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

Bertrand Fouks, Mark C. Harrison, Alina A. Mikhailova, Elisabeth Marchal, Sinead English, Madeleine Carruthers, Emily C. Jennings, Martin Pippel, Geoffroy M. Attardo, Joshua B. Benoit, Erich Bornberg-Bauer, and Stephen S. Tobe

Westfälische Wilhelms-Universität, Institute for Evolution and Biodiversity, Molecular Evolution and Bioinformatics, Hüberstraße 1, 48149 Münster, Germany; Department of Biology, Molecula...
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 towards 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 immuno-repression roles in the uterus and placenta (13).

Most of our knowledge on the molecular evolution of viviparity stems from studies in vertebrates (15). The expansion of genomic resources for insects represents an ideal opportunity for investigating more general insights into the emergence of viviparity and comparing distantly related taxa for convergence. By sequencing and assembling the genome of the only known truly viviparous cockroach, *D. punctata*, we investigated the genomic signatures of insect viviparity comparing three origins of insect viviparity, the obligate viviparous cockroach, the obligate viviparous tsetse (*Glossina morsitans*), and two cyclically viviparous aphids (*Acyrthosiphon pisum, Rhopalosiphum maidis*). These three systems use strikingly different form of viviparity (5), which are even more divergent than those in other vertebrate systems. Furthermore, we analyzed the transcriptomes of *D. punctata* and multiple *Glossina* species during different pregnancy stages to unravel patterns of specific gene expression before and during pregnancy. We detected genes and pathways under positive selection and also experiencing variation of selection at each of these transitions from oviparity to viviparity. From the results obtained with our comparative genomic and transcriptomic analyses, we selected candidate genes, whose strong effects on pregnancy were validated with RNA interference experiments in *D. punctata*. Our analyses shed light on the biological bases of the emergence of viviparity in insects, which to a large extent mirror convergent viviparous adaptations in vertebrates despite broad physiological and morphological differences.

### Results

**Genome of the viviparous cockroach.** We sequenced and assembled the genome of the viviparous cockroach, *Diploptera punctata* (Blaberidae), with a combination of long- (PacBio, 6x) and short-read (Illumina, 45x) sequencing data. We obtained a highly contiguous (contig N50: 1.4 Mb) and complete (97.6% of insect BUSCOs) genome assembly of length 3.13 Gb (estimated genome size based on k-mer distribution of Illumina data: 3.07 Gb). This genome size is considerably larger than that of the closest related blattodean species with a sequenced genome, *Blattella germanica* (Ectobiidae, 2.0 Gb) (16), and in fact is closer in size to the more distantly related American cockroach, *Periplaneta americana* (Blattidae, 3.4 Gb) (17). The differences in genome size do not seem to have been aided by variation in transposable element content as all three cockroach genome assemblies exhibit similar proportions of repetitive elements: 54.3% in *D. punctata*, 54.7% in *B. germanica* and 57.8% in *P. americana*. Similarly, we find no evidence for genome size being driven by proteome expansion as we identified 27,939 protein-coding genes similar to the number of proteins first reported for *B. germanica* (29,216), both greater than in *P. americana* (21,336, Fig. 1).

**Gene family evolution related to insect viviparity.** To understand mechanisms underlying adaptations to viviparity in insects, we compiled a data set comprising genomes and proteomes of 18 insect species from 3 different insect orders (Blattodea, Diptera, and Hemiptera), in which viviparity has arisen independently (Fig. 1). An analysis of gene family size variation from orthogroups sizes defined by OrthoFinder (18) using CAFE (19) across these species revealed several significant gene family expansions and contractions that were shared between independent origins of viviparity. While none of these were shared among all 3 origins of viviparity, 5 expanding gene families were shared between *D. punctata* and *G. morsitans*, 3
for expanded gene families in both the Pacific beetle-mimic cockroach and tsetse (Table S1). Shared functional categories enriched for expanded gene families among *D. punctata* and aphids were mainly related with immunity, while the tsetse and aphids shared GO terms related with neurogenesis, pole cells, and response to hypoxia. However, these GO terms were no longer significant after False Discovery Rate (FDR) correction at 20%, as the 11 significant functional categories after FDR were only found in aphids, mainly involved in transposition and chromosome organization (Table S1).

