

SCIENTIFIC REPORTS



OPEN

Comparative transcriptomics reveal developmental turning points during embryogenesis of a hemimetabolous insect, the damselfly *Ischnura elegans*

Sabrina Simon^{1,2}, Sven Sagasser³, Edoardo Saccenti⁴, Mercer R. Brugler^{2,5}, M. Eric Schranz¹, Heike Hadrys^{2,6,7}, George Amato² & Rob DeSalle²

Identifying transcriptional changes during embryogenesis is of crucial importance for unravelling evolutionary, molecular and cellular mechanisms that underpin patterning and morphogenesis. However, comparative studies focusing on early/embryonic stages during insect development are limited to a few taxa. *Drosophila melanogaster* is the paradigm for insect development, whereas comparative transcriptomic studies of embryonic stages of hemimetabolous insects are completely lacking. We reconstructed the first comparative transcriptome covering the daily embryonic developmental progression of the blue-tailed damselfly *Ischnura elegans* (Odonata), an ancient hemimetabolous representative. We identified a “core” set of 6,794 transcripts – shared by all embryonic stages – which are mainly involved in anatomical structure development and cellular nitrogen compound metabolic processes. We further used weighted gene co-expression network analysis to identify transcriptional changes during Odonata embryogenesis. Based on these analyses distinct clusters of transcriptional active sequences could be revealed, indicating that embryos at different development stages have their own transcriptomic profile according to the developmental events and leading to sequential reprogramming of metabolic and developmental genes. Interestingly, a major change in transcriptionally active sequences is correlated with katatrepsis (revolution) during mid-embryogenesis, a 180° rotation of the embryo within the egg and specific to hemimetabolous insects.

During embryogenesis, the central life cycle, the embryonic body plan is laid out, starting with blastoderm formation, germ band formation, followed by elongation, segmentation, and appendage formation. Most of our knowledge about developmental gene networks during insect embryogenesis is built on the *Drosophila* paradigm, which is far from being universal¹. In addition, the involvement of genes in specific developmental processes is usually determined on a small scale by comparing expression patterns of specific key genes across species by means of *in situ* hybridization or quantitative RT-PCR. This approach has identified genes with deep conservation of expression patterns, that have also been shown to underlie developmental similarities on unexpectedly large evolutionary scales^{2,3}. However, given that genes commonly function together, concerted expression changes of distinct sets of genes may often be phenotypically relevant. In this context, transcriptomic developmental time courses

¹Biosystematics Group, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands. ²Sackler Institute for Comparative Genomics, American Museum of Natural History, Central Park West and 79th St., New York, NY, 10024, USA. ³Ludwig Institute for Cancer Research, Karolinska Institutet, 17177, Stockholm, Sweden. ⁴Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeng 4, 6708 WE, Wageningen, The Netherlands. ⁵Biological Sciences Department, NYC College of Technology, City University of New York, 300 Jay Street, Brooklyn, New York, 11201, USA. ⁶ITZ, Ecology&Evolution, University of Veterinary Medicine Hannover, Buenteweg 17d, D-30559, Hannover, Germany. ⁷Yale University, Department of Ecology & Evolutionary Biology, 165 Prospect Street, New Haven, CT, 06511, USA. Correspondence and requests for materials should be addressed to S.S. (email: sabrina.simon@wur.nl)

have already demonstrated the use of *de novo* assembled transcriptomes spanning various developmental stages to identify developmental genes and members of signalling pathways and to explore genome-level questions^{4–6}. However, comparative molecular studies focusing on early/embryonic stages during insect development are limited to a few taxa, mainly holometabolous insects, especially the model systems like the fruitfly (*Drosophila melanogaster*), the red flour beetle (*Tribolium castaneum*), or the parasitic wasp (*Nasonia vitripennis*)^{7–10}. In contrast, a few hemimetabolous insect species, e.g. *Oncopeltus* and *Gryllus*, have become more widely used, but have been investigated only for selected key *Drosophila* homologs^{11–16}. Although several studies have extensively examined morphological changes during hemimetabolous embryogenesis^{17–19}, large-scale embryonic transcriptomic studies are still missing.

Here, we attempt to fill in this gap and present the first comparative embryonic transcriptome for the blue-tailed damselfly *Ischnura elegans*. *I. elegans* belongs to the family Coenagrionidae of the suborder Zygoptera (damselfly) within the order Odonata. Odonata have become a model organism for studies in ecology and evolutionary biology and currently serves different research aspects like assessing the impact of global warming^{20,21}, trait-dependent diversification patterns^{22,23}, colour vision^{24,25} and colour polymorphism evolution^{26–28} (for a review see also Bybee *et al.*²⁹ and references therein). There is also an increasing source of Odonata molecular studies^{30–33} and recently a study comprising the first draft genome of an Odonata species was published³⁴. In addition, Odonata represent a promising system for future evo-devo research. They represent one of the two earliest pterygote (winged) insect orders^{35–37}. Consequently, studying the evolution of developmental processes in an Odonata representative would provide crucial insights in key mechanisms underlying the origin and diversification of insect wings.

In the present study, we generated expression data throughout all embryonic developmental stages covering germ band formation, elongation, segmentation, and appendage formation, by performing comprehensive RNA sequencing on single *I. elegans* embryos. Based on this RNA-seq data we developed a novel *I. elegans* reference transcriptome and examined gene expression divergence across all embryonic stages to provide novel insights in the genetics of embryogenesis of a hemimetabolous insect. The *de novo* reference transcriptome is undoubtedly valuable for further ecological and evolutionary studies in Odonata. Furthermore, our comparative data will provide insights into the extent of gene expression variation during embryogenesis in more “primitive” hemimetabolous lineages.

Methods

Insect Sampling. A mating wheel of *Ischnura elegans* was collected in Southern-France in June 2012. To obtain the egg clutch, the mating wheel was placed in an oviposition chamber that consisted of a vial containing only wet filter paper. After termination of copulation, the male was released and the female was kept overnight in the vial for egg oviposition. The wet filter paper in the vials is known to be sufficient to elicit oviposition in some odonate species^{38,39}. On nine subsequent days starting the day after oviposition, approximately 20 eggs of the egg clutch were preserved in RNAlater once at the same time of the day and stored at -80°C . On the 10th day, no embryos were further preserved as the first nymphs of the egg clutch hatched.

454-Sequencing Approach. For RNA extraction, several embryos spanning two to three days were pooled together (day 1–3, day 4–5, day 6–7 and day 8–9, Supplementary Table S1). Total RNA extraction and cDNA synthesis was conducted as described in Kvist *et al.*⁴⁰. In total, four cDNA Rapid Libraries (RL) with different indexed barcodes were prepared using a Roche 454 GS RL Prep Kit by following manufacturer’s protocols as outlined in the Roche 454 RL Preparation Method Manual (Roche Applied Sciences, Indianapolis, IN, USA). Emulsion-based clonal amplification (PCR) was carried out using the GS Junior Titanium emPCR (Lib-L) Kit and following manufacturer’s protocols as outlined in the emPCR Amplification Method Manual (Lib-L). This manual was also used for bead recovery, DNA library bead enrichment, and sequence primer annealing. Enriched beads were prepared for sequencing on a GS Junior Titanium PicoTitrePlate Device using the GS Junior Titanium Sequencing Kit and following manufacturer’s protocols as outlined in the Sequencing Method Manual. Massively parallel single-end pyrosequencing was conducted by one multiplexed run on a 454 GS Junior at the Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY, USA.

Post-sequencing processing was conducted as described in Kvist *et al.*⁴⁰ followed by trimming of low quality regions; only bases between positions 59–500 and those with a Phred quality score ≥ 25 and a minimum length of 20 base pairs (`-v -t 25 -l 20 -Q 33`) were retained in the data set using FASTX_trimmer and FASTQ_quality_trimmer (both part of the FASTX toolkit; http://hannonlab.cshl.edu/fastx_toolkit/). Before assembly the raw reads were further checked for potential contamination through local Blast against UniVec (<ftp://ftp.ncbi.nlm.nih.gov/pub/UniVec/>, accessed Oct 7, 2014) using BLASTN (`-reward 1 -penalty -3 -evalue 700 -searchsp 1750000000000 -dust yes -gapopen 3 -gapextend 3`). Raw sequences were considered to contain potential contamination if the alignment length of the query with the target exceeded 25 base pairs (bp) and were filtered out using custom perl scripts (VecScreenFilter.pl, compare_2Files.pl, bad_data_uniq.pl; available upon request) and seqtk (<https://github.com/lh3/seqtk>, accessed Oct 8, 2014). Afterwards, iAssembler tool (v1.3.2.) (`-a 10 -b 10 -d`) was used to cluster and assembly contigs to obtain unigene sequences⁴¹. Raw sequence reads can be found in the SRA database under BioProject PRJNA401426.

Illumina Sequencing Approach. Following the Smart-seq 2 protocol⁴², we prepared 15 different Nextera indexed RNA Seq libraries each representing a single embryo and including a replicate of each embryonic developmental stage (expect day 1, 2 and 5; Supplementary Table S1). These developmental stages were defined according to the day after oviposition. Libraries were sequenced on two lanes of $2 \times 150\text{bp}$ on a Illumina HiSeq 2500 at the NY Genome Center, New York, NY, USA.

The raw Illumina reads for each of the 15 libraries were delivered as individual fastq files. The Illumina reads were quality-filtered, and sequencing and indexing adapters were removed using Trimmomatic (0.32)⁴³ (PE; Final_Adapter-Trim.txt:2:30:10; LEADING:3; TRAILING:3; SLIDINGWINDOW:4:20; MINLEN:35). Only reads with a minimum length of 35 bp were further kept. Overlapping paired-end reads were merged using Flash (1.2.11)⁴⁴ setting max-overlap to 135 bp. Before assembly, the raw reads were further checked for potential contamination through local Blast against UniVec (<ftp://ftp.ncbi.nlm.nih.gov/pub/UniVec/>, accessed Oct 7, 2014) using search parameters and filtering criteria as described above. Raw sequence reads can be found in the SRA database under BioProject PRJNA401426.

