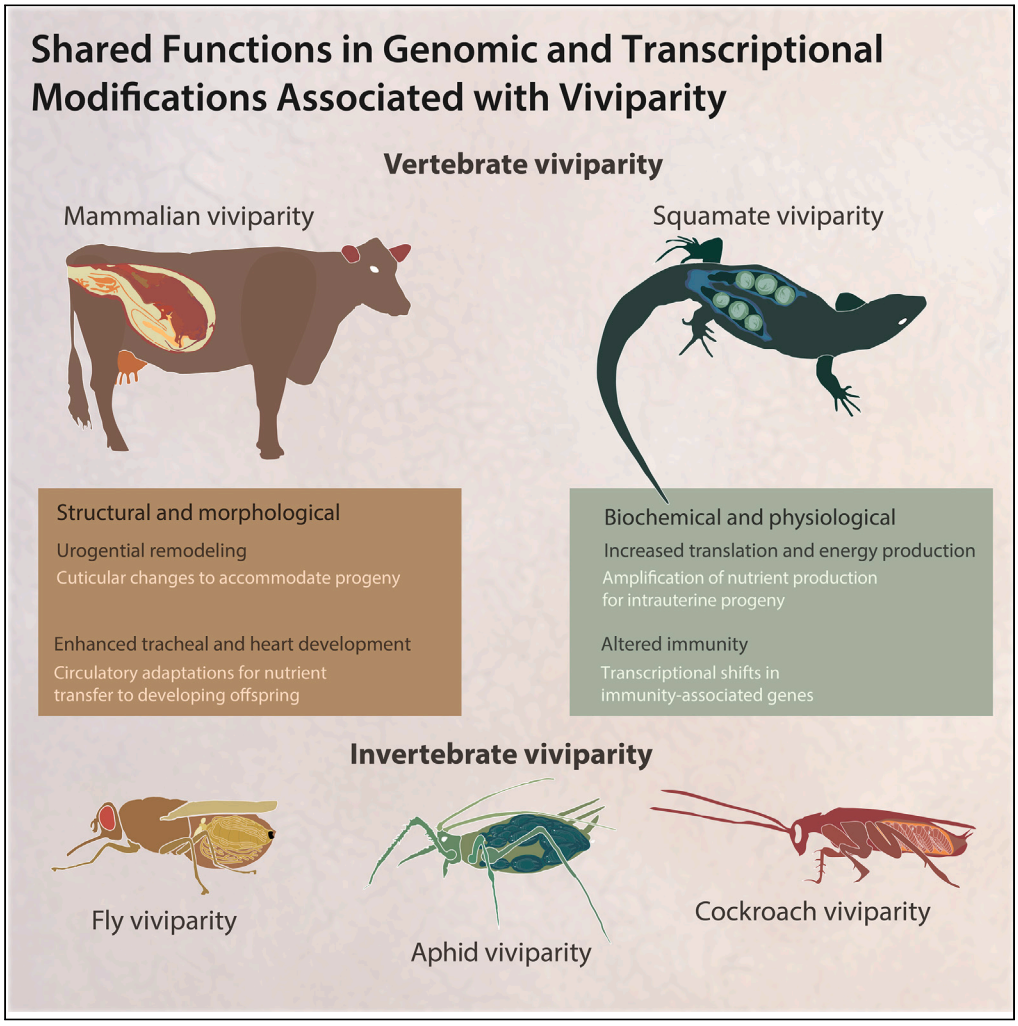
iScience

CellPress
OPEN ACCESS

Article

Live-bearing cockroach genome reveals convergent evolutionary mechanisms linked to viviparity in insects and beyond



Bertrand Fouks, Mark C. Harrison, Alina A. Mikhailova, ..., Joshua B. Benoit, Erich Bornberg-Bauer, Stephen S. Tobe

joshua.benoit@uc.edu (J.B.B.)
ebb.admin@uni-muenster.de (E.B.-B.)

Highlights

A *Diploptera* cockroach genome is now available, 1st with long-read sequencing

No gene targeted by selection is shared among independent insect viviparous origins

Similar pathways are selected during the independent animal viviparous transitions

Viviparity is linked with urogenital remodeling, protein secretion, and reduced immunity

Fouks et al., iScience *26*, 107832
October 20, 2023
https://doi.org/10.1016/j.isci.2023.107832

iScience

CellPress
OPEN ACCESS

Article

Live-bearing cockroach genome reveals convergent evolutionary mechanisms linked to viviparity in insects and beyond

Bertrand Fouks,[1,9] Mark C. Harrison,[1,9] Alina A. Mikhailova,[1] Elisabeth Marchal,[2] Sinead English,[3] Madeleine Carruthers,[3] Emily C. Jennings,[4] Ezemuoka L. Chiamaka,[4] Ronja A. Frigard,[4] Martin Pippel,[5] Geoffrey M. Attardo,[6] Joshua B. Benoit,[4,11,*] Erich Bornberg-Bauer,[1,7,*] and Stephen S. Tobe[2,8,10]

SUMMARY

Live birth (viviparity) has arisen repeatedly and independently among animals. We sequenced the genome and transcriptome of the viviparous Pacific beetle-mimic cockroach and performed comparative analyses with two other viviparous insect lineages, tsetse flies and aphids, to unravel the basis underlying the transition to viviparity in insects. We identified pathways undergoing adaptive evolution for insects, involved in urogenital remodeling, tracheal system, heart development, and nutrient metabolism. Transcriptomic analysis of cockroach and tsetse flies revealed that uterine remodeling and nutrient production are increased and the immune response is altered during pregnancy, facilitating structural and physiological changes to accommodate and nourish the progeny. These patterns of convergent evolution of viviparity among insects, together with similar adaptive mechanisms identified among vertebrates, highlight that the transition to viviparity requires changes in urogenital remodeling, enhanced tracheal and heart development (corresponding to angiogenesis in vertebrates), altered nutrient metabolism, and shifted immunity in animal systems.

INTRODUCTION

Insecta is one of the most diverse animal classes with the highest number of living species, which have colonized most habitats spanning terrestrial, freshwater, and aerial environments.[1] Insects have adapted to numerous ecological niches and display a wide range of phenotypic traits. Insect biodiversity is a valuable resource for ecosystems and the source of many new scientific discoveries.[1] For instance, insects exhibit a broad spectrum of complex traits such as sociality (solitary, gregarious, sub-to eusociality), metamorphosis (a-, hemi-, pauro- and holometabolous development), and reproductive modes (ovi- to viviparity). While the majority of insects are oviparous (egg laying), viviparity (live birth), both facultative (including ovoviviparity) and obligate, has emerged independently over 65 times across insect evolution.[2–4] Among all viviparous insects, the Pacific beetle-mimic cockroach, *Diploptera punctata*, and tsetse fly (*Glossina*), stand out by their evolutionary adaptations to have yielded specific organs that house developing progeny and produce protein-rich nutrition, which are functionally equivalent to placental structures in vertebrate.[5,6]

True viviparity is a reproductive mode in which females harbor developing embryos and other juvenile stages within their reproductive tracts until giving birth to live and active offspring.[7] In contrast, oviparity describes the reproductive mode whereby females lay eggs, while embryogenesis as well as other early development stages occur outside the female body.[7] Viviparity has gradually evolved from oviparity repeatedly and independently across the Tree of Life, for instance, within reptiles, mammals, fish, and insects,[3] suggesting this is one of the most common types of convergent evolution among animals.

[1]University of Münster, Institute for Evolution and Biodiversity, Molecular Evolution and Bioinformatics, Hüfferstrasse 1, 48149 Münster, Germany
[2]Department of Biology, Molecular Developmental Physiology and Signal Transduction Lab., Division of Animal Physiology and Neurobiology, Naamsestraat 59-Box 2465, B-3000 Leuven, Belgium
[3]School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, UK
[4]Department of Biological Sciences, University of Cincinnati, Cincinnati, OH 45221, USA
[5]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany
[6]Department of Entomology and Nematology, College of Agriculture and Environmental Sciences, University of California, Davis, Davis, CA, USA
[7]Department of Protein Evolution, Max Planck Institute for Biology, Max-Planck-Ring 5, 72076 Tübingen, Germany
[8]Department of Cell and Systems Biology, University of Toronto, Toronto, Canada
[9]These authors contributed equally
[10]Deceased
[11]Lead contact
*Correspondence: joshua.benoit@uc.edu (J.B.B.), ebb.admin@uni-muenster.de (E.B.-B.)
https://doi.org/10.1016/j.isci.2023.107832

Despite broad physiological and morphological differences among viviparous animal clades, the emergence of viviparity has led to similar physiological, morphological, and immunological changes to the female reproductive tract for vertebrate systems.[8–10] This transition requires numerous adaptations, observed in both mammalian and reptile lineages, including eggshell reduction, delayed oviposition, enhanced supply of water and nutrition to the embryo by the mother, enhanced gas exchange, and suppression of maternal immune rejection of the embryo.[8,11] The adaptation to viviparity requires acceptance of a developing non-self organism by the mother. In mammals, repression of maternal immunity toward placental cells is essential for successful pregnancy,[12] while in reptiles some viviparous squamates display reduced immunocompetence during pregnancy.[13] In both mammals and viviparous reptiles, genes and gene families share similar immunorepression roles in the uterus and placenta.[13]

The repeated and independent evolution of viviparity from an egg-depositing reproductive strategy has occurred across large taxonomic branches including reptiles, mammals, fish, and insects.[3] While the transition to viviparity has been extensively studied in vertebrates, in insects the evolutionary processes underlying such a transition remain understudied. Elucidating common patterns associated with the transition to viviparity across distantly related taxa will lead to a deeper evolutionary understanding of this complex reproductive strategy. In this context, we sequenced and assembled the genome of the viviparous cockroach, the Pacific beetle-mimic cockroach, *Diploptera punctata* (Blaberidae). With this new genome available, we performed comparative genomics and transcriptomics using 18 insect species spanning over 3 independent origins of viviparity to unravel patterns of convergent evolution and molecular mechanisms underlying this reproductive mode.

RESULTS AND DISCUSSION

Genome of the viviparous cockroach

The genome of *D. punctata* has been sequenced and made publicly available (NCBI, BioSample: SAMN25610536). We obtained a highly contiguous (contig N_{50}: 1.4 Mb) and complete (97.6% of insect BUSCOs[14]) genome assembly of length 3.13 Gb (Table S1). We then compiled a dataset comprising genomes - including this new genome - and proteomes of 18 insect species of three insect orders (Blattodea, Diptera, and Hemiptera), in which viviparity has arisen independently in *D. punctata*, *Glossina morsitans*, and in the aphid branch (*Acyrthosiphon pisum*, *Rhopalosiphum maidis*) (Figure 1A). This dataset, including a total of 22,466 orthologs of which 4,671 are single-copy, allowed us to identify genes (and related pathways) under positive selection in viviparous species (aBSREL, see STAR Methods), genes experiencing variation of selection pressure in branches where viviparity has arisen (RELAX, see STAR Methods), genes with variations of evolutionary rates correlated with viviparous origins (RERc, see STAR Methods) and gene duplication and loss in viviparous species (CAFE, see STAR Methods). To determine if transcriptional patterns related to pregnancy were shared among independent origins of viviparity, we compared the transcript levels of single-copy orthologs among six *Glossina* species[15] and *D. punctata*[6] across pregnancy using a gene co-expression analysis. To confirm that specific genes identified by our evolutionary selection and transcriptome analyses are involved in the process of viviparity, we performed RNA interference (RNAi) in *D. punctata* to suppress transcript levels of implicated genes based on previously developed methods.[6] Full comparative analyses is included in the supplement as Figures S1–S5 and Tables S1–S13.