Detecting common patterns of gene family evolution among phylogenetically distant taxa may prove difficult due to a reduced ability to detect sequence homology. Hence, we carried out a more robust estimate of duplication and loss events by comparing protein domains rather than genes (21, 22). While no protein domains were found to be significantly expanded or contracted among all origins of viviparity, we found two domains to be significantly expanded during two transitions to viviparity. The transcription initiation factor IID, 31KD subunit domain (PF02291) was expanded in *D. punctata* and aphids, while the proton-conducting membrane transporter domain (PF00361) was expanded in *D. punctata* and *G. morsitans*. Expansion of transcription factors during viviparous transitions might have aided changes of gene expression over the course of pregnancy, similar to the transposon-mediated increase of transcription factor binding sites during mammal evolution (23, 24).

We manually annotated three classes of chemoreceptors in *D. punctata*, the odorant (ORs), gustatory (GRs) and ionotropic receptors (IRs), which are known to be highly abundant in two cockroach species (17, 25) and are reduced in *Glossina* sp. (26) compared to oviparous flies. Compared to numbers in both the German and the American cockroach, we found strongly reduced numbers of each of these chemoreceptor classes. We annotated 434 IRs, 69 ORs and 261 GRs, which were 32-55% and 28-46% lower than in *B. germanica*, and *P. americana*, respectively. Similarly, the predicted number of cuticle proteins in the genome was lower in *D. punctata* when compared to the German (22% increase) and American (27% increase) cockroaches, which also are reduced in viviparous flies (26) relative to oviparous counterparts.

**Table 1.** Gene family experiencing either expansion or contraction shared among insect viviparous origins.

<table>
<thead>
<tr>
<th>Orthogroup</th>
<th>Reference</th>
<th>Putative function</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPANSIONS SHARED BETWEEN <em>D. PUNCTATA</em> &amp; <em>G. MORSITANS</em></td>
<td>OG0000835</td>
<td>polybromo</td>
</tr>
<tr>
<td></td>
<td>OG0001065</td>
<td>NADH dehydrogenase</td>
</tr>
<tr>
<td></td>
<td>OG0001123</td>
<td>NADH dehydrogenase</td>
</tr>
<tr>
<td></td>
<td>OG0001271</td>
<td>NADH dehydrogenase</td>
</tr>
<tr>
<td></td>
<td>OG0001573</td>
<td>ref(2)P</td>
</tr>
<tr>
<td>EXPANSIONS SHARED BETWEEN <em>D. PUNCTATA</em> &amp; APHIDS</td>
<td>OG0000078</td>
<td>His3</td>
</tr>
<tr>
<td></td>
<td>OG0000160</td>
<td>His1</td>
</tr>
<tr>
<td></td>
<td>OG0000345</td>
<td>Poly(ADP-ribose) poly-merase catalytic domain</td>
</tr>
<tr>
<td>EXPANSIONS SHARED BETWEEN <em>G. MORSITANS</em> &amp; APHIDS</td>
<td>OG0000062</td>
<td>DNA polymerase type B</td>
</tr>
<tr>
<td></td>
<td>OG0000395</td>
<td>Ulp1</td>
</tr>
<tr>
<td></td>
<td>OG0000630</td>
<td>Collagen triple helix repeat</td>
</tr>
<tr>
<td>CONTRACTIONS SHARED BETWEEN <em>D. PUNCTATA</em> &amp; APHIDS</td>
<td>OG0000737</td>
<td>Reverse transcriptase</td>
</tr>
<tr>
<td>CONTRACTIONS SHARED BETWEEN <em>G. MORSITANS</em> &amp; APHIDS</td>
<td>OG0000100</td>
<td>Chk2v</td>
</tr>
<tr>
<td></td>
<td>OG0000117</td>
<td>Ser12</td>
</tr>
<tr>
<td></td>
<td>OG0000333</td>
<td>Jon25Bi</td>
</tr>
</tbody>
</table>

Fewer overlaps were found between contracting gene families with no overlap between *D. punctata* and *G. morsitans*, only 1 contracting gene family containing a reverse transcriptase domain was shared between *D. punctata* and aphids, and 3 contracting gene families involved in serine enzymatic pathways between *G. morsitans* and aphids (Table 1). In addition, 2 gene families that were expanded in *D. punctata* were contracted in both *G. morsitans* and aphids. These 2 gene families are part of the serine protease and ecdysteroid kinase-like protein families. Interestingly, serine protease and peptidase inhibitors have been associated with placental growth in lizards (13). The contraction of serine protease gene families in viviparous insects might serve a similar role as inhibitors in reptiles allowing placenta formation.

