

Conserved roles of Osiris genes in insect development, polymorphism and protection

C. R. SMITH* , C. MORANDIN†, M. NOUREDDINE* & S. PANT*

*Department of Biology, Earlham College, Richmond, IN, USA

†Centre of Excellence in Biological Interactions, Department of Biosciences, University of Helsinki, Helsinki, Finland

Keywords:

chitin;
cuticle;
evolution of gene expression;
gene family;
phenotypic plasticity;
social insect.

Abstract

Much of the variation among insects is derived from the different ways that chitin has been moulded to form rigid structures, both internal and external. In this study, we identify a highly conserved expression pattern in an insect-only gene family, the Osiris genes, that is essential for development, but also plays a significant role in phenotypic plasticity and in immunity/toxicity responses. The majority of Osiris genes exist in a highly syntenic cluster, and the cluster itself appears to have arisen very early in the evolution of insects. We used developmental gene expression in the fruit fly, *Drosophila melanogaster*, the bumble bee, *Bombus terrestris*, the harvester ant, *Pogonomyrmex barbatus*, and the wood ant, *Formica exsecta*, to compare patterns of Osiris gene expression both during development and between alternate caste phenotypes in the polymorphic social insects. Developmental gene expression of Osiris genes is highly conserved across species and correlated with gene location and evolutionary history. The social insect castes are highly divergent in pupal Osiris gene expression. Sets of co-expressed genes that include Osiris genes are enriched in gene ontology terms related to chitin/cuticle and peptidase activity. Osiris genes are essential for cuticle formation in both embryos and pupae, and genes co-expressed with Osiris genes affect wing development. Additionally, Osiris genes and those co-expressed seem to play a conserved role in insect toxicology defences and digestion. Given their role in development, plasticity, and protection, we propose that the Osiris genes play a central role in insect adaptive evolution.

Introduction

The chitin-based insect cuticle is a key ingredient in the spectacular ecological and evolutionary success of the insects. Chitin forms parts of both the external and internal anatomy of an insect, forming a hard skeleton that has been modified in nearly endless ways and facilitated anatomical, physiological and functional divergence in insect form (Emlen & Nijhout, 2000). On the outside, chitin is a barrier to the external environment, but also light and flexible enough to form the skeletal basis of locomotory appendages (legs and wings). On the inside, rigid invaginations of the

exoskeleton (apodemes) are muscle attachment points that have allowed insects to evolve amazing strength and speed (Larabee & Suarez, 2014), but also form the respiratory tracheal system and the fore and hindgut (Terra, 1990). Chitin is also secreted in the midgut of insects as part of digestion and immunity – the peritrophic matrix/membrane is a chitinous structure that surrounds ingested material and is a barrier between the digestive tract and surrounding tissues (Lehane, 1997). The peritrophic matrix thus enables the encapsulation and isolation of toxins and infectious agents.

Where as a great deal is known about the structure and development of insect cuticle (Wigglesworth, 1957; Andersen, 1979; Hopkins & Kramer, 1992; Merzendorfer, 2011; Tomoyasu & Fujiwara, 2017), some studies using next-generation transcriptomic techniques, such as RNA sequencing (RNAseq), find genes of unknown

Correspondence: Chris R. Smith, Department of Biology, Earlham College, 801 National Road West, Richmond, IN 47374, USA.
Tel.: +1 765 983 1377; fax: +1 765 983 1497; e-mail: crsmith.ant@gmail.com

function coincident with the timing of cuticle deposition (Ren *et al.*, 2005; Sobala & Adler, 2016). Some of these genes are the most highly expressed genes during particular developmental stages, such as several Osiris family genes during stages of pupal development (Graveley *et al.*, 2011; Gelbart & Emmert, 2013). Intriguingly, very little is known about many of these highly expressed genes, including those in the Osiris family.

The Osiris genes are a family of genes that have a high degree of synteny across all studied insects and appear to have radiated via gene duplications near the origin of the Insecta, approximately 400MYA (Shah *et al.*, 2012; Misof *et al.*, 2014). The majority of Osiris genes, ~20, lie in an 160-kbp region (on chromosome 3R in *D. melanogaster*). The locus containing these genes has been known for several decades in *D. melanogaster* due to its embryonic lethality when present in haploid or triploid copy (the triplo-lethal locus, *Tpl*; Lindsley *et al.*, 1972; Dorer *et al.*, 1995, 2003). Although this result clearly suggests dosage-dependent lethality, it is unclear how or why the lethality occurs. Previous work established that embryos with an abnormal copy number of *tpl* fail to form midgut and die as late embryos or upon hatching from the egg (Smoyer *et al.*, 2003). The function of only one Osiris gene is known, *Osi21*, and this particular gene is not in the main cluster. *Osi21* in *D. melanogaster*, also known as *diehard4*, inhibits the recycling of rhodopsin in the retina as part of the endocytic pathway (Lee *et al.*, 2013).

Despite a longstanding knowledge of their existence, their precise role remains unclear. These genes are primarily characterized by a protein domain of unknown function (DUF1676) and have no gene ontology (GO) terms associated with their molecular function. In addition to DUF1676, Osiris genes contain a putative signal peptide, a hydrophobic region that is likely a transmembrane region, and an AQXLAY motif at their 3' end (Shah *et al.*, 2012). This structure is compatible with the known role of *Osi21* in endosomal trafficking (Lee *et al.*, 2013).

We used comparative transcriptomics to shed light on this enigmatic family of genes. We predicted that the conservation of synteny in the Osiris cluster has a functional basis through gene regulation and that genes with physical proximity would share similar expression patterns across species. Due to their high degree of sequence conservation, we also hypothesized that the timing of expression of these genes would be consistent even across clades that diverged hundreds of millions of years ago and that their expression would coincide with suites of other genes, forming coregulated expression modules. By leveraging species producing multiple divergent phenotypes, the social insect castes (queen, worker, and male), we were also able to test for differences in Osiris gene expression responsiveness to the developmental environment (Anderson *et al.*, 2008). In the context of studies on insect chemical protection,

our data suggest that Osiris genes may be involved in the regulation of insect cuticle, both in the generation of external morphology as well as in forming a defensive barrier.

Materials and methods

Ortholog detection

The *D. melanogaster* set of Osiris genes is the most curated set and was used as a basis for detection of orthologs and paralogs. The methods used in this research was similar to that of Shah *et al.* (2012), but for consistency was repeated for all genomes included in this paper. A combination of approaches was used to detect Osiris genes. First, reciprocal best hit BLAST (rbhBLAST), with an e-value threshold of $e < 0.0001$, was performed between each genome and that of *D. melanogaster* (using the 24 *D. melanogaster* protein models as queries). To widen this search to paralogs, all identified Osiris genes from rbhBLAST were used as BLAST queries into their own genome, and hits with $e < 1E-5$ were then used as reciprocal queries into *D. melanogaster* (kept if their best hit was an Osiris gene). As an independent, and complementary, approach, hidden Markov models were used to search proteins. A model was built, function hmmbuild, using all 24 *D. melanogaster* Osiris proteins and used as a template to search all other genomes using the function hmmsearch in HMMER3.1b2 (Wheeler & Eddy, 2013). Evidence from all of these approaches was combined to determine likely Osiris genes.

Phylogeny

Using orthologs as determined from the above methods, evolutionary relationships were determined using BEAST2 (Drummond *et al.*, 2012; Bouckaert *et al.*, 2014) on the CIPRES Science Gateway server (Miller *et al.*, 2015). Protein evolution models were tested using ProtTest (Abascal *et al.*, 2005; Darriba *et al.*, 2011).

First, a phylogeny including all orthologs from the following taxa was made: centipede (*Strigamia maritima*; Chipman *et al.*, 2014), dipluran (*Catajapyx aquilonaris*; Baylor College of Medicine Human Genome Sequencing Center, BCM-HGSC), mayfly (*Ephemera danica*; BCM-HGSC), dragonfly (*Ladona fulva*; BCM-HGSC), German cockroach (*Blattella germanica*; BCM-HGSC), termite (*Zootermopsis nevadensis*; Terrapon *et al.*, 2014), pea aphid (*Acyrtosiphon pisum*; The International Aphid Genomics Consortium 2010), jewel wasp (*Nasonia vitripennis*; Werren *et al.*, 2010), bumble bee (*Bombus terrestris*; Sadd *et al.*, 2015), honeybee (*Apis mellifera*; Weinstock *et al.*, 2006), jumping ant (*Harpegnathos saltator*; Bonasio *et al.*, 2010), carpenter ant (*Camponotus floridanus*; Bonasio *et al.*, 2010), harvester ant (*Pogonomyrmex*

barbatus; Smith *et al.*, 2011), fruit fly (*Drosophila melanogaster*; Adams *et al.*, 2000), flour beetle (*Tribolium castaneum*; Richards *et al.*, 2008), tobacco hornworm moth (*Manduca sexta*; Kanost *et al.*, 2016) and silkworm moth (*Bombyx mori*; Xia *et al.*, 2004). The parameters for the best model for ProtTest were as follows: a gamma site model with a category count of 4 and a shape parameter of 2.677, using the WAG substitution matrix. Additionally, we used a relaxed log-normal clock model, a coalescent constant population tree prior and an MCMC chain length of 10 million; a burn-in of 10% was chosen to maximize ESS values and the consistency of likelihood values. The outgroup was chosen as the top BLAST hit, other than an Osiris gene, in a search of the *D. melanogaster* genome using *D. melanogaster* annotated Osiris genes – the 1:1 orthologs of this gene was located in several other insect genomes. The resulting tree is Fig. S1.