Urogenital remodeling and early development underlie viviparous transition in insects

The transition to viviparity is accompanied by the development of novel structures that allow enhanced gas and nutrient exchange between the mother and fetus.[3,4,8,9] In vertebrates, the adaptive apex of live-birth is the development of a novel organ, the placenta,[3,4,8,9] facilitated by uterine vascular development.[9] While placenta-like structures differ in many aspects among viviparous insects,[4,5,20,21] they share similar characteristics, such as a dense tracheal and nervous innervation surrounding them, as well as the development of trophic capabilities and secretory structures.[5,20–22] Accordingly, an enhanced oxygenation associated with the transition to viviparity is reflected in our results on variation of selection pressure (see STAR Methods), which indicate a strengthened selection pressure on the tracheal system and heart contractions in viviparous origins (Figure 1B). Moreover, we find genes involved in hemolymph circulation to be mainly under relaxed selection pressure in viviparous species (Figure S1). Urogenital remodeling is also necessary in viviparous insect species to accommodate the developing fetus, with the placenta-like structure unfolding during early pregnancy in the viviparous cockroach[20] and expanding in tsetse.[5] Our results indicate that developmental processes (notably tissue development and reproduction) in viviparous species show enrichment in genes under strengthened selection pressure (Figure 1B; Tables S10 and S11), under positive selection (see STAR Methods, Figure S2; Table S7 and S8) and expanded gene families (see STAR Methods, Figure S3; Table S6), which could underlie the structural adaptations of the urogenital system of viviparous insects.[5,20,21] In addition, in both tsetse and the viviparous cockroach, the up-regulation of genes involved in chitin metabolism and related processes during early pregnancy likely underlie the process enabling structural changes of the urogenital system (Figures 2A and 2B). The involvement of protein-coding genes involved in chitin metabolism and muscle system during the unfolding (remodeling) of the placenta-like structure in the viviparous cockroach are determinant for a successful pregnancy cycle. This was demonstrated by our silencing of the gene encoding the myosin heavy chain protein essential for muscle contraction[23] resulting in disrupted or delayed pregnancy outcomes (Figure 3A). The role of such genes could be related, to some extent, to tube development (trachea and secretory structures, which are both associated with the brood sac), as those are composed of a chitinous membrane and are expanded during early pregnancy in viviparous insects.[24]

Expression of genes involved in chitin metabolism during early pregnancy might also be important in tsetse, as the first-instar larva undergoes the sclerotization (chitin metabolism) of 2 spiracles needed for respiration.[25] Another shared characteristic among the viviparous insects studied is a change in early development compared to oviparous species. Both aphids and the Pacific beetle-mimic cockroach undergo a shorter embryogenesis compared to their oviparous counterparts,[26,27] while the tsetse reproduction cycle has been adapted to produce an individual oocyte at a time.[5] Our results suggest these shifts in early development in viviparous insects were accompanied by an expansion of

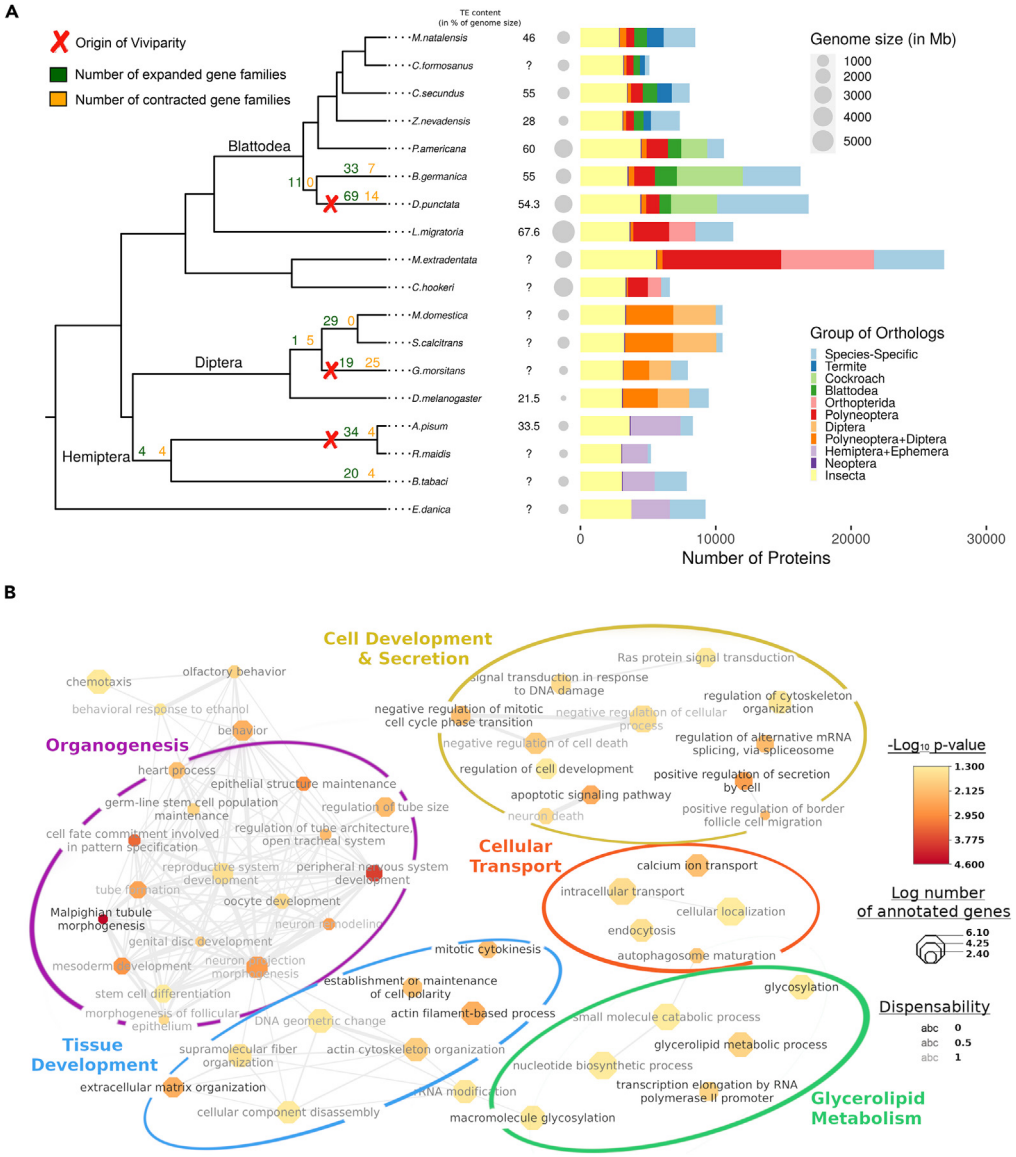


Figure 1. Comparative genomics of the origin of viviparity in 3 insect groups

(A) Genomic features associated with viviparity across 3 insect groups. The phylogenetic tree (on the left, constructed with PhyML[16] and ggtree,[17] all node supports > 0.999) depicts the evolutionary history of 18 insect species along which viviparity originated 3 times independently: in aphids, flies and cockroaches (red crosses). Green and yellow numbers represent the numbers of expanded and contracted gene families, respectively, at selected nodes, as determined with a CAFE analysis.[18] Right of the phylogenetic tree, the numbers represent the proportions of transposable elements within each species genome (in %), the gray dots depict their genome size, and the barplot the number of orthologs shared among all 18 insect species (in yellow), each clade, Blattodea (dark green), cockroaches (light green), termites (dark blue), and those that are species-specific (light blue).

(B) Functional categories enriched for genes under strengthened selection pressure in all branches where viviparity emerged (Table S10 and S11). Enriched GO terms are shown with a p value < 0.05 and are clustered into broader functions whenever possible, using REVIGO[19] and Cytoscape for visualization.

gene families (in *D. punctata* and aphids) and a strengthened selection pressure on genes involved in oocyte development (Figures 1B and S3). In particular, we observed strong purifying selection during the transition to viviparity on genes involved in oocyte and early embryo development, through pole plasm oskar mRNA localization as evidenced by our correlation of genes' relative evolutionary rate with viviparity (RERc, see STAR Methods).

Nutrient production and secretion altered in viviparous insects

The transitions to viviparity could not have been achieved in insects without the development of trophic capabilities from placenta-like structures. This translates at the physiological level into the development of an epithelial tissue with a high density of organelles involved in nutrient

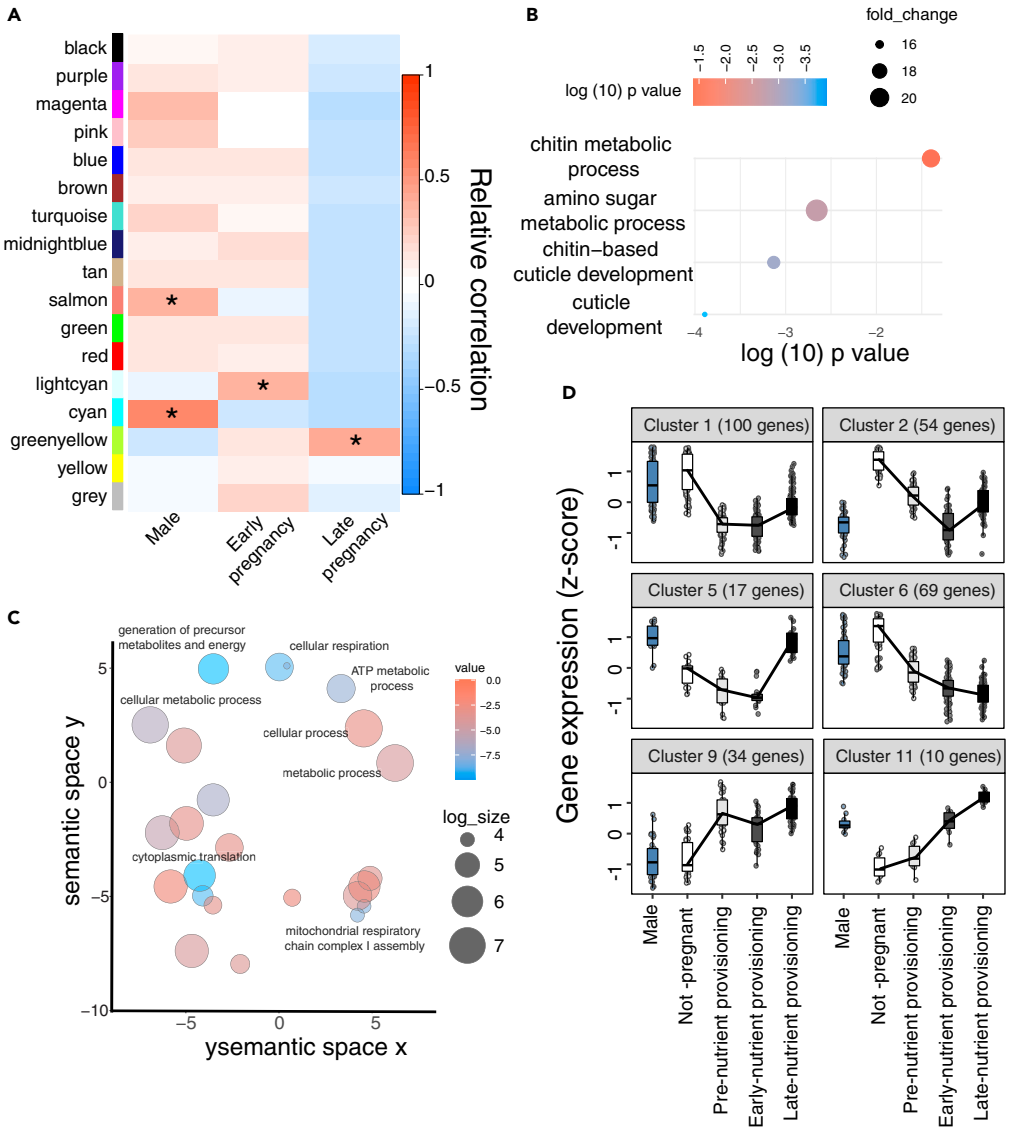


Figure 2. Transcriptional changes associated with single-copy orthologs during the course of pregnancy in Pacific beetle-mimic cockroaches, *Diploptera punctata*, and tsetse, *Glossina* spp

(A) WGCNA-based assessment examining males, early pregnancy, and late pregnancy for *Glossina* and *D. punctata*. * denotes significant correlation between transcript expression for specific genes in groups (color) and specific trait (male or stage of pregnancy).

