48hAEL, Fig. 1D) [30]. While the embryo further elongates, segments and appendages become discernible (60hAEL, Fig. 1E). Appendages include the prothoracic (l1), mesothoracic (l2), and metathoracic legs (l3), the antennae (ant) and the parts that form the future mouth, the maxilla (mx), mandible (md) and labium (lb). At 72 hAEL, these appendages are in the process of articulation and the embryo noticeably thickens (72hAEL, Fig. 1F). At around 84 hAEL the amnio-serosal fold tears, initiating katatrepsis. The serosa begins to migrate toward the anterior pole of the egg and the amnion, followed by the embryo, appear outside of the yolk (not shown; 84hAEL, Fig. 1G) [30]. With the progressive migration of the serosa towards the anterior, the embryo moves along the ventral side of the egg with its head ahead. Finally, katatrepsis is completed when the embryo reverses its anteroposterior axis and positions itself on the ventral side of the egg (108hAEL, Fig. 1H). At this stage, the compound eye (ce), cerci (cc) and caudal filament (cf.) become discernible as well (Fig. S6). The serosa condenses just dorsally to the embryo's head to form the dorsal organ (do), while the amnion spreads over the area previously occupied by the serosa to cover the dorsal yolk as a provisional dorsal closure (132hAEL, Fig. 1I). As the embryo continues to grow, the lateral body walls expand dorsally to establish the definitive dorsal closure, replacing the provisional dorsal closure formed by the amnion (156hAEL, Fig. 1J). The definitive dorsal closure completes at around 168 hAEL, and the yolk is fully enclosed (168hAEL, Fig. 1K). At the last stage of development, the abdomen largely grows, being tucked to the left side of the head and the insect is ready to hatch (180hAEL, Fig. 1L).

Processing and validation of developmental gene expression data

A total of 3.2 billion 150 bp-long paired end (PE) Illumina seq reads were assembled into a raw transcriptome comprising 580,443 contigs with an overall read alignment rate of 97.57%. Quality and redundancy filtering resulted in a transcriptome consisting of 87,767 transcript isoforms having an N50 of 3,467 with 50,547 putative genes. Of these, 35,091 putative genes passed the minimal expression filtering step. The resulting set of transcripts has a BUSCO completeness of 96.9% (of which 1021 complete single-copy, 304 complete duplicated, 22 fragmented and 20 missing BUSCOs), marking it as a relatively complete and high-quality transcriptome that can be used as reference for DGE analyses. A Principal Component Analysis (PCA) plot that depicts the sample and replicate relationships based on expression count matrices is shown in Fig. 2A (and Fig. S11). Except for samples belonging to 180hAEL, biological replicates group

together while developmental timepoints separate, which is supported by correlation matrix data (Fig. S12). Therefore, we averaged replicate data and excluded 180hAEL from subsequent analyses.

Pairwise gene expression level comparisons between samples throughout development determined that 6,091 out of 35,091 genes were significantly differentially expressed. Stage-specific differentially expressed gene (DEG) count comparisons are shown in Table S1. Notably, the first stage (12hAEL) displayed the highest number of DEGs compared to other samples, followed by the last stage (168hAEL). This reflects a U-shaped relationship between developmental time and the comparative stage-specific sum of DEGs (Fig. S13). Hierarchical clustering on centred and normalized expression data identified 23 clusters, each representing a distinct developmental gene expression pattern. A complete overview of DEG-specific expression levels and their clustering is shown in the heatmap in Fig. 2B. Next, we analysed the biological processes related to these patterns using a gene ontology (GO) enrichment analysis and summarised the terms to relevant GOslims. A selection of clusters with notable expression patterns and functional information of interest to this study are presented in Figs. 3, 4 and 5, while a complete overview of all clusters is displayed in Figs. S14–17. Normalized gene expression patterns of genes of interest are visualized in separate heatmaps (Figs. 3, 4 and 5) or added to the expression patterns of clusters (Fig. 5).

The maternal to zygotic transition

Two distinct clusters appeared as particularly interesting candidates to identify the MZT (Fig. 3A, B). Genes belonging to cluster 14 are up-regulated only at 12 hAEL and neutrally regulated during all subsequent stages. Conversely, genes in cluster 3 display the opposite expression pattern, reflecting a shift in expression between 12 hAEL and 24 hAEL. Indeed, these clusters show a disparity in the presence of maternally deposited transcripts, with cluster 14 containing the highest number of maternal genes of all clusters at a proportion of 42%, while just 13% of the transcripts in cluster 3 are deposited maternally. The expression pattern associated with the maternally deposited transcripts specifically within cluster 14 also reflects the cluster's global expression pattern (Fig. S18). Cluster 14 is mostly enriched for genes associated with RNA metabolic process (GO:0016070). Other significantly enriched GO terms associated with cluster 14 involve those expected during the cleavage stage, where rapid nuclear division is important, including but not limited to cellular metabolic process (GO:0044237), protein catabolic process (GO:0030163), organic substance metabolic process (GO:0071704), cell cycle (GO:0007049), and DNA metabolic process (GO:0006259). In contrast,

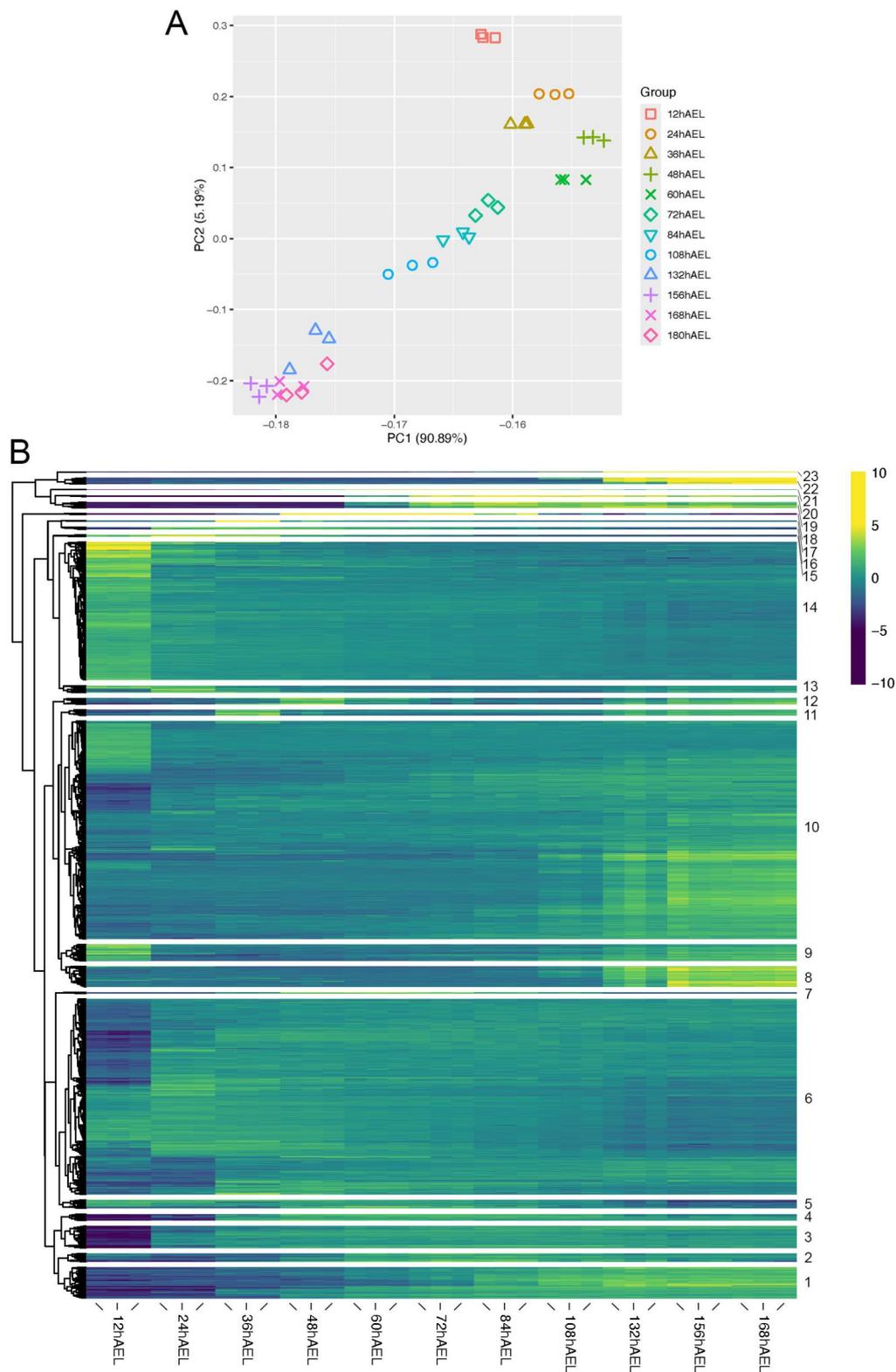


Fig. 2 Embryonic differential gene expression of *Ephemera vulgata*. **(A)** Principal components 1 and 2 of a principal component analysis depicting sample relatedness of biological replicates within sampled timepoints in hours after egg laying (hAEL). **(B)** Hierarchically clustered heatmap with dendrogram (left) depicting 6,091 significantly differentially expressed genes (fold change > 4 and FDR of < 1e-3). Successive sampled timepoints are presented from left to right, with three biological replicates for each timepoint (n=450 eggs) separately included. Dendrogram is cut with K=23 and resulting clusters are separated and identified by number (right).