For a simplified tree, we used a subset of the above taxa, but still representative of major insect clades, only 1:1 orthologs, and the same parameters as above. To help resolve internal nodes, we forced monophyly of Osiris ortholog groups based on the results of the larger tree. The resulting tree is Fig. 1.

DNA and RNA source and processing

RNA-sequencing data from *D. melanogaster* were from Graveley *et al.* (2011, Table S9) for 12 embryonic stages, six larval stages, six pupal stages and six adult stages. Data from *P. barbatus* were from Smith *et al.* (2015) (NCBI project PRJDB3493), and consisted of 16 libraries spread across four developmental stages. These were as follows: early last (4th-instar) instar larvae (prior to differences in mass), late last instar larvae (queens and workers are distinguishable by mass), pupae and adults. In each developmental stage, there were two worker and two queen libraries. The raw RNA-sequencing reads were mapped to the newer NCBI annotation (Thibaud-Nissen *et al.*, 2013) of the *P. barbatus* genome, Pbar_UMD_V03, at NCBI (annotation release 100), using tophat2 (Kim *et al.*, 2013), and bowtie2 for genome indexing (Langmead & Salzberg, 2012). Feature counting was performed using HTseq (Anders *et al.*, 2015). *B. terrestris* RNA-sequencing data were from Harrison *et al.* (2015; NCBI project PRJEB9366), and remapped as above to the assembly Bter_1.0 and annotation release 102 performed at NCBI (Thibaud-Nissen *et al.*, 2013). The *B. terrestris* libraries consisted of larval, pupal and adult stages spread over queen, male and worker castes. We only used nonreproductive adult workers when comparing adults. Each stage by caste library set had three biological replicates, except for queens, where there was only a single library per stage. See the supplement for sample code used in RNA mapping.

RNA-sequencing data from *Formica exsecta* were from Morandin *et al.* (2015) (GenBank Biosample

SAMN02046301–SAMN02046306) and consisted of 12 libraries spread across three developmental stages (pupae, emerging adult and old adult). In each developmental stage, there were two biological replicates of queen and worker. The raw RNA-sequencing reads were mapped to the *Formica exsecta* genome (NCBI project) using tophat2 (Kim *et al.*, 2013) and Cufflinks (Trapnell *et al.*, 2010); RSEM was used for transcript quantification (Li & Dewey, 2011). Gene expression data, as RPKM, are included in Data S1.

Gene expression correlation over space and time

Osiris gene expression data in each species, including all stages and castes, was used to calculate expression similarity. A correlation matrix of Osiris gene expression was performed in R using the cor function in the base package (R Core Team 2017). Visualization of correlation was performed with the corrplot package (Wei *et al.*, 2016), using AOE clustering. Gene expression over space and evolutionary history was examined only in *D. melanogaster*. To examine correlation over space, the distance between pairwise gene starting positions was calculated and was correlated with correlation coefficients of Osiris developmental gene expression. Similarly, pairwise branch length distances between all genes in *D. melanogaster* were calculated using the cophenetic.phylo function from the ape package (Paradis *et al.*, 2004) using a tree with all other species omitted. These pairwise distances were then correlated with Osiris developmental gene expression correlation coefficients.

Similarity in Osiris gene expression among species was calculated for each species using all libraries – as above, (dis)similarity distance matrices were calculated using Bray–Curtis distance. Pairwise Mantel tests compared the similarity of Osiris gene expression between species pairs.

Analysis of expression and co-expression

Only features with at least one count per million (CPM) in at least four or three libraries (i.e. the number of developmental stages in each data set), for *P. barbatus* and *B. terrestris*, respectively, were kept for downstream analysis. Downstream analysis and graphing were performed using the R statistical environment (R Core Team 2017), with graphics produced in ggplot2 (Wickham, 2016). Differential expression analysis between castes, within developmental stages, was performed using edgeR (Robinson *et al.*, 2010), using SVA (Leek & Storey, 2007; Leek *et al.*, 2017) first to remove possible unwanted variation; caste was used as a factor in the model using the function svseq. Differences among libraries were adjusted using TMM normalization (Robinson & Oshlack, 2010), and comparisons between groups were performed using a linear model

using the sveq result as a covariate and caste × developmental stage as a factor.

Co-expression modules were calculated using WGCNA on CPM expression count (Langfelder & Horvath, 2008) with one-step network construction and module detection. All modules were correlated against developmental stages (in the case of *D. melanogaster*) or developmental stage × caste (for the social insects). The input data set was filtered using WGCNA cleaning step that removes genes with too many missing values

or zero variance. Due to the interest of this study, we mapped Osiris genes to modules and looked most closely at modules loaded with Osiris genes, which were also positively correlated with pupae of the worker caste. Gene overlap lists and Venn diagrams were calculated with gplots (Warnes *et al.*, 2009) and VennDiagram (Chen & Boutros, 2011).

WGCNA recommend the use of filtered normalized expression count or FPKM, RPKM, without it altering the final results. To verify this assumption, the above