(B) Gene ontology (GO) assessment of modules associated with early pregnancy in *Glossina* and *D. punctata*.

(C) GO categories associated with late pregnancy in *Glossina* and *D. punctata*. Dimensional space derived by applying multidimensional scaling to a matrix of the GO terms' semantic similarities.

(D) Transcriptional changes associated with immune categories that have correlated expression over the course of pregnancy in *Diploptera punctata*. Complete patterns are shown as Figure S5 Similar immune-pregnancy interactions have been observed in *Glossina* during pregnancy.[15] RNA-seq datasets were acquired from[6] for *D. punctata* and[15] for *Glossina* spp. Early pregnancy stages are before nutrient provisioning and late pregnancy stages are after nutrient provisioning. The different classification allow for *Glossina* and *D. punctata* to be grouped together. Boxplots show upper and lower bound with the bottom of the boxes representing the bottom 25th percentile and the top 75th percentile.

and energy production, as well as cellular and extra-cellular structures allowing the transfer of nutrients to the fetus.[5,20,21] The development of structures facilitating nutrient transfer will likely be aided by a change in the regulation of genes involved in chitin metabolism observed during early pregnancy in the viviparous cockroach and tsetse, favoring tube development as previously mentioned. Furthermore, the structural changes enhancing nutrient transfer associated with viviparous transition in insects may have been driven by a strengthened selection pressure on genes involved in the cytoskeleton and extracellular matrix organization (Figure 1B; Tables S10 and S11). The transport of nutrients from the placenta-like structure to the fetus requires enhanced cellular secretion,[20] this adaptation was accompanied by a strengthened

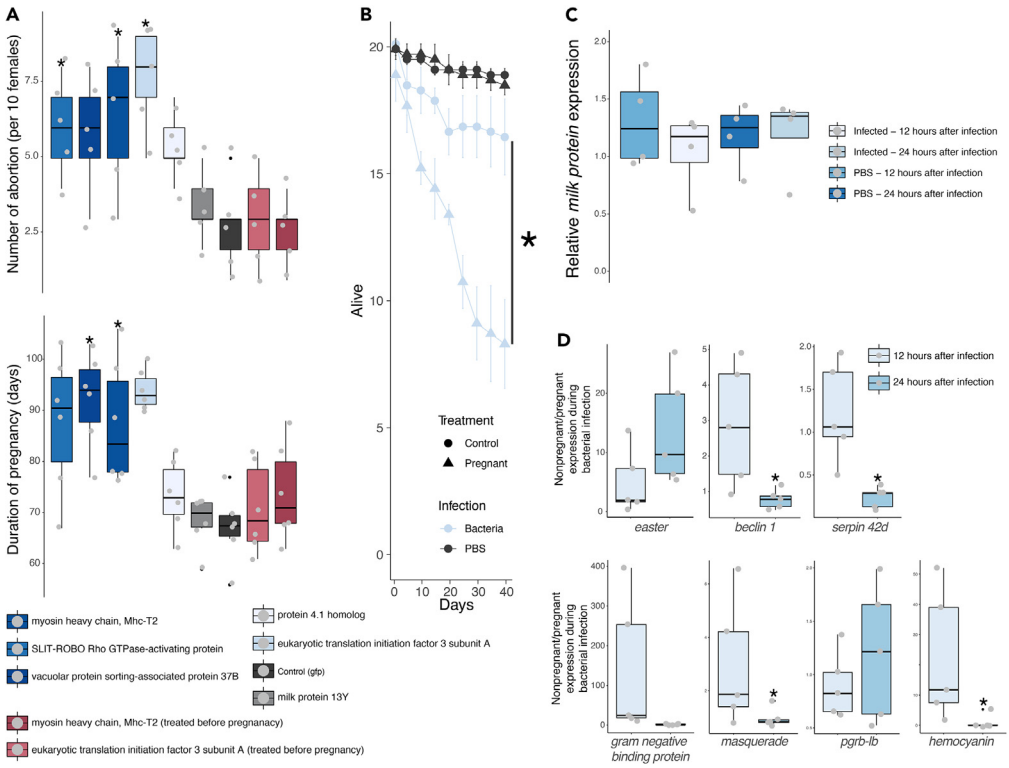


Figure 3. Functional validation of genes and the immune response during viviparity in *Diploptera punctata*

(A) RNA interference of genes of interest, selected based on our genomic and transcriptomic analyses, at each of the different pregnancy stages in *D. punctata* (before or after progeny is developing in brood sac) confirmed the role of specific genes during pregnancy leading to disruption of pregnancy and/or delays in birth. Suppression of transcript expression of target genes before pregnancy had a more muted impact on pregnancy phenotypes.

(B) Immune challenge in pregnant compared to non-pregnant females confirmed the reduced immunity during the course of pregnancy associated with a reduction of survival rate.

(C) Transcripts levels for a milk protein are not reduced during bacterial infection ($p>0.05$).

(D) Differences in the transcriptional response of immune proteins highlights that the response between non-pregnant and pregnant cockroaches is different. * denotes significance based on a linear model or t-test at $P<0.05$. Boxplots whiskers show upper and lower bound with the bottom of the boxes representing the bottom 25th percentile and the top 75th percentile.

selection pressure and positive selection of genes involved in cell secretion and vesicle formation in all branches where viviparity has arisen (Figures 1B and S2; Tables S7, S8, S10, and S11). Moreover, a strong conservation of the gene sequence encoding the vacuolar protein sorting-associated protein 37B, predicted to be involved in multivesicular body sorting pathway,[28] is correlated with the transition to viviparity in insects (RERc, see STAR Methods, Table S12). When we silenced the *Vps37B* gene during pregnancy of the viviparous cockroach, this led to delays in birth (Figure 3A). Such birth outcomes are likely due to a reduction of nutrient transport efficiency to the embryo.

The nutrients produced by mothers for their offspring during pregnancy have been well characterized in the viviparous cockroach and tsetse and are mainly composed of lipids, proteins and lower levels of carbohydrates.[20,29–31] Several adaptive mechanisms are implicated in changes enhancing nutrient production during pregnancy in insects, as we observed several molecular signatures related to nutrient production during viviparous transition. Specifically, genes involved in the glycerolipid metabolic pathway display either signals of strengthened and relaxed selection pressure associated with transitions to viviparity (Figures 1B and S1; Tables S10 and S11). Lipoprotein and/or carbohydrate pathways are enriched in genes under positive selection in all viviparous insects (Figure S2; Tables S7 and S8). Genes involved in translation undergo fast evolution associated with viviparous transition, as reflected by high relative evolutionary rates in viviparous branches. In addition to rapid evolution, genes involved in translation are up-regulated during late pregnancy in the viviparous cockroach and tsetse (Figure 2). Nevertheless, in the viviparous cockroach, the suppression of a single milk gland protein, a key component of nutrients for the developing embryo, did not impact pregnancy (Figure 3A). This is likely due to the presence of over 20 similar milk protein genes that can compensate to feed the embryos.[6,15] Gene families related to lipid and energy metabolism have expanded in the viviparous cockroach and tsetse respectively (Figure S3; Table S5). In tsetse, the transition to viviparity is associated with a loss of genes involved in aminoglycan metabolic pathway (Figure S4; Table S5). Hence, the production of lipids, proteins and carbohydrates necessary to feed the developing fetus in viviparous insects is likely to have been adapted through diverse evolutionary mechanisms mainly targeting translation, lipid and protein metabolism pathways.

The regulation of nutrient production and secretion is mainly governed by maternal hormonal control in viviparous insects, notably by the juvenile hormone.[6,27,32] It is hypothesised that the transition to viviparity led to evolutionary conflicts between females and males, and between mothers and offspring, over resource allocation.[3] Our analyses indicate the relaxation of selection pressure on hormone metabolism pathways in viviparous branches, which could highlight such evolutionary conflict over resource allocations, as these pathways represent the main source by which to regulate partitioning of nutrients for the embryo (Figure S1; Tables S10 and S11). Traits associated with evolutionary conflict over resource allocation are predicted to be under relaxed selection, leading to genetic polymorphism in populations.[33] Such relaxation of selection originates from genes undergoing selection pressure in antagonistic directions, due to maternal-offspring conflict.

Changes in immune and stress responses in viviparous insects enhance fetus acceptance

Aside from major structural and physiological changes, a necessary adaptation enabling the transition to viviparity in both invertebrates and vertebrates is the acceptance of the developing fetus within a mother's womb. We found pathways related to stress and immune responses to be enriched in genes under relaxed selection pressure during viviparous transitions (Figure S1; Tables S10 and S11). In many insects, immune pathways are expected to be the target of strong positive selection.[34,35] This denotes the importance of reduced immune activity during the transition to viviparity to minimize embryo rejection, as evidenced by our observed patterns of gene expression and responses to infection in pregnant *D. punctata* females (Figure 2D). Indeed, closer examination of expression of immune-related genes in *D. punctata*, both across pregnancy and in comparison to males (comparative control) and non-pregnant females, revealed immune changes associated with pregnancy (Figure 2D). Multiple immune genes are differentially expressed between pregnant and non-pregnant cockroaches during bacterial infection (Figure 3). These immune changes during pregnancy likely lead to a reduced disease resistance of pregnant female cockroaches, as they died more rapidly following bacterial infection compared to non-pregnant individuals (Figure 3B). Moreover, transcript levels for a milk protein are not reduced during infection. This suggests that the females are not increasing investment into the developing embryos during infection, but instead are trying to simultaneously complete pregnancy and respond to the infection (Figure 3C). These studies confirm that genomic and transcriptomic factors identified by our analyses are directly linked to cockroach viviparity that are similar to those in vertebrate systems. Lastly, the viviparous cockroach and aphids experienced gains of genes involved in stress and immune response (mainly the Toll signaling pathway), while in tsetse those pathways undergo a reduction of gene number (Figure S4; Table S5). This suggests that convergent evolutionary trajectories to minimize embryo rejection in insects are undertaken by diametrically opposed adaptive mechanisms across species (i.e., gain versus loss of genes), but targeted studies need to be conducted to confirm if specific differences are linked to viviparity.

Gene gain and loss of the Toll signaling pathway in viviparous species could also underlie adaptations for embryo development rather than solely immunity, as the Toll signaling pathway plays a major role in both early embryo development and immune response.[36,37] Similarly, the evolutionary signals of genes involved in the stress response could result from adaptations during the larval stage to mitigate stress related to lack of oxygen and accumulation of uric acid.[22,31] Our results support these general findings with evidence of gain of genes related to hypoxia in aphids (Figure S3C; Table S5), gain and loss of genes involved in nitrogen compound metabolism in tsetse and in the viviparous cockroach, respectively (Figures S3B and S4A; Table S5), and loss of genes involved in response to nitrogen compounds in tsetse (Figure S4B; Table S5).