To link broad functions with gene duplication and loss events during the adaptation to viviparity, we performed enrichment analyses of Gene Ontology (GO) terms among all expanded and contracted gene families, using Fisher’s exact test within topGO package (20) (see Methods). More GO terms were found to be significantly enriched for expanded gene families (*P* < 0.05), with 72, 95, and 142 significant GOs in *D. punctata*, *G. morsitans*, and aphids respectively, compared to GO terms enriched for contracted gene families, with 44, 94, and 4 significant GOs in *D. punctata*, *G. morsitans*, and aphids respectively. The functional categories “negative regulation of chromatin silencing” and “nucleosome assembly/organisation” were found to be significantly enriched for expanded gene families in all origins of viviparity. In addition, functional categories involved in eggshell formation, oxidative metabolism, and vitamin metabolism were enriched between *D. punctata* and aphids and 3 between *G. morsitans* and aphids. Interestingly, most of these gene families experiencing significant expansions across pairs of viviparity origins were either related to protein ubiquitination or chromatin remodelling (Table 1).
Fig. 2. Functional categories involved in the emergence of viviparity in insects. A) Enriched functional categories in genes under positive selection in the viviparous cockroach, *D. punctata*. B) Enriched functional categories in genes under positive selection in the tsetse, *G. morsitans*. C) Enriched functional categories in genes under positive selection in the aphid branch. D) Functional categories enriched for genes under strengthened selection pressure in all branches where viviparity emerged. E) Functional categories enriched for genes under relaxed selection pressure in all branches where viviparity emerged. In each case enriched GO-terms are shown with a p-value < 0.05 and are clustered into broader functions whenever possible, using REVIGO (30) and Cytoscape for visualization.

To identify whether positive selection among viviparous origins quantitatively relates to particular functions, we classified orthogroups based on their GO term annotations from *D. melanogaster* orthologs and protein domains from a pfam library (see Methods). Using SUMSTAT (31) with the topGO R package (20) to test for gene set enrichment, we identified (P < 0.05), 59, 64, and 104 functional categories that were enriched among positively selected genes in the viviparous cockroach, the tsetse, and the aphid branches, respectively (Fig. 2 A-C). While no specific, enriched GO terms were shared among the 3 viviparous origins, all the branches, where viviparity has arisen, had functional categories enriched among genes under positive selection linked with embryogenesis and organ development (Fig. 2 A-C). Furthermore, all branches of viviparous origins shared positively evolving functional categories related to the alleviation of multiple stressors, cell fate and metabolism (Fig. 2 A-C). We also detected signals of positive selection on genes associated with oogenesis and chitin metabolism in both *D. punctata* and *G. morsitans* (Fig. 2A & B). After FDR correction three functional categories remained significant (FDR < 20%) among genes under positive selection on the aphid branch: “mitochondrial electron transport, NADH to ubiquinone”, “vascular endothelial growth factor receptor signaling pathway”, and “imaginal disc-derived wing margin morphogenesis” (Table S3).

**Genes with increased and relaxed selection along transitions to viviparity.** With the transition to viviparity, it is predicted that there will be stronger selection for adaptations that facilitate live-bearing and relaxed selection on traits associated with evolutionary conflict over resource allocation. To test this prediction, we identified genes experiencing variation of selection during the transition to viviparity, using the RELAX method in Hyphy (27, 32). This assigns a value determining relaxation or strengthening of selection pressure as well as estimates its significance using maximum likelihood methods, on 4,671 single-copy orthogroups. We identified 215 and 112 orthogroups to be under relaxed and strengthened selection, respectively, in the viviparous branches at 10% FDR. From those
lists, the highest significant orthogroups (N = 15) experiencing strengthened selection are mostly involved in development, protein metabolism and the excretory system, while those under relaxed selection (N = 11) are mainly involved in cell-cell signalling and cell metabolism (Table S4).