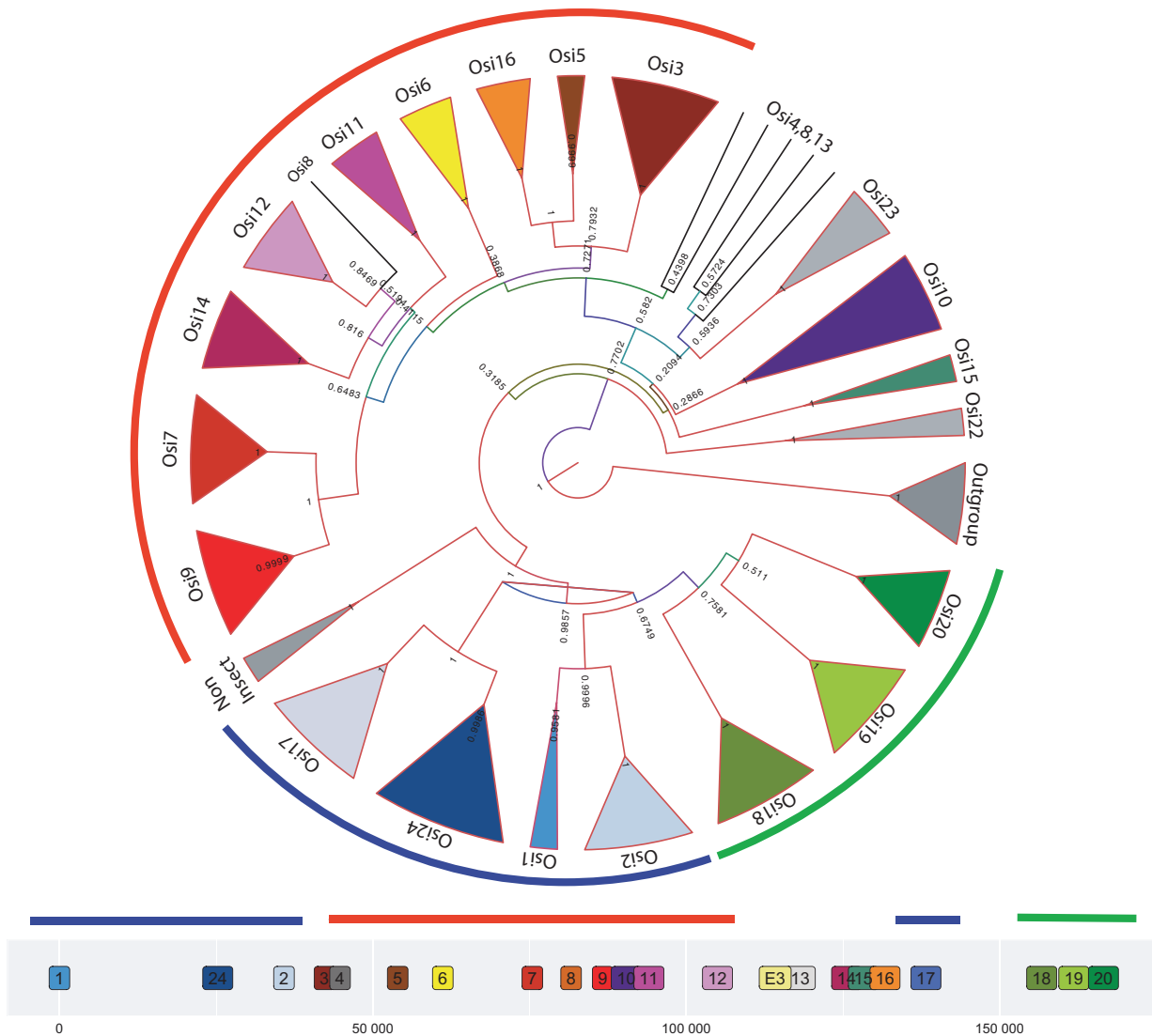


Fig. 1 Bayesian phylogeny of the Osiris proteins across insects. Collapsed nodes all have a posterior probability > 0.95 and are colour coded by clade. Branches are colour coded by their posterior probability support. Semi-circular lines represent larger clades and are mapped to the synteny of *D. melanogaster* below. Sequences in this phylogeny are 1:1 orthologs from insects representative of most major lineages: mayfly (*Ephemera danica*), termite (*Zootermopsis nevadensis*), pea aphid (*Acyrtosiphon pisum*), honeybee (*Apis mellifera*), flour beetle (*Tribolium castaneum*) and fruit fly (*Drosophila melanogaster*). Singleton noninsect Osiris genes are from a centipede (*Strigamia maritima*) and dipluran (*Catajapyx aquilonaris*). Outgroup sequences are orthologs of *D. melanogaster* CG6236.

WGCNA analysis was repeated with RPKM values as input data set, and similar results were obtained. Gene-module associations are available in Data S2, and for Osiris genes (with their Osiris ortholog names) in Data S3.

Gene ontology (GO) analysis

InterProScan5 (Jones *et al.*, 2014) was used to search all proteins in each annotation for matching protein domains Pfam database, (Finn *et al.*, 2016) and GO terms. The package topGO (Alexa *et al.*, 2006; Alexa & Rahnenfuhrer, 2016) was used to detect enrichment of terms in co-expression modules using the weight01 algorithm to take into account GO term relatedness.

Quantitative real-time PCR

Colonies of German cockroaches, *Blattella germanica*, were housed in plastic containers and fed *ad-lib* commercial cat food and water. Individuals were anaesthetized with wet ice, and heads were measured using a calibrated stereo-microscope camera to assign instars (Tanaka & Hasegawa, 1979). After head measurements, the samples were immediately deep-frozen. RNA was extracted by homogenizing whole bodies in TRIzol (Life Technologies) using a pestel, and then, we followed the manufacturer's protocol for the Direct-zol™ RNA Mini-prep kit (Zymo Research). Purified RNA (0.25–1 µg) was reverse-transcribed using the QuantiTect RT kit (Qiagen). Quantitative real-time PCR (qRT-PCR) was performed using the QuantiTect SYBR Green PCR Kit (Qiagen) with a total volume of 10 µL with 10 µM each forward and reverse primer and 1 ng cDNA. Primers were designed using NCBI primer designing tool (Ye *et al.*, 2012) for four Osiris genes (Osi7, two paralogs of Osi9 – a and b, and Osi20) as well as two control genes, GAPDH and Actin. Protein sequences were obtained from Baylor College of Medicine Human Genome Sequencing Center (NCBI project PRJNA203136). The primers used were (5'–3', forward then reverse): GAPDH (BGER005687-RA, F: TGGCCGTATTGGTCGCTTTCG, R: TGCCTTACTTCGCCCTTGA), Actin (BGER004121-RA, F: GATGATGCACCCAGAGCAGT, R: AGTTTGT CACAATGCCGTGC), Osi 7 (BGER018074-RA and BGER018075-RA, these two gene models were merged, F: GACTTGGTGGCGTGTATTGC, R: ATGAGGTCGTCGCTGTTCTG), Osi9a (BGER014472-RA, F: GCTCTTCTGTTG GCTGGTCT, R: GTTCTCGGAGTCAGGAAGGG), Osi9b (BGER018072-RA, F: AGCTGTGGCGAGAAAGACAT, R: CTCTGCCGTTGAGTTCCTG), Osi20 (BGER026568-RA, F: CGCTATCTCAAGACGCACCA, R: CATCCCACGGG CTTTCAGAAA). All PCRs were carried out in a Mastercycler® ep realplex (Eppendorf) with the following cycle parameters: 95 °C for 15 min, then 40 cycles of 95 °C for 15 s, 59 °C for 30 s and 72 °C for 30 s followed by a measurement each cycle; after cycling, melt

curves were run to confirm the appropriate product. All primer pairs had efficiency ~0.95, and it was verified that GAPDH and Actin transcript abundance did not vary predictably with development. All reactions were performed in triplicate. The difference in cycle threshold (ΔC_T) between each gene and each control gene was calculated and averaged. Results were graphed in ggplot2 (Wickham, 2016).

Results and discussion

Ortholog identification and phylogeny

We used methods similar to Shah *et al.* (2012) to find putative orthologs and establish orthology among Osiris genes across divergent insect species. We used a combination of protein BLAST (Altschul *et al.*, 1990) and hidden Markov model searches with HMMER 3.1.2b (Eddy & Wheeler, 2015) using all 24 annotated *D. melanogaster* Osiris genes. Once putative Osiris genes were found, we performed BLAST searches for additional paralogs within each genome. To establish orthology relationships, we used reciprocal best hit BLAST (rbhBLAST) between putative Osiris genes and *D. melanogaster*. Finally, once lists were compiled for each species, we constructed phylogenies of both all potential orthologs and 1:1:1 orthologs across insect taxa using BEAST2 (Drummond *et al.*, 2012; Bouckaert *et al.*, 2014); to help resolve internal nodes, we forced monophyly of groups with posterior probability support > 0.9.

The phylogeny of representative orthologs across insects (Fig. S1) supports monophyly of nearly all Osiris ortholog groups found in *D. melanogaster*, in agreement with the results of Shah *et al.* (2012). Figure 1 shows the relationships of a subset of 1:1:1 orthologs across the major insect clades. The major radiation of Osiris genes dates at least to the origin of insects, ~400 MYA (Misof *et al.*, 2014). Many internal nodes in the phylogeny lack robust support. There appear to be two major ancient lineages of Osiris genes, the red cluster and the blue-green cluster of Fig. 1. As the noninsect Osiris orthologs group with Osi 17/24, this group is the putative ancestral Osiris gene. The green cluster (Osi 18–20) resulted from an internal duplication within the blue cluster. The duplications in the blue-green group likely gave rise to the genes at the beginning and end of the current cluster (Figs 1 and S2). The red group is duplications that filled in the middle of the cluster (Fig. 1).

Spatial/phylogenetic regulation of osiris genes

To explore the dynamics of Osiris gene expression, we used three primary gene expression data sets, the *D. melanogaster* modENCODE developmental transcriptome that spans 12 embryonic, six larval, six pupal and

six adult developmental stages (Graveley *et al.*, 2011), a data set from the bumble bee, *Bombus terrestris*, with larvae, pupae and adults of each males, workers and queens (Harrison *et al.*, 2015) and a data set from the red harvester ant, *Pogonomyrmex barbatus*, with two larval, pupal and adult stages of both workers and queens (Smith *et al.*, 2015). A fourth data set on the wood ant, *Formica exsecta* (Morandin *et al.*, 2015), was included for comparative purposes, but only includes pupae and adult stages of queens and workers. Downstream analysis was conducted on normalized gene expression after filtering out genes with low transcript abundance; all analyses were performed using the R statistical framework version 3.3.3 (R Core Team 2017).

Transcript abundances of Osiris genes tend to be positively correlated within each species, though there are clear ‘blocks’ (Figs 2–3, Figs S3–S6, Data S1–S3) where expression among genes of a block is very highly correlated ($r = 0.8–1$). Among pairwise correlations of Osiris genes in *D. melanogaster*, 82% were positive and 51% were statistically significant ($P < 0.05$, Fig. S6). To assess similarity of expression among species, a similarity matrix of expression was made for each species using the vegan package in R (Oksanen *et al.*, 2013). Genes had similar patterns of developmental expression across species. The gene expression matrices were all highly correlated (Mantel test, $P < 0.05$ in all comparisons, Fig. S7). Despite these species diverging > 300 MYA (Misof *et al.*, 2014), the Osiris genes have

maintained similar within-cluster patterns of gene expression, suggesting that they are coregulated. Precise temporal regulation of these genes is likely necessary given their putative function in the regulation of chitin because chitin deposition is periodic during the development of insects.