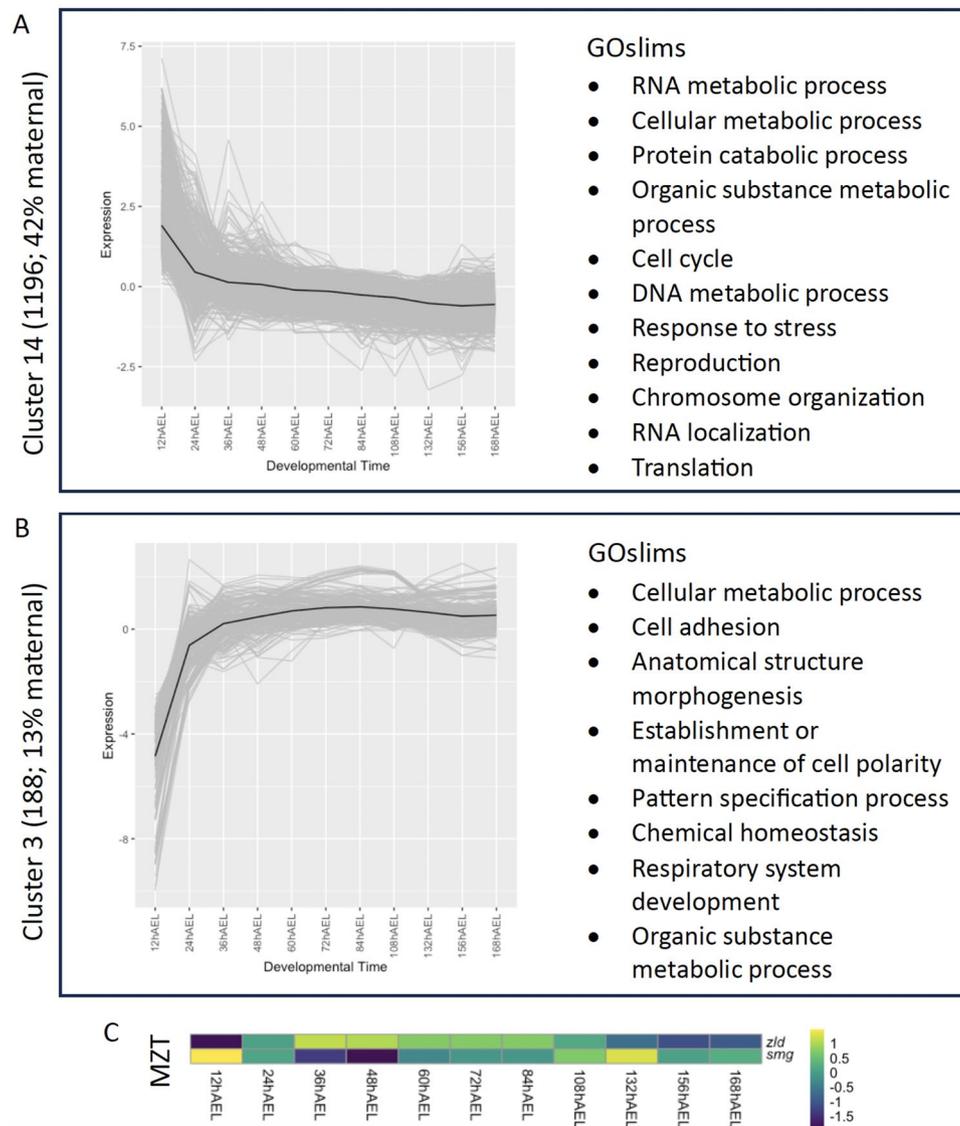


Fig. 3 Maternal to zygotic transition-associated gene expression patterns. **(A, B)** Expression patterns throughout development of and GOSlims associated with clusters 14 ($n = 1196$, of which 42% is deposited maternally) and 3 ($n = 188$, of which 13% is deposited maternally), respectively. Each timepoint shows mean values of biological replicates ($n = 1350$ eggs). These clusters reflect a shift in expression at 24 hours after egg laying (hAEL) between expression of maternal and zygotic genes. Total mean values are depicted in black. **(C)** Developmental expression levels of MZT related genes. *zld*, *zelda*; *smg*, *smaug*. Expression data presented in log₂-centered TMM (+1) values.

genes belonging to cluster 3 are associated with processes indicative of germband development and subsequent stages like cell adhesion (GO:0007155), anatomical structure morphogenesis (GO:0009653), pattern specification process (GO:0007389) and respiratory system development (GO:0060541).

This observation is substantiated with the gene expression patterns belonging to *smg* and *zld* (Fig. 3C). Consistent with their functions, expression of *smg* and *zld* display a strong negative correlation, $r^2 = -0.90$, $p < .001$, during the first 36 hAEL. At 12 hAEL, *smg* is highly up-regulated while *zld* is down-regulated. A shift in expression occurs at 24 hAEL, where both are neutrally

regulated, which evolves to *zld* being highly up-regulated at 36 hAEL while *smg* is down-regulated. Because of their expression patterns during the later stages of development, *smg* and *zld* are not part of either cluster 14 or 3.

Developmental timing of anatomical structures

Next, we aimed to verify and enhance the developmental overview gathered from our fluorescence images by associating differential expression data to the development of anatomical structures, specifically that of muscle structures, the respiratory system and eyes. Analysis of clusters 8 and 12 provide a clear image of muscle structure and respiratory system development (see Fig. 4A & B).