To identify broader functions under differential selection pressure during the transition of viviparity, we tested for GO-term enrichment among genes under relaxed and strengthened selection separately. Only the functional category “heart process” was found to be enriched for genes under stronger selection at 20% FDR. However, using an uncorrected p-value < 0.05 the functional categories enriched among genes under strengthened selection can be grouped into 4 main categories, namely: embryogenesis and organ development; negative regulation of mitotic cell cycle; chromosome organisation; and translation (Fig. 2).

Furthermore, several functional categories under strengthened selection pressure in viviparous species seem to be linked with cell-cell adhesion and signalling (i.e., “establishment of protein localization to membrane”, “endocytosis” and “autophagosome maturation”) and the regulation of gene expression (i.e. “non-coding RNA process”). While functional categories enriched among genes under relaxed selection (P < 0.05) do not cluster as well, they can still be grouped within 4 main functions: DNA replication and repair; hormone and pheromone metabolism; cell replication; and circulatory system. In addition to these broad functions, the immune effector pathway is under relaxed selection pressure in branches of live-bearing origin (Table S5).

Expression patterns related to pregnancy are similar among remotely related viviparous insects. To determine if transcriptional patterns related to pregnancy were shared among independent origins of viviparity, we compared the expression of single copy orthologs between six tsetse species (G. morsitans, G. palpalis, G. fuscipens, G. austeni, G. pallidipes, G. brevipalpis) (26) and D. punctata (6). Specifically, the expression in males, early pregnancy, and late pregnancy were compared using a gene co-expression analysis (Fig. 3). These analyses identified that there were specific groups of genes associated with early and late pregnancy in both insect groups (Fig. 3). Early pregnancy showed an enrichment for cuticle changes and associated processes, suggesting that structural changes are a key factor early in insect viviparity (Fig. 3B). The genes
associated with late pregnancy are linked with energy and protein production, which highlights the necessary increase in factors to provide nourishment for developing progeny (Fig. 3). Lastly, closer examination of expression of immune-related genes in *D. punctata*, both across pregnancy and in comparison to males and non-pregnant females, revealed immune changes associated with pregnancy (Fig. 3D). These transcriptional shifts associated with pregnancy show functional similarities to those in vertebrate systems (8, 9).

![Figure 4](image)

**Fig. 4.** Functional validation of genes and pathways involved in the adaptation to live-bearing in the viviparous cockroach. 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*, confirmed the role of specific genes during pregnancy leading to disruption of pregnancy and/or delays in birth. B) Immune challenge in pregnant females compared to non-pregnant females confirmed the reduced immunity during the course of pregnancy associated with a reduction of survival rate.

**RNA interference and immune studies confirm the critical role of viviparity-associated genes.** To confirm that specific genes identified by our earlier selection and transcriptome analyses are involved in the process of live birth, we performed RNA interference (RNAi) in *D. punctata* to suppress transcript levels based on previously developed methods (6). These results confirmed that reduction in transcript levels of targeted genes potentially associated with viviparity could increase the rate of abortion and extend the duration of the pregnancy cycle when compared to controls (Fig. 4). These specific genes are involved in protein synthesis and structural aspects, and would not directly be identified as critical to viviparity without our selection and transcriptome analyses. Of interest is the suppression of single milk gland protein, a key component of nutrients for the developing embryo, did not impact pregnancy, likely due to the presence of over 20 similar genes that can compensate to feed the embryos (6, 26). Lastly, immune function was assessed by injection of a bacterium, *Pseudomonas aeruginosa*, which revealed that pregnant females died more rapidly following infection compared to non-pregnant ones. These studies confirm that genomic and transcriptomic factors identified by our analyses are directly linked to cockroach viviparity.

**Discussion**

Increasing the number of sequenced insect genomes represents a major step towards improving our understanding of the molecular basis underlying adaptive radiation. Comparative genomics of such a diverse animal class provide insights into the key genomic changes along the evolution of insects and also sheds light on the mechanisms by which certain genes and pathways enable the emergence of specific phenotypes. Our genome-wide analyses reveal that convergent adaptations to viviparity in insects are driven by strong positive selection on specific pathways and functional categories, as well as the regulation of specific gene expression patterns during the different stages of pregnancy. Most intriguingly, our results parallel vertebrate adaptations to viviparity with strengthened selection targeting embryogenesis, reproductive system development, tracheal system, and heart development, as well as gene expression patterns during pregnancy linked with reduced immunity and uterine remodelling (Fig. 5).