Trinity *in silico* read normalization (trinityrnaseq_r20140413p1)⁴⁵ was applied to remove redundant reads before assembly. Here, the remaining reads of the 15 libraries were normalized together with published Illumina reads from an adult male of *Ischnura elegans*⁴⁶ using default commands with a max coverage of 50. Orphan reads that resulted due to the trimming and merging step were separately normalized (only left orphans (trimmed R1 reads and merged reads) and right orphans (only R2 orphans)) using the same commands except the paired reads options. *De novo* assembly was conducted using Trinity (trinityrnaseq_r20140413p1)⁴⁵ using default parameters with a minimum kmer coverage of 2 and with the paired modus including the orphans to left and right reads, respectively.

Building of the Reference Transcriptome for Gene Expression Analyses. Bacterial genomic contamination is common in eukaryotic samples⁴⁷. Therefore, the pre-assemblies were checked for human and bacterial sequence contamination using DeconSeq⁴⁸, with an alignment identity threshold of 97% ($-i$ 97) and an alignment coverage threshold of 90% ($-c$ 90). Both pre-assemblies were analysed separately against the Human Reference (GRCh37; ftp://ftp.ncbi.nih.gov/genomes/Homo_sapiens/ARCHIVE/BUILD.37.2/Assembled_chromosomes/seq/; accessed July 25, 2014), and 5,242 unique bacterial genomes (<ftp://ftp.ncbi.nih.gov/genomes/Bacterial/>; accessed Jan 28, 2015) with a cross-check ($-dbs_retain$) against *Drosophila melanogaster* (ftp://ftp.flybase.net/genomes/Drosophila_melanogaster/dmel_r6.01_FB2014_04/fastq/; accessed July 27, 2014) and *Acyrtosiphon pisum* (aphidbase_2.1b_mRNA; <https://www.aphidbase.com/aphidbase>; accessed July 28, 2014). In addition, in order to reduce the redundancy of the pre-assemblies, they were first processed by CD-HIT-EST (v4.6.1-2012-08-27)⁴⁹ with 95% identity to remove identical fragments.

The resulting contigs of both pre-assemblies (contamination-reduced and assembly improved) were merged using CAP3 (VersionDate: 12/21/07)⁵⁰ to reduce potential redundancy. To improve the overall quality of the hybrid assembly likely coding regions with a minimum open reading frame (ORF) length of 200 bp were extracted from the transcripts using TransDecoder from the Trinity package⁴⁵. The hybrid assembly was used as a reference transcriptome for the weighted gene correlation network analyses (WGCNA). For these transcripts, the base-level coverage was calculated using bowtie2 (v2.2.5)⁵¹ and aligning all Illumina reads against the hybrid assembly. To calculate the mean coverage per base genomeCoverageBed of bedtools⁵² was applied. Transcripts with a mean coverage per base of less than 5 bp were removed from the final reference transcriptome. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GFWX000000000. The version described in this paper is the first version, GFWX010000000. The final reference transcriptome is available in the TSA database under BioProject PRJNA401426. The completeness of the reference transcriptome was assessed using CEGMA (v2.5)⁵³ and BUSCO (v1.1b1)⁵⁴. Functional annotation and analysis of the reference transcriptome was conducted using the Trinotate pipeline (v2.0)⁴⁵. All transcripts and transdecoder-predicted proteins with a minimum length of 200 bp were used as query for BLASTX and BLASTP search, respectively, against the SwissProt non-redundant and the Uniref90 database (both accessed March 2016). Protein domains were predicted using HMMER (v. 3.1b2)⁵⁵ against the Pfam-A database (v.28)⁵⁶, signal peptides were predicted using the SignalP 4.1 server⁵⁷, and transmembrane regions were predicted using the TMHMM server v2.0⁵⁸. RNAMMER (v.1.2)⁵⁹ was used to identify rRNA genes.

Transcript Quantification and Co-Expression Analyses. Illumina-reads from each embryonic sample were separately aligned to our *de novo* reference transcriptome using bowtie2 (v2.2.5)⁵¹ and the isoform/gene abundances were estimated using *express* (v1.5.1)⁶⁰. The resulting count matrix was filtered by abundance based on count-per-million (CPM) values as converted with *edgeR* (3.8.6)⁶¹ (R version 3.1.3). Here, differences in library sizes between samples are taken into account and only genes with at least 5 counts in one of the libraries were kept. Following the common approach when constructing gene correlation networks, genes with variance smaller than twice the observed overall variance were also removed since low variance genes represent noise and may hamper the reconstruction of co-expression networks. The resulting filtered count matrix of 27,027 genes was normalized by the trimmed-mean of M values (TMM) method implemented in *edgeR* and \log_2 transformed. Using these 27,027 genes, a step-by-step signed hybrid co-expression network was built using WGCNA (v. 1.49) R package⁶². The adjacency matrix was created by calculating the biweight mid-correlation between all genes and by restricting the number of excluded outliers ($\text{maxPOutliers} = 0.1$). These settings have less sensitivity to outliers⁶³ as compared to Pearson's correlation but also takes into account the potential risk of unwanted results when the data have a bi-modal distribution⁶⁴. Outliers are expected due to the high biological heterogeneity in our samples (e.g. long time-span, not an inbred culture, see also Results & Discussion).

Based on the scale-free topology criterion⁶⁵, the power for calculating the adjacency matrix was set to 22 resulting in an $R^2 = 0.86$ for the scale-free fit). Genes were hierarchical clustered based on the TOM-based dissimilarity (Topological Overlap Measure (TOM)) and modules (clusters of highly correlated genes) were detected using *DynamicTreeCut*⁶⁶ with a minimum module size of 50. The resulting 78 identified modules were further merged when their eigengenes (the first principal component of module expression pattern) showed a correlation of 0.9⁶⁷. The correlation coefficients between the resulting 34 merged modules and different 'traits' were calculated.

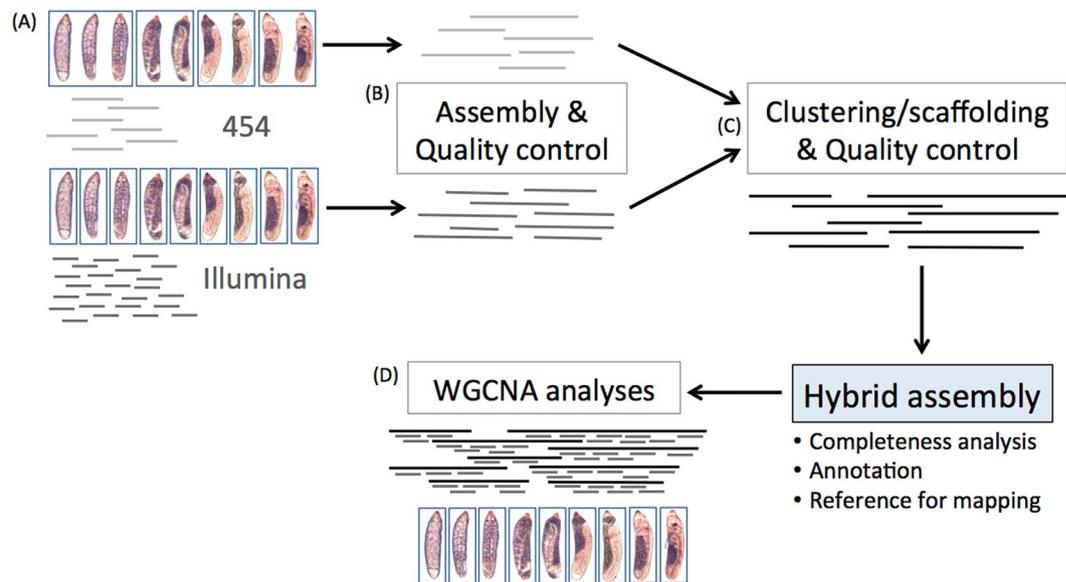


Figure 1. Workflow for the *de novo* hybrid assembly and differential expression (DE) analyses. **(A)** For the 454 sequencing approach several embryos (~20) from several days were pooled together (day 1–3, day 4–5, day 6–7 and day 8–9). For the Illumina sequencing approach one single embryo of each embryonic developmental stage was successfully prepared using the Smart-seq 2 protocol. **(B)** 454 and Illumina reads were separately pre-assembled using iAssembler and Trinity, respectively. Both *de novo* pre-assemblies were separately checked for contamination and redundancy. **(C)** Filtered pre-assemblies were clustered and resulted into the reference hybrid assembly. **(D)** For the WGCNA analyses the Illumina reads from the separate embryonic stages were mapped against the resulting reference hybrid assembly.

Results and Discussion

De Novo Hybrid Assembly of The Transcriptome of *Ischnura Elegans*. To establish the first gene expression profiles during embryonic development of a hemimetabolous insect, two sequencing approaches were conducted for *I. elegans* (454 & Illumina) (see Fig. 1 for an overview of the workflow). To prepare the cDNA for both approaches the same egg clutch were used and all embryonic life stages were included (in total 9 days until nymphs hatched at day 10).

A total of 111,393 454-sequence reads and 149,254,447 Illumina-sequence reads were obtained. The number of raw reads for each library and resulting reads after trimming and cleaning is provided in Supplementary Table S2. The 454 data was *de novo* pre-assembled into 58,271 contigs and (including 3,550 singletons) with a total number of 2,2145,630 bp and a sequence length range from 20 bp to 4,549 bp (Supplementary Table S3). Before assembly of the newly generated embryonic 149,254,447 Illumina sequence reads, in addition to the adult male Illumina reads⁴⁶, Trinity *in silico* read normalization⁴⁵ was applied for removing redundant reads. The normalized Illumina data was assembled *de novo* into 820,838 contigs with a total number of 327,796,547 bp and a sequence length range from 201 bp to 17,100 bp (Supplementary Table S3).