Conclusions

Overall, our study reveals that similar pathways are shared among the three origins of viviparity in insects, but there is little overlap among specific genes involved in these adaptations. This is not surprising, as differences between insect lineages are substantial, and pathways necessary for viviparity development likely vary. Moreover, our study suggests that similar pathways can undergo different and even opposite adaptive mechanisms depending on viviparity origins. The transition to viviparity in insects is associated with morphological and physiological adaptations during early developmental stages, enabled by a combination of gene duplication and a strong selection pressure on genes involved in oogenesis and embryogenesis. In addition, our analyses highlight uterine remodeling and enhanced maternal-fetal gas exchange associated with viviparity. Other changes include alterations in gene expression associated with cuticular metabolism along with positive selection and strengthened selective pressure on genes associated with urogenital, tissue, tracheal, and heart development. In addition to pathways involved in uterine remodeling, we found that the development of placental-like structures in viviparous insects are associated with nutrient production and secretion, driven mainly by a strong positive selection and strengthened selection pressure on genes related to protein and lipid metabolism and cellular secretion, and the fast evolution and shift in expression during pregnancy of genes involved in translation. Furthermore, immunity and maternal tolerance of the embryo during the transition to viviparity is corroborated by altered immune responses during early pregnancy in the viviparous cockroach, and relaxed selection on the immune response in all three studied viviparous insect branches. Despite broad physiological and morphological differences, the adaptation to viviparity appears to be common among animals and involves similar pathways, but not necessarily common genes and adaptive mechanisms.[9]

Our study highlights that convergent evolution of viviparity among animals displays similar adaptations through the co-option of different genetic toolkits. In both insects and vertebrates, pathways with common functions including urogenital remodeling, circulatory/respiratory development, nutrient provisioning and immunity have undergone rapid evolutionary shifts (Figure 4). These shifts are undertaken by identical aspects of evolutionary adaptive mechanisms including functional co-option, duplication and changes in tissue and temporal expression patterns. These evolutionary mechanisms are associated with pregnancy and represent corresponding functions across animal taxa.[38] The evolution of extended parental care, such as viviparity and prolonged nutrient provisioning, has been suggested to prevent predation and increase survival in uncertain environments.[39,40] Thus, the selection pressures associated with the evolution of viviparity is likely similar between vertebrates and invertebrates and acts to yield functionally similar changes. Comparative genomics of insects represents an

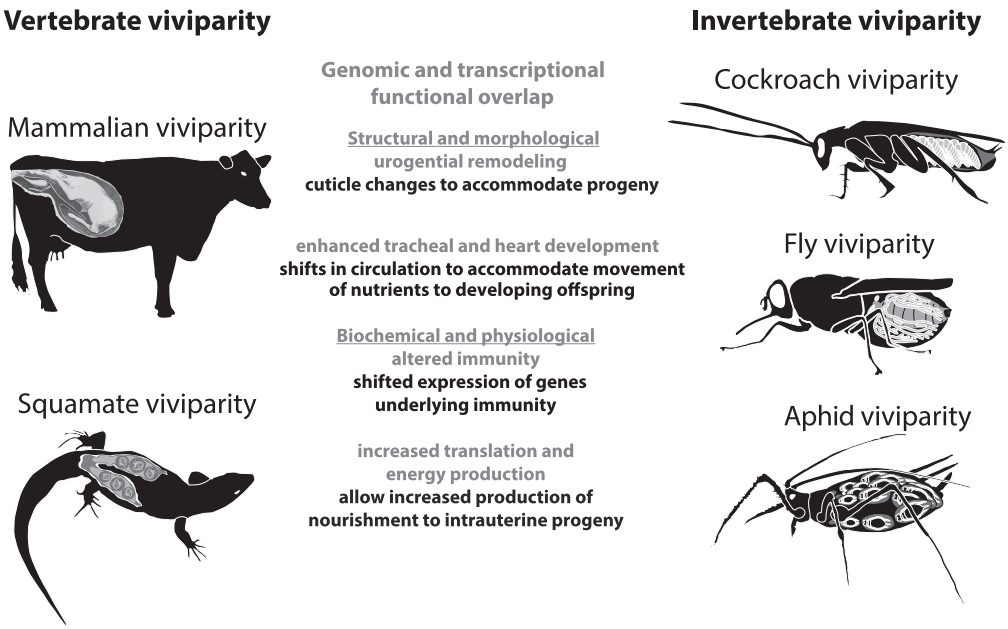


Figure 4. Summary of factors that overlap between vertebrate and invertebrate viviparity

The specific aspects identified were established by a combination of genomic and transcriptomic analyses. Gray indicates common factor between invertebrate and vertebrate viviparity and black is the specific aspect identified in this study for insect systems. Aspects associated with vertebrate viviparity are based upon previous studies.[8,9,42]

excellent avenue to study in depth the genomic basis of the gradual emergence of viviparity, since different forms of viviparity as well as intermediate stages (ovoviviparity) are well represented in this group.[2,41]

Limitations of the study

This study shows that there are similar genomic and transcriptional changes in invertebrates compared to vertebrates in relation to viviparity. Although our studies provide strong evidence, sequencing of additional viviparous insect lineages could lead to increased evidence for functional convergence between invertebrate and vertebrate viviparity. Moreover, our selection analyses, like all genome-wide analyses for detection of selection, are restricted to a subset of highly conserved ortholog families. Our analyses were, indeed, restricted to only single-copy ortholog families, and multiple sequence alignments were filtered to remove unreliable sites and gaps limiting spurious detection of selection. Nevertheless, our conservative method ensures the robustness of our results and avoids false positives. The use of gene ontology (GO) terms and enrichment analyses to identify key functions under various selection pressure, like all studies using GO terms and similar analyses, is subject to caution for two main reasons. The association of a GO term with a gene to predict its function often relies on few animal models or predictions and therefore the gene function in other species cannot be ascertained. The interpretation of functional enrichment analyses is almost unlimited, due to the lack of *a priori* hypothesis.[43] Therefore, further investigation of individual and groups of genes highlighted in our study is needed to assess their function in each viviparous species. In our study, however, the highlighted functions underlying the transition to viviparity (1) were precise, such as 'reproductive system development' or 'positive regulation of secretion by cell', limiting over-interpretation, (2) were shared among our several selection analyses and viviparous species, (3) clearly corresponded to major adaptations underlying viviparous transition, and (4) were tested against specific genes of interest, to assess their impact on viviparity. Another limitation of our study is the lack of certainty that pathways found to be under strengthened or relaxed selection pressure during viviparous transition are not experiencing similar selection pressure in oviparous species. Nevertheless, those analyses were performed to understand which functions were undergoing similar selection pressure during independent viviparous origins. Therefore, the occurrence of similar selection pressures on functions associated with viviparity in an oviparous species does not nullify that these functions share similar selection pressure during all viviparous origins studied.

Furthermore, our RNAi studies targeted genes that are likely to be core functional genes, which could have more general effects that make females less successful independent of viviparity. This is partially alleviated as early interference through dsRNA injection did not have the same effect. Similarly, immune pressure due to bacterial infection may have an impact that reduces overall female health and success. Future studies could examine more nuanced effects in relation to immune dynamics and specific genes that are critical during pregnancy. Another factor that could increase the understanding of insect viviparity would be the genomic and transcriptome sequencing and analysis of other lineages that give live birth, such as other viviparous flies and dermapterans (earwigs). These additional resources would allow for examining convergent factors among more lineages, which could provide more conclusive evidence for changes associated with insect viviparity.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Experimental model: *Diploptera punctata*
- METHOD DETAILS
  - Genome sequencing and assembly
  - Repeat annotation
  - Gene annotation
  - Ortholog detection and phylogeny
  - Duplication and loss events
  - Multiple sequence alignment
  - Identifying selection pressures
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Functional enrichment of genes with selection signals
  - RNA-seq analyses
  - RNAi, immune and survival assays

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107832.

ACKNOWLEDGMENTS

We thank David L. Denlinger for comments and suggestions on an early draft of this manuscript and Carsten Kemena, Elias Dohmen, and Anna Grandchamp for their help in troubleshooting the comparative genomics pipeline. This work was supported by a Discovery grant, no. 9408-08 from the Natural Sciences and Engineering Research Council of Canada to S.S.T, National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R01AI148551 and National Science Foundation DEB1654417 (to J.B.B. for shared computational resources), B.F. was supported by an EU-H2020 MSCA-IF-2020 fellowship (101024100, TEEPI), M.C.H. was supported by a DFG grant BO2544/11-2 to E.B.B, A.M. was supported by a DFG grant HA8997/1-1 to M.C.H. and S.E. was supported by a Royal Society Dorothy Hodgkin Fellowship (DH140236). Stephen S. Tobe sadly passed away during the final production of this manuscript. Dr. Tobe made critical contributions to this article and his work on invertebrate biology has been foundational.

AUTHOR CONTRIBUTIONS

B.F. performed the multiple sequence alignments, detection of selection, and functional categories enrichment analyses, wrote the first version of the manuscript; M.C.H. performed the genome assembly and annotations, ortholog clustering, gene family evolution analyses, wrote the first version of the manuscript; A.A.M. performed the annotation of chemoreceptors; E.M. participated in the organization of the project; S.E. participated in the RNA-seq analysis; M.C. performed the RNA-seq analysis; E.C.J. performed the RNA-seq analysis and collected samples for DNA sequencing; G.M.A. contributed to figure development; M.P. assisted in the genome assembly; J.B.B. provided funding, performed the RNA-seq analysis, and functional gene suppression studies; E.L.C. and R.A.F. performed functional studies on the immune response and collected RNA; E.B.B. provided funding and supervised bioinformatics analyses and organization of the project; S.S.T. provided funding and participated in early organization of the project. All authors contributed to revisions of the manuscript with the exception of SST †.

DECLARATION OF INTERESTS

The authors declare no competing interest.

Received: March 8, 2022
Revised: February 13, 2023
Accepted: September 1, 2023
Published: September 9, 2023