**Morphological and physiological adaptation to viviparity.** Our study of gene family evolution revealed a duplication of genes involved in eggshell formation in both the viviparous cockroach and the tsetse, which may have favored viviparous transitions similar to eggshell reduction observed in viviparous reptiles (8, 9). Our genome-wide selection analyses reveal that the adaptation to viviparity in insects is linked with strong positive selection of genes involved in oogenesis in both cockroaches and flies. The transition from oviparity to viviparity is accompanied by a reduction of offspring production per cycle (5). Therefore, the regulation of egg production to reduce offspring numbers might have been attained through amino-acid changes in proteins involved in the development of oocytes. Moreover, in tsetse, alternate oocyte production between left and right ovaries is associated with viviparity, as a consequence of resource constraints (5). In addition to oogenesis and its regulation, our results, for the first time, highlight the rapid evolution of embryogenesis and early development along with the urogenital system development in insects during live-bearing adaptation. These are key adaptations to viviparity, which enable the uterus to host the developing embryo (11, 15). The morphological changes of the uterus associated with live-bearing is reflected in our results with rapid evolution of genes involved in reproductive system development in all viviparous branches and gene expression changes during pregnancy of both cockroach and tsetse associated with chitin metabolism.

In addition, the strengthened selection pressure of genes involved in cell-cell adhesion and membrane trafficking we observed in viviparous branches is likely linked with the emergence of the pseudo-placenta or a placenta-like structure adapted in viviparous insects, which enables enhanced gas exchange and nutrient supply (4, 33, 34). Our results suggest that the rapid evolution of genes linked to the development of the tracheal system and heart underlies the adaptation
Vertebrate viviparity

Mammalian viviparity

Squamate viviparity

Invertebrate viviparity

Cockroach viviparity

Fly viviparity

Aphid viviparity

**Fig. 5.** 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 the specific aspect identified in this study for insect systems. Aspects associated with vertebrate viviparity are based upon previous studies (8, 9, 37).

**Vertebrate viviparity**

- Genomic and transcriptional functional overlap
- Structural and morphological urogenital remodeling
- Cuticle changes to accommodate progeny
- Oocyte and embryo development changes
- Reduction in egg structure as early development occurs in reproductive tract
- Enhanced tracheal and heart development
- Shifts in circulation to accommodate movement of nutrients to developing offspring
- Biochemical and physiological altered immunity
- Shifted expression of genes underlying immunity
- Increased translation and energy production
- Allow increased production of nourishment to intrauterine progeny

**Invertebrate viviparity**

- Cockroach viviparity
- Fly viviparity
- Aphid viviparity

**Immunity and maternal tolerance of the embryo.** Another major change linked with live-bearing adaptation in vertebrates is a reduced immunity necessary for not rejecting the developing embryo. In support, during *D. punctata* pregnancy, we observed differential regulation of immune genes especially during early pregnancy, in line with findings in viviparous vertebrates (12, 13). Furthermore, we find immune effectors to be under relaxed selection in viviparous branches, while in many insects immune pathways are expected to be the target of strong positive selection (35, 36), denoting the importance of reduced immune efficiency during live-bearing transition to minimize embryo rejection. Reduced immunity during pregnancy is not only inferred from our transcriptomic and genomic analyses but also revealed by a reduced survival rate of pregnant females after an immune challenge. This difference in survival could also be due to nutritional trade-offs, with the high energetic demands of provisioning developing progeny resulting in reduced immune function, whereas such trade-offs would be less pronounced in oviparous counterparts. The tolerance of the developing embryo within a female’s body might have led to the strong positive selection on pathways related to stress response, which we inferred for all three viviparous insect lineages. Indeed, in all viviparous insect branches, pathways involved in response to chemicals or toxic substances were enriched among positively selected genes, which might indicate adaptations to sustaining embryo growth and development.

**Molecular basis of evolutionary conflict over resource allocations.** While profound changes are needed to enable the transition to viviparity as seen above, evolutionary conflicts over resource allocation arise among females against males and offspring (3). Such evolutionary conflicts should result in divergent evolution which could produce similar signals to relaxed selection. One of the major pathways enriched for genes evolving under relaxed selection comprises hormone and pheromone metabolism. In the viviparous cockroach, as well as in tsetse, developing embryos are directly fed with specific nutrients (5, 6). The production of these nutrients is governed by the regulation of Juvenile Hormone (38). Considering the evolutionary conflict over resource allocations, the regulation of nutritional secretions should be the main source of conflict as it represents the main source of nutrients for the embryo. The relaxation of selection on these pathways in viviparous branches could highlight such evolutionary conflict, with balancing selection between females minimizing the production of pregnancy-associated secretions while male and offspring maximize its intake.