Before the two pre-assemblies were combined into a hybrid assembly, potential contamination and redundancy were removed (Supplementary Table S3). These improved contigs of both pre-assemblies were clustered into combined into hybrid contigs using CAP3. To improve the overall quality of the hybrid assembly and to remove potential assembly artefacts, open reading frames with at least 5 bp mean coverage per base were only selected for the final reference assembly (Supplementary Table S3). This final reference assembly comprised 105,664 transcript isoforms and 92,284 unique transcripts, with an N50 of 1,571. The completeness analysis revealed 235 complete CEG's (94.76%) and 10 partial CEG's, resulting in an estimated gene completeness of 98.79% (245/248) and a BUSCO completeness of 82.02% (1,128 complete single-copy BUSCOs, 549 complete duplicated BUSCOs, 517 fragmented BUSCOs, 481 missing BUSCOs), thereby indicating a very complete representation of expressed genes which could be used as a reference.

Annotation. The hybrid assembly was further annotated using the Trinotate Pipeline (v3.0.0) (<https://trinotate.github.io/>), including 1) capturing Blast homologies (BLASTX and BLASTP) against Uniprot-uniref90 database and Swissprot database (https://data.broadinstitute.org/Trinity/Trinotate_v2.0_RESOURCES/, both accessed March 2016), 2) protein domain identification using PfamA database (https://data.broadinstitute.org/Trinity/Trinotate_v2.0_RESOURCES/, accessed March 2016), 3) prediction of signal peptides using SignalP (v4), 4) prediction of transmembrane regions using tmHMM (v2), and 5) identification of rRNA transcripts using RNAMMER. Trinotate further retrieves various Kegg, GO, and Egglog annotations from the Swissprot database.

A total of 122,769 annotations were retrieved for the hybrid assembly, of which 49,352 unique gene IDs have retrieved at least one annotation (Supplementary Table S4). We further sorted the annotations according to BLASTX homologies against the Uniprot-uniref90 database and analysed which species were most highly

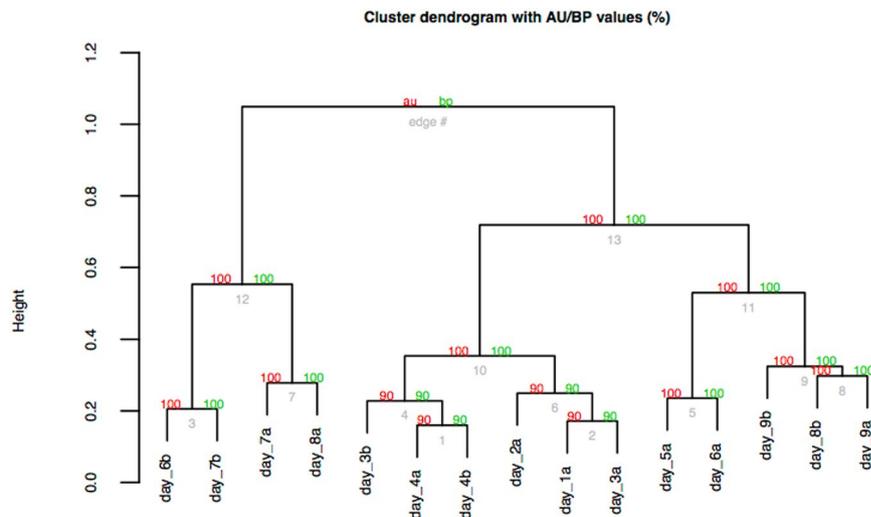


Figure 2. Hierarchical clustering of *I. elegans* embryonic expression patterns via multiscale bootstrap resampling. Analysis has been conducted on the filtered and normalized count matrix (27,027 unique genes). Bootstrap Probability (BP) and Approximately Unbiased (AU) values were computed for each of the clusters using complete linkage clustering based on Pearson correlation.

represented. Here, out of these 26,586 unique gene IDs with a Uniprot-uniref90 annotation based on BLASTX, *Zootermopsis nevadensis* proteins dominated these BLASTX results (7,182 contigs), which also reflects the phylogenetic distance to the other proteomes³⁷ (Supplementary Fig. S1).

Gene Ontology (GO) analysis was further performed using the GSeq package adjusting for transcript length bias in deep sequencing data⁶⁸ and using the GO annotation retrieved from the Trinotate annotation pipeline. GO terms were further summarized to generic GOSlim categories using the R package GOSTats⁶⁹.

Read Abundance and Stage Specific Expression and Similarity. Transcript quantification revealed 105,665 (isoforms)/92,285 (genes) reference transcriptional active sequences, of which 6,794 unique genes were expressed throughout all embryonic stages (Supplementary Table S5). The major represented GO terms according to GOSlim of these “core” embryonic genes were genes associated with (1) anatomical structure development, (2) cellular nitrogen compound metabolic process, (3) biosynthetic process, (4) transport, and (5) small molecule metabolic process (Supplementary Fig. S2, Supplementary Table S5). For all downstream analyses, only read counts at the putative gene level were used. Distribution of expression patterns across the embryonic stages was further evaluated by dividing RPKM values into six bins and defining gene expression into low-(>0–5), moderate-(>5–50) and high-expression (>50). This revealed that in all embryonic stages the majority of transcripts are expressed at a low level (see Supplementary Fig. S3). In addition, starting from day 6 in the embryonic development more transcriptional active sequences could be detected.

To measure the similarity of the samples covering all embryonic stages, the filtered and normalized count matrix (see Methods) was used for cluster bootstrapping analyses (10,000 iterations) using the R package PVclust (v.1.3-2)⁷⁰ (Fig. 2). The bootstrap analysis provided statistical support for the sample relationships based on their gene expression and that the samples were differentiated according to embryonic stage status (*i.e.* day 1–4 versus day 5–6). The same differentiation between the samples and embryonic stage was revealed by multi-dimensional scaling (MDS) analyses (Supplementary Fig. S4). Here the samples were differentiated according to embryonic stage status (*i.e.* day 1–5/6 versus day 6/7–9) along dimension 1, while dimension 2 further separated the samples from day 6–8. In addition, both analyses revealed that individuals from the same developmental stages – as determined according to the days after oviposition – do not necessarily closely cluster together, indicating that there is variation amongst the individuals from the same embryonic stage. The ‘b’ sample of day 6 has a more similar gene expression to day 7 and 8 than to the ‘a’ sample of day 6, which clusters together with day 5. The same holds true for sample ‘b’ of day 3 which is more similar to day 4 while the other day three sample is more similar to day 1–2 (Fig. 2). The female damselfly’s oviposition behaviour could be an explanation for this variance between individuals of the same developmental stage according to oviposition. It was observed that the female in captivity lays the individual eggs during a long time span (>12 hours) that already accounts for a natural high variance in the development. This could be also further observed in the variable hatching times (9–10 days) of the embryos although kept under the same conditions which is known to strongly influence hatching times in general^{71,72}. Furthermore, the comparable long time span and the relative low number of collected samples could not cover this existing natural high variance in the development between the individual embryos.

Consequently, our samples could not be treated as biological replicates and another complementary approach to compare expression between embryonic stages was adopted because differential expression studies require biological replicates for accuracy⁷³. We therefore employed WGCNA, which is a topological-similarity based hierarchical clustering method that has been widely used in transcriptome studies^{74–76}.

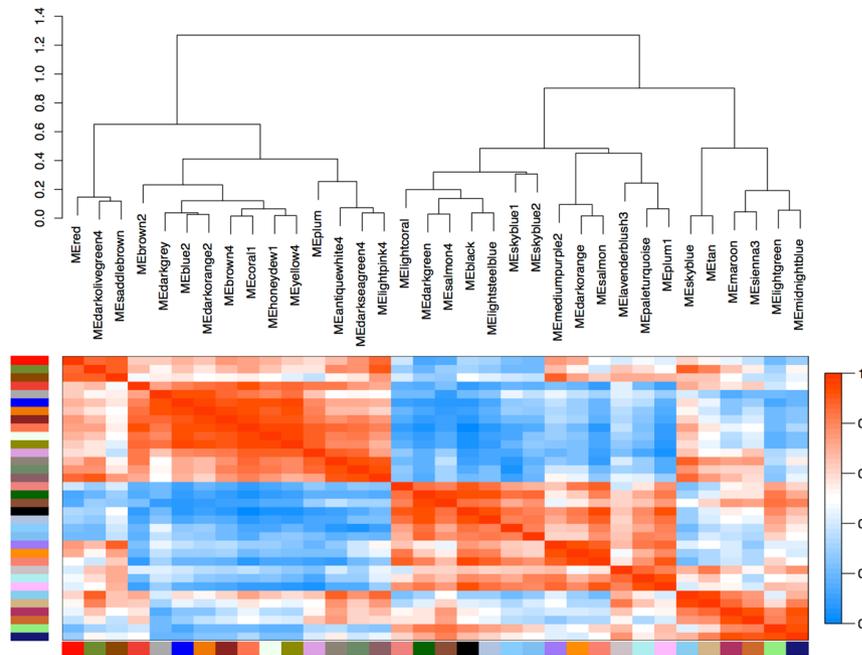


Figure 3. Module similarity by eigengene correlation of merged modules. Modules were merged based on a dissimilarity of module eigengenes (MEDissThres) of 0.1. This resulted into a total of 34 modules.