REFERENCES

1. Scudder, G.G.E. (2017). The Importance of Insects. In Insect Biodiversity, R.G. Foottit and P.H. Adler, eds. (Oxford, UK: John Wiley & Sons), pp. 9–43.
2. Meier, R., Kotrba, M., and Ferrar, P. (2007). Ovoviviparity and viviparity in the Diptera. Biol. Rev. *74*, 199–258. https://doi.org/10.1111/j.1469-185X.1999.tb00186.x.
3. Kalinka, A.T. (2015). How did viviparity originate and evolve? Of conflict, co-option, and cryptic choice. Bioessays *37*, 721–731. https://doi.org/10.1002/bies.201400200.
4. Ostrovsky, A.N., Lidgard, S., Gordon, D.P., Schwaha, T., Genikhovich, G., and Ereskovsky, A.V. (2016). Matrotrophy and placentation in invertebrates: a new paradigm. Biol. Rev. Camb. Philos. Soc. *91*, 673–711. https://doi.org/10.1111/brv.12189.
5. Benoit, J.B., Attardo, G.M., Baumann, A.A., Michalkova, V., and Aksoy, S. (2015). Adenotrophic viviparity in tsetse flies: Potential for population control and as an insect model for lactation. Annu. Rev. Entomol. *60*, 351–371. https://doi.org/10.1146/annurev-ento-010814-020834.
6. Jennings, E.C., Korthauer, M.W., Hendershot, J.M., Bailey, S.T., Weirauch, M.T., Ribeiro, J.M.C., and Benoit, J.B. (2020). Molecular mechanisms underlying milk production and viviparity in the cockroach, *Diploptera punctata*. Insect Biochem. Mol. Biol. *120*, 103333. https://doi.org/10.1016/j.ibmb.2020.103333.
7. Lodé, T. (2012). Oviparity or viviparity? That is the question. Reprod. Biol. *12*, 259–264. URL. https://doi.org/10.1016/j.repbio.2012.09.001. https://linkinghub.elsevier.com/retrieve/pii/S1642431X12000022.
8. Gao, W., Sun, Y.B., Zhou, W.W., Xiong, Z.J., Chen, L., Li, H., Fu, T.T., Xu, K., Xu, W., Ma, L., et al. (2019). Genomic and transcriptomic investigations of the evolutionary transition from oviparity to viviparity. Proc. Natl. Acad. Sci. USA *116*, 3646–3655. https://doi.org/10.1073/pnas.1816086116.
9. Recknagel, H., Carruthers, M., Yurchenko, A.A., Nokhbatolfoghahai, M., Kamenos, N.A., Bain, M.M., and Elmer, K.R. (2021). The functional genetic architecture of egg-laying and live-bearing reproduction in common lizards. Nat. Ecol. Evol. *5*, 1546–1556. https://doi.org/10.1038/s41559-021-01555-4.
10. Blackburn, D.G. (2015). Viviparous placentotrophy in reptiles and the parent-offspring conflict. J. Exp. Zool. B Mol. Dev. Evol. *324*, 532–548. https://doi.org/10.1002/jez.b.22624.
11. Murphy, B.F., and Thompson, M.B. (2011). A review of the evolution of viviparity in squamate reptiles: The past, present and future role of molecular biology and genomics. J. Comp. Physiol. B *181*, 575–594. https://doi.org/10.1007/s00360-011-0584-0.
12. Moffett, A., and Loke, C. (2006). Immunology of placentation in eutherian mammals. Nat. Rev. Immunol. *6*, 584–594. https://doi.org/10.1038/nri1897.
13. Van Dyke, J.U., Brandley, M.C., and Thompson, M.B. (2014). The evolution of viviparity: Molecular and genomic data from squamate reptiles advance understanding of live birth in amniotes. Reproduction *147*, R15–R26. https://doi.org/10.1530/REP-13-0309.
14. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). Busco: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.
15. Attardo, G.M., Abd-Alla, A.M.M., Acosta-Serrano, A., Allen, J.E., Bateta, R., Benoit, J.B., Bourtzis, K., Caers, J., Caljon, G., Christensen, M.B., et al. (2019). Comparative genomic analysis of six *Glossina* genomes, vectors of African trypanosomes. Genome Biol. *20*, 187–231. https://doi.org/10.1186/s13059-019-1768-2.
16. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of phyml 3.0. Syst. Biol. *59*, 307–321.
17. Yu, G. (2020). Using ggtree to visualize data on tree-like structures. Curr. Protoc. Bioinformatics *69*, e96. https://doi.org/10.1002/cpbi.96.
18. De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: A computational tool for the study of gene family evolution. Bioinformatics *22*, 1269–1271. https://doi.org/10.1093/bioinformatics/btl097.
19. Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800. https://doi.org/10.1371/journal.pone.0021800.
20. Stay, B., and Coop, A.C. (1974). 'Milk' secretion for embryogenesis in a viviparous cockroach. Tissue Cell *6*, 669–693. https://doi.org/10.1016/0040-8166(74)90009-3.
21. Bermingham, J., and Wilkinson, T.L. (2009). Embryo nutrition in parthenogenetic viviparous aphids. Physiol. Entomol. *34*, 103–109. https://doi.org/10.1111/j.1365-3032.2008.00669.x.
22. Jaglarz, M.K., Tworzydlo, W., Rak, A., Kotula-Balak, M., Sekula, M., and Bilinski, S.M. (2019). Viviparity in the dermapteran *Arixenia esau*: respiration inside mother's body requires both maternal and larval contribution. Protoplasma *256*, 1573–1584. https://doi.org/10.1007/s00709-019-01402-1.
23. Viswanathan, M.C., Tham, R.C., Kronert, W.A., Sarsoza, F., Trujillo, A.S., Cammarato, A., and Bernstein, S.I. (2017). Myosin storage myopathy mutations yield defective myosin filament assembly in vitro and disrupted myofibrillar structure and function in vivo. Hum. Mol. Genet. *26*, 4799–4813. https://doi.org/10.1093/hmg/ddx359.
24. Roth, L.M., and Stay, B. (1958). The occurrence of para-quinones in some arthropods, with emphasis on the quinone-secreting tracheal glands of *Diploptera punctata* (Blattaria). J. Insect Physiol. *1*, 305–318. https://doi.org/10.1016/0022-1910(58)90049-0.
25. Žďárek, J., Weyda, F., Chimtawi, M.M., and Denlinger, D.L. (1996). Functional morphology and anatomy of the polypneustic lobes of the last larval instar of tsetse flies, *Glossina* spp. (Diptera: Glossinidae). Int. J. Insect Morphol. Embryol. *25*, 235–248. https://doi.org/10.1016/0020-7322(96)00008-6.
26. Le Trionnaire, G., Hardie, J., Jaubert-Possamai, S., Simon, J.C., and Tagu, D. (2008). Shifting from clonal to sexual reproduction in aphids: physiological and developmental aspects. Biol. Cell *100*, 441–451. https://doi.org/10.1042/bc20070135.
27. Marchal, E., Hult, E.F., Huang, J., Stay, B., and Tobe, S.S. (2013). *Diploptera punctata* as a model for studying the endocrinology of arthropod reproduction and development. Gen. Comp. Endocrinol. *188*, 85–93. https://doi.org/10.1016/j.ygcen.2013.04.018.
28. Li, Z., and Blissard, G. (2015). The vacuolar protein sorting genes in insects: A comparative genomeview. Insect Biochem. Mol. Biol. *62*, 211–225. https://doi.org/10.1016/j.ibmb.2014.11.007.
29. Benoit, J.B., Attardo, G.M., Michalkova, V., Krause, T.B., Bohova, J., Zhang, Q., Baumann, A.A., Mireji, P.O., Takáč, P., Denlinger, D.L., et al. (2014). A novel highly divergent protein family identified from a viviparous insect by RNA-seq analysis: a potential target for tsetse fly-specific abortifacients. PLoS Genet. *10*, e10038744. https://doi.org/10.1371/journal.pgen.1003874.
30. Williford, A., Stay, B., and Bhattacharya, D. (2004). Evolution of a novel function: Nutritive milk in the viviparous cockroach, *Diploptera punctata*. Evol. Dev. *6*, 67–77. https://doi.org/10.1111/j.1525-142X.2004.04012.x.
31. Stay, B., and Coop, A. (1973). Developmental stages and chemical composition in embryos of the cockroach, *Diploptera punctata*, with observations on the effect of diet. J. Insect Physiol. *19*, 147–171. https://doi.org/10.1016/0022-1910(73)90230-8.
32. Attardo, G.M., Guz, N., Strickler-Dinglasan, P., and Aksoy, S. (2006). Molecular aspects of viviparous reproductive biology of the tsetse fly (*Glossina morsitans morsitans*): Regulation of yolk and milk gland protein synthesis. J. Insect Physiol. *52*, 1128–1136. https://doi.org/10.1016/j.jinsphys.2006.07.007.
33. Champagne, F.A., and Curley, J.P. (2012). The Evolution of Parental Care, Chapter Genetics and Epigenetics of Parental Care (Oxford University Press).
34. Sackton, T.B. (2019). Comparative genomics and transcriptomics of host–pathogen interactions in insects: evolutionary insights and future directions. Curr. Opin. Insect Sci. *31*, 106–113. https://doi.org/10.1016/j.cois.2018.12.007.
35. Viljakainen, L., Evans, J.D., Hasselmann, M., Rueppell, O., Tingek, S., and Pamilo, P. (2009). Rapid evolution of immune proteins in social insects. Mol. Biol. Evol. *26*, 1791–1801. https://doi.org/10.1093/molbev/msp086.
36. Anthoney, N., Foldi, I., and Hidalgo, A. (2018). Toll and toll-like receptor signalling in development. Dev *145*. dev156018–6. https://doi.org/10.1242/dev.156018.
37. Valanne, S., Wang, J.H., and Rämet, M. (2011). The *Drosophila* Toll signaling pathway. J. Immunol. *186*, 649–656. https://doi.org/10.4049/jimmunol.1002302.
38. Griffith, O.W., and Wagner, G.P. (2017). The placenta as a model for understanding the origin and evolution of vertebrate organs. Nat. Ecol. Evol. *1*, 72. https://doi.org/10.1038/s41559-017-0072.
39. Benoit, J.B., Kölliker, M., and Attardo, G.M. (2019). Putting invertebrate lactation in

context. Science *363*, 593. https://doi.org/10.1126/science.aaw2732.
40. Klug, H., and Bonsall, M.B. (2010). Life history and the evolution of parental care. Evolution *64*, 823–835. https://doi.org/10.1111/j.1558-5646.2009.00854.x.
41. Gavrilov-Zimin, I.A. (2021). Egg retention, viviparity and ovoviviparity in paraneoptera. Comp. Cytogenet. *15*, 239–252. https://doi.org/10.3897/compcytogen.v15.i3.70216.
42. Warren, W.C., and Grutzner, F. (2021). The Evolution of Viviparity in Vertebrates. In Placentation in Mammals, pp. 7–19. https://doi.org/10.1007/978-3-030-77360-12.
43. Pavlidis, P., Jensen, J.D., Stephan, W., and Stamatakis, A. (2012). A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. Mol. Biol. Evol. *29*, 3237–3248. https://doi.org/10.1093/molbev/mss136.
44. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
45. Marcais, G., and Kingsford, C. (2012). Jellyfish: A fast k-mer counter. Tutorialis e Manuais *1*, 1–8.
46. Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). Genomescope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. Nat. Commun. *11*, 1432–1510.
47. Amgarten, D., Braga, L.P.P., da Silva, A.M., and Setubal, J.C. (2018). MARVEL, a tool for prediction of bacteriophage sequences in metagenomic bins. Front. Genet. *9*, 304. https://doi.org/10.3389/fgene.2018.00304.
48. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. *5*, R12. https://doi.org/10.1186/gb-2004-5-2-r12.
49. Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. *27*, 737–746.
50. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
51. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One *9*, e112963.
52. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. *38*, 4647–4654. https://doi.org/10.1093/molbev/msab199.
53. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117.
54. Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). Ltrharvest, an efficient and flexible software for de novo detection of ltr retrotransposons. BMC Bioinf. *9*, 18.
55. Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics *26*, 680–682. https://doi.org/10.1093/bioinformatics/btq003.
56. Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. *31*, 5654–5666.
57. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using evidencemodeler and the program to assemble spliced alignments. Genome Biol. *9*, R7–R22.
58. Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics *Chapter 4*, 4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25.
59. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.
60. Keilwagen, J., Hartung, F., and Grau, J. (2019). Gemoma: homology-based gene prediction utilizing intron position conservation and rna-seq data. In Gene Prediction (Springer), pp. 161–177.
61. Iwata, H., and Gotoh, O. (2012). Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. Nucleic Acids Res. *40*, e161. https://doi.org/10.1093/nar/gks708.
62. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., and Bateman, A. (2021). Pfam: the protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. https://doi.org/10.1093/nar/gkaa913.
63. Vizueta, J., Sánchez-Gracia, A., and Rozas, J. (2020). bitacora: A comprehensive tool for the identification and annotation of gene families in genome assemblies. Mol. Ecol. Resour. *20*, 1445–1452. https://doi.org/10.1111/1755-0998.13202.
64. Slater, G.S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics *6*, 31. https://doi.org/10.1186/1471-2105-6-31.
65. Katoh, K., and Standley, D.M. (2013). Mafft multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.
66. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321. https://doi.org/10.1093/sysbio/syq010.
67. Han, M.V., Thomas, G.W., Lugo-Martinez, J., and Hahn, M.W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol. Biol. Evol. *30*, 1987–1997. https://doi.org/10.1093/molbev/mst100.
68. Löytynoja, A. (2021). Phylogeny-aware alignment with PRANK and PAGAN. Methods Mol. Biol. *2231*, 17–37. https://doi.org/10.1007/978-1-0716-1036-7_2.
69. Sela, I., Ashkenazy, H., Katoh, K., and Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. *43*. W7–14. https://doi.org/10.1093/nar/gkv318.
70. Pond, S.L., Frost, S.D., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. Bioinformatics *21*, 676–679. https://doi.org/10.1093/bioinformatics/bti079.
71. Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S.L. (2015). Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol. Biol. Evol. *32*, 1342–1353. https://doi.org/10.1093/molbev/msv022.
72. Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., and Scheffler, K. (2015). RELAX: Detecting relaxed selection in a phylogenetic framework. Mol. Biol. Evol. *32*, 820–832. https://doi.org/10.1093/molbev/msu400.
73. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589. https://doi.org/10.1038/nmeth.4285.
74. Emms, D.M. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.
75. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.
76. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80. https://doi.org/10.1186/gb-2004-5-10-r80.
77. Alexa, A., and Rahnenfuhrer, J. (2023). topGO: Enrichment Analysis for Gene Ontology (R package version 2.52.0).
78. Kowalczyk, A., Meyer, W.K., Partha, R., Mao, W., Clark, N.L., and Chikina, M. (2019). RERconverge: An R package for associating evolutionary rates with convergent traits. Bioinformatics *35*, 4815–4817. https://doi.org/10.1093/bioinformatics/btz468.
79. Rimal, R., Almøy, T., and Sæbø, S. (2018). A tool for simulating multi-response linear model data. Chemometr. Intell. Lab. Syst. *176*, 1–10. https://doi.org/10.1016/j.chemolab.2018.02.009.
80. Langfelder, P., and Horvath, S. (2008). Wgcna: an r package for weighted correlation network analysis. BMC Bioinf. *9*, 1–13.
81. Pantano, L. (2017). DEGreport: Report of DEG analysis. https://doi.org/10.18129/B9.bioc.DEGreport.
82. Patro, R., Mount, S.M., and Kingsford, C. (2014). Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat. Biotechnol. *32*, 462–464. https://doi.org/10.1038/nbt.2862.
83. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. Genome biology *15*, 1–21.

84. Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R.H.B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., et al. (2009). Package 'lme4'. http://lme4.r-forge.r-project.org.
85. Nowoshilow, S., Schloissnig, S., Fei, J.F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., et al. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature *554*, 50–55.
86. Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). Mummer4: A fast and versatile genome alignment system. PLoS Comput. Biol. *14*, e1005944.
87. Langdon, W.B. (2015). Performance of genetic programming optimised bowtie2 on genome comparison and analytic testing (gcat) benchmarks. BioData Min. *8*, 1–7.
88. Vassetzky, N.S., and Kramerov, D.A. (2013). Sinebase: a database and tool for sine analysis. Nucleic Acids Res. *41*, D83–D89.
89. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.
90. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from rna-seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.
91. Hoff, K., Lomsadze, A., Borodovsky, M., and Stanke, M. (2019). Whole-genome annotation with braker. Methods in Molecular Biology *1962*, 65.
92. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). Augustus: ab initio prediction of alternative transcripts. Nucleic Acids Res. *34*, W435–W439.
93. Lukashin, A.V., and Borodovsky, M. (1998). Genemark. hmm: new solutions for gene finding. Nucleic Acids Res. *26*, 1107–1115.
94. Gotoh, O. (2008). A space-efficient and accurate method for mapping and aligning cdna sequences onto genomic sequence. Nucleic Acids Res. *36*, 2630–2638.
95. Harrison, M.C., Jongepier, E., Robertson, H.M., Arning, N., Bitard-Feildel, T., Chao, H., Childers, C.P., Dinh, H., Doddapaneni, H., Dugan, S., et al. (2018). Hemimetabolous genomes reveal molecular basis of termite eusociality. Nat. Ecol. Evol. *2*, 557–566.
96. Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth, W., Chen, Z., Childers, C.P., Glastad, K.M., Gokhale, K., et al. (2014). Molecular traces of alternative social organization in a termite genome. Nat. Commun. *5*, 3636–3712.
97. Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S., Nygaard, S., Nobre, T., Klaubauf, S., Schindler, P.M., et al. (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. Proc. Natl. Acad. Sci. USA *111*, 14500–14505.
98. Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinf. *6*, 31. https://doi.org/10.1186/1471-2105-6-31.
99. Robertson, H.M., Baits, R.L., Walden, K.K.O., Wada-Katsumata, A., and Schal, C. (2018). Enormous expansion of the chemosensory gene repertoire in the omnivorous German cockroach *Blattella germanica*. J. Exp. Zool. B Mol. Dev. Evol. *330*, 265–278.
100. Emms, D.M., and Kelly, S. (2019). Orthofinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238–314.
101. Letunic, I., and Bork, P. (2021). Interactive tree of life (itol) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296.
102. Löytynoja, A., and Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science *320*, 1632–1635. https://doi.org/10.1126/science.1158395.
103. Penn, O., Privman, E., Landan, G., Graur, D., and Pupko, T. (2010). An alignment confidence score capturing robustness to guide tree uncertainty. Mol. Biol. Evol. *27*, 1759–1767. https://doi.org/10.1093/molbev/msq066.
104. Jordan, G., and Goldman, N. (2012). The effects of alignment error and alignment filtering on the sitewise detection of positive selection. Mol. Biol. Evol. *29*, 1125–1139. https://doi.org/10.1093/molbev/msr272.
105. Gouveia-Oliveira, R., Sackett, P.W., and Pedersen, A.G. (2007). MaxAlign: Maximizing usable data in an alignment. BMC Bioinf. *8*, 312. https://doi.org/10.1186/1471-2105-8-312.
106. Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. *34*, 609–612. https://doi.org/10.1093/nar/gkl315.
107. Markova-Raina, P., and Petrov, D. (2011). High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. Genome Res. *21*, 863–874. https://doi.org/10.1101/gr.115949.110.
108. Hambuch, T.M., and Parsch, J. (2005). Patterns of synonymous codon usage in *Drosophila melanogaster* genes with sex-biased expression. Genetics *170*, 1691–1700. https://doi.org/10.1534/genetics.104.038109.
109. Kosakovsky Pond, S.L., Poon, A.F.Y., Velazquez, R., Weaver, S., Hepler, N.L., Murrell, B., Shank, S.D., Magalis, B.R., Bouvier, D., Nekrutenko, A., et al. (2020). HyPhy 2.5—A customizable platform for evolutionary hypothesis testing using phylogenies. Mol. Biol. Evol. *37*, 295–299. –5. https://doi.org/10.1093/molbev/msz197.
110. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B *57*, 289–300.
111. Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.
112. Chikina, M., Robinson, J.D., and Clark, N.L. (2016). Hundreds of genes experienced convergent shifts in selective pressure in marine mammals. Mol. Biol. Evol. *33*, 2182–2192. https://doi.org/10.1093/molbev/msw112.
113. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: Tool for the identification of biology. Nat. Genet. *25*, 25–29.
114. Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics *22*, 1600–1607. https://doi.org/10.1093/bioinformatics/btl140.
115. Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80. https://doi.org/10.1186/gb-2004-5-10-r80.
116. Fouks, B., Brand, P., Nguyen, H.N., Herman, J., Camara, F., Ence, D., Hagen, D.E., Hoff, K.J., Nachweide, S., Romoth, L., et al. (2021). The genomic basis of evolutionary differentiation among honey bees. Genome Res. *31*, 1203–1215. https://doi.org/10.1101/gr.272310.120.
117. Tintle, N.L., Borchers, B., Brown, M., and Bekmetjev, A. (2009). Comparing gene set analysis methods on single-nucleotide polymorphism data from Genetic Analysis Workshop 16. BMC Proc. *3*, S96. https://doi.org/10.1186/1753-6561-3-s7-s96.
118. Canal, L. (2005). A normal approximation for the chi-square distribution. Comput. Stat. Data Anal. *48*, 803–808. https://doi.org/10.1016/j.csda.2004.04.001.
119. Kelley, M., Uhran, M.R., Herbert, C., Yoshida, G., Watts, E., Limbach, P.A., and Benoit, J.B. (2021). Abundances of transfer RNA modifications and transcriptional levels for tRNA-modifying enzymes are sex-specific in mosquitoes. Preprint at bioRxiv. https://doi.org/10.1101/2021.08.03.454936.
120. Marchal, E., Hult, E.F., Huang, J., Pang, Z., Stay, B., and Tobe, S.S. (2014). Methoprene-tolerant (met) knockdown in the adult female cockroach, *Diploptera punctata* completely inhibits ovarian development. PLoS One *9*, e1067377. https://doi.org/10.1371/journal.pone.0106737.
121. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Software *67*, 1–48. URL. https://doi.org/10.18637/jss.v067.i01. https://www.jstatsoft.org/index.php/jss/article/view/v067i01.
122. Pantano, L., Hutchinson, J., Barrera, V., Piper, M., Khetani, R., Daily, K., Perumal, T., Kirchner, R., and Steinbaugh, M. (2022), Degreport: Report of Deg Analysis. R Package version 1.34. 0.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Pseudomonas aeruginosa	Ward's Science	470179–204
Biological samples		
Diploptera punctata	The Ohio State University Insectary	N/A
Glossina morsitans morsitans	Yale strain	N/A
Critical commercial assays		
MEGAscript RNAi Kit	Ambion	AM1626
Deposited data and code		
Diploptera genome	This Study	PRJNA803029
Illumina RNA-seq data - *Diploptera*	Jennings et al. 2020[6]	PRJNA577484
Illumina RNA-seq data - *Glossina*	Benoit et al. 2014[29]Attardo et al. 2019[15]	PRJNA486170PRJNA205861
Code availability		10.5281/zenodo.8250677
Oligonucleotides		
Customized oligonucleotides for RT-PCR and qRT-PCR	Integrated DNA Technologies	N/A
Customized oligonucleotides for dsRNAi synthesis	Integrated DNA Technologies	N/A
See Table S2 for list of primers	N/A	N/A
Software and algorithms		
Illumina HiSeq	Illumina	https://www.illumina.com/systems/sequencing-platforms/hiseq-x.html
Trimmomatic version 0.38	Bolger et al.[44]	http://www.usadellab.org/cms/index.php?page=trimmomatic
Jellyfish version 2.3.0	Marcais and Kingsford[45]	https://github.com/gmarcais/Jellyfish
GenomeScope version 2.0	Ranallo-Benavidez et al.[46]	http://qb.cshl.edu/genomescope/genomescope2.0/
PacBio Sequel system	PacBio	https://www.pacb.com/technology/hifi-sequencing/sequel-system/
MARVEL	Amgarten et al.[47]	https://github.com/LaboratorioBioinformatica
nucmer, MUMmer suite version 4.0.0beta2	Kurtz et al.[48]	https://github.com/mummer4/mummer/releases/tag/v4.0.0beta2
racon	Vaser et al.[49]	https://github.com/isovic/racon
Bowtie2, version 2.3.4.3	Langmead and Salzberg[50]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Pilon, version 1.23	Walker et al.[51]	https://github.com/broadinstitute/pilon
Pseudohaploid	Schatz Lab	https://github.com/schatzlab/pseudohaploid
BUSCO version 3.0.2 pipeline	Manni et al.[52]	https://busco-archive.ezlab.org/v3/
Repeat modeler	Flynn et al.[53]	https://www.repeatmasker.org/RepeatModeler/
LTRharvest	Ellinghaus et al.[54]	https://www.zbh.uni-hamburg.de/en/forschung/gi/software/ltrharvest.html
TransposonPSI	NBI Sweden	https://github.com/NBISweden/TransposonPSI
cd-hit-est	Huang et al.[55]	https://sites.google.com/view/cd-hit
CLC Genomics Workbench, version 12.0	Qiagen	qiagen.com