**Universal pathways to viviparity.** Overall, our study reveals that the viviparity transition in insects is associated with strong positive selection or strengthened selection pressure of genes involved in oogenesis, embryogenesis, tracheal system, and heart development. In addition, our analyses highlighted uterus remodelling associated with viviparity in insects detected in a change of gene expression related to cuticle metabolism, as well as a strong positive selection pressure on the urogenital development. Along with pathways involved in uterus remodelling, we found the development of placental-like structures in viviparous insects to be associated with strengthened selection pressure on genes involved in maternal control over embryo development. The viviparous cockroach displays reduced immunity during pregnancy with a reduction of immune gene expression during early pregnancy, and the evolutionary transitions to viviparity seem to have led to the relaxation of selection on immune effectors in all three studied viviparous insect branches. Hormonal changes were also noted in the genomic analyses of invertebrate viviparity, which could be as critical as the hormone shifts and changes that are necessary for vertebrate viviparity (8–10).

Moreover, we found that in all viviparous insect branches non-coding RNA processes, involved in the regulation of gene expression (39), are under strengthened selection pressure. Likewise, the basal evolution of eutherian mammals is associated with bursts of regulatory miRNAs regulating the expression of genes involved in placentaion (40), non-coding...
RNA pathways in viviparous insects might therefore play a similar role. In addition, the formation of the placental tissue in mammals is associated with chromosomal maintenance pathways (41), which could explain the strengthened selection pressure of genes involved in chromosomal organisation in viviparous insect branches. Despite similar pathways shared among the three origins of live-in bearing in insects, little overlap was found among genes involved in theses adaptations among the different viviparous insect species. Our results highlight that the transition to viviparity involves similar pathways in both insects and vertebrates, but not necessarily common functional genes (9).

Strikingly, despite broad physiological and morphological differences, the adaptation to live-bearing seems to be universal among animals with pathways. In both arthropods and vertebrates, pathways related to eggshell formation, urogenital and heart development corresponding to angiogenesis in vertebrates, reduced immunity have undergone fast evolutionary changes and gene expression changes during the course of pregnancy (Fig. 5). Even more remarkably, not only similar pathways but similar evolutionary mechanisms underlie the transition to viviparity in animals, with fast evolution, co-option, gene duplication, and expression changes during pregnancy of genes involved in corresponding functions across different animal taxa (15). Different forms of viviparity as well as ovoviviparity, which can be considered intermediate along the transition from oviparity to viviparity, are well represented in insects (2, 42). Comparative genomics of insects, therefore, represents a great avenue to study in depth the genomic basis of the gradual emergence of live-bearing reproduction mode.

Materials and Methods

Specimen Collection. Cockroaches were acquired from the Ohio State University Insectary and maintained according to Jennings et al. (6). DNA was extracted from testes and sequenced at the Centre d’expertise et de services Génome Québec.

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) (43), resulting in 466.4M read pairs and 136.5Gb. We used these trimmed Illumina reads to estimate the genome size by first calculating k-mer frequencies with Jellyfish (v2.3.0) (44), 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 (45), 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 (46). A database was created using blocksize 250. Then to reduce run times, prior to the first alignment step of MARVEL (daligre), 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 nucler from the MUMmer suite (v4.0.0beta2) (47), then a consensus was created with raxon (48). This improved assembly was further polished using the Illumina reads, which were first mapped to the assembly with bowtie2 (v2.3.4.3) (49). The resulting bam file was then used to polish the assembly using Pilon (v1.23) (50). 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 (51). We identified single-copy orthologs based on the insecta-d9. Each of the correction steps improved the assembly quality, especially with regard to BUSCO completeness scores (Table 2).