A filtered count matrix comprising 27,027 genes was used for a step-by-step signed hybrid co-expression network approach. We also used BLASTX to locally compare the 27,027 filtered genes against all Arthropoda protein sequences (NCBI non-redundant protein (nr) database February 2016; E-value cutoff $\leq 10^{-3}$) (Supplementary Table S6). For the signed hybrid co-expression network the minimum module size was adjusted to 50 as we expected high biological variance between our samples as already indicated by the clustering and MDS analyses (Fig. 2 and Supplementary Fig. S4). However, subsequent quantification of module similarity revealed that DynamicTreeCut⁶⁶ might have identified modules which are very similar (Supplementary Fig. S5). Therefore modules were merged based on module eigengene correlations of 0.9 (MEDissThres = 0.1). Although module similarity of the 34 merged modules based on eigengene correlation is for some modules still high, the dissimilarity of module eigengenes (MEDissThres) was set to a small value (0.1) because the samples are fairly biologically different and consequently we expected a large number of resulting modules (Fig. 3). The gene expression of the merged modules covering the embryonic development is shown in Supplementary Fig. S6. The 34 module eigengenes for the 34 merged modules were correlated with specific sample ‘traits’ (Fig. 4). These ‘traits’ were defined as 1) day: the embryonic stage as defined as the day after oviposition, 2) clade: clade definition according to the sample relationships based on the bootstrap and the MDS analysis, and 3) individual: ‘a’ or ‘b’ of the biological replicate (see Supplementary Table S7).

Notably, 15 out of the 34 co-expression modules were significant correlated for day and clade (FDR adjusted p-values < 0.05 (Supplementary Table S8)). Day was most strongly correlated with the darkgreen and the red module, although with opposite directions ($r = -0.89$, $r = 0.87$ and both with FDR adjusted p-value $< 4 \times 10^{-4}$, respectively) (Fig. 5). For these two modules the 30 most highly expressed genes were identified because they might provide insights into important processes during these developmental stages. Notably, in the darkgreen module, with a high eigengene expression during developmental stages day 1–day 4, the most abundant transcripts were ribosomal proteins, further reflecting the fact that ribosome formation is a significant activity during the earliest stages of insect embryogenesis⁷⁷ (Supplementary Table S9). Additionally highly expressed genes were related to DNA replication (Mcm7), transcription regulation (Hrp65), and mRNA processing (Protein DEK). We further identified Geminin among the 30 most highly expressed in the darkgreen module, which plays a role in DNA replication, in anaphase and in neural differentiation⁷⁸. In contrast, in the red module, with a high eigengene expression during developmental stages day 8–day 9, most of the highly expressed transcripts were muscle function related proteins such as Muscle LIM protein (Mlp), Troponin (Tpn) and Tropomyosin (Tm) and Actin (Act). Also detected were proteins involved in the formation of cuticle (cuticle protein 21-like, endochitinase).

A recent study has revealed significant sex-biased gene expression in *I. elegans* adults⁷⁹. And although it has been shown that the amount of sex-biased gene expression tends to increase during development, with low levels in embryonic stages and high levels in sexually mature adults^{80,81}, we explored if the identified modules could be a result of sex-biased gene expression. Under the assumption that the embryos would developed in m males and n females (with m, n, >0 and m + n = 15), we evaluated if specific modules were related to the ‘trait’ sex by considering all 16,383 possible partitions of the 15 sample in two groups (1–2, male-female, female-male respectively) and testing their correlation with the identified gene expression modules. We found only the skyblue2 module to be significantly correlated at the 0.01 confidence level after Bonferroni correction (p-value = 2.82×10^{-8}) to one of these sex combinations after Bonferroni-correction on a 0.01 nominal p-value correction (p-value = 2.92×10^{-11})

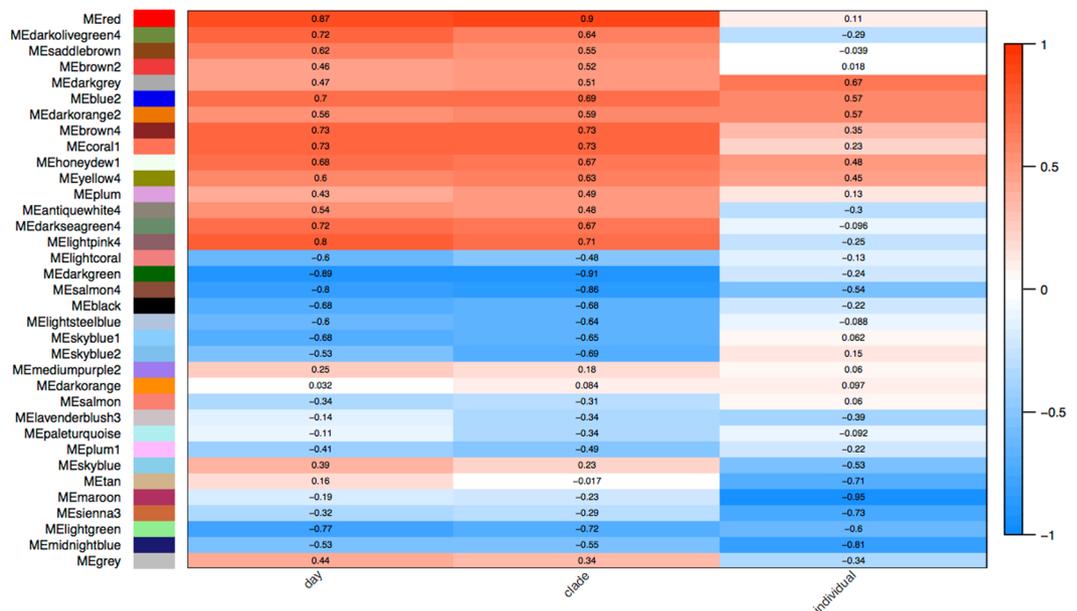


Figure 4. Module-trait association. Each row corresponds to a module. The correlation of the module with the corresponding 'trait' is provided in the associated cell. The FDR adjusted p-values as well as the gene number per module is provided in Supplementary Table S7.

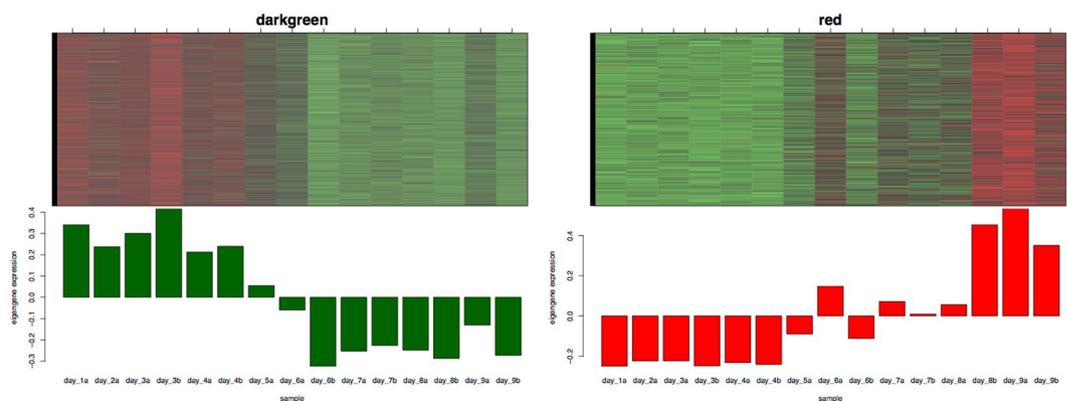


Figure 5. Heatmap and eigengene expression profiles of highest day significant modules. The y axis indicates the value of the module eigengene, the x axis the development and sample type. (A) The darkgreen module comprises 1,303 genes and a correlation coefficient of $r = -0.89$ with a FDR adjusted p-value 0.00032. (B) The red module comprises 635 genes and a correlation coefficient of $r = 0.87$ with a FDR adjusted p-value 0.00039.

(Supplementary Fig. S7). The skyblue2 module comprises 107 genes and the annotation against Arthropoda protein sequences (NCBI non-redundant protein (nr) database, assessed February 2016) (Supplementary Table S10) and the GO analysis revealed overrepresented genes involved into the structural constituent of cuticle (GO:0042302) and serine-type exopeptidase activity (GO:0070008). Interestingly, previous studies have shown sex-dependent differential expression of proteins involved in the structural constituent of cuticle, e.g. cuticle composition^{82,83}. Nevertheless, based on these analyses we concluded that the identified clusters significant correlated for day and clade within *I. elegans* embryogenesis were not a result of sex-biased gene expression.

We further used the WGCNA measure of intramodular connectivity (kME) to identify intramodular hub genes in all 15 significantly day- and clade-related modules. Expression profiles of hub genes represent that of the entire module⁸⁴ and has been found to have more biologically relevant information than whole-network hub genes when considering gene co-expression networks⁸⁵. In total, 3,452 hub genes in 15 modules were identified ($kME > 0.9$, p-value $< 10^{-6}$) (Supplementary Table S11).