We calculated both pairwise linear chromosomal distances and pairwise phylogenetic distances for *D. melanogaster* to examine potential mechanisms of regulation. Osiris gene expression similarity (pairwise gene correlations) was correlated with physical gene distance ($P < 0.0001$, Fig. 3), and phylogenetic distance ($P = 0.004$, Figs S8–S11), suggesting either co-expression due to locality or related ancestral regulatory mechanisms. Co-regulation of neighbouring genes is a pervasive pattern in genomes (Boutanaev *et al.*, 2002; Spellman & Rubin, 2002; Michalak, 2008). Intriguingly, there are Hymenoptera-specific ultraconserved elements (UCEs) in the 5′ UTR of six Osiris genes (Osi7,8,9,16,20; Davies *et al.*, 2015). These elements do not appear to have a conserved RNA secondary structure, but may play a role in translational regulation. Because we detect both a pattern of phylogenetic conservation of expression, as well as spatial co-expression, it is unclear whether the observed co-regulation of genes in this cluster is due to neighbourhood effects, inherited sequence patterns or a combination of the two. A previous analysis (Morandin *et al.*, 2016) examining signatures of selection in ants found that all of

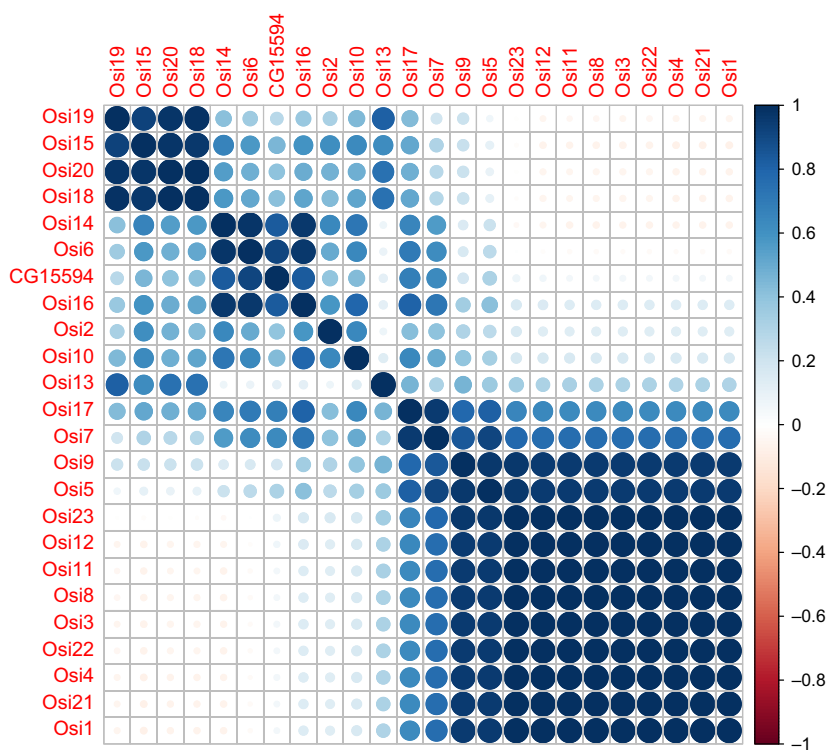


Fig. 2 Correlation matrix of developmental gene expression for *D. melanogaster* Osiris genes. Most Osiris genes are positively correlated with each other (Fig. S6). Colour coding is by correlation r -values (heat-scale on right). All 30 developmental stages from Graveley *et al.* (2011) were included in this analysis. The large bottom block largely corresponds to the gene co-expression module used in other analyses.

the ten Osiris genes examined have a ratio of nonsynonymous to synonymous nucleotide changes less than one, suggesting that they have likely evolved (at least in the ants) under purifying selection. In summary, Osiris genes are highly conserved at the coding and noncoding sequence level, and conservation extends to both synteny and developmental gene expression – over hundreds of millions of years.

Temporal expression patterns

There are three clearly distinct times of Osiris upregulation in *D. melanogaster* development, 14- to 18-h embryos, second-instar larvae and 48-h-old pupae (Fig. 4). During these times, some Osiris genes are among the most expressed genes in the transcriptome. Osi6 is the sixth highest in 14- to 16-h embryos and Osi3, 7, 9 and 12 are first, third, ninth and second during 48-h-old pupae, at well over 100× mean gene expression in those developmental stages. Similarly, Osiris gene expression peaks in pupal development in both *B. terrestris* and *P. barbatus*. There is a global pattern across data sets, ubiquitously higher expression in workers (and males) compared to queens – Osiris gene expression in worker pupae was > 1000-fold higher than in queens for some genes (e.g. Osi3, 5, 6). Due to low power to detect significant differences in expression between castes, only a few of the Osiris genes (Osi 3, 6, and 9) were statistically different after correction for false discovery rate (FDR < 0.05, Table S1). The pattern of greater expression in workers relative to queens suggests a possible role of these genes in generating, or regulating, caste differences. Differences in dispersal and reproduction are among the most defining differences among social insect castes. In *B. terrestris* and both ant species, there are differences in the reproductive capacities of queens and workers, as well as body size differences. There are also wing polymorphisms in all three species. In *B. terrestris*, in which all castes have functional wings, worker and male wings are more similar to each other than either are to queen wings in both shape and size (Medler, 1962; Gerard *et al.*, 2015). Worker castes in both ant species lack wings. Interestingly, Osiris genes are among the most highly expressed genes in the early stages of *D. melanogaster* wing development (Sobala & Adler, 2016), when their expression coincides with the timing of cuticle envelope formation.

Whereas temporal peaks of most Osiris genes are relatively consistent, Osi18-20 (the green group, Fig. 1) are temporally different relative to the others, peaking later in both embryonic and pupal development in *D. melanogaster* (Fig. 4, Figs S8–S11), and peaking in larvae of male and worker *B. terrestris* (Fig. 4). This difference in developmental timing, coupled with their physically tight linkage, suggests differential regulation. Interestingly, despite both minor and major inversions in the Osiris cluster of some insects, Osi18-20 remained

physically proximate. The largest rearrangement of the Osiris cluster can be seen in the flour beetle, *Tribolium castaneum*, where the entire middle of the cluster was inverted and translocated to the 5' of Osi1 – the tail end of this inversion, Osi14-16, was further translocated over 2Mbp away (Fig. S2).

Co-expression modules

To gain insight into the function of the Osiris genes, we examined which other genes have correlated expression using weighted gene co-expression analysis using the WGCNA package in R (Langfelder & Horvath, 2008). This analysis was performed for all four species, and in each, we correlated module eigengenes with pooled developmental stages, or caste × development for the social insects. There were five co-expression modules that were positively correlated with the pupal stage of *D. melanogaster*, two of these had Osiris genes and one had the majority of the cluster (11 Osiris genes of 220 total genes) and included those with the highest levels of expression (Fig. S12). There was only a single co-expression module significantly correlated with worker pupae in both *B. terrestris* (Fig. S13) and *P. barbatus* (Fig. S14). Osiris genes were abundant in each of these worker-pupa-correlated modules; there were 11 Osiris genes of 88 total genes in the *B. terrestris* worker-pupa module, and seven Osiris genes of 182 total, with another seven Osiris genes (of 243 total) in a second module weakly correlated with worker pupae in *P. barbatus*. No modules were significantly correlated with worker pupae in *F. exsecta* (Fig. S15), likely due to limited power with few distinct developmental stages, and fewer samples. Nonetheless, *F. exsecta* Osiris genes were divided into two modules (seven and 11 genes in each) with both modules having a weak association with worker pupae.

We annotated gene ontology (GO) terms for all genes in the above modules, concentrating on those modules enriched with Osiris genes. The 'structural constituent of cuticle' gene ontology term was enriched in both ants and *D. melanogaster*, but was not statistically significant in *B. terrestris* ($P = 0.17$; Fig. 5). Gene ontology terms that were enriched in at least two species were 'chitin binding' and 'serine endopeptidase activity.' The Osiris genes do not have annotated molecular function GO terms, and thus did not bias this analysis. Overall, the enriched GO terms strongly suggest that these modules play a role in chitin and cuticle formation. The term for serine endopeptidase activity is also consistent with a role in chitin processes as these enzymes are known to cleave chitin synthases, which are secreted as zymogens, and increase the rate of chitin synthesis (Merzendorfer & Zimoch, 2003; Broehan *et al.*, 2007, 2010).