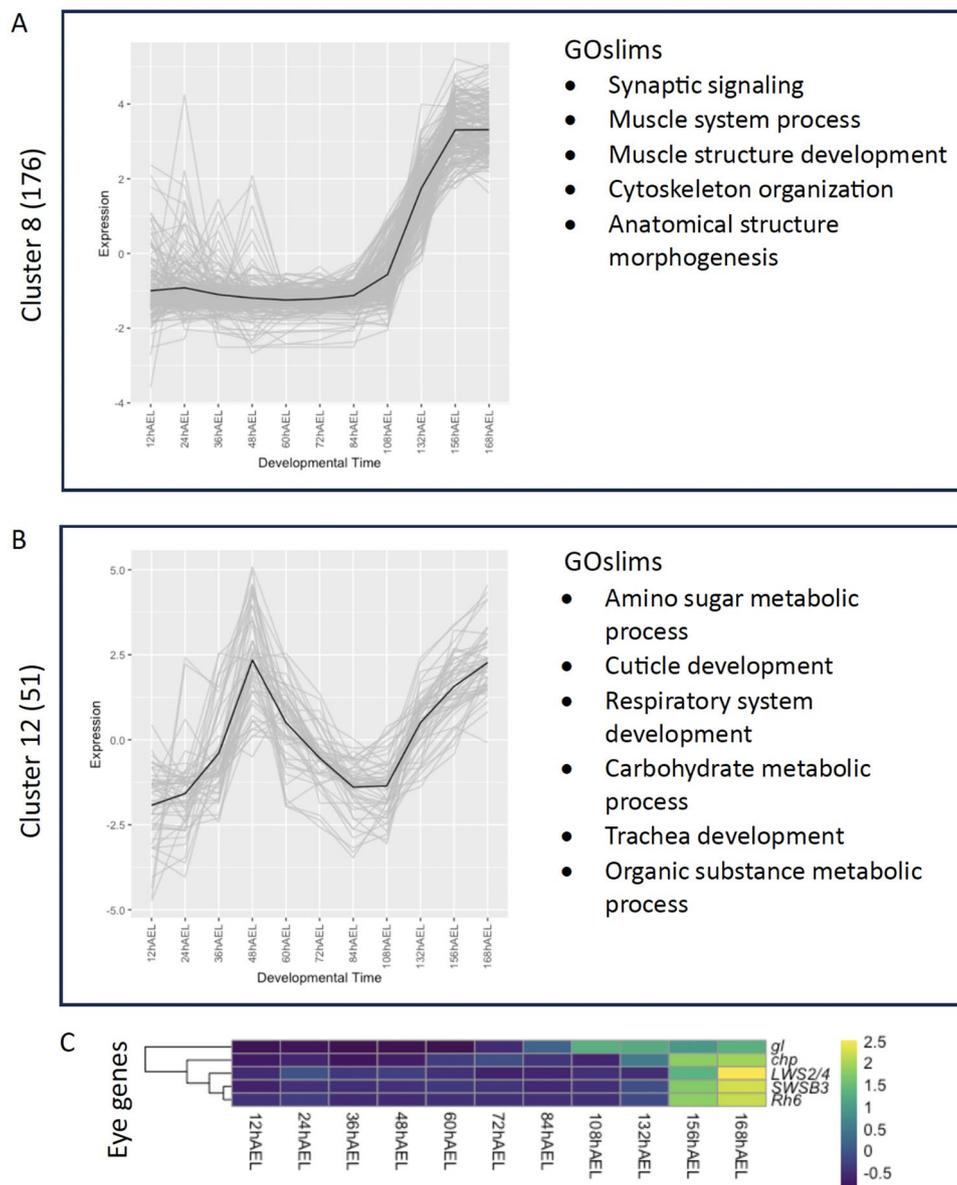


Fig. 4 Anatomical structure development-associated gene expression patterns. **(A, B)** Expression patterns throughout development of and GOslims associated with clusters 8 ($n=176$) and 12 ($n=51$), respectively. GOslim results from cluster 8 reveal that muscle structure development occurs from 132 hours after egg laying (hAEL) onwards and cluster 12 reveals a peak in respiratory system development at 48 hAEL and from 132 hAEL onwards. Each timepoint shows mean values of biological replicates ($n=1350$ eggs). Total mean values are depicted in black. **(C)** Developmental expression levels of eye related genes. *gl*, *glass*; *chp*, *choaptin*; *LWS2/4*, *Long wavelength-sensitive cone opsin 2/4*; *SWSB3*, *Short wavelength-sensitive B opsin 3*; *Rh6*, *Rhodopsin 6*. Expression data presented in log₂-centered TMM (+1) values.

Genes in cluster 8 are associated with muscle development, being uniquely enriched for GO terms related to muscle system process (GO:0003012) and enriched for muscle structure development (GO:0061061). The remaining terms enriched in this cluster, namely synaptic signalling (GO:0099536), cytoskeleton organization (GO:0007010) and anatomical structure morphogenesis (GO:0009653), could also be considered as part of this category. Given the expression pattern associated with

cluster 8, muscle development seems to occur mostly as of 132 hAEL onwards.

Cluster 12 is uniquely enriched for GO terms associated with trachea development (GO:0060438) and genes associated with it show an isolated peak of up-regulation at 48 hAEL and are again up-regulated from 132 hAEL onwards. In addition, respiratory system development (GO:0060541), cuticle development (GO:0042335) and several metabolic processes, including amino sugar (GO:0006040), carbohydrate (GO:0005975) and organic

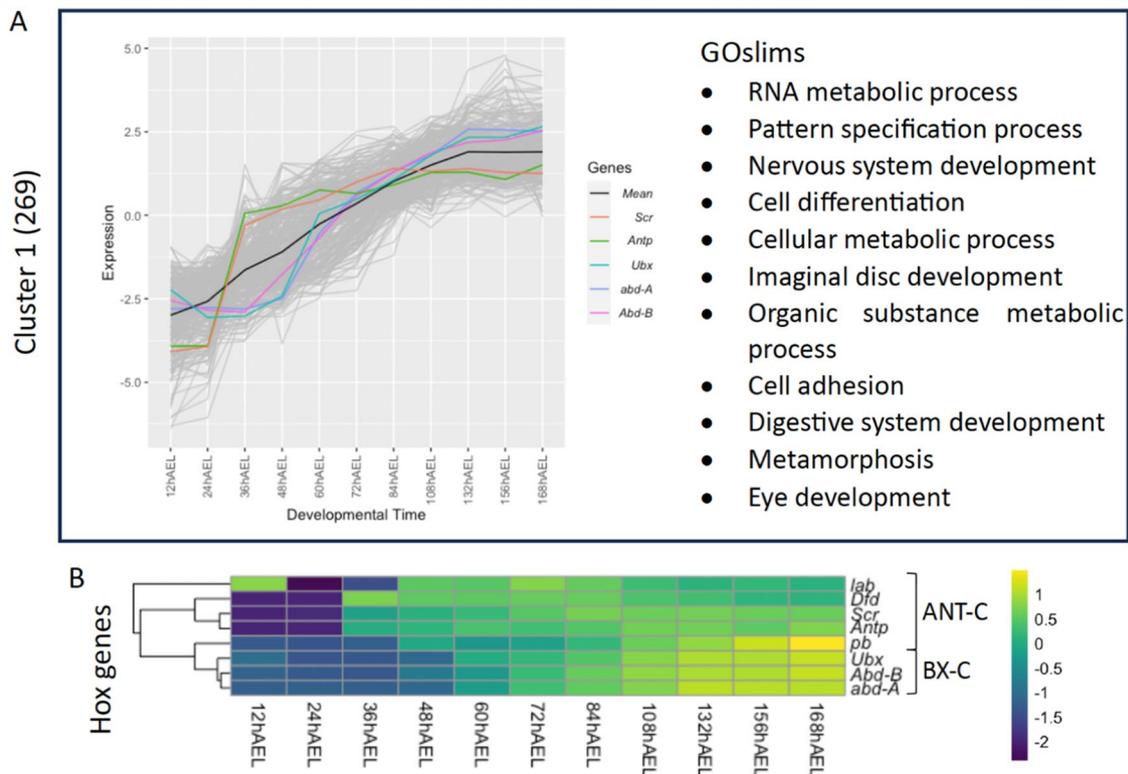


Fig. 5 Hox gene-associated gene expression patterns. **(A)** Expression patterns throughout development of and GOslims associated with clusters 1 ($n = 269$). Each timepoint shows mean values of biological replicates ($n = 1350$ eggs). Coloured lines of the expression levels of Hox genes belonging to cluster 1 are drawn in the graph with the overall mean value depicted in black. **(B)** Developmental expression levels of Hox genes. *abd-A*, abdominal A; *Abd-B*, Abdominal B; ANT-C, Antennapedia complex; *Antp*, Antennapedia; BX-C, Bithorax complex; *Dfd*, Deformed; *lab*, labial; *pb*, proboscipedia; *Scr*, Sex combs reduced; *Ubx*, Ultrabithorax. Expression data presented in log₂-centered TMM (+1) values

substance (GO:0071704) related terms were found to be significantly enriched.

Finally, we determined the timepoint associated with the development of the eye using the expression of a homolog to a gene used previously as a marker for larval photoreceptor differentiation in *Tribolium castaneum*, *glass (gl)*; Fig. 4C) [43]. Homologs for several opsins described by Almudi et al. (2020) were also included [25]. Expression of *gl* is slightly up-regulated at 86 hAEL during katarptosis, which increases over time, and most of the identified opsins follow with up-regulation later during development as of 156 hAEL.

Hox gene activation reflects sequential mode of segmentation in short germ insects

Individual transcripts could be assigned to all insect Hox genes using an arthropod gene tree (Fig. S19). Hox genes were fully assembled except for *lab* and *Scr*, which were partially assembled. Hox genes display two main expression patterns (Fig. 5B). Most members of ANT-C are upregulated at 36 hAEL, prior to members of BX-C at around 60 hAEL, reflecting the temporal disparity between the determination of the identity of anterior and posterior segments in short germ insects. Furthermore,

the ANT-C cluster displays a sudden increase in expression, while BX-C gene up-regulation is more gradual, consistent with the more progressive quality of thoracic and abdominal patterning. The expression patterns of *pb* and *lab* diverge from the main expression patterns, both seeing up-regulation as of 48 hAEL with *lab* displaying an initial peak at 12 hAEL.