A heatmap of the identified hub genes is shown in Fig. 6. Based on this, three clusters of similar gene expression could be observed (see also Fig. 7):

- Cluster 1: modules darkgreen, salmon4, black, lightsteelblue, skyblue1, lightgreen; gene expression up-regulated early during embryogenesis (day 1-day 4/5), followed by a down-regulation after

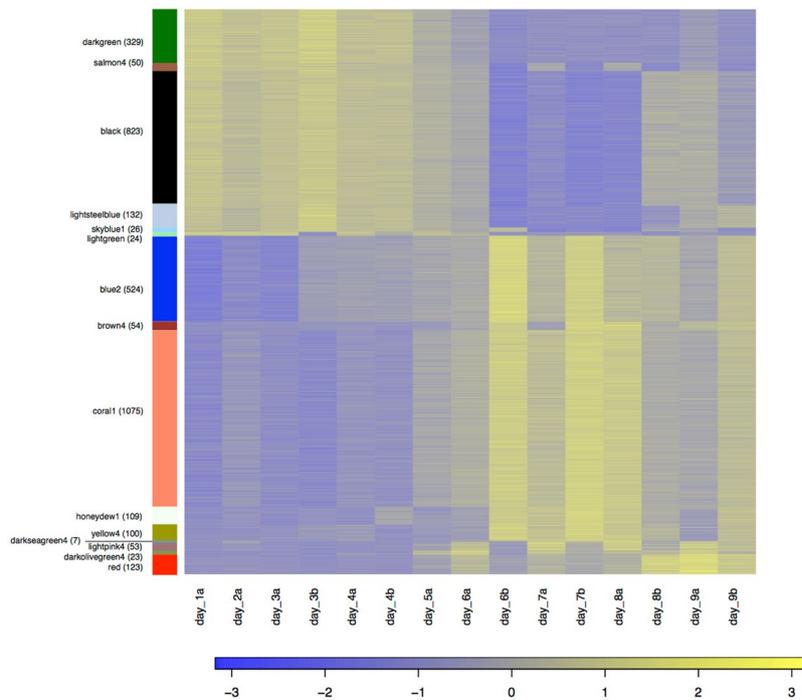


Figure 6. Heatmap of the 3,452 identified hub genes in the significant day and clade module. Each row of the heatmap represents a single gene and genes are grouped according to their module membership. The number of hub genes in each module is given in parenthesis. Modules are sorted according to their module similarity (see also Fig. S5). Expression patterns of all hub genes are visualized as Z-scores using the log₂ TMM-normalized FPKM values.

- mid-embryogenesis (day 6–day 7) and an up-regulation again during late embryogenesis (day 8–day 9).
- (b) Cluster 2: modules blue2, brown4, coral1, honeydew1, yellow4, darkseagreen4, lightpink4; gene expression antagonistic to cluster 1. Gene expression down-regulated early in embryogenesis (day 1–day 4/5), followed by an up-regulation after mid-embryogenesis (day 6–day 7) and a down-regulation again during late embryogenesis (day 8–day 9).
- (c) Cluster 3: modules red and darkolivegreen4; gene expression up-regulated from day 5 on and the highest genes expression during late embryonic stages (day 8–day 9).

For the identified hub genes, statistically over-represented GO terms in a given gene list were identified using the Benjamini-Hochberg correction (p -value < 0.05) relative to the reference set of the 27,027 genes. These statistically over-represented GO terms were further summarized to generic GOSlim categories (Fig. 7 and Supplementary Table S12).

We further analysed the expression dynamics of conserved signalling pathways as well as key developmental genes. A list of *D. melanogaster* genes from Flybase according to the signalling pathways and developmental processes as assigned by QuickGO (<http://www.ebi.ac.uk/QuickGO/>, assessed February 2016) was used as a query to identify homologous sequences in the *I. elegans* transcriptome. The developmental pathways included embryonic axis formation (GO:0000578), regulation of JAK-STAT cascade (GO:0046425), TGFbeta receptor signalling pathway (GO:0007179), Notch signalling pathway (GO:0007219), hedgehog signalling pathway (GO:0007224), sex determination (GO:0007530), Wnt signalling pathway (GO:0016055), and segmentation (GO:0035282). Transcripts with a blast hit to *Drosophila* (E-value cutoff $\leq 10^{-3}$) were then used in a reciprocal blast analysis using BLASTX against all Arthropoda protein sequences (NCBI non-redundant protein (nr) database, assessed February 2016) to establish orthology. Blast results were manually selected for ortholog matches and tabulated (Supplementary Table S13). The expression dynamics across the embryonic development of selected developmental genes are further shown in Supplementary Figs S8–S15.

Gene Expression Divergence In Relation to Embryonic Developmental Stages. The embryogenesis of hemimetabolous insects can be broadly divided into germ band formation, antatrepsis, intertrepsis (or germband stage) and katrepsis^{18,86,87}. During the earliest embryonic stages, proliferation of the germ band is followed by penetration into the yolk mass and differentiation of protocephalon (wide anterior portion of the embryo) and protocorm (narrow posterior region) occur. In long germ types, which are only found in multiple clades within the Holometabola, all segments develop simultaneously at the blastoderm stage⁸⁸. Contrary, Odonata display an intermediate germ type⁸⁹ where an anterior stretch of the germ anlage subdivides rapidly to yield the anterior segments (protocephalon), whereas the remaining segments are added successively⁹⁰. With abdomen elongation and segmentation, antatrepsis – invagination of the embryo into the yolk and posterior

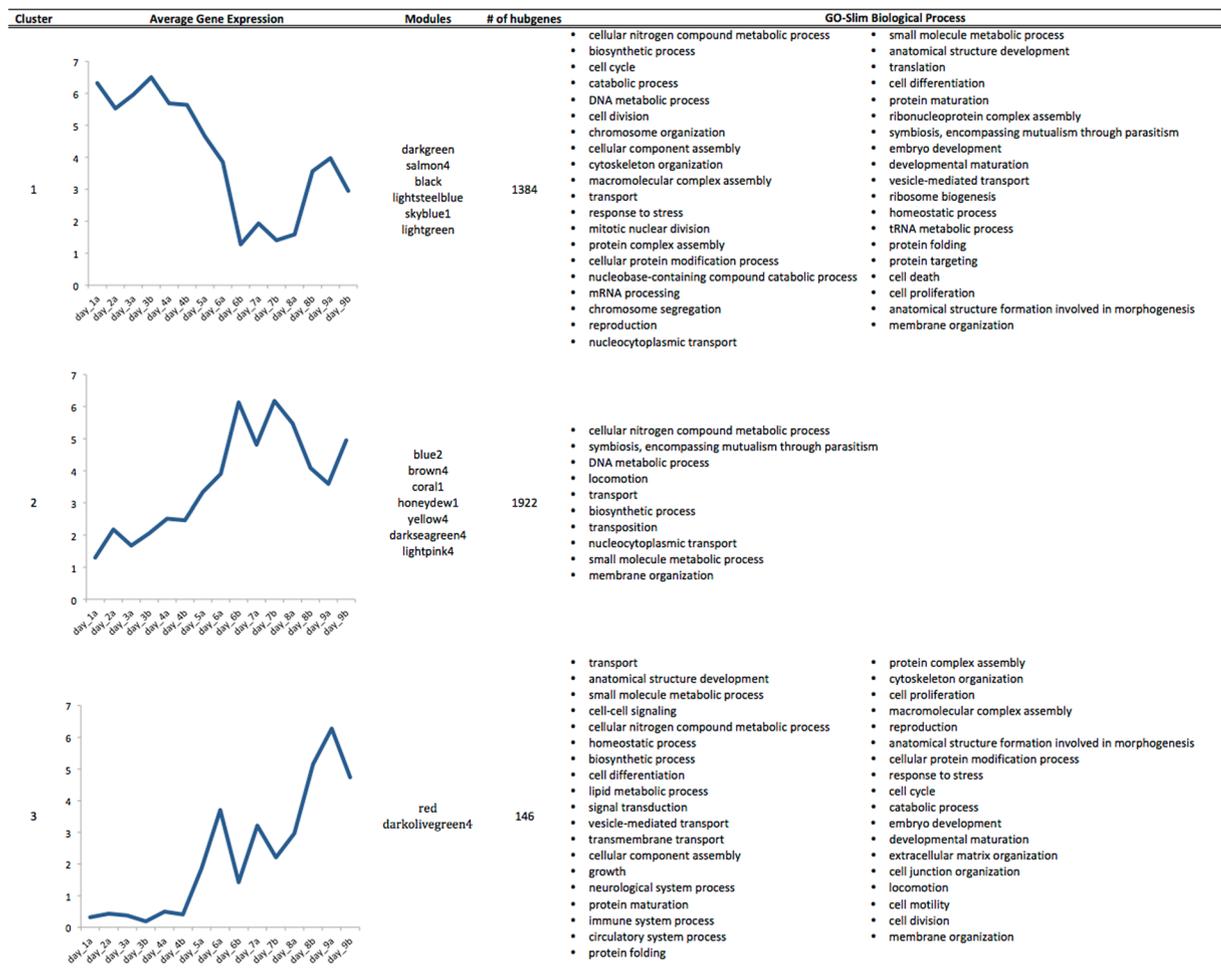


Figure 7. GOSlim categories for over-represented biological processes in each of the three clusters. The average gene expression (TMM log₂) is based on the identified hub genes in the corresponding modules.

movement of the head – starts⁸⁶. Following anatrepsis, the abdomen further elongates during intertrepsis and has to curl back towards the head of the embryo. During this stage the appendage formation starts and thoracic segments are more clearly defined¹⁸. Intertrepsis is followed by katatrepsis, a 180° rotation of the whole insect embryo within the egg by reorganization of the extraembryonic membranes that repositions the embryo⁸⁶. The entire process of movement during embryonic development within the egg is also summarized as blastokinesis in concert with morphogenetic movements of the two extraembryonic membranes and occurs only in hemimetabolous insects: for review see also Panfilio⁸⁶. In all Odonata, katatrepsis takes place midway in embryonic development and lasts only a few hours¹⁷. In previous detailed histological observation the same was also observed in *I. elegans* if kept under different temperature conditions⁹¹ (Simon *et al.*, unpublished data). Ando¹⁷ also described the stages before katatrepsis (revolution) as pre-revolutionary stages and after katatrepsis as post-revolutionary stages.