Among the primary co-expression modules for all species, there was a four-species overlap of three genes,

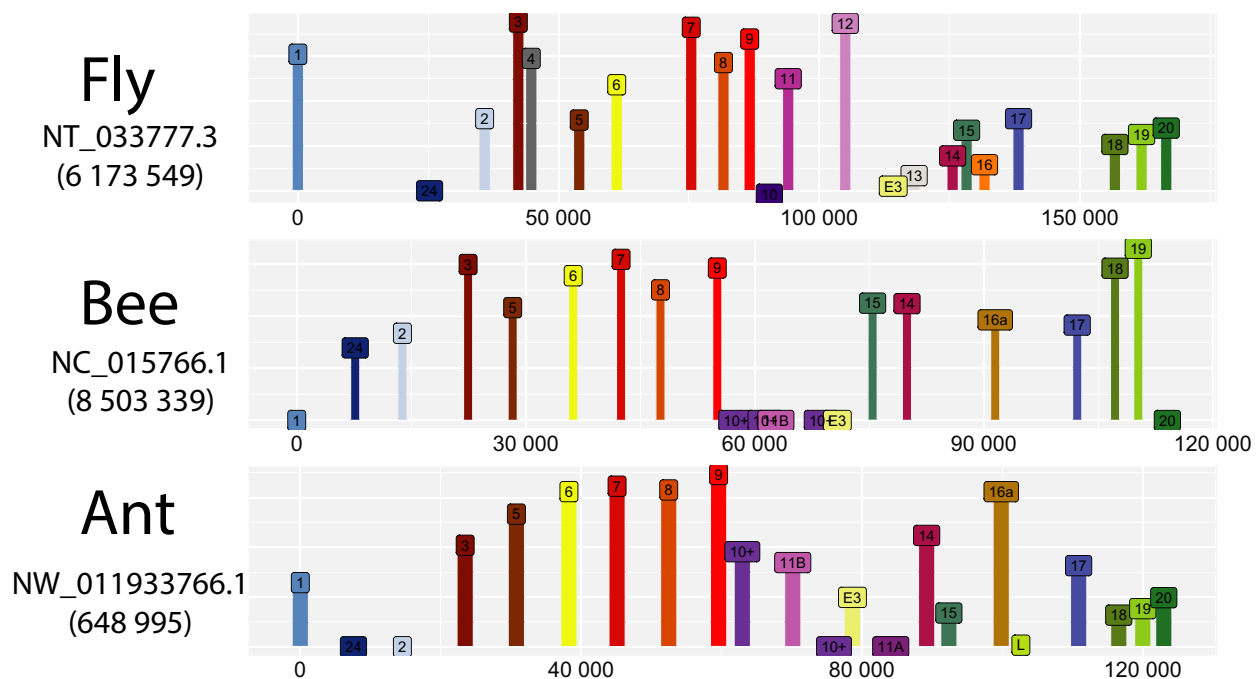


Fig. 3 Osiris gene relative expression in the context of physical location on the chromosome/scaffold. Gene expression values were normalized to the transcriptome mean. Only gene expression from pupal samples was included here – presented here is the mean of all pupal samples. From top to bottom are *D. melanogaster*, *B. terrestris* and *P. barbatus*. Osiris genes are colour coded as in Fig. 1. The x-axis scale starts at the location of Osi1, and scales are different for each species; the identity and total length of the genome region pictured is on the left.

Osi8-9 and CG6055 (Fig. 5). Three-species overlaps included Osi3, 6, 7, as well as five genes of unknown function (CG7031, CG12964, CG14237, CG14636 and CG32816), and multiple genes implicated in cuticle/chitin formation and assembly (Fig. 4). The function of *dyl* has been established as the deposition of chitin, both around bristles (Nagaraj & Adler, 2012) and in the expansion of wings (Ren *et al.*, 2005). The relevance of St2 sulfotransferase is not clear, although sulfotransferase activity is required for proper epidermis formation in nematodes (Kim *et al.*, 2005). Genes CG6055 and CG14866 encode C-type lectin-like proteins, which in other species are known to bind chitin as part of antifungal immune responses (Bueter *et al.*, 2013). The genes *dusky* (*dy*), *Mind-the-Gap* (*mtg*), *Gasp* and *obstructor A* (*obst-A*) all bind chitin, with the latter two known to effect epidermal cuticle integrity (Tiklová *et al.*, 2013).

Interestingly, a butterfly with a seasonal wing colour polymorphism upregulates several Osiris genes under long-day conditions (Vilcinskas & Vogel, 2016). Upregulated along with the Osiris genes is *dyl*, but upregulated under the alternative short-day condition are multiple cuticle proteins, C-type lectins, a carboxypeptidase and *Gasp*. This result (and see Osiris gene expression in toxicology and immunity, below) suggests that genes within these modules can be decoupled for different

purposes, though acting on the same phenotypes (in this case, wings).

Osiris gene expression in the hemimetabola

The only currently available nonholometabolous insect gene expression developmental series is in the termite, *Zootermopsis nevadensis* (Terrapon *et al.*, 2014). Osiris gene expression is high in the egg sample, but is otherwise very low. To gain a better understanding of the evolution of gene expression in Osiris genes, we conducted quantitative real-time PCR (qRT-PCR) on the hemimetabolous cockroach, *Blattella germanica*. We designed primers to amplify Osi7, Osi20 and two copies of Osi9 (named a and b), along with control genes Actin and GAPDH. We sampled 56 individuals representing all nymphal instars from a laboratory colony. Osiris gene expression was undetectable, or nearly so, in all but the last larval instar, along with some Osi9a expression in one adult sample (Fig. 6). When Osiris gene expression peaked, it had very high abundance relative to GAPDH and Actin, ubiquitously expressed genes. These data suggest a conserved timing of expression for Osiris genes in the penultimate instar – the last nymphal instar for the hemimetabola, and the pupa for the holometabola. In both insect groups, the penultimate stage coincides with