Despite the dissimilarities in expression patterns between members of ANT-C and BX-C, *Scr*, *Antp*, *Ubx*, *abd-A* and *Abd-B*, all clustered within cluster 1 of the DGE study, albeit in two distinct subclusters (Fig. S20). Leveraging our transcriptome-wide analysis, we identified genes grouping within these subclusters. Some notable genes are the homeobox genes *LIM homeobox 1 (Lim1)* and *HGTX*, *mab-21* and *disco-related (disco-r)* clustering with members of ANT-C, and homeobox gene *cut*, paired box gene *shaven*, *scratch (scrt)* and *pannier (pnr)* clustering with members of BX-C.

Discussion

Using precisely timed morphological and transcriptomic data, we uncovered a wide spectrum of developmental intricacies of the early-diverging short germ insect, *E. vulgata*. Using large-scale DGE analysis and genetic

markers, we identified the developmental timing of key anatomical structures and events, including short germ specific hox gene expression, providing a novel perspective on this conserved gene set.

Sampling and creating a reference of *E. vulgata* development

While the embryonic development of *E. vulgata* is similar to that of *E. japonica*, described in Tojo & Machida (1997), the timing of its developmental stages had to be established for accurate sampling and data interpretation [30]. Consequently, we created a clear developmental reference overview. Developmental speed fluctuates based on several factors caused by biological variation. The speed of development of individual eggs can differ due to the degree of oxygenation [44]. During *E. vulgata* oviposition, a mucous is deposited along with the eggs, clumping them together. We found that eggs positioned on the inside developed slower than eggs positioned on the outside of the egg clump, likely caused by differences in available oxygen. This was considered during sampling by collecting eggs along the outer ridges of the egg clump. The PCA shows that expression signatures of later time-points were less differentiated than earlier time-points, which is likely caused by both a build-up of variability in developmental speed and the accumulation of developmental processes, resulting in less distinction between stages towards the end of development (Fig. 2A). Overall, however, developmental time points exhibited distinct separation, and biological replicates clustered accurately.

Global developmental DGE patterns

Based on the PCA, we found that developmental time points can be broadly aggregated into four groups: [1] 12hAEL; [2] 24hAEL, 36hAEL, 48hAEL, 60hAEL; [3] 72hAEL, 84hAEL, 108hAEL; [4] 132hAEL, 156hAEL, 168hAEL, 180hAEL (Fig. 2A). In accordance with observations from the large-scale developmental transcriptional analysis of *I. elegans*, this suggests that a transcriptional turning point occurs around katarrepsis, between 72 hAEL and 108 hAEL [23].

We observed a U-shaped pattern in pairwise comparisons of the number of significantly DEGs at successive developmental timepoints, with the highest deviation at 12 hAEL and a minimum at 72 hAEL (Fig. S13). This pattern is reminiscent of the hourglass model of evolutionary development, where, during the phylotypic period, a mid-developmental transition of conserved gene expression occurs [45–47]. However, the hourglass model is a phenomenon that becomes evident when performing inter-species temporal differential gene expression analysis and is therefore likely unrelated to our observations. In *D. melanogaster* embryogenesis, inter-embryo variability caused by stochastic perturbation was found to

be at a minimum during the phylotypic period, further demonstrating its conserved nature [48]. Whether our observations represent a similar conservation in gene expression at 72hAEL or is another illustration of the transcriptional turning point occurring around katarrepsis, is uncertain. The embryonic morphology at 72 hAEL (Fig. 1F) aligns more with the morphology of the phylotypic period, characterised in arthropods as the fully segmented germband stage, than the timepoint at which katarrepsis occurs at 84 hAEL in *E. vulgata* (Fig. 1G) [49]. However, all Hox genes see up-regulation as of 60 hAEL, which has been implicated with the phylotypic period in bilaterians (Fig. 5B) [46, 49]. Based on this, we can ascertain that the phylotypic period is likely to occur somewhere around 60 hAEL and 72 hAEL, but highlights discrepancies between the phenotypic and molecular definition of the phylotypic stage in our dataset.

Timing of key developmental processes and anatomical development

Specifying the onset of developmental processes can be challenging, as development is a dynamic biological process with single regulatory genes often having effect on the development of many structures. Nevertheless, literature based genetic markers can be used to indicate when a transition to active development of such structures has occurred. A chronological outline of the development of *E. vulgata*, including the timing of all processes observed here, is presented in Fig. 6.

During the MZT in the hemimetabolous insect *B. germanica*, *smg* peaks in expression at 8 hAEL (2% of embryo development), followed by *zld* peaking at 24 hAEL (6% of embryo development), showing a negative correlation with each other during the first 48 h (12%) of development [17]. This is followed by an abrupt decrease in expression of both genes that remains until the end of development. *D. melanogaster* displays a similar pattern during early embryogenesis but sees a maintained upregulation of both *smg* and *zld* during late embryogenesis. In *E. vulgata*, we observe the same pattern of strong negative correlation during early embryogenesis with initial up-regulation of *smg* at 12hAEL (6.7% of embryo development) followed by *zld* at 36hAEL (20% of embryo development). However, *zld* expression is maintained until mid-embryogenesis when it decreases and is again replaced with a resurgence of *smg* expression. Thus, in contrast to findings from *B. germanica* and *D. melanogaster*, negative correlation between *zld* and *smg* is maintained throughout embryogenesis. To ascertain if this pattern is ancestral, similarly detailed transcriptomic information from other early-diverging insect species is needed. Expression patterns and GO enrichment analysis of clusters 14 and 3, coupled with their associated proportions of maternally deposited transcripts based on

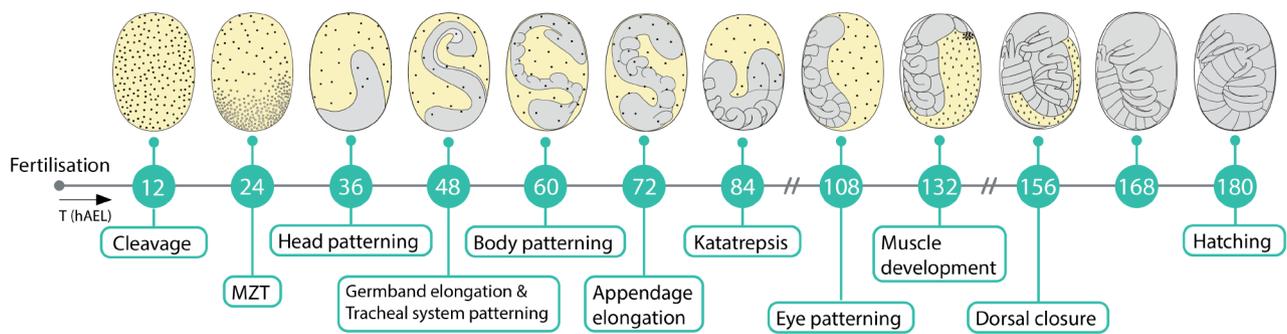


Fig. 6 An overview of the embryonic development of *Ephemera vulgata*. Graphical representations of the embryonic morphology are correlated with the development of anatomical structures and key events on a timeline in hours after egg laying (hAEL). Time between samples is 12 h, except between the intersected lines where there are 24 h between samples

annotations from *D. melanogaster* and expression patterns of *smg* and *zld*, pinpoint the MZT to 24 hAEL.

Cluster 12 specified a peak in tracheal system development at 48 hAEL. In Odonata, the initial visual indication of tracheal system development appears during early pre-revolution after 10 days of development [50]. This morphologically coincides with around 60 hAEL in *E. vulgata*, just after germband elongation at 48 hAEL (Fig. 1D, E). However, specification of tracheal placodes precedes morphological visibility, beginning at 5 hAEL in *D. melanogaster* as the germband elongates, which is consistent with our observations [35, 51]. In addition, cuticle formation was significantly associated with cluster 12 and cluster 18 (Fig. 4B and Fig. S16). The serosal cuticle was found to be secreted upon the completion of anatrepsis in *E. japonica*, occurring at 48 hAEL in *E. vulgata*, aligning with expectations and marking 48 hAEL as serosal cuticle formation [30].