<table>
<thead>
<tr>
<th>Assembly</th>
<th>MARVEL</th>
<th>+Racon</th>
<th>+Pilon</th>
<th>+Pseudohaploid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sequence</td>
<td>3.371 Gb</td>
<td>3.380 Gb</td>
<td>3.376 Gb</td>
<td>3.127 Gb</td>
</tr>
<tr>
<td>Number of contigs</td>
<td>14,530</td>
<td>14,450</td>
<td>14,450</td>
<td>10,704</td>
</tr>
<tr>
<td>Longest contig</td>
<td>8.328 Mb</td>
<td>8.350 Mb</td>
<td>8.341 Mb</td>
<td>8.341 Mb</td>
</tr>
<tr>
<td>Shortest contig</td>
<td>2189</td>
<td>2383</td>
<td>2139</td>
<td>2139</td>
</tr>
<tr>
<td>BUSCO</td>
<td>1.266 Mb</td>
<td>1.268 Mb</td>
<td>1.265 Mb</td>
<td>1.407 Mb</td>
</tr>
<tr>
<td>BUSCO - complete single</td>
<td>85.9%</td>
<td>87.0%</td>
<td>89.4%</td>
<td>92.0%</td>
</tr>
<tr>
<td>BUSCO - complete duplicated</td>
<td>7.2%</td>
<td>7.1%</td>
<td>8.3%</td>
<td>5.6%</td>
</tr>
<tr>
<td>BUSCO - fragmented</td>
<td>3.9%</td>
<td>3.4%</td>
<td>1.4%</td>
<td>1.4%</td>
</tr>
<tr>
<td>BUSCO - missing</td>
<td>3.0%</td>
<td>2.5%</td>
<td>0.9%</td>
<td>1.0%</td>
</tr>
</tbody>
</table>

Table 2. Assembly statistics and BUSCO scores for each assembly stage.

Repeat annotation. Repetitive elements from D. punctata genome assembly were categorised with repeat modeler (http://www.repeatmasker.org/), LTRHarvest (52) and TransposonPSI (http://transposonpsi.sourceforge.net). The resulting libraries were merged together with the SINEbank repeat data base, specific to Insecta (53). The merged repeat library was filtered for redundancy using cd-hit-est (parameters: -<e-value 1e-5) against the RepeatPeps library contained in the RepeatMasker data set. We then blasted 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: Braier (56), which combines Augustus (57) and GeneMark (58), and GeMoMa (59). Both were trained with the Blattella germanica genome and D. punctata RNAseq (60). We additionally used two programs of evidence-based gene prediction. With Spanil (v2.4.6) (61) we aligned a large database of proteins against our genome assembly. The protein database contained the Uniprot arthropod proteomes library (6) using Trinity (55) at default settings. Nucleotide coding transcripts were merged together with the SINEbank repeat data base, specific to Insecta (53). The merged repeat library was filtered for redundancy using cd-hit-est (parameters: -<e-value 1e-5) against the RepeatPeps library contained in the RepeatMasker data set. We then blasted 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.
ing equivalent. All putative genes with an annotation score $\leq$ 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 considered if evidence from protein alignments, transcript alignments (Pasa), or homology within Metazoa, resulting in a gene set of 27,940 protein coding genes.

Chemosensor receptors are notoriously difficult to predict with standard tools and were therefore annotated manually with bitacora (65) and exonerate (66) in two rounds. For the first round, the chemoreceptors from Blattella germanica (35), Drosophila melanogaster, Apis mellifera and Apolypus 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 protein in other insect system (26).

Ortholog detection and phylogeny. The 18 insect species have been carefully chosen to investigate viviparous transition as it encompasses diverse genecomes origins and viviparity in different insect orders, along with, at least, one 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 (Acrystosiphon pisum, Rhopalosiphum maidis) representing origin of viviparity in Hemiptera, 1 species of Diptera (Glossina morsitans as viviparous dipteron, and Musca domestica, Drosophila melanogaster, Stomoxys calcitrans as dipteran outgroups), 2 species of stick insects (Medauviroidea extradentata, Citharchus hookeri) and locust (Locusta migratoria) as outgroup of Blattodea, 3 species of cockroaches (Blattella germanica and Periplaneta americana as outgroups, Diplodopa punctata as viviparous blattodean), and 4 species of termites (Zootermopsis nevadensis, Cryptotermes secundus, Coptotermes formosanus, Macrotermes natalensis); were discovered using Orthofinder (v2.5.2) (18). To optimize the number of single-copy orthologs, We categorized them as such if they were single-copy or absent in viviparous species and oviparous species with multiple copies were dropped. Ortholog families including at least one viviparous species and three oviparous species were retained for further analyses, culminating at 5463 single-copy ortholog families.