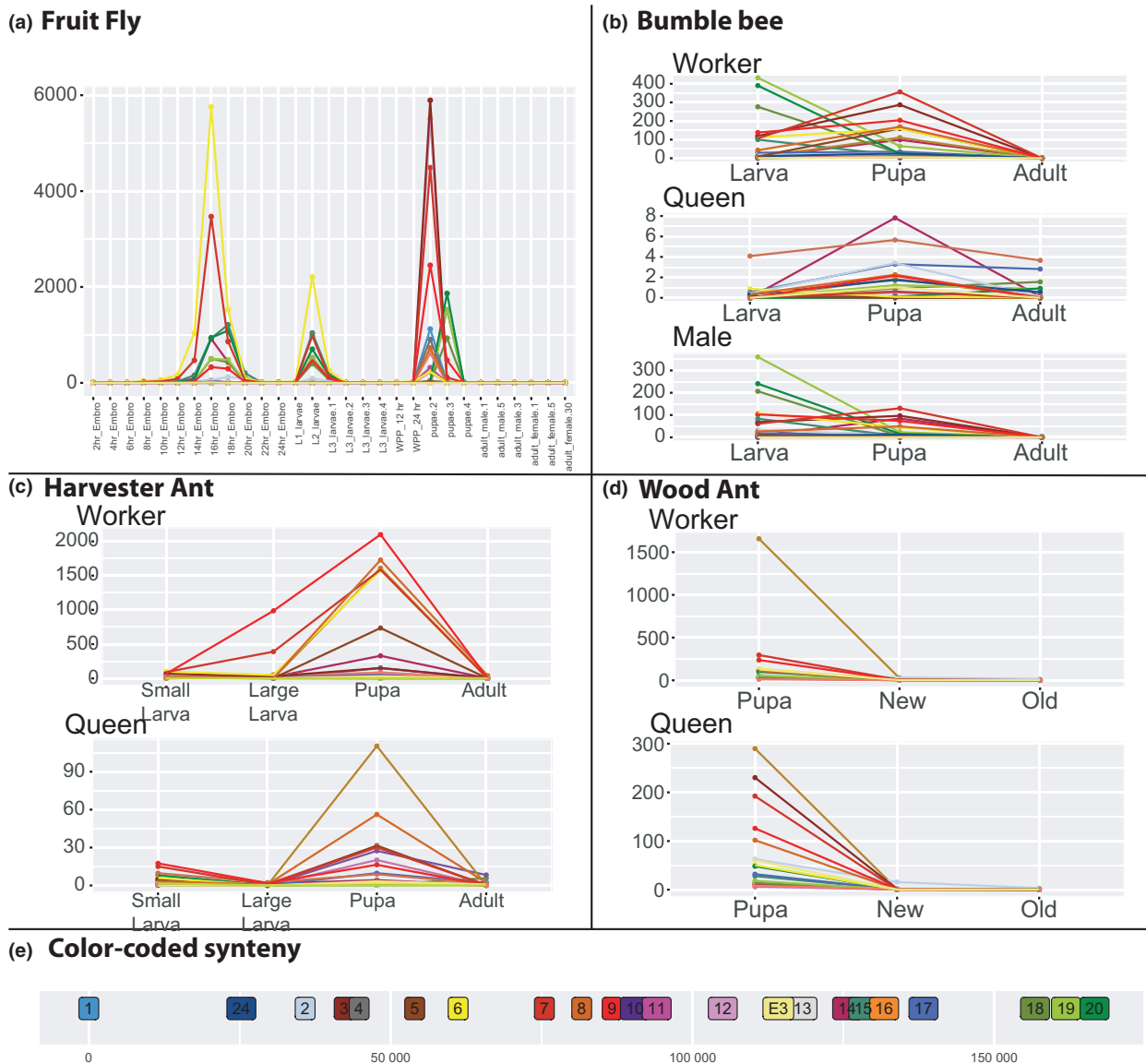


Fig. 4 Line plots of developmental relative (normalize within species) gene expression of Osiris genes in four species, colour coded by gene. The panels are as follows: (a) 30 developmental stages of *D. melanogaster*, (b) larvae, pupae and adults from worker, male and queen *B. terrestris*, (c) two larval stages, pupae and adults from worker and queen *P. barbatus*, (d) pupae, new adults and old adults from worker and queen *F. exsecta* and (e) a key to the colour coding of genes mapped to the syntenic region of *D. melanogaster*.

the final transition to adulthood – the maturation/acquisition of traits associated with dispersal and reproduction.

Osiris gene knock-down

Despite their unknown function, Osiris genes are known to be essential for proper development. Abnormal copy number of the *tpl* locus (encompassing the Osiris cluster) in *D. melanogaster* results in embryonic or post-hatching lethality. The iBeetle knock-down screen (Schmitt-Engel *et al.*, 2015) has embryonic and

pupal phenotypes for ten Osiris genes. Most of these phenotypes are lethal in late-larval injections, with lethality up to 100% for Osi7, 17 and 24. The larval knock-down of Osi7 fails to metamorphose and arrests as a prepupa, while others have increased lethality in pupae or early adults. Embryonic knock-down phenotypes indicate major developmental errors, from segment loss to appendage malformation, and of course death (Schmitt-Engel *et al.*, 2015). Larval knock-down of Osi6 and Osi7 produced adults with visible wing phenotypes that impaired flight (Andrade López *et al.*, 2017).

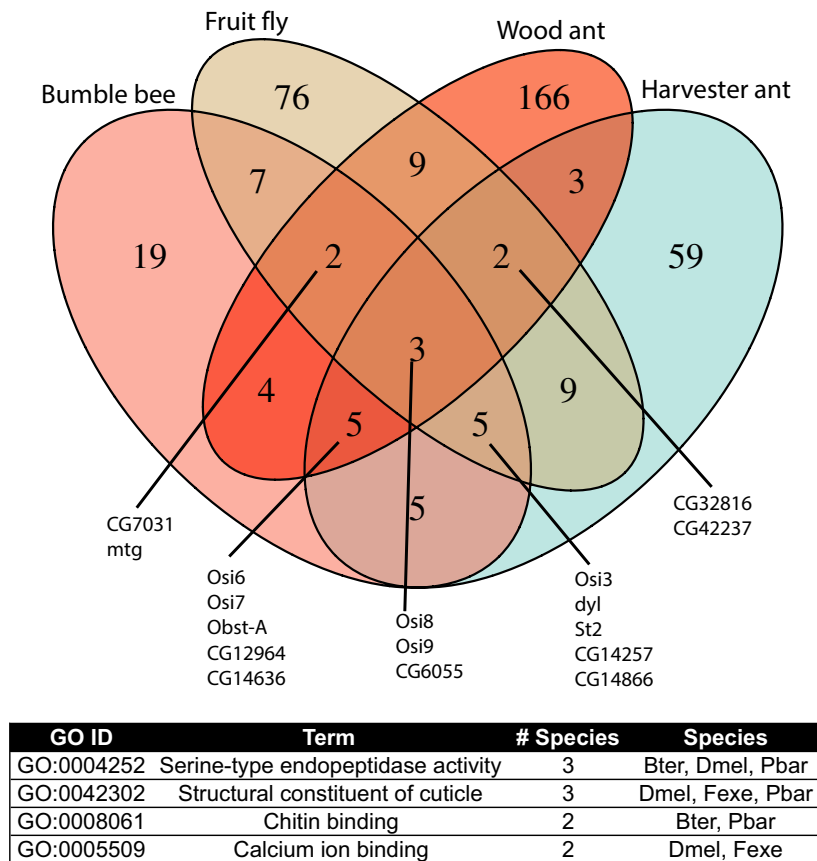


Fig. 5 Four-species overlap of genes in gene co-expression modules correlated with pupae/worker pupae, and enriched in Osiris genes. Below are gene ontology (GO) terms enriched in these modules for at least two species, with the species column indicating which species module was enriched for the term. Dmel = *Drosophila melanogaster*, Bter = *Bombus terrestris*, Pbar = *Pogonomyrmex barbatus*, Fexe = *Formica exsecta*.

Osiris gene expression in toxicology and immunity

Given the chitinous nature of the peritrophic matrix and its role as an infection barrier by isolating food and toxins, and the possible role of Osiris genes in chitin regulation, it is not surprising that the Osiris genes are upregulated in response to immune challenge. When fed supernatant from cultured *Bacillus thuringiensis* (*Bt*), the flour beetle, *T. castaneum*, upregulated seven Osiris genes (Greenwood *et al.*, 2017). Interestingly, these same genes were down-regulated when challenged with *Bt* (regardless of whether they were immune primed). Also of interest from this study is a decoupling of molecular function, as noted from GO term enrichment, in *Bt* primed and *Bt* challenged individuals. The structural constituent of cuticle GO term is enriched with *Bt* priming, and the serine endopeptidase activity term is under-represented with priming, although these terms reverse upon *Bt* challenge. This is in contrast to our study where these terms were enriched in pupae/worker pupae, again suggesting a possible decoupling of developmental gene co-expression modules in immunity.

Several studies in honeybees, *Apis mellifera*, have made associations with Osiris genes and immunity.

When challenged with chalkbrood fungus, *Ascosphaera apis*, Osi6 was down-regulated (Aronstein *et al.*, 2010). On the other hand, a study infecting *A. mellifera* larvae with foulbrood bacteria, *Paenibacillus*, led to upregulation of 17 Osiris genes (Cornman *et al.*, 2013). The authors of this last study suggested that this effect could be confounded by development, as other genes like dusky (*dy*) and dusky-like (*dyl*) were also upregulated. However, in the light of the frequency with which Osiris genes are associated with infection protection, at different life stages, it is likely that the observed pattern is biologically relevant. Interestingly, this effect of Osiris genes may not be localized to the midgut, as much of their expression is in other tissues, especially in the pupal fat body (Gelbart & Emmert, 2013). A proteomic study of the sexually transmitted honeybee fungal pathogen, *Nosema apis*, found Osi7 protein upregulated in drone sperm infected with the pathogen (Grassl *et al.*, 2017). Also among the upregulated proteins were chitinases, which act to break down chitin in fungi, suggesting a logical connection to the upregulation of Osi7 in this study.