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Trinity	Haas et al.[56,57]	https://github.com/trinityrnaseq/trinityrnaseq/wiki
TransDecoder	Haas et al.[56,57]	https://github.com/TransDecoder/TransDecoder
RepeatClassifier	Flynn et al.[53]	https://github.com/Dfam-consortium/RepeatModeler/blob/master/RepeatClassifier
RepeatMasker	Tarailo-Graovac and Chen[58]	https://www.repeatmasker.org/
Braker	Kim et al.[59]	https://github.com/Gaius-Augustus/BRAKER
GeMoMa	Keilwagen et al.[60]	http://www.jstacs.de/index.php/GeMoMa
Spaln, version 2.4.6	Iwata and Gotoh[61]	https://github.com/ogotoh/spaln
Pasa	Haas et al.[56,57]	https://github.com/PASApipeline/PASApipeline/blob/master/docs/index.asciidoc
EVidenceModeler	Haas et al.[56,57]	https://github.com/EVidenceModeler
pfamscan	Mistry et al.[62]	https://www.ebi.ac.uk/Tools/pfa/pfamscan/
bitacora	Vizueta et al.[63]	https://github.com/molevol-ub/bitacora
exonerate	Slate and Birney[64]	https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate
mafft version 7.397	Katoh et al.[65]	https://github.com/GSLBiotech/mafft
PhyML	Guindon et al.[66]	https://github.com/stephaneguindon/phyml
CAFE version 4.2.1	Han et al.[67]	https://github.com/hahnlab/CAFE
PRANK version 0.0.150803	Löytynoja[68]	https://anaconda.org/bioconda/prank/files?version=v.150803
GUIDANCE2	Sela et al.[69]	http://guidance.tau.ac.il/source
HyPhy, version 2.5.41	Pond et al.[70]	https://github.com/veg/hyphy
aBSREL, version 2.5.41	Smith et al.[71]	https://github.com/veg/hyphy
RELAX, version 2.5.41	Wertheim et al.[72]	https://github.com/veg/hyphy
ModelFinder	Kalyaanamoorthy et al.[73]	http://www.iqtree.org/ModelFinder/
Orthofinder	Emms[74]	https://github.com/davidemms/OrthoFinder
baseml	Yang[75]	http://abacus.gene.ucl.ac.uk/software/paml.html
Bioconductor	Genlemen et al.[76]	https://www.bioconductor.org/
topGO, version 2.4	Alexa and Rahnenfuher[77]	https://topgo.bioinf.mpi-inf.mpg.de/download.php
RERconverge	Kowalczyk et al.[78]	https://github.com/nclark-lab/RERconverge
SimRel	Rimal et al.[79]	https://github.com/simulatr/simrel
Weighted correlation network analysis	Langfelder and Horvath[80]	https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/WGCNA_1.70-3.tar.gz
DEGreports, version 1.25.1	Pantano[81]	https://www.bioconductor.org/packages/release/bioc/src/contrib/DEGreport_1.36.0.tar.gz
REVIGO	Supek et al.[19]	http://revigo.irb.hr/
Sailfish	Patro et al.[82]	https://github.com/kingsfordgroup/sailfish/archive/v0.9.2.tar.gz
DeSeq2	Love et al.[83]	https://bioconductor.org/packages/release/bioc/src/contrib/DESeq2_1.40.2.tar.gz
lme4 R package	Bates et al.[84]	http://cran.nexr.com/web/packages/lme4/index.html
Adobe Photoshop and Illustrator	Adobe	https://www.adobe.com/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Joshua B. Benoit joshua.benoit@uc.edu.

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All generated genomic resources have been deposited in the National Center for Biotechnology Information (NCBI) Genome Archive (BioProject accession no. PRJNA803029, BioSample accession no. SAMN25610536). Previously published RNA-seq datasets are available under the the NCBI Bioproject accession no. PRJNA577484 and PRJNA486170. https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_030220185.1/.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs (https://doi.org/10.5281/zenodo.8250677) are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Experimental model: *Diploptera punctata*

Cockroaches were acquired from the Ohio State University Insectary and maintained according to Jennings et al.[6] DNA was extracted from testes of males 20-30 days after adult emergence and sequenced at the Centre d'expertise et de services Génome Québec.

METHOD DETAILS

Genome sequencing and assembly

The genome was sequenced with a combination of long- and short-read technologies. Using Illumina HiSeq, we generated 147Gb of 150bp paired-end reads (486.8M read pairs), with 500bp fragment size. These reads were quality and adapter trimmed with Trimmomatic (v0.38),[44] resulting in 466.4M read pairs and 136.5Gb. We used these trimmed Illumina reads to estimate the genome size by first calculating kmer-frequencies with Jellyfish (v2.3.0),[45] with a kmer size of 21 and a hash size of 10^9. The resulting histogram of kmer distribution was then used to model genome size with GenomeScope 2.0,[46] which was predicted at 3.07Gb, with an estimated heterozygosity level of 0.4% and repetitive content of 64.2%.

With 38 SMRT cells on a PacBio Sequel system, we generated 15.4M reads and a total of 164.5Gb of subread sequence data (mean read length: 10 712bp). The PacBio sequences were assembled with MARVEL.[85] A database was created using blocksize 250. Then to reduce run times, prior to the first alignment step of MARVEL (daligner), raw reads were masked for repeat regions. This was first carried out only on diagonal blocks (e.g. DB.1 vs DB.1, DB.2 vs DB.2 etc.), then subsequently on a broader diagonal of ten blocks, setting the coverage threshold at 10 and 15, respectively. MARVEL was then run with standard settings on these patched reads. The resulting assembly was polished with the patched PacBio reads that were produced within the MARVEL assembly. For this reads were first aligned against the assembly using nucmer from the MUMmer suite (v4.0.0beta2),[86] then a consensus was created with racon.[49] This improved assembly was further polished using the Illumina reads, which were first mapped to the assembly with bowtie2 (v2.3.4.3).[87] The resulting bam file was then used to polish the assembly using Pilon (v1.23).[51] Finally, we removed duplicate contigs with Pseudohaploid (https://github.com/schatzlab/pseudohaploid). After each of these correction steps, completeness of the assembly was assessed by identifying Benchmarking Universal Single-Copy Orthologs (BUSCOs) using the BUSCO (v3.0.2) pipeline in genome mode.[14] We identified single-copy orthologs based on the insecta–db9. Each of the correction steps improved the assembly quality, especially with regard to BUSCO completeness scores (Table S1).

Repeat annotation

Repetitive elements from *D. punctata* genome assembly were categorised with repeat modeler (http://www.repeatmasker.org/), LTRharvest[54] and TransposonPSI (http://transposonpsi.sourceforge.net). The resulting libraries were merged together with the SINEbank repeat data base, specific to Insecta.[88] The merged repeat library was filtered for redundancy using cd-hit-est (parameters: -c 0.8 -n 5)[89] and for true proteins by blasting against a *de novo* assembled *Diploptera punctata* transcriptome. Specifically, we generated the *de novo* transcriptome assembly with 29 previously published RNAseq libraries[6] using Trinity[90] at default settings. Nucleotide coding and protein sequences were generated from the Trinity assembly with TransDecoder (http://transdecoder.github.io/). Sequences were removed from this transcriptome if they received a significant blast hit (e-value < 1e-5) against the RepeatPeps library contained in the RepeatMasker data set. We then performed a BLAST of our merged repeat library against this reduced set of transcripts using blastn. Any hits with an e-value < 1e-10 were removed from the library. The repeat library was classified with RepeatClassifier. The genome assembly was then soft masked with RepeatMasker.

Gene annotation

We used two programs to predict *ab initio* gene models: Braker,[91] which combines Augustus[92] and GeneMark,[93] and GeMoMa.[60] Both were trained with the *Blattella germanica* genome and *D. punctata* RNAseq.[6] We additionally used two methods of evidence-based gene prediction. With Spaln (v2.4.6)[94] we aligned a large database of proteins against our genome assembly. The protein database contained the Uniprot arthropod database (version April 2018) and all available Blattodea proteomes: *B. germanica*,[95] *Periplaneta americana*, *Cryptotermus secundus*,[95] *Zootermopsis nevadensis*[96] and *Macrotermes natalensis*.[97] Finally gene models with predicted by aligning the RNAseq data to the genome assembly with Pasa.[56] EVidenceModeler[57] was then used to combine the different gene sets. The following weights were applied to each gene set; Augustus and GeneMark: 1; GeMoMa: 2; Spaln: 5; Pasa: 10. This produced a GFF containing 61,692 putative protein coding genes, which was further filtered to remove contamination and repetitive elements using blast against the NCBI nr database and our repeat database, respectively. Annotation scores from EVM output were compared to noncoding equivalent. All putative genes with an annotation score = ¸ noncoding equivalent were removed. Furthermore, to detect true positives, PFAM domains were annotated on translated sequences with pfamscan and RNAseq reads were mapped against the putative gene regions. All gene models with at least one significant PFAM domain or to which at least 10 reads mapped in at least one sample were considered true positives and retained. All further genes were only kept if supported by evidence from protein alignments (Spaln), transcript alignments (Pasa), or homology within Metazoa, resulting in a gene set of 27,940 protein coding genes.

Chemoreceptor genes are notoriously difficult to predict with standard tools and were therefore annotated manually with bitacora[63] and exonerate[98] in two rounds. For the first round, the chemoreceptors from *Blattella germanica*,[99] *Drosophila melanogaster*, *Apis melifera* and *Apolygus lucorum* species were taken as a database for bitacora and exonerate. Predicted gene models were filtered for the presence of domains of interest and length (85% of domain length average) and used as a database for the second round. The filtered predictions were merged between bitacora and exonerate and with the previous annotations. Predicted cuticle proteins were identified with BLAST comparison to known cuticle protein in other insect system.[15]

Ortholog detection and phylogeny

The 18 insect species have been carefully chosen to investigate viviparous transition as it encompasses the independent origins of viviparity in different insect orders along with, at least, 1 outgroup species per order. Orthologs among the insect species selected; including mayfly (*Ephemera danica*) used as a general insect outgroup, whitefly (*Bemisia tabaci*) as outgroup for the Hemiptera branch, 2 species of aphids (*Acyrthosiphon pisum*, *Rhopalosiphum maidis*) representing origin of viviparity in Hemiptera, 4 species of Diptera (*Glossina morsitans* as viviparous dipteran, and *Musca domesticus*, *Drosophila melanogaster*, *Stomoxys calcitrans* as dipteran outgroups), 2 species of stick insects (*Medauroidea extradentata*, *Clitarchus hookeri*) and locust (*Locusta migratoria*) as outgroup of Blattodea, 3 species of cockroaches (*Blattella germanica* and *Periplaneta americana* as outgroups, *Diploptera punctata* as viviparous blattodean), and 4 species of termites (*Zootermopsis nevadensis*, *Cryptotermes secundus*, *Coptotermes formosanus*, *Macrotermes natalensis*); were discovered using Orthofinder (v2.3.1).[100] The phylogenetic tree was constructed by aligning all single-copy orthologs detected by OrthoFinder[100] using mafft v7.397[65] with the E-INS-i algorithm. The alignments were subsequently concatenated and the tree was constructed with PhyML[16] at default settings. The output tree was rooted on the branch of *Ephemera danica* in iTOL[101] and subsequently saved as a newick file for further analyses. To optimize the number of single-copy orthologs for the selection analyses, ortholog families were retained if genes were present in single-copy in at least one viviparous species and absent in the others, removing any ortholog family with multiple-copy genes in viviparous species. Within those ortholog families, genes of oviparous species were kept if present in single-copy, otherwise they were removed from the ortholog family. Ortholog families including at least one viviparous species and three oviparous species were retained for further analyses.