Phylogenetic tree reconstruction, including all species described above, was undertaken by OrthoFinder (18).

Duplication and loss events. The variation of gene family size across the phylogenetic tree was assessed with CAFE (v4.2.1) (19) to unravel the expansion and contraction of gene families in all branches of the tree. Moreover, the variation of domain numbers in across the phylogeny tree was assessed with CAFE (v4.2.1) (19).

Multiple 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 (v1.150803) (67) and unreliable aligned residues and sequences were masked with GUIDANCE (v2.02) (68). This combination was shown to perform the best on simulated data (69). To optimize alignment length without gaps, we ran maxalign script (70) 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 (71). Alignments regions, where gapped positions were present, were removed with a custom python script (see Supplementary Code S1), as these are the most problematic for positive selection inference (72). We further filtered by evidences from protein alignment, transcript alignments or homology within Metazoa, resulting in a gene set of 27,940 protein coding genes.

Identifying selection pressures.

Branch-Site tests to detect positive selection. 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 branch-site models assess under relaxed selection (branch) in comparison to a background branch (90). The aBSREL (28) branch-site tests were implemented in Hyphy software package (27) 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 only on the branches leading to the origin of viviparity, namely the Diplodopa punctata branch, the Glossina morsitans branch, and the aphid branch. Results from the adaptive branch-site random effects model were considered as those resulting from False Discovery Rate (FDR) (74) and set up our significant threshold at 10% (31).

Variation of Selection pressure. 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 (32) 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, and towards neutrality) versus overall intensification of selection (i.e. strengthening of both purifying 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) using the K statistics ($K < 1$) or intensification ($K > 1$). Eight alignments failed to run due to removal of reference species during filtering. Results from RELAX models were corrected for multiple testing as one series using (FDR) (74) and set up our significant threshold at 10% (31).

Test for functional category enrichment. Gene Ontology (GO) (75) 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 (20) of Bioconductor (76). To identify functional categories enriched for expanded and contracted gene families, the Fisher exact test with the ‘elmir’ 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. The results were then corrected with the FDR to account for multiple testing (74) and set up our significant threshold at 20% (31). Gene Ontology categories mapped to less than 10 genes were discarded.

To identify functional categories enriched for genes under positive selection, strengthened selection, relaxed selection pressure, the SUMSTAT test was used as described in (31). The SUMSTAT test is more sensitive than other methods, and minimizes the rate of false positives (77). 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 (31). This transformation conserves the ranks of gene families (78). 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. Gene Ontology categories mapped to less than 10 genes were discarded.

The list of significant gene sets resulting from enrichment tests is usually highly redundant. We therefore implemented the ‘elim’ algorithm from the Bioconductor package topGO, to decorate the graph structure of the Gene Ontology (20). To account for multiple testing, the final list of p-values resulting from this test was corrected with the FDR (74) and set up our significant threshold at 20% (31). To cluster the large list of significant functional categories before FDR correction, we used REVIGO (30) with the SimRel semantic similarity algorithm and medium size (0.7) result list.

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, 26).
In specific, this allowed comparison between males, non-pregnant females, early pregnancy, and late pregnancy. Transcripts per million (TPM) was determined using SAILfish (79). The expression changes were compared with the weighted correlation network analysis (WGCNA). In specific, orthologs that were identified through the use of Orthofinder (18) between Dipteroptera and Glossina sp. with less than 100% identity (6, 26). 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. The WGCNA was conducted as a signed analyses with a soft power of 12. Modules that were significantly associated with the adaptation to viviparity. RNA interference was conducted with a dsRNA target-

Functional validation of factors identified in genomics and转录组学研究。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 (6) and (81). Briefly, dsRNA was generated with a MEGAscript RNAi Kit (Ambion). Following preparation of the dsRNA, each pregnancy female was injected with 2.5 µg of dsRNA at 30–40 days into the pregnancy 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.

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 × 10^6 CFU of P. aeruginosa in 3 PBS in pregnant or virgin females. Survival was monitored for 40 days.

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 past, present and future role of molecular biology and genomics. J. Comp. Physiol. B: Biochem. Syst. Environ. Physiol. 181, 575–594 (2011).