The pre-revolutionary stages are mainly covered by cluster 1 which show highest expression levels from day 1 to day 4/5, followed by a later moderate up-regulation of the genes during late maturation of the embryo before hatching of the nymphs (day 8–day 9). This cluster was dominated by signatures of cell cycle (GO:0007049), cell division (GO:0051301), cell differentiation (GO:0030154) and mitotic nuclear division (GO:0007067) (Fig. 7 and Supplementary Table S12). This likely reflects an extensive reproduction of embryonic cell mass, pattern formation and regional specification that occurs during early embryonic stages until katatrepsis. This was further reflected in the expression dynamics of conserved signalling pathways as well as key developmental genes. For example, genes involved in axis formation and segmentation showed a clear down-regulation around mid-embryogenesis (Supplementary Figs S8 and S12). Here, for example Delta, which has a role in the proper morphogenesis of body segments and posterior elongation¹⁵ and hunchback, which plays a role in segment patterning⁹², could be identified (Supplementary Fig. S8). Several homeobox genes, e.g. homothorax, proboscipedia, ultrabithorax, LIM/homeobox protein Lhx9; and essential transcription factors, e.g. Transcription factor SOX-2, Transcription factor Sox-6, POU domain class 6 transcription factor 2, were found in the final transcripts (Supplementary Table S13), however they were not included in the identified hub genes due to their low expression levels. Cluster 1 was also enriched for genes involved in mRNA processing (GO:0006397), translation (GO:0006412), ribonucleoprotein complex assembly (GO:0022618) and ribosome biogenesis (GO:0042254) and

highlight the rapid succession of cell cycles associated with chromatin replication and initiation of transcription and translation for embryo patterning⁹³.

Cluster 3 harbours transcriptional active genes during post-revolutionary stages with sequential up-regulation from day 5 on and the highest gene expression during late embryonic stages (day 8/9). This cluster includes markers for neurological system process (GO:0050877), immune system process (GO:0002376), and circulatory system process (GO:0003013). Combination of specific activity of cell cycle markers different from the ones identified in cluster 1 indicates the final differentiation processes for maturation of the embryo prior to hatching. In addition, transcriptional active genes involved in cell-cell signaling (GO:0007267), homeostatic process (GO:0042592), and signal transduction (GO:0007165) reflect the peak time of organogenesis, in accordance with the observation of formation of the compound eye, differentiation of the tracheal system, and completion of heart development and muscle formation¹⁷. For example we identified Slit, an important regulator of axon guidance⁹⁴, and Cubilin for functional maturation of nephrocytes and intestines⁹⁵. High expression of genes involved in muscle structure and function such as Muscle LIM protein, Troponin, Tropomyosin, Myosin and Actin further indicates the maturation of the muscular system for active movement shortly before hatching. This is in agreement with the observation of Ando¹⁷ that the formation of musculature first takes place shortly before katatrepsis and makes rapid progress after the dorsal closure.

Interestingly, a major shift in gene expression was detected around mid-embryogenesis, presumably after katatrepsis (~day 6) and during the early post-revolutionary stages. This was reflected by cluster 2, which contrary to cluster 1, showed elevated expression levels of marker genes from day 6 on. Peak expression appeared at days 6–8 and these stages were also clearly separated from the other developmental stages based on the cluster bootstrap analysis (Fig. 2) and the MDS analyses (Supplementary Fig. S4). The up-regulated genes during these stages comprised markers for locomotion (GO:0040011), transport (GO:0006810) and membrane organization (GO:0061024) and harbours mainly hypothetical and uncharacterised proteins (Supplementary Table S6). In addition, this cluster comprised several transposable elements (TEs), like DNA transposon Mariner and piggyback, and several newly expressed reverse transcriptases. TEs are known to occupy different portions of insect genomes and account for the huge variety in insect genome sizes⁹⁶. In recent years there is increasing evidence that TEs play vital roles in regulation of gene expression by remodeling the chromatin conformation, by inserting into promoters or enhancers and providing binding sites for transcription factors^{97–99}. In addition, TEs are known to play a role in insect embryonic development¹⁰⁰, in phenotypic plasticity¹⁰¹, and diapause¹⁰². Indeed, active transcription of TEs has been detected at various stages of development and have a major role in generating intraspecies variation¹⁰³. Recently, high transcriptional activity of TEs in the egg stage of the migratory locust, *Locusta migratoria*, was detected¹⁰¹. We detected the increase of TE expression around day 6 after down regulation of the DNA (cytosine-5)-methyltransferase 1 (DNMT1). DNMT1 was detected in module salmon4, which is part of cluster 1 (Supplementary Table S6). Down-regulation of TEs occurred when DNA (cytosine-5)-methyltransferase 1 was again up-regulated after day 7/8. This observation is in agreement with previous studies where TE suppression is directly linked to increased DNA methylation activity¹⁰³, although there is also recently increasing evidence for self-regulation of TEs¹⁰⁴. The functional role of TE activity in mid-embryonic stages can only be speculated on and so far no comparable data exists to further verify an up-regulation of transposable elements after mid-embryogenesis (katatrepsis). One possible explanation for TE usage during embryonic development could be the inactivation of genomic regions, important for early embryonic regulation, by insertions and deletions of TEs as an alternative silencing mechanism, other than DNA methylation. This would be in agreement with our observation that cluster 1 and cluster 2 are exactly contrary transcriptionally active. So far, these data remain preliminary and the clarification of the overall role of TEs in insect embryonic development demands more detailed research.

In summary, the identified temporal cluster activities mirror the timeline for developmental progression. In contrast to Holometabola, *I. elegans* embryos develop directly into the final patterned pterygote Bauplan. Early axial/spatial progenitor establishment is mediated through cluster 1 transcript activity and after the appendages are established in adaptation to the aquatic- and terrestrial life cycle, the embryo undergoes katatrepsis/revolution. Differentiation of the muscular system, organs and outgrowth of appendages is later on governed by overlapping activities of cluster 2 and cluster 3 from mid embryonic stages on. While cluster 3 activity peaks during late embryonic stages when the embryo undergoes maturation, we detect a second onset of cluster 1 activity, shortly before hatching of the individuals. This indicates that final differentiation processes of *I. elegans* depend on early embryonic genes for very late embryonic developmental specification.

Conclusion

In this study, we present the first comprehensive embryonic transcriptome of a hemimetabolous insect, the damselfly *I. elegans*. Using a single-embryo sequencing approach we were able to elucidate the transcriptional divergence of pre- and post-revolutionary embryonic stages highlighting the transcriptional complexity during insect embryogenesis. During pre-revolutionary stages, reflecting early embryogenesis until katatrepsis, transcriptional active genes were characterised for their biological functions in cell cycle, mitosis and differentiation. In addition, genes involved in signalling pathways and key development processes were enriched during these early embryonic stages. This is indicative for cell mass production for germ-band elongation and subsequent early pattern formation. During post-revolutionary stages, we identified up-regulated genes related to late embryonic development such as active movement and signal transduction for sensory perception. For the transit to the nymphal stage, we further observed activation of genes involved in circulatory, immune and neurological system maturation. The increased activity of transposable elements of different classes during mid-embryogenesis could indicate a previously unknown mechanism for developmental gene regulation. Evidently, more comprehensive embryonic transcriptomic studies of hemimetabolous insect are needed for elucidating a potential role of transposable elements and their correlation to post-revolutionary embryogenesis.