In *T. castaneum*, *gl* expression was used to detect eye development initiation at the fully extended germband stage [43]. In *E. vulgata* embryogenesis, the germband reaches full extension at around 72 hAEL, whereas *gl* sees up-regulation at 84 hAEL. However, *gl* expression was found to be very low initially in *T. castaneum*, which is in line with our observations, showing neutral regulation at 72 hAEL. Expectedly, *gl* expression is succeeded by that of several opsins.

Finally, our GO analysis found that genes in cluster 8 are significantly associated with muscle development, which see up-regulation from 132 hAEL onwards. During Odonata embryogenesis, the first signs of muscle formation are observed at the onset of katatrepsis, which occurs at 84 hAEL in *E. vulgata* and therefore doesn't align with our findings [50]. We therefore investigated the specific functions of the genes annotated with GO terms related to muscle development in cluster 8 and found them to be associated with maturation and specialization of existing muscle structures, rather than patterning of muscle founder cells. The genes in cluster 8 include *wings up A*

(*wupA*) and *Ca²⁺-channel protein α_1 subunit D* (*Ca- α 1D*), associated with regulation of muscle contraction, *sallimus* (*sls*) and *Z band alternatively spliced PDZ-motif protein 52* (*Zasp52*), associated with myofibril assembly and *myomesin and myosin binding protein* (*MnM*) and *flightin* (*fln*), which are expressed in adult muscle structures [52–57]. Analysing the expression of genes known to function in the patterning of muscle founder cells, namely *slouch* (*slou*), *apterous* (*ap*), *myoblast city* (*mbc*) and *twist* (*twi*), we find up-regulation between 72 hAEL and 84 hAEL, aligning with morphological observations from Odonata (Fig. S21) [58–61]. Therefore, we associate 132 hAEL with a transition to muscle development and maturation rather than patterning.

Sequential Hox gene activation in short germ insects

Hox gene expression patterns in the short germ insect *E. vulgata* contrast that of long germ insects such as *D. melanogaster*, as the sequential nature of short germ associated segmentation is translated to a sequential Hox gene complex activation (Fig. 5B) [17, 62, 63]. While earlier studies using in situ hybridizations qualitatively imply this pattern in other short germ insects, our large-scale gene expression analysis is the first to quantifiably reveal the temporality of all Hox genes combined [28, 64]. Similar to our research, Ylla et al. (2018) studied the developmental expression patterns of the short germ Blattodea, *B. germanica*, and compared this to *D. melanogaster* (based on data from the modENCODE project) [17, 20], highlighting Hox gene expression. They found no sequential activation between the expression of the ANT-C and BX-C genes in either insect. Rather, most Hox genes collectively see up-regulation at the same developmental timepoint of 144 hAEL in *B. germanica*, although the presence of sequential activation cannot be concluded due to a gap of 96 h in sampling of the previous timepoint.

In vertebrates, Hox gene expression is generally observed to follow the same sequence temporally as their

position in the genome from 5' to 3', referred to as temporal colinearity, although this appears not to be the case for *D. melanogaster* [62, 65, 66]. The resolution of our sampling is not sufficient to clearly order the sequence of activation for each Hox gene individually. Also, due to the incomplete assembly and divergent expression pattern of *lab*, its result cannot be considered conclusive. Nevertheless, the general order of activation of the remaining Hox genes is in accordance with our morphological observations and can be summarized as follows: [1] at 36 hAEL: *Dfd*, *Scr*, *Antp*; [2] at 48 hAEL: *pb*; [3] at 60 hAEL: *Ubx*, *abd-A*, *Abd-B*. According to Hox gene positions in the genome of the close relative *Ephemera danica*, in sequence: *pb*, *Dfd*, *Scr*, *Antp*, *Ubx*, *abd-A*, *Abd-B*, *lab*; our observed order of activation is inconsistent with their position [28]. Similarly, in the short germ insect *T. castaneum*, *mxp* (ortholog of *pb*), is positioned upstream of *Dfd* as well and is first detected in embryos that have formed four to five segments, while *Dfd* transcripts are already identified during the very late blastoderm stage [67, 68]. This leads us to conclude that, like in *D. melanogaster* and *T. castaneum*, Hox gene expression does not display temporal collinearity in *E. vulgata*.

Based on the observed order of Hox gene activation, we find a discrepancy in our expectations for short germ insects, specifically regarding the timing of activation of the thoracic patterning gene *Antp* (Fig. 5B). In the apterygote *Thermobia domestica*, *Antp* expression is first detected in three thoracic stripes [64]. Since *Antp* is highly likely to have the same function in *E. vulgata*, this suggests that their development should be classified as intermediate germ type. However, the 12-hour intervals in our dataset lack the resolution needed to confirm this with certainty. Additionally, the relatively strong upregulation of *Dfd* at 36hAEL could imply earlier activation compared to *Antp* and *Scr*. Consequently, we adhere to the morphology-based classification proposed by Tojo & Machida (1997).

Conclusion

We demonstrate the versatility of our global temporal expression atlas by revealing both overarching and distinct patterns of developmental expression of the short germ insect *E. vulgata*. By performing RNAseq and morphology studies on meticulously sampled developmental timepoints we unravelled intricacies of the embryogenesis of a field-collected non-model organism. We provide valuable information on a crucial evolutionary step relevant to testing hypotheses about changes in the developmental expression of genes throughout evolution, particularly those related to key evolutionary innovations like metamorphosis, wing development, or short germ development. In addition, our sequence data has the capacity to contribute significantly to phylogenetic

insights in early-diverging insect developmental biology. These data can be deployed in both molecular applications, such as RNAi and ISH and genomic applications for future research. Including similar studies from representatives of other insect orders in a comparative manner can infer evolutionary context and assess the evolution and conservation of developmental processes in insects.

Methods

Egg collection

After observing copulation, all fertilised female *E. vulgata* imagos were collected at once on 31-05-2021 between 19:45 and 21:15 in a field adjacent to the Valleikanaal near Wageningen, The Netherlands (GPS coordinates 51.966288, 5.618384). Within two hours after copulation, females were placed ventral surface down in tap water to induce oviposition. During oviposition, a mucous was deposited that hardened upon contact with water. *E. vulgata* females laid a roughly estimated number of 3,000 eggs per clutch that hatched after approximately 8 days at room temperature (RT), lacking diapause. Nine egg clutches were each sampled in tandem, starting at 12 hAEL at intervals of 12 h, except for samples 108hAEL, 132hAEL, 156hAEL, which were sampled at 1-day intervals, for a total of 12 timepoints and 108 samples. For each RNA sample, a corresponding and separate fluorescence nuclear staining sample from the same clutch was collected and processed according to the fluorescence staining protocol before storage. RNA samples each contained 150 eggs which were stored in RNAlater® (Qiagen, Valencia, CA, USA) at -80 °C until RNA extraction was performed.

Fluorescence nuclear staining and visualization

Eggs were heated on a heat block for five minutes at 99°C to stop development, followed by overnight fixation in 4% paraformaldehyde (PFA) – phosphate-buffered saline (PBS)/Tween-20 (0.1%; PBT) under constant agitation. After three rounds of washing with PBT for five minutes, the fixed mucous membrane was removed manually. Eggs were fixed in a Heptane/12% PFA-PBT emulsion for 25 minutes on an orbital shaker at 300 rpm. The emulsion was replaced with fresh Heptane and MeOH cracking was performed by adding MeOH at a 1:1 ratio, followed by immediate vigorous shaking for three rounds of 30 seconds. The emulsion was replaced with fresh MeOH and stored at -20°C until further use. The following steps for DNA-specific fluorescent dye, 4',6-diamidino-2-phenylindole dihydrochloride (DAPI, Sigma-Aldrich, St. Louis, MO, USA) staining of *E. vulgata* eggs were adjusted for development time as described below because the eggshell matures and becomes impermeable to DAPI after two days. After serially diluting MeOH in four steps with decreasing increments of 25% in PBT and

three rounds of PBT washing, eggs up to two days old were post-fixed overnight in 4% PFA-PBT. The PFA solution was washed out in three rounds of 5 minutes using PBT and stained with 300ng/mL DAPI-PBT for 20 minutes. The eggshell of eggs older than two days was punctured manually prior to overnight post-fixation using Carnoy's fixative. Subsequently, the Carnoy's fixative was washed out in four rounds using PBT and the eggshell was removed manually using fine forceps. A one hour 20 µg/ml proteinase K treatment was used to permeabilise the embryos prior to DAPI staining with 300ng/mL DAPI-PBT for 20 minutes. Microscopic images were taken using a Zeiss AxioImager.Z2 microscope with a Zeiss AxioCam 506 mono camera and a 60 N-C 1" 1.0x camera adapter (Carl Zeiss AG, Oberkochen, Germany). To visualise DAPI fluorescence, excitation and emission filters were set to 358nm and 461nm, respectively. Image stacking and processing were performed with CombineZP (v1.0) using Soft stacking or Pyramid stacking and FIJI [69], respectively.