Duplication and loss events

The variation of gene family size across the phylogenetic tree was assessed with CAFE (v4.2.1),[18] to unravel the expansion and contraction of gene families in all branches of the tree. Moreover, the variation of domain numbers in across the phylogenetic tree was assessed with CAFE (v4.2.1).[18]

Multiple sequence alignment

For each single-copy ortholog family, the longest protein isoforms for each of the species gene were used in multiple sequence alignment with PRANK (- F) (v.150803)[102] and unreliably aligned residues (- - colCutoff 0.93) and sequences (- - seqCutoff 0.6) were masked with GUIDANCE (- - bootstraps 100) (v2.02).[103] This combination was shown to perform the best on simulated data.[104] To optimize alignment length without gaps, we ran maxalign script with the heuristic algorithm[105] and removed subsequent sequences leading to more than 30% of gapped alignment as long as it did not result in the removal of a viviparous species' sequence, and an alignment of less than 4 sequences. The protein sequences were replaced with coding sequences in the multiple alignments using pal2nal script.[106] Alignments regions, where gapped positions were present, were removed with a custom python script (https://github.com/Dr-ShinyRaven-Mr-Fox/MSA_filtering), as these are the most problematic for positive selection inference.[107] Finally, CDS shorter than 100 nucleotides were eliminated.[108] After filtering, our dataset included 4,671 gene families. The mean length of filtered alignment was 614 nucleotides (median = 471 nucleotides), ranging from a minimum of 102 nucleotides to a maximum of 7836 nucleotides and included on average 10 sequences (median = 11), ranging from 4 to 18.

Identifying selection pressures

Branch-site tests to detect positive selection - aBSREL

Phylogenetic tests of positive selection in protein-coding genes usually contrast substitution rates at non-synonymous sites to substitution rates at synonymous sites taken as a proxy to neutral rates of evolution. The adaptive branch-site random effects model (aBSREL,[71]) from Hyphy software package[109](v2.5.41) was used to detect positive selection experienced by a gene family in a subset of sites in a specific branch of its phylogenetic tree. Test for positive selection was run on the branches leading to the origin of viviparity, namely the *Diploptera punctata* branch, the *Glossina morsitans* branch, and the aphid branch, and as a control on the parent and sister branches of the above-mentioned branches. Test for positive selection accounted for variation in synonymous mutations among branches (- - srv Yes - - syn-rates 3) and the non-independence of mutations in neighbouring nucleotides (- - multiple-hits Double+Triple). Corrected p-values given by aBSREL,[71] using Holm-Bonferroni correction, were used to evaluate genes under positive selection in both viviparous and control branches. Genes were considered under positive selection if corrected p-values were below 0.05. Genes experiencing positive selection in both a viviparous and its corresponding control branches were not considered for further analyses and results' interpretations.

Variation of selection pressure - RELAX

While elevated dN/dS can be caused by increased positive selection, it can also be the result of relaxed purifying selection, or a combination of both. We used RELAX[72] to categorize if shifts in the distribution of dN/dS across individual genes are caused by overall relaxation of selection (i.e. weakening of both purifying selection and positive selection, towards neutrality) versus overall intensification of selection (i.e. strengthening of both purifying selection and positive selection, away from neutrality). Specifically, RELAX models the distribution of three categories of dN/dS (i.e. positive selection, neutral evolution, purifying selection) across a phylogeny and compares the distributions for foreground branches (here, the branches of viviparous origins) to background branches (here, the ancestral and sister branches of viviparous origins) and estimates a parameter K that indicates overall relaxation ($K < 1$) or intensification ($K > 1$). Test detecting variation of selection pressure accounted for variation in synonymous mutations among branches (- - srv Yes - - syn-rates 3) and the non-independence of mutations in neighbouring nucleotides (- - multiple-hits Double+Triple), using three initial random rates (- - starting-points 3). Eight alignments failed to run due to removal of reference species during filtering. As RELAX models do not provide correction for multiple testing, RELAX p-values were corrected as one series using False Discovery Rate (FDR)[110] and the significant cut-off was 0.05.

Convergence of relative evolutionary rates with the emergence of viviparity - RERc

To identify shared genes among insect viviparous origins under similar evolutionary rate changes, we estimated relative evolutionary rates for each of our filtered 4,671 single-copy orthologs. First, we estimated branch lengths using the baseml program[111] under nucleotide substitution model chosen from results of ModelFinder[73] and constraining tree topology to our species tree from Orthofinder.[100] RERconverge[78] was then used to calculate relative evolutionary rates from branch lengths obtained by baseml[111] and correlate them with viviparity origins, as described in.[112] This correlation was done using weighted relative evolutionary rates of gene families with a minimum of 4 species and 2 viviparous origins and unidirectional transition. Correlation slopes (Rho) and corrected p-values given by RERconverge,[78] using Benjamini-Hochberg correction, were used to evaluate genes experiencing higher or lower relative evolutionary rates during insect viviparous transition.

Test for functional category enrichment

Gene Ontology (GO)[113] annotations for our gene families were taken from pfam annotations and from orthologs of *Drosophila melanogaster* and the enrichment of functional categories was evaluated with the package topGO version 2.4[114] of Bioconductor[115] and the RERconverge package for evolutionary rate results.[78] For all topGO analyses, we implemented the "elim" algorithm to decorrelate the graph structure of the Gene Ontology.[114]

To identify functional categories enriched for expanded and contracted gene families, the Fisher exact test with the 'elim' algorithm of topGO was run separately for the significantly expanded and contracted gene families which were given the score of 1 while other gene families were given the score of 0. Enrichment analyses were performed on both viviparous branches and their respective control branches. Gene Ontology categories mapped to less than 10 genes were discarded. Furthermore, Gene Ontology categories found to be significant in both a viviparous and its respective control branches or to be enriched of only one gene family were not considered for further analyses and results' interpretations.

To identify functional categories enriched for genes under positive selection, strengthened, and relaxed selection pressure, the SUMSTAT test was used as described in.[116] The SUMSTAT test is more sensitive than other methods, and minimizes the rate of false positives.[117] To be able to use the distribution of log-likelihood ratios of the aBSREL and RELAX tests as scores in the SUMSTAT test, a fourth root transformation was used.[116] This transformation conserves the ranks of gene families.[118] In addition, we assigned a log-likelihood ratios of zero for genes under relaxed selection ($K < 1$) when testing for enrichment of functional categories with genes under strengthened selection and vice-versa (0 for genes with $K > 1$) when testing for enrichment from genes under relaxed selection. Furthermore, genes found to be under positive selection (aBSREL) in both viviparous and corresponding control branches were removed from the functional annotation data-set for the enrichment analyses. Gene Ontology categories mapped to less than 10 genes were discarded. Furthermore, Gene Ontology categories found to be enriched of only one gene with a non-zero log-likelihood ratio were not considered for further analyses and results' interpretations.

The list of significant gene sets resulting from enrichment tests is usually highly redundant. We therefore clustered the long list of significant functional categories using REVIGO[19] with the SimRel semantic similarity algorithm and medium size (0.7) result list. Small size (0.5) result list was used for functional enrichment in expanded gene families in all branches and contracted gene families only for tsetse.

To identify functional categories enriched for genes experiencing change of evolutionary rates during viviparous transitions, the Wilcoxon Rank-Sum statistics was used with GO terms and genes ranked according to their negative log-transformed correlation slope (Rho) between relative evolutionary rate and viviparity (RERc), as described in.[112]

RNA-seq analyses

To assess if transcriptional changes are similar between viviparous insects, previously available RNA-seq data sets of multiple *Glossina* species and *D. punctata* were analyzed.[6,15] In specific, this allowed comparison between males, early pregnancy (not-pregnant and pre-nutrient provisioning), and late pregnancy (early-nutrient provisioning and early-nutrient provisioning). These different description types allow for assigning *Glossina* species and *D. punctata* into similar groups even though there are developmental and lineage-specific differences. Transcripts per million (TPM) was determined using Sailfish.[119] The expressional changes were compared with the weighted correlation network analysis (WGCNA). In specific, orthologs that were identified through the use of Orthofinder[100] between *Diploptera* and *Glossina* sp. with sequenced genomes.[6,15] The single-copy orthologs obtained from Orthofinder were used for WGCNA to identify groups of genes with similar expression profiles males, during early pregnancy, and late pregnancy. WGCNA was conducted as a signed analyses with a soft power of 12. Modules that were significantly associated with early and late pregnancy were analyzed for enriched GOs following a false detection rate detection. Immune-related genes were analyzed through the use of Deseq2 based on recommended methods of previous data,[6] where the putative immune genes were identified. Clustering was performed on normalised counts using the R package DEGreports v. 1.25.1,[81] with a minimum cluster size of 20.

Functional validation of factors identified in genomics and transcriptomic studies

Together with our results of detection of selection in viviparous species and gene's expression at different life-stages in *D. punctata*, we identified genes of interest, which could be linked with the adaptation to viviparity. RNA interference was conducted according to Jenning et al.[6] and Marchal et al. 2014.[120] Briefly, dsRNA was generated with a MEGAscript RNAi Kit (Ambion). Following preparation of the dsRNA, each pregnancy female was injected with 2–3 μg of dsRNA either at 5-6 (before pregnancy) or 30-40 days (during pregnancy) in the reproductive cycle with a pulled glass capillary needle. Control individuals were injected with a dsRNA targeting green fluorescent protein.[6] Individuals were monitored for abortions and the duration of pregnancy (N = 6). RNA was extracted for a subset of cockroaches (N=10, pooled into five groups) and qPCR was used to confirm the reduction in transcript levels for three groups according to previous methods.[6] The sample was measured 5-6 days after injection and a t-test was used to examine differences based on expression determined with a delta-delta ct method. All primers used in these studies are available in Table S2 qPCR validation results is in Table S3.

Immune functionality was assessed through the use of injection of *Pseudomonas aeruginosa* in pregnant females to confirm potential altered immunity during this state. To do so, bacteria was grown until a log-phase and injected 1.0×10^5 CFU of *P. aeruginosa* in 3 μl PBS in pregnant or virgin females. Survival was monitored for 40 days. One and two days following injection with *P. aeruginosa* samples were collected to obtain RNA and qPCR was used to examine transcript levels for a milk protein and six immune-associated genes. A linear regression model was applied to the data (lm function in the lme4 R package[121]) to assess significance in the RNAi and immune assays. t-tests were used to compare the expression levels of the immune genes in nonpregnant and pregnant cockroaches.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical tests were performed with R (version >3.6).

Functional enrichment of genes with selection signals

Functional enrichment tests were performed with the topgo package[114] and the 'elim' algorithm, using Sumstat and Fisher exact tests, and the RERconverge package,[112] using a Wilcoxon Rank-Sum statistics. All information related to sample size and results of those analyses are present in the Supplementary Information.

RNA-seq analyses

The RNA-seq statistical analyses were performed with the WGCNA,[80] Deseq,[83] and DEGreports[122] packages.

RNAi, immune and survival assays

The linear regression analyses were performed with the lme4 package[84] and t-test with core R programs (version >3.6).