References

1. Peel, A. D., Chipman, A. D. & Akam, M. Arthropod segmentation: beyond the *Drosophila* paradigm. *Nat Rev Genet* **6**, 905–916, <https://doi.org/10.1038/nrg1724> (2005).
2. Grenier, J. K., Garber, T. L., Warren, R., Whittington, P. M. & Carroll, S. Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. *Curr Biol* **7**, 547–553 (1997).
3. Janssen, R. & Budd, G. E. Gene expression suggests conserved aspects of Hox gene regulation in arthropods and provides additional support for monophyletic Myriapoda. *EvoDevo* **1**, 4, <https://doi.org/10.1186/2041-9139-1-4> (2010).
4. Ewen-Campen, B. *et al.* The maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*. *BMC Genomics* **12**, 61, <https://doi.org/10.1186/1471-2164-12-61> (2011).
5. Zeng, V. *et al.* Developmental gene discovery in a hemimetabolous insect: de novo assembly and annotation of a transcriptome for the cricket *Gryllus bimaculatus*. *PLoS ONE* **8**, e61479, <https://doi.org/10.1371/journal.pone.0061479> (2013).
6. Chen, S. *et al.* De Novo Analysis of Transcriptome Dynamics in the Migratory Locust during the Development of Phase Traits. *PLoS One* **5**, <https://doi.org/10.1371/journal.pone.0015633> (2010).
7. Graveley, B. R. *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479, <https://doi.org/10.1038/nature09715> (2011).
8. Tomancak, P. *et al.* Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol* **8**, R145, <https://doi.org/10.1186/gb-2007-8-7-r145> (2007).
9. Arbeitman, M. N. *et al.* Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**, 2270–2275, <https://doi.org/10.1126/science.1072152> (2002).
10. Peel, A. D. The evolution of developmental gene networks: lessons from comparative studies on holometabolous insects. *Philos Trans R Soc Lond B Biol Sci* **363**, 1539–1547, <https://doi.org/10.1098/rstb.2007.2244> (2008).
11. Liu, P. Z. & Patel, N. H. giant is a bona fide gap gene in the intermediate germ band insect, *Oncopeltus fasciatus*. *Development* **137**, 835–844, <https://doi.org/10.1242/dev.045948> (2010).
12. Angelini, D. R. & Kaufman, T. C. Functional analyses in the milkweed bug *Oncopeltus fasciatus* (Hemiptera) support a role for Wnt signaling in body segmentation but not appendage development. *Dev Biol* **283**, 409–423, <https://doi.org/10.1016/j.ydbio.2005.04.034> (2005).
13. Mito, T. *et al.* Divergent and conserved roles of extradenticle in body segmentation and appendage formation, respectively, in the cricket *Gryllus bimaculatus*. *Dev Biol* **313**, 67–79, <https://doi.org/10.1016/j.ydbio.2007.09.060> (2008).
14. Mito, T. *et al.* Kruppel acts as a gap gene regulating expression of hunchback and even-skipped in the intermediate germ cricket *Gryllus bimaculatus*. *Dev Biol* **294**, 471–481, <https://doi.org/10.1016/j.ydbio.2005.12.057> (2006).
15. Mito, T. *et al.* Ancestral functions of Delta/Notch signaling in the formation of body and leg segments in the cricket *Gryllus bimaculatus*. *Development* **138**, 3823–3833, <https://doi.org/10.1242/dev.060681> (2011).
16. Hadrys, H. *et al.* Isolation of hox cluster genes from insects reveals an accelerated sequence evolution rate. *PLoS ONE* **7**, e34682, <https://doi.org/10.1371/journal.pone.0034682> (2012).
17. Ando, H. *The Comparative Embryology of Odonata with Special Reference to a Relic Dragonfly Epiophlebia Superstes Selys*. (Japan Society for the Promotion of Science, 1962).
18. Donoughe, S. & Extavour, C. G. Embryonic development of the cricket *Gryllus bimaculatus*. *Dev Biol* **411**, 140–156, <https://doi.org/10.1016/j.ydbio.2015.04.009> (2016).
19. Bentley, D., Keshishian, H., Shankland, M. & Toroianraymond, A. Quantitative Staging of Embryonic-Development of the Grasshopper, *Schistocerca-Nitens*. *J. Embryol. Exp. Morphol.* **54**, 47–74 (1979).
20. Córdoba-Aguilar, A. *Dragonflies and Damselflies: Model Organisms for Ecological and Evolutionary Research*. (Oxford University Press, Oxford, 2009).
21. Lancaster, L. T. *et al.* Gene expression under thermal stress varies across a geographical range expansion front. *Mol Ecol* **25**, 1141–1156, <https://doi.org/10.1111/mec.13548> (2016).
22. Letsch, H., Gottsberger, B. & Ware, J. L. Not going with the flow: a comprehensive time-calibrated phylogeny of dragonflies (Anisoptera: Odonata: Insecta) provides evidence for the role of lentic habitats on diversification. *Mol Ecol* **25**, 1340–1353, <https://doi.org/10.1111/mec.13562> (2016).
23. Damm, S., Dijkstra, K. D. & Hadrys, H. Red drifters and dark residents: the phylogeny and ecology of a Plio-Pleistocene dragonfly radiation reflects Africa's changing environment (Odonata, Libellulidae, Trithemis). *Mol Phylogenet Evol* **54**, 870–882, <https://doi.org/10.1016/j.ympev.2009.12.006> (2010).
24. Futahashi, R. *et al.* Extraordinary diversity of visual opsin genes in dragonflies. *Proc Natl Acad Sci USA* **112**, E1247–1256, <https://doi.org/10.1073/pnas.1424670112> (2015).
25. Bybee, S. M., Johnson, K. K., Gering, E. J., Whiting, M. F. & Crandall, K. A. All the better to see you with: a review of odonate color vision with transcriptomic insight into the odonate eye. *Organisms Diversity & Evolution* **12**, 241–250, <https://doi.org/10.1007/s13127-012-0090-6> (2012).
26. Cooper, I. A., Brown, J. M. & Getty, T. A role for ecology in the evolution of colour variation and sexual dimorphism in Hawaiian damselflies. *J Evolution Biol* **29**, 418–427, <https://doi.org/10.1111/jeb.12796> (2016).
27. Fincke, O. M. Trade-offs in female signal apparency to males offer alternative anti-harassment strategies for colour polymorphic females. *J Evolution Biol* **28**, 931–943, <https://doi.org/10.1111/jeb.12623> (2015).
28. Sanmartín-Villar, I. & Cordero-Rivera, A. The inheritance of female colour polymorphism in *Ischnura* genei (Zygoptera: Coenagrionidae), with observations on melanism under laboratory conditions. *PeerJ* **4**, e2380, <https://doi.org/10.7717/peerj.2380> (2016).
29. Bybee, S. *et al.* Odonata (dragonflies and damselflies) as a bridge between ecology and evolutionary genomics. *Front Zool* **13**, 46, <https://doi.org/10.1186/s12983-016-0176-7> (2016).
30. Shanku, A. G., McPeck, M. A. & Kern, A. D. Functional Annotation and Comparative Analysis of a Zygopteran Transcriptome. *G3 (Bethesda)* **3**, 763–770, <https://doi.org/10.1534/g3.113.005637> (2013).
31. Feindt, W., Osigus, H. J., Herzog, R., Mason, C. E. & Hadrys, H. The complete mitochondrial genome of the neotropical helicopter damselfly *Megaloprepus caerulatus* (Odonata: Zygoptera) assembled from next generation sequencing data. *Mitochondrial DNA Part B* **1**, 497–499, <https://doi.org/10.1080/23802359.2016.1192504> (2016).
32. Feindt, W., Herzog, R., Osigus, H. J., Schierwater, B. & Hadrys, H. Short read sequencing assembly revealed the complete mitochondrial genome of *Ischnura elegans* Vander Linden, 1820 (Odonata: Zygoptera). *Mitochondrial DNA Part B* **1**, 574–576, <https://doi.org/10.1080/23802359.2016.1192510> (2016).
33. Herzog, R., Osigus, H. J., Feindt, W., Schierwater, B. & Hadrys, H. The complete mitochondrial genome of the emperor dragonfly *Anax imperator* LEACH, 1815 (Odonata: Aeshnidae) via NGS sequencing. *Mitochondrial DNA Part B* **1**, 783–786, <https://doi.org/10.1080/23802359.2016.1186523> (2016).
34. Ioannidis, P. *et al.* Genomic Features of the Damselfly *Calopteryx splendens* Representing a Sister Clade to Most Insect Orders. *Genome Biol Evol* **9**, 415–430, <https://doi.org/10.1093/gbe/evx006> (2017).
35. Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767, <https://doi.org/10.1126/science.1257570> (2014).
36. Simon, S., Strauss, S., von Haeseler, A. & Hadrys, H. A phylogenomic approach to resolve the basal pterygote divergence. *Mol Biol Evol* **26**, 2719–2730 (2009).

37. Simon, S., Narechania, A., Desalle, R. & Hadrys, H. Insect phylogenomics: exploring the source of incongruence using new transcriptomic data. *Genome Biol Evol* **4**, 1295–1309, <https://doi.org/10.1093/gbe/evs104> (2012).
38. Hadrys, H., Schierwater, B., Dellaporta, S. L., DeSalle, R. & Buss, L. W. Determination of paternity in dragonflies by Random Amplified Polymorphic DNA fingerprinting. *Mol Ecol* **2**, 79–87 (1993).
39. Fincke, O. M. & Hadrys, H. Unpredictable offspring survivorship in the damselfly, *Megaloprepus coeruleus*, shapes parental behavior, constrains sexual selection, and challenges traditional fitness estimates. *Evolution* **55**, 762–772 (2001).
40. Krivist, S., Brugler, M. R., Goh, T. G., Giribet, G. & Siddall, M. E. Pyrosequencing the salivary transcriptome of *Haemadipsa interrupta* (Annelida: Clitellata: Haemadipsidae): anticoagulant diversity and insight into the evolution of anticoagulation capabilities in leeches. *Invertebrate Biology* **133**, 74–98, <https://doi.org/10.1111/ivb.12039> (2014).
41. Zheng, Y., Zhao, L., Gao, J. & Fei, Z. iAssembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences. *BMC Bioinformatics* **12**, 453, <https://doi.org/10.1186/1471-2105-12-453> (2011).
42. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq 2. *Nature protocols* **9**, 171–181, <https://doi.org/10.1038/nprot.2014.006> (2014).
43. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
44. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963, <https://doi.org/10.1093/bioinformatics/btr507> (2011).
45. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* **8**, 1494–1512, <https://doi.org/10.1038/nprot.2013.084> (2013).
46. Chauhan, P. *et al.* De novo transcriptome of *Ischnura elegans* provides insights into sensory biology, colour and vision genes. *BMC Genomics* **15**, 808, <https://doi.org/10.1186/1471-2164-15-808> (2014).
47. Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M. & Blaxter, M. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Frontiers in genetics* **4**, 237, <https://doi.org/10.3389/fgene.2013.00237> (2013).
48. Schmieder, R. & Edwards, R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS ONE* **6**, e17288, <https://doi.org/10.1371/journal.pone.0017288> (2011).
49. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659, <https://doi.org/10.1093/bioinformatics/btl158> (2006).
50. Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res* **9**, 868–877 (1999).
51. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359, <https://doi.org/10.1038/nmeth.1923> (2012).
52. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842, <https://doi.org/10.1093/bioinformatics/btq033> (2010).
53. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061–1067, <https://doi.org/10.1093/bioinformatics/btm071> (2007).
54. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212, <https://doi.org/10.1093/bioinformatics/btv351> (2015).
55. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29–37, <https://doi.org/10.1093/nar/gkr367> (2011).
56. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res* **40**, D290–301, <https://doi.org/10.1093/nar/gkr1065> (2012).
57. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785–786, <https://doi.org/10.1038/nmeth.1701> (2011).
58. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567–580, <https://doi.org/10.1006/jmbi.2000.4315> (2001).
59. Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**, 3100–3108, <https://doi.org/10.1093/nar/gkm160> (2007).
60. Roberts, A. & Pachter, L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods* **10**, 71–73, <https://doi.org/10.1038/nmeth.2251> (2013).
61. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, <https://doi.org/10.1093/bioinformatics/btp616> (2010).
62. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, <https://doi.org/10.1186/1471-2105-9-559> (2008).
63. Zheng, C. H., Yuan, L., Sha, W. & Sun, Z. L. Gene differential coexpression analysis based on biweight correlation and maximum clique. *BMC Bioinformatics*, S3, <https://doi.org/10.1186/1471-2105-15-S15-S3> (2014).
64. Langfelder, P. & Horvath, S. Fast R Functions for Robust Correlations and Hierarchical Clustering. 2012 **46**, 17, doi:<https://doi.org/10.18637/jss.v046.i11> (2012).
65. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* **4**, Article17, <https://doi.org/10.2202/1544-6115.1128> (2005).
66. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720, <https://doi.org/10.1093/bioinformatics/btm563> (2008).
67. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC systems biology* **1**, 54, <https://doi.org/10.1186/1752-0509-1-54> (2007).
68. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**, R14, <https://doi.org/10.1186/gb-2010-11-2-r14> (2010).
69. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258, <https://doi.org/10.1093/bioinformatics/btl567> (2007).
70. Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542, <https://doi.org/10.1093/bioinformatics/btl117> (2006).
71. Waringer, J. A. & Humpesch, U. H. Embryonic-Development, Larval Growth and Life-Cycle of Coenagrion-Puella (Odonata, Zygoptera) from an Austrian Pond. *Freshwater Biol* **14**, 385–399, <https://doi.org/10.1111/j.1365-2427.1984.tb00162.x> (1984).
72. Koch, K. Influence of temperature and photoperiod on embryonic development in the dragonfly *Sympetrum striolatum* (Odonata: Libellulidae). *Physiological Entomology* **40**, 90–101, <https://doi.org/10.1111/phen.12091> (2015).
73. Liu, Y. W., Zhou, J. & White, K. P. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* **30**, 301–304, <https://doi.org/10.1093/bioinformatics/btt688> (2014).
74. Mikheyev, A. S. & Linksvayer, T. A. Genes associated with ant social behavior show distinct transcriptional and evolutionary patterns. *Elife* **4**, doi:ARTN e04775 10.7554/eLife.04775 (2015).
75. Wright, R. M., Aglyamova, G. V., Meyer, E. & Matz, M. V. Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. *BMC Genomics* **16**, doi:Artn 371 10.1186/s12864-015-1540-2 (2015).
76. Brekhman, V., Malik, A., Haas, B., Sher, N. & Lotan, T. Transcriptome profiling of the dynamic life cycle of the scyphozoan jellyfish *Aurelia aurita*. *BMC Genomics* **16**, 74, <https://doi.org/10.1186/s12864-015-1320-z> (2015).