RNA extraction and sequencing

Samples were thawed on ice and eggs were transferred to 50 µl of TRI Reagent® (Zymo Research, Irvine, CA, USA). A pestle was used to homogenize the eggs, followed by adding 500 µl TRI Reagent®. The contents were triturated using a syringe (25G; BD, Franklin Lakes, NJ, USA) to further homogenize the tissue. After vortexing and letting the contents settle for 5 min, 150 µl chloroform was added and mixed by inverting the tube. The mixture was left to settle for 3 min, followed by centrifuging at 11,000 RPM for 15 min at 4 °C. The supernatant was then collected and used for step 2 of the RNeasy® Mini Kit (Qiagen, Valencia, CA, USA) protocol, continuing onwards. For each time point, the 9 RNA samples were randomly grouped into 3 groups of 3, thus producing 3 pooled samples ($n=450$ eggs per replicate), by centrifuging lysate mixtures on the same column during step 3 of the RNeasy® Mini Kit protocol. A total of 36 RNA extraction samples were used for mRNA sequencing. cDNA library preparation and short-read (150 bp) PE sequencing using the Illumina NovaSeq 6000 system were performed by Novogene (Cambridge, UK).

Transcriptome assembly and functional annotation

FastQC v0.11.9 [70] Quality assessment indicated no need for read quality filtering or trimming (see supplementary information). SortMeRNA v4.3.4 [71] with provided rRNA databases was used to filter out rRNA sequences. Reads needed re-pairing afterwards which was accomplished using the BMap, Repair.sh release 9/11/2016 [72] script. FastQ Screen v0.14.0 [73] was used to assess contamination of which no significant amount was found. *de novo* transcriptome assembly was

conducted using Trinity v2.8.5 [74], applying in silico read normalization and default parameters. The read alignment rate was assessed by mapping the RNAseq reads to the transcriptome using bowtie2 v2.4.5 [75]. Post-processing consisted of removing transcripts with an open reading frame encoding less than 66 amino acids using TransDecoder.LongOrfs v5.5.0 [76]. Redundancy was removed by means of evidence-based filtering, using the tr2aacds.pl tool from EvidentialGene v2022.01.20 [77]. Lowly expressed genes were filtered out using edgeR's v3.42.4 [78] filterByExpr tool (min.count=10, min.total.count=15). Vector contamination was filtered out using the Univec database [79] and following the default BLAST [80] parameters of the VecScreen protocol. Quality of the transcriptome was assessed using BUSCO v5.3.0 [81] with the insecta_odb10 database and transcriptome mode and TrinityStats.pl from the Trinity package [74]. Sample correlation of the biological replicates was assessed using Trinity's PtR tool [74], creating a PCA plot and a distance matrix only including the top 5,000 variable genes. Homology-based functional annotation of the transcriptome was performed using BLASTX and BLASTP [80] results against protein databases from FlyBase [82] (dmel-all-translation-r6.48.fasta, accessed 13 Apr 2023), Swissprot [83], Uniref90 [84] and a curated arthropod database from NCBI [84] (settings: txid6656[Organism: exp] AND (refseq[filter] OR swissprot[filter] OR pdb[filter])). Results were imported by Trinotate [85] into an SQLite database. GO [86, 87] terms associated with annotation results were manually retrieved from the FlyBase [82] and Swissprot [83] online web interfaces. Each transcript was assigned a single annotation with the order of priority as the database order described above, obtaining 79.6% of annotations from FlyBase, 13.5% from SwissProt and 7% from Uniref90. Finally, transcripts were labelled as maternally deposited based on FlyBase entries containing "Comment: maternally deposited" in their description.

DGE analysis

Gene abundance estimation and DGE analyses were performed using Trinity's [74] Transcript Quantification and Differential expression analysis pipelines, respectively. Alignment free abundance estimation was performed using Salmon v1.7.0 [88]. Raw counts, trimmed mean of M (TMM)- and transcripts per million (TPM)-normalised expression matrices were generated using the abundance_estimates_to_matrix.pl tool from Trinity [74]. The voom function from the limma R package v3.46.0 [89] was incorporated to conduct pairwise sample gene expression level comparisons on TMM normalised transcript counts and select significant DEGs based on a fold change of 4 and a false discovery rate (FDR) p-value of $<1e-3$. The resulting selection of DEGs were clustered

hierarchically with a euclidean distance-based measure and complete-linkage method using the `analyze_diff_expr.pl` script as part of the Trinity package [74]. Clusters were defined with the `define_clusters_by_cutting_tree.pl` tool by cutting the resulting dendrogram with a K of 23, which is based on a manual assessment of separation of expression patterns and number of genes per cluster. Gene expression results were log2 normalized and centered prior to visualization using a heatmap created with the R `heatmap` package v1.0.12 [90] or cluster-based line graphs using `ggplot2` v3.4.3 [91].

GO enrichment

GOseq v1.42.0 [92] was utilized by the `_runGOseq.R` script as part of Trinity [74] to perform cluster specific GO enrichment analysis, incorporating all ancestral terms of biological process related GO terms associated with each transcript. GO terms with an over-represented p-value and false discovery rate (FDR) of both < 0.05 were selected as significantly enriched within each cluster. Retrieved GO terms were summarized using an adapted GOslim script [93] with a GOslim [86, 87] database manually curated to include embryonic developmental terms, modified from `goslim_drosophila.obo` release 2023-01-01.

Gene trees

To verify the identity of transcripts for genes of interest (Table S2), a BLAST [80] database was made using protein sequences from reference genomes of several arthropods, including *Orchesella cincta* (GCA_001718145.1), *Folsomia candida* (GCF_002217175.1), *Ladona fulva* (GCA_000376725.2), *B. germanica* (GCA_003018175.1), *Zootermopsis nevadensis* (GCF_000696155.1), *Diuraphis noxia* (GCF_001186385.1), *Thrips palmi* (GCF_012932325.1), *Frankliniella occidentalis* (GCF_000697945.2), *Apis mellifera* (GCF_003254395.2), *T. castaneum* (GCF_000002335.3), *Chrysoperla carnea* (GCF_905475395.1), *B. mori* (GCF_014905235.1) and *D. melanogaster* (GCF_000001215.4). Sequences of the two mayfly species, *E. danica* (GCA_000507165.2) and *Cloeon dipterum* (GCA_902829235.1) were included as well. BLASTP with broad settings (e-value < 1e⁻¹) was used on translations of transcripts of interest and the results were used for a MAFFT v7.017 [94] alignments (setting algorithm selection to “Auto”, gap open penalty to 2 and offset value to 0.123). Subsequently gene trees were built using FastTree v2.1.11 [95], utilising the Jones-Taylor-Thornton (JTT) [96] model and the results were analysed manually (Figs. S22–S29). All steps were performed using Geneious 8.1.9 (<https://www.geneious.com>).

Abbreviations

ANT-C	Antennapedia complex
BX-C	Bithorax complex

DEG	Differentially expressed gene
DGE	Differential gene expression
FDR	False discovery rate
GO	Gene ontology
hAEL	Hours after egg laying
IHC	Immunohistochemistry
ISH	in situ hybridisation
MZT	Maternal to zygotic transition
PCA	Principal Component Analysis
PE	paired end
RT-qPCR	Quantitative reverse transcription PCR
RNAi	RNA interference
TMM	Trimmed mean of M
TPM	Transcripts per million

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10934-7>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

The authors thank the owners of V.O.F. Schimmel for allowing field work for mayfly collection to be conducted on their premises; Sivasubramani Selvanayagam for the significant aid in the bioinformatics analyses; and Daan Drukker for sharing information on mayfly collection.