77. Santon, J. B. & Pellegrini, M. Rates of Ribosomal-Protein and Total Protein-Synthesis during Drosophila Early Embryogenesis. *Dev. Biol.* **85**, 252–257, [https://doi.org/10.1016/0012-1606\(81\)90255-4](https://doi.org/10.1016/0012-1606(81)90255-4) (1981).
78. Quinn, L. M., Herr, A., McGarry, T. J. & Richardson, H. The Drosophila Geminin homolog: roles for Geminin in limiting DNA replication, in anaphase and in neurogenesis. *Genes Dev.* **15**, 2741–2754, <https://doi.org/10.1101/Gad.916201> (2001).
79. Chauhan, P., Wellenreuther, M. & Hansson, B. Transcriptome profiling in the damselfly *Ischnura elegans* identifies genes with sex-biased expression. *BMC Genomics* **17**, 985, <https://doi.org/10.1186/s12864-016-3334-6> (2016).
80. Perry, J. C., Harrison, P. W. & Mank, J. E. The ontogeny and evolution of sex-biased gene expression in *Drosophila melanogaster*. *Mol Biol Evol* **31**, 1206–1219, <https://doi.org/10.1093/molbev/msu072> (2014).
81. Grath, S. & Parsch, J. Sex-Biased Gene Expression. *Annu. Rev. Genet.* **50**, 29–44, <https://doi.org/10.1146/annurev-genet-120215-035429> (2016).
82. Xie, W. *et al.* Transcriptomic dissection of sexual differences in *Bemisia tabaci*, an invasive agricultural pest worldwide. *Scientific reports* **4**, 4088, <https://doi.org/10.1038/srep04088> (2014).
83. Eads, B. D., Colbourne, J. K., Bohuski, E. & Andrews, J. Profiling sex-biased gene expression during parthenogenetic reproduction in *Daphnia pulex*. *BMC Genomics* **8**, 464, <https://doi.org/10.1186/1471-2164-8-464> (2007).
84. Horvath, S. & Dong, J. Geometric interpretation of gene coexpression network analysis. *PLoS computational biology* **4**, e1000117, <https://doi.org/10.1371/journal.pcbi.1000117> (2008).
85. Langfelder, P., Mischel, P. S. & Horvath, S. When is hub gene selection better than standard meta-analysis? *PLoS ONE* **8**, e61505, <https://doi.org/10.1371/journal.pone.0061505> (2013).
86. Panfilio, K. A. Extraembryonic development in insects and the acrobatics of blastokinesis. *Dev Biol* **313**, 471–491, <https://doi.org/10.1016/j.ydbio.2007.11.004> (2008).
87. Masumoto, M. & Machida, R. Development of embryonic membranes in the silverfish *Lepisma saccharina* linnaeus (insecta: Zygentoma, Lepismatidae). *Tissue Cell* **38**, 159–169, <https://doi.org/10.1016/j.tice.2006.01.004> (2006).
88. Tautz, D., Friedrich, M. & Schroder, R. Insect Embryogenesis - What Is Ancestral and What Is Derived. *Development*, 193–199 (1994).
89. Sander, K. In *Advances in Insect Physiology* Vol. Volume 12 (eds M. J. Berridge J.E. Treherne & V. B. Wigglesworth) 125–238 (Academic Press, 1976).
90. Sander, K. Pattern formation in insect embryogenesis: The evolution of concepts and mechanisms. *Int J Insect Morphol* **25**, 349–367, [https://doi.org/10.1016/S0020-7322\(96\)00021-9](https://doi.org/10.1016/S0020-7322(96)00021-9) (1996).
91. Khadjeh, S. *Establishment of the damselfly Ischnura elegans (VAND. 1823) as a new model organism: Hox gene and complex life cycle studies*, Leibniz Universität Hannover, (2008).
92. Mito, T. *et al.* Non-canonical functions of hunchback in segment patterning of the intermediate germ cricket *Gryllus bimaculatus*. *Development* **132**, 2069–2079, <https://doi.org/10.1242/dev.01784> (2005).
93. Koutsos, A. C. *et al.* Life cycle transcriptome of the malaria mosquito *Anopheles gambiae* and comparison with the fruitfly *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **104**, 11304–11309, <https://doi.org/10.1073/pnas.0703988104> (2007).
94. Brose, K. & Tessier-Lavigne, M. Slit proteins: key regulators of axon guidance, axonal branching, and cell migration. *Curr. Opin. Neurobiol.* **10**, 95–102 (2000).
95. Zhang, F., Zhao, Y., Chao, Y., Muir, K. & Han, Z. Cubilin and amnionless mediate protein reabsorption in *Drosophila* nephrocytes. *J. Am. Soc. Nephrol.* **24**, 209–216, <https://doi.org/10.1681/ASN.2012080795> (2013).
96. Maumus, F., Fiston-Lavier, A.-S. & Quesneville, H. Impact of transposable elements on insect genomes and biology. *Current opinion in insect science* **7**, 30–36, <https://doi.org/10.1016/j.cois.2015.01.001> (2015).
97. Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476, <https://doi.org/10.1038/nature02651> (2004).
98. Lunyak, V. V. *et al.* Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317**, 248–251, <https://doi.org/10.1126/science.1140871> (2007).
99. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* **9**, 397–405, <https://doi.org/10.1038/nrg2337> (2008).
100. Ding, D. & Lipshitz, H. D. Spatially regulated expression of retrovirus-like transposons during *Drosophila melanogaster* embryogenesis. *Genet. Res.* **64**, 167–181 (1994).
101. Jiang, F., Yang, M., Guo, W., Wang, X. & Kang, L. Large-scale transcriptome analysis of retroelements in the migratory locust, *Locusta migratoria*. *PLoS ONE* **7**, e40532, <https://doi.org/10.1371/journal.pone.0040532> (2012).
102. Kankare, M., Parker, D. J., Merisalo, M., Salminen, T. S. & Hoikkala, A. Transcriptional Differences between Diapausing and Non-Diapausing *D. montana* Females Reared under the Same Photoperiod and Temperature. *PLoS ONE* **11**, e0161852, <https://doi.org/10.1371/journal.pone.0161852> (2016).
103. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**, 272–285, <https://doi.org/10.1038/nrg2072> (2007).
104. Bire, S. *et al.* Mariner Transposons Contain a Silencer: Possible Role of the Polycomb Repressive Complex 2. *PLoS genetics* **12**, e1005902, <https://doi.org/10.1371/journal.pgen.1005902> (2016).

Acknowledgements

We would like to thank Rickard Sandberg and Gosta Winberg for providing recombinant hyperactive TN5 and the NY Genome Center for sequencing. We thank Sara Khadjeh for her previous work on the embryogenesis of *I. elegans* conducted in Hannover. S.Si acknowledges funding of the German Academic Exchange Service (DAAD). MRB acknowledges the Gerstner Family Foundation for providing support. HH acknowledges funding by the German Science Foundation (DFG HA 1947/5).

Author Contributions

S.Si, S.Sa, H.H. and R.D. designed the study. H.H. provided the embryo collection. S.Si, S.Sa and M.R.B. performed the molecular analyses, and the data was analysed by S.Si and E.S. The manuscript was written by S.Si and S.Sa with comments from E.S., H.H., M.E.S., M.R.B., R.D. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-13176-8>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017