Author contributions

WPDM, SSi and MES designed the study. Insect and egg collection was done by WPDM and IB. The molecular analyses were performed by WPDM, IB and PV. Data analyses was done by WPDM with supervision from RvV and SSi. RM assessed the embryonic morphology. The manuscript was written by WPDM with comments from RvV, SSi, RM and MES.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Data availability

This Transcriptome Shotgun Assembly project has been deposited at DDBJ/ENA/GenBank under the accession GKXC00000000. The version described in this paper is the first version, GKXC01000000.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Biosystematics group, Wageningen University & Research, Droevendaalsesteeg 1, Wageningen, The Netherlands 6708PB
²Sugadaira Research Station, Mountain Science Center, University of Tsukuba, Sugadaira Kogen, Ueda, Nagano 386-2204, Japan

Received: 24 June 2024 / Accepted: 22 October 2024

Published online: 04 December 2024

References

- Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, et al. The developmental transcriptome of *Drosophila melanogaster*. *Nature*. 2011;471(7339):473–9.
- Tolwinski NS. Introduction: *Drosophila*—A Model System for Developmental Biology. *J Dev Biol*. 2017;5(3):9.
- Peel AD, Chipman AD, Akam M. Arthropod Segmentation: beyond the *Drosophila* paradigm. *Nat Rev Genet*. 2005;6(12):905–16.
- Panfilio KA. Extraembryonic development in insects and the acrobatics of blastokinesis. *Dev Biol*. 2008;313(2):471–91.
- Liu PZ, Kaufman TC. Short and long germ segmentation: unanswered questions in the evolution of a developmental mode. *Evol Dev*. 2005;7(6):629–46.
- Rosenberg MI, Lynch JA, Desplan C. Heads and tails: Evolution of antero-posterior patterning in insects. *BBA - Gene Regul Mech*. 2009;1789(4):333–42.
- Krause G. Die Eitypen der Insekten. *Biol Zbl*. 1939;59:495–536.
- Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science*. 2014;346(6210):763–7.
- Simon S, Blanke A, Meusemann K. Reanalyzing the Palaeoptera problem – The origin of insect flight remains obscure. *Arthropod Struct Dev*. 2018;47(4):328–38.
- Oppenheim SJ, Baker RH, Simon S, DeSalle R. We can't all be supermodels: the value of comparative transcriptomics to the study of non-model insects. *Insect Mol Biol*. 2015;24(2):139–54.
- Cheng H, Wang Y, Sun MA. Comparison of gene expression profiles in nonmodel eukaryotic organisms with RNA-Seq. *Methods Mol Biol*. 2018;1751:3–16.
- Chalifa-Caspi V. RNA-Seq in nonmodel organisms. *Methods Mol Biol*. 2021;2243:143–67.
- Simon S, Breeschoten T, Jansen HJ, Dirks RP, Schranz ME, Ros VID. Genome and transcriptome analysis of the beet armyworm *Spodoptera exigua* reveals targets for pest control. *G3 (Bethesda)*. 2021;11(11):jkab311.
- Rago A, Werren JH, Colbourne JK. Sex biased expression and co-expression networks in development, using the hymenopteran *Nasonia vitripennis*. *PLOS Genet*. 2020;16(11):e1008518.
- Ewen-Campen B, Shaner N, Panfilio KA, Suzuki Y, Roth S, Extavour CG. The maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*. *BMC Genomics*. 2011;12(1):61.
- Fan XB, Pang R, Li WX, Ojha A, Li D, Zhang WQ. An Overview of Embryogenesis: External Morphology and Transcriptome Profiling in the Hemipteran Insect *Nilaparvata lugens*. *Front Physiol*. 2020;11:106.
- Ylla G, Piulachs MD, Belles X. Comparative Transcriptomics in Two Extreme Neopterans Reveals General Trends in the Evolution of Modern Insects. *iScience*. 2018;4:164–79.
- Chen S, Yang P, Jiang F, Wei Y, Ma Z, Kang L. De Novo Analysis of Transcriptome Dynamics in the Migratory Locust during the Development of Phase Traits. *PLoS ONE*. 2010;5(12):e15633.
- Zeng V, Ewen-Campen B, Horch HW, Roth S, Mito T, Extavour CG. Developmental Gene Discovery in a Hemimetabolous Insect: *De Novo* Assembly and Annotation of a Transcriptome for the Cricket *Gryllus bimaculatus*. *PLoS ONE*. 2013;8(5):e61479.
- Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, et al. Identification of Functional Elements and Regulatory Circuits by *Drosophila* modENCODE. *Science*. 2010;330(6012):1787–97.
- Tomančák P, Berman BP, Beaton A, Weiszmann R, Kwan E, Hartenstein V, et al. Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol*. 2007;8(7):R145.
- Xu GF, Gong CC, Lyu H, Deng HM, Zheng SC. Dynamic transcriptome analysis of *Bombyx mori* embryonic development. *Insect Sci*. 2022;29(2):344–62.
- Simon S, Sagasser S, Saccenti E, Brugler MR, Schranz ME, Hadrys H, et al. Comparative transcriptomics reveal developmental turning points during embryogenesis of a hemimetabolous insect, the damselfly *Ischnura elegans*. *Sci Rep*. 2017;7(1):13547.
- Almudi I, Martín-Blanco CA, García-Fernández IM, López-Catalina A, Davie K, Aerts S, et al. Establishment of the mayfly *Cloeon dipterum* as a new model system to investigate insect evolution. *EvoDevo*. 2019;10(1):6.
- Almudi I, Vizueta J, Wyatt CDR, de Mendoza A, Marlétaz F, Firas PN, et al. Genomic adaptations to aquatic and aerial life in mayflies and the origin of insect wings. *Nat Commun*. 2020;11(1):2631.
- O'Donnell BC, Jockusch EL. The expression of *wingless* and *Engrailed* in developing embryos of the mayfly *Ephoron leukon* (Ephemeroptera: Polymitarcyidae). *Dev Genes Evol*. 2010;220(1–2):11–24.
- Niwa N, Akimoto-Kato A, Niimi T, Tojo K, Machida R, Hayashi S. Evolutionary origin of the insect wing via integration of two developmental modules: New hypothesis of insect wing evolution. *Evol Dev*. 2010;12(2):168–76.
- Gonzalez CJ, Hildebrandt TR, O'Donnell B. Characterizing Hox genes in mayflies (Ephemeroptera), with *Hexagenia limbata* as a new mayfly model. *EvoDevo*. 2022;13(1):15.
- Si Q, Luo JY, Hu Z, Zhang W, Zhou CF. *De novo* transcriptome of the mayfly *Cloeon viridulum* and transcriptional signatures of Prometabola. *PLoS ONE*. 2017;12(6):e0179083.
- Tojo K, Machida R. Embryogenesis of the mayfly *Ephemera japonica* McLachlan (Insecta: Ephemeroptera, Ephemeridae), with special reference to abdominal formation. *J Morphol*. 1997;234(1):97–107.
- Davis GK, Patel NH. Short, Long, and Beyond: Molecular and Embryological Approaches to Insect Segmentation. *Annu Rev Entomol*. 2002;47(1):669–99.
- Xiang J, Forrest IS, Pick L. *Dermestes maculatus*: an intermediate-germ beetle model system for evo-devo. *EvoDevo*. 2015;6(1):32.
- Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature*. 1978;276(5688):565–70.
- Kaufman TC, Seeger MA, Olsen G. Molecular and genetic organization of the Antennapedia gene complex of *Drosophila melanogaster*. *Adv Genet*. 1990;27:309–62.
- Campos-Ortega JA, Hartenstein V. The embryonic development of *Drosophila melanogaster*. 2nd ed. Springer; 1997.
- O'Farrell PH, Stumpff J, Su TT. Embryonic Cleavage Cycles: How Is a Mouse Like a Fly? *Curr Biol CB*. 2004;14(1):R35–45.
- Merrill PT, Sweeton D, Wieschaus E. Requirements for autosomal gene activity during precellular stages of *Drosophila melanogaster*. *Development*. 1988;104(3):495–509.
- Benoit B, He CH, Zhang F, Votruba SM, Tadros W, Westwood JT, et al. An essential role for the RNA-binding protein Smaug during the *Drosophila* maternal-to-zygotic transition. *Development*. 2009;136(6):923–32.
- Vastenhout NL, Cao WX, Lipshitz HD. The maternal-to-zygotic transition revisited. *Development*. 2019;146(11):dev161471.
- Tadros W, Goldman AL, Babak T, Menzies F, Vardy L, Orr-Weaver T, et al. SMAUG Is a Major Regulator of Maternal mRNA Destabilization in *Drosophila* and Its Translation Is Activated by the PAN GU Kinase. *Dev Cell*. 2007;12(1):143–55.
- Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature*. 2008;456(7220):400–3.
- Tojo K, Machida R. Early embryonic development of the mayfly *Ephemera japonica* McLachlan (Insecta: Ephemeroptera, Ephemeridae). *J Morphol*. 1998;238(3):327–35.
- Liu Z, Friedrich M. The *Tribolium* homologue of *glass* and the evolution of insect larval eyes. *Dev Biol*. 2004;269(1):36–54.
- Fremling CR. Methods for Mass-Rearing *Hexagenia* Mayflies (Ephemeroptera:Ephemeridae). *Trans Am Fish Soc*. 1967;96(4):407–10.
- Kalinka AT, Varga KM, Gerrard DT, Corcoran DL, Jarréls J, et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature*. 2010;468(7325):811–4.
- Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, et al. The mid-developmental transition and the evolution of animal body plans. *Nature*. 2016;531(7596):637–41.
- Sander K. The evolution of patterning mechanisms: gleanings from insect embryogenesis. *Dev Evol*. 1983. pp. 137–59.
- Liu J, Frochoux M, Gardeux V, Deplancke B, Robinson-Rechavi M. Inter-embryo gene expression variability recapitulates the hourglass pattern of evo-devo. *BMC Biol*. 2020;18(1):129.
- Slack J, Holland P, Graham C. The zootype and the phylotypic stage. *Nature*. 1993;361:490–2.
- Andó H. The comparative embryology of Odonata with special reference to a relic dragonfly *Epiophlebia superstes* Selys. Volume 3. JSPS; 1962.
- Hayashi S, Kondo T. Development and Function of the *Drosophila* Tracheal System. *Genetics*. 2018;209(2):367–80.
- Casas-Tintó S, Ferrús A. The haplolethality paradox of the *wupA* gene in *Drosophila*. *PLoS Genet*. 2021;17(3):e1009108.
- Hsu IU, Linsley JW, Zhang X, Varineau JE, Berkhoudt DA, Reid LE, et al. Stac protein regulates release of neuropeptides. *Proc Natl Acad Sci*. 2020;117(47):29914–24.
- Bullard B, Linke WA, Leonard K. Varieties of elastic protein in invertebrate muscles. *J Muscle Res Cell Motil*. 2002;23(5–6):435–47.
- Jani K, Schöck F. Zasp is required for the assembly of functional integrin adhesion sites. *J Cell Biol*. 2007;179(7):1583–97.

56. Auxerre-Plantié E, Nielsen T, Grunert M, Olejniczak O, Perrot A, Özcelik C, et al. Identification of *MYOM2* as a candidate gene in hypertrophic cardiomyopathy and Tetralogy of Fallot, and its functional evaluation in the *Drosophila* heart. *Dis Model Mech*. 2020;13(12):dmm045377.
57. Henkin JA, Maughan DW, Vigoreaux JO. Mutations that affect *flightin* expression in *Drosophila* alter the viscoelastic properties of flight muscle fibers. *Am J Physiol-Cell Physiol*. 2004;286(1):C65–72.
58. Knirr S, Azpiazu N, Frasch M. The role of the NK-homeobox gene *slouch* (*S59*) in somatic muscle patterning. *Development*. 1999;126(20):4525–35.
59. Bourgouin C, Lundgren SE, Thomas JB. *apterous* is a *Drosophila* LIM domain gene required for the development of a subset of embryonic muscles. *Neuron*. 1992;9(3):549–61.
60. Rushton E, Drysdale R, Abmayr SM, Michelson AM, Bate M. Mutations in a novel gene, *myoblast city*, provide evidence in support of the founder cell hypothesis for *Drosophila* muscle development. *Development*. 1995;121(7):1979–88.
61. Baylies MK, Bate M. twist: A Myogenic Switch in *Drosophila*. *Science*. 1996;272(5267):1481–4.
62. Duboule D. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development*. 1994;1994(Supplement):135–42.
63. Gaunt SJ. Hox cluster genes and collinearities throughout the tree of animal life. *Int J Dev Biol*. 2018;62(11–12):673–83.
64. Peterson MD, Rogers BT, Popadić A, Kaufman TC. The embryonic expression pattern of *labial*, posterior homeotic complex genes and the *teashirt* homologue in an apterygote insect. *Dev Genes Evol*. 1999;209(2):77–90.
65. Dollé P, Izpisua-Belmonte JC, Falkenstein H, Renucci A, Duboule D. Coordinate expression of the murine *Hox-5* complex homeobox-containing genes during limb pattern formation. *Nature*. 1989;342(6251):767–72.
66. Izpisua-Belmonte Jc, Falkenstein H, Dollé P, Renucci A, Duboule D. Murine genes related to the *Drosophila AbdB* homeotic genes are sequentially expressed during development of the posterior part of the body. *EMBO J*. 1991;10(8):2279–89.
67. Shippy TD, Brown SJ, Denell RE. maxillopedia is the *Tribolium* ortholog of proboscipedia. *Appl Dev*. 2000;2(3):145–51.
68. Brown S, Holtzman S, Kaufman T, Denell R. Characterization of the *Tribolium Deformed* ortholog and its ability to directly regulate *Deformed* target genes in the rescue of a *Drosophila Deformed* null mutant. *Dev Genes Evol*. 1999;209(7):389–98.
69. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676–82.
70. Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. FASTQC: A quality control tool for high throughput sequence data. Babraham, UK. 2012. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
71. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*. 2012;28(24):3211–7.
72. Bushnell B, BBMap: A Fast, Accurate, Splice-Aware Aligner. 2014. <https://www.osti.gov/biblio/1241166>
73. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research*. 2018;7:1338.
74. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52.
75. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
76. Haas BJ, TransDecoder. Find Coding Regions Within Transcripts. 2018. <https://github.com/TransDecoder/TransDecoder>
77. Gilbert D. Gene-omes built from mRNA-seq not genome DNA. 7th Annu Arthropod Genomics Symp. 2013. <http://arthropods.eugenics.org/EvidentialGene/about/EvigeneRNA2013poster.pdf>
78. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.
79. UniVec. National Library of Medicine (US), National Center for Biotechnology Information. 2017. <ftp://ncbi.nlm.nih.gov/pub/UniVec/>. Accessed 29 Sept 2022.
80. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
81. Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38(10):4647–54.
82. Jenkins VK, Larkin A, Thurmond J, Consortium F. Using FlyBase: a database of *Drosophila* genes and genetics. *Drosophila: Methods and protocols*. Springer; 2022. pp. 1–34. Accessed 13 Apr 2023.
83. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. *Methods Mol Biol*. 2007;406:89–112. Accessed 18 Apr 2023.
84. Gene. National Library of Medicine (US), National Center for Biotechnology Information. 2004. <https://www.ncbi.nlm.nih.gov/gene/>. Accessed 18 Apr 2023.
85. Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, et al. A tissue-mapped axolotl *de novo* transcriptome enables identification of limb regeneration factors. *Cell Rep*. 2017;18(3):762–76.
86. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25–9.
87. Consortium TGO, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. *Genetics*. 2023;224(1):iyad031.
88. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417–9.
89. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47–47.
90. Kolde R. pheatmap. Pretty Heatmaps. 2019. <https://CRAN.R-project.org/package=pheatmap>
91. Wickham H. Data Analysis. ggplot2. Use R! Cham: Springer; 2016. pp. 189–201. <https://ggplot2.tidyverse.org>.
92. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*. 2010;11:R14.
93. Girke T. R Script to Perform goSlim Counts on Vector of GO IDs. 2005. https://aculty.ucr.edu/~tgirke/Documents/R_BioCond/My_R_Scripts/goSlim.txt
94. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
95. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE*. 2010;5(3):e9490.
96. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci CABIOS*. 1992;8(3):275–82.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.