The Osiris genes are also upregulated upon challenge with plant secondary chemical defences. In response to feeding on plants producing glucosinolates, the

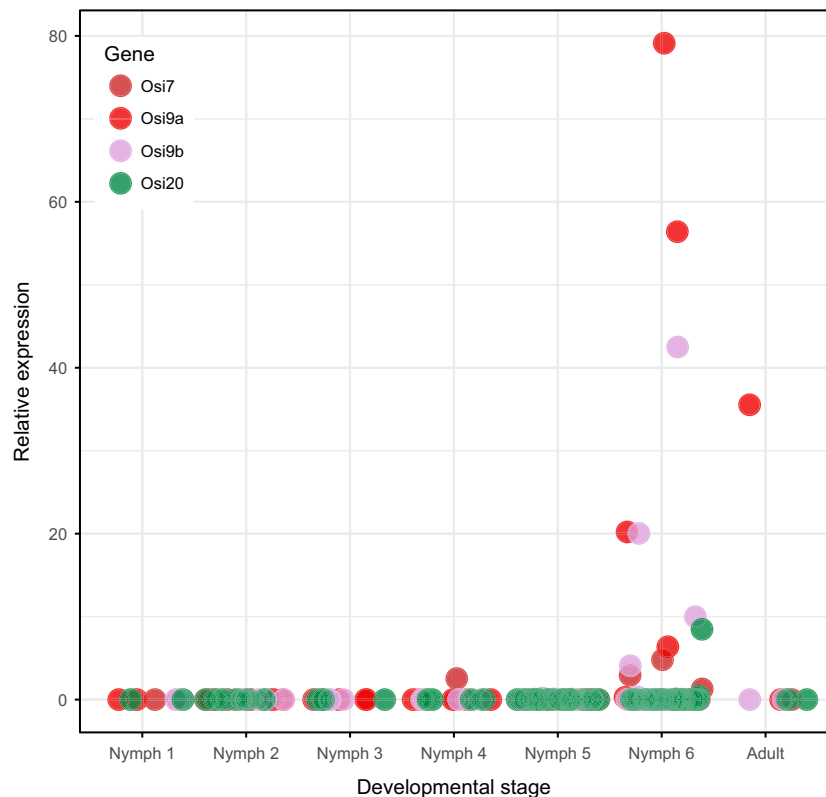


Fig. 6 Gene expression of four Osiris genes (average ΔC_t , relative to Actin and GAPDH) across nymphal and adult stages of development in the German cockroach, *Blattella germanica*. Points are jittered to prevent overlap, but chronology is not known within stages. Genes are colour coded as in Fig. 1, although *B. germanica* has two copies of Osi9. For contrast, one paralog (Osi9a) retained the original colour while the other (Osi9b) was assigned a new colour.

leaf-rolling fly, *Scaptomyza flava*, upregulated eight Osiris genes along with various genes involved in chitin binding and that are structural constituents of cuticle (Whiteman *et al.*, 2012). Both *D. sechellia* and some populations of *D. yakuba* have independently evolved host specificity for the noni fruit, *Morinda citrifolia*, which produces a potent insecticide, octanoic acid. Both species show evidence of strong selection on the Osiris cluster (Yassin *et al.*, 2016). Moreover, several Osiris knock-downs (Osi6-8) lose resistance to octanoic acid (Andrade López *et al.*, 2017). Interestingly, the opposite effect was found in adult tissue-specific knock-downs of these same genes, particularly in fat body and salivary gland.

Additionally, *D. melanogaster* upregulated several Osiris genes upon chemical challenges, including at several life stages. For example, Osi6 has high to very high expression when exposed to heavy metals, caffeine and the Sindbis virus (Gelbart & Emmert, 2013).

Conclusions

With this study, we aim to address a major knowledge gap regarding the importance of Osiris genes for insect development, ecology and evolution. Osiris genes have seemingly contrary function. They are upregulated in social insect castes that have no wings (or smaller wings), yet are also upregulated in wing development

of nonsocial species. They are upregulated and down-regulated in immune responses, and similarly in the presence of a toxin. However, the essential nature of these genes in insect development is apparent, with lethality at multiple developmental time-points in their absence/knock-down. Furthermore, their role in toxin resistance is also becoming clear, as is their role in the generation of alternative phenotypes. Given the breadth of their action, these genes could have very important roles in insect adaptation and success, from host-parasite and herbivore-plant coevolution to morphological novelty and phenotypic plasticity. The expansion of these genes at the root of the insect tree is further suggestive that they may have played an important role in wing evolution and/or the radiation of early insects.

Acknowledgments

The authors are very grateful to J. Gadau, the Bornberg-Bauer and Kurtz laboratory groups, and the entire Institute of Evolutionary Biology at U. Münster, for being very generous with their ideas and time. H. Lerner advised on phylogenetic methods, thank you. Joe Coolon provided helpful discussions and input on the manuscript. Funding for this project was provided as a fellowship by the Evolution Think Tank (U. Münster) to CRS and grant support from the Indiana Academy of

Science to CRS. Additional support from the Gerald Bakker Collaborative Research Endowed Fund and a GLCA Innovation Grant at Earlham College helped support this project.

References

- Abascal, F., Zardoya, R. & Posada, D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**: 2104–2105.
- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G. *et al.* 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- Alexa, A. & Rahnenfuhrer, J. 2016. topGO: Enrichment analysis for gene ontology. R package version 2.26.0.
- Alexa, A., Rahnenfuhrer, J. & Lengauer, T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**: 1600–1607.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Anders, S., Pyl, P.T. & Huber, W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.
- Andersen, S.O. 1979. Biochemistry of insect cuticle. *Annu. Rev. Entomol.* **24**: 29–59.
- Anderson, K.E., Linksvayer, T.A. & Smith, C.R. 2008. The causes and consequences of genetic caste determination in ants (Hymenoptera: Formicidae). *Myrmecol. News* **11**: 119–132.
- Andrade López, J.M., Lanno, S.M., Auerbach, J.M., Moskowitz, E.C., Sliagar, L.A., Wittkopp, P.J. *et al.* 2017. Genetic basis of octanoic acid resistance in *Drosophila sechellia*: functional analysis of a fine-mapped region. *Mol. Ecol.* **26**: 1148–1160.
- Aronstein, K.A., Murray, K.D. & Saldivar, E. 2010. Transcriptional responses in Honey Bee larvae infected with chalkbrood fungus. *BMC Genom.* **11**: 391.
- Bonasio, R., Zhang, G., Ye, C., Mutti, N.S., Fang, X., Qin, N. *et al.* 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science* **329**: 1068–1071.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D. *et al.* 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**: e1003537.
- Boutanaev, A.M., Kalmykova, A.I., Shevelyov, Y.Y. & Nurminsky, D.I. 2002. Large clusters of co-expressed genes in the *Drosophila* genome. *Nature* **420**: 666–669.
- Broehan, G., Zimoch, L., Wessels, A., Ertas, B. & Merzendorfer, H. 2007. A chymotrypsin-like serine protease interacts with the chitin synthase from the midgut of the tobacco hornworm. *J. Exp. Biol.* **210**: 3636–3643.
- Broehan, G., Arakane, Y., Beeman, R.W., Kramer, K.J., Muthukrishnan, S. & Merzendorfer, H. 2010. Chymotrypsin-like peptidases from *Tribolium castaneum*: a role in molting revealed by RNA interference. *Insect Biochem. Mol. Biol.* **40**: 274–283.
- Bueter, C.L., Specht, C.A. & Levitz, S.M. 2013. Innate sensing of chitin and chitosan. *PLoS Pathog.* **9**: e1003080.
- Chen, H. & Boutros, P.C. 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**: 35.
- Chipman, A.D., Ferrier, D.E.K., Brena, C., Qu, J., Hughes, D.S.T., Schröder, R. *et al.* 2014. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. *PLoS Biol.* **12**: e1002005.
- Cornman, R.S., Lopez, D. & Evans, J.D. 2013. Transcriptional response of honey bee larvae infected with the bacterial pathogen *Paenibacillus* larvae. *PLoS ONE* **8**: e65424.
- Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**: 1164–1165.
- Davies, N.J., Krusche, P., Tauber, E. & Ott, S. 2015. Analysis of 5' gene regions reveals extraordinary conservation of novel non-coding sequences in a wide range of animals. *BMC Evol. Biol.* **15**: 227.
- Dorer, D.R., Ezekiel, D.H. & Christensen, A.C. 1995. The Triplo-lethal locus of *Drosophila*: reexamination of mutants and discovery of a second-site suppressor. *Genetics* **141**: 1037–1042.
- Dorer, D.R., Rudnick, J.A., Moriyama, E.N. & Christensen, A.C. 2003. A family of genes clustered at the Triplo-lethal locus of *Drosophila melanogaster* has an unusual evolutionary history and significant synteny with *Anopheles gambiae*. *Genetics* **165**: 613–621.
- Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**: 1969–1973.
- Eddy, S.R. & Wheeler, T.J., the HMMER Development Team. 2015. HMMER 3.1.2b. <http://hmmerr.org/>.
- Emlen, D.J. & Nijhout, H.F. 2000. The development and evolution of exaggerated morphologies in insects. *Annu. Rev. Entomol.* **45**: 661–708.
- Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L. *et al.* 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**: D279–D285.
- Gelbart, W.M. & Emmert, D.B.. 2013. FlyBase high throughput expression pattern data. <http://flybase.org/reports/FBFB0221009.html>.
- Gerard, M., Michez, D., Fournier, D., Maebe, K., Smagghe, G., Biesmeijer, J.C. *et al.* 2015. Discrimination of haploid and diploid males of *Bombus terrestris* (Hymenoptera: Apidae) based on wing shape. *Apidologie* **46**: 644–653.
- Grassl, J., Peng, Y., Baer-Imhoof, B., Welch, M., Millar, A.H. & Baer, B. 2017. Infections with the sexually transmitted pathogen *Nosema apis* trigger an immune response in the seminal fluid of honey bees (*Apis mellifera*). *J. Proteome Res.* **16**: 319–334.
- Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L. *et al.* 2011. The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**: 473–479.
- Greenwood, J.M., Milutinović, B., Peuß, R., Behrens, S., Esser, D., Rosenstiel, P. *et al.* 2017. Oral immune priming with *Bacillus thuringiensis* induces a shift in the gene expression of *Tribolium castaneum* larvae. *BMC Genom.* **18**: 329.
- Harrison, M.C., Hammond, R.L. & Mallon, E.B. 2015. Reproductive workers show queenlike gene expression in an intermediately eusocial insect, the buff-tailed bumble bee *Bombus terrestris*. *Mol. Ecol.* **24**: 3043–3063.
- Hopkins, T.L. & Kramer, K.J. 1992. Insect cuticle sclerotization. *Annu. Rev. Entomol.* **37**: 273–302.

- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C. *et al.* 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.
- Kanost, M.R., Arrese, E.L., Cao, X., Chen, Y.-R., Chellapilla, S., Goldsmith, M.R. *et al.* 2016. Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. *Insect Biochem. Mol. Biol.* **76**: 118–147.
- Kim, T.H., Hwang, S.B., Jeong, P.-Y., Lee, J. & Cho, J.W. 2005. Requirement of tyrosylprotein sulfotransferase-A for proper cuticle formation in the nematode *C. elegans*. *FEBS Lett.* **579**: 53–58.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. & Salzberg, S.L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**: R36.
- Langfelder, P. & Horvath, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559.
- Langmead, B. & Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357–359.
- Larabee, F.J. & Suarez, A.V. 2014. The evolution and functional morphology of trap-jaw ants (Hymenoptera: Formicidae). *Myrmecol News* **20**: 25–36.
- Lee, J., Song, M. & Hong, S. 2013. Negative regulation of the novel norpAP24 suppressor, *diehard4*, in the endo-lysosomal trafficking underlies photoreceptor cell degeneration. *PLoS Genet.* **9**: e1003559.
- Leek, J.T. & Storey, J.D. 2007. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**: e161.
- Leek, J.T., Johnson, W.E., Parker, H.S., Fertig, E.J., Jaffe, A.E. & Storey, J.D. 2017. sva: surrogate variable analysis.
- Lehane, M.J. 1997. Peritrophic matrix structure and function. *Annu. Rev. Entomol.* **42**: 525–550.
- Li, B. & Dewey, C.N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**: 323.
- Lindsley, D.L., Sandler, L., Baker, B.S., Carpenter, A.T.C., Denell, R.E., Hall, J.C. *et al.* 1972. Segmental aneuploidy and the genetic gross structure of the *Drosophila* genome. *Genetics* **71**: 157–184.
- Medler, J.T. 1962. Morphometric studies on bumble bees. *Ann. Entomol. Soc. Am.* **55**: 212–218.
- Merzendorfer, H. 2011. The cellular basis of chitin synthesis in fungi and insects: common principles and differences. *Eur. J. Cell Biol.* **90**: 759–769.
- Merzendorfer, H. & Zimoch, L. 2003. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *J. Exp. Biol.* **206**: 4393–4412.
- Michalak, P. 2008. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* **91**: 243–248.
- Miller, M.A., Schwartz, T., Pickett, B.E., He, S., Klem, E.B., Scheuermann, R.H. *et al.* 2015. A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol. Bioinforma.* **11**: 43.
- Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C. *et al.* 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**: 763–767.
- Morandin, C., Dhaygude, K., Paviola, J., Trontti, K., Wheat, C. & Helanterä, H. 2015. Caste-biases in gene expression are specific to developmental stage in the ant *Formica exsecta*. *J. Evol. Biol.* **28**: 1705–1718.
- Morandin, C., Tin, M.M.Y., Abril, S., Gómez, C., Pontieri, L., Schjøtt, M. *et al.* 2016. Comparative transcriptomics reveals the conserved building blocks involved in parallel evolution of diverse phenotypic traits in ants. *Genome Biol.* **17**: 43.
- Nagaraj, R. & Adler, P.N. 2012. Dusky-like functions as a Rab11 effector for the deposition of cuticle during *Drosophila* bristle development. *Development* **139**: 906–916.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. *et al.* 2013. Package 'vegan.' Community Ecol. Package Version.
- Paradis, E., Claude, J. & Strimmer, K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ren, N., Zhu, C., Lee, H. & Adler, P.N. 2005. Gene expression during *Drosophila* wing morphogenesis and differentiation. *Genetics* **171**: 625–638.
- Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W. *et al.* 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**: 949–955.
- Robinson, M.D. & Oshlack, A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**: R25.
- Robinson, M.D., McCarthy, D.J. & Smyth, G.K. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Sadd, B.M., Barribeau, S.M., Bloch, G., de Graaf, D.C., Dearn, P., Elisk, C.G. *et al.* 2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol.* **16**: 76.
- Schmitt-Engel, C., Schultheis, D., Schwirz, J., Ströhlein, N., Troelenberg, N., Majumdar, U. *et al.* 2015. The iBeetle large-scale RNAi screen reveals gene functions for insect development and physiology. *Nat. Commun.* **6**: 7822.
- Shah, N., Dorer, D.R., Moriyama, E.N. & Christensen, A.C. 2012. Evolution of a large, conserved, and syntenic gene family in insects. *Genes Genomes Genet.* **2**: 313–319.
- Smith, C.R., Smith, C.D., Robertson, H.M., Helmkamp, M., Zimin, A., Yandell, M. *et al.* 2011. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc. Natl. Acad. Sci. USA* **108**: 5667–5672.
- Smith, C.R., Cahan, S.H., Kemena, C., Brady, S.G., Yang, W., Bornberg-Bauer, E. *et al.* 2015. How do genomes create novel phenotypes? Insights from the loss of the worker caste in ant social parasites. *Mol. Biol. Evol.* **32**: 2919–2931.
- Smoyer, L.K., Dorer, D.R., Nickerson, K.W. & Christensen, A.C. 2003. Phenotype of the Triplo-lethal locus of *Drosophila melanogaster* and its suppression by hyperoxia. *Genet. Res.* **82**: 163–170.
- Sobala, L.F. & Adler, P.N. 2016. The gene expression program for the formation of wing cuticle in *Drosophila*. *PLoS Genet.* **12**: e1006100.
- Spellman, P.T. & Rubin, G.M. 2002. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* **1**: 5.
- Tanaka, A. & Hasegawa, A. 1979. Nymphal development of the German cockroach, *Blattella germanica* Linne (Blattaria: Blattellidae), with special reference to instar determination and intra-instar staging. *Kontyu Entomol. Soc. Jpn.* **47**: 225–238.

- Terra, W.R. 1990. Evolution of digestive systems of insects. *Annu. Rev. Entomol.* **35**: 181–200.
- Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth, W. *et al.* 2014. Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* **5**: 3636.
- The International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* **8**: e1000313.
- Thibaud-Nissen, F., Souvorov, A., Murphy, T., DiCuccio, M. & Kitts, P. 2013. Eukaryotic genome annotation pipeline. <https://www.ncbi.nlm.nih.gov/books/n/handbook2e/EukaryoticAnnotationPipeline/?report=reader>
- Tiklová, K., Tsarouhas, V. & Samakovlis, C. 2013. Control of airway tube diameter and integrity by secreted chitin-binding proteins in *Drosophila*. *PLoS ONE* **8**: e67415.
- Tomoyasu, Y. & Fujiwara, H. 2017. Editorial overview: development and regulation: the diverse traits that have facilitated the successful radiation of insects. *Curr. Opin. Insect Sci.* **19**: vi–ix.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J. *et al.* 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**: 511–515.
- Vilcinskis, A. & Vogel, H. 2016. Seasonal phenotype-specific transcriptional reprogramming during metamorphosis in the European map butterfly *Araschnia levana*. *Ecol. Evol.* **6**: 3476–3485.
- Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A. *et al.* 2009. gplots: Various R programming tools for plotting data. R Package Version.
- Wei, T., Simko, V. & Wei, M.T. 2016. Package ‘corrplot’. *Statistician* **56**: 316–324.
- Weinstock, G.M., Robinson, G.E., Gibbs, R.A., Worley, K.C., Evans, J.D., Maleszka, R. *et al.* 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**: 931–949.
- Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., *et al.* 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**: 343–348.
- Wheeler, T.J. & Eddy, S.R. 2013. nhmmer: DNA homology search with profile HMMs. *Bioinforma. Oxf. Engl.* **29**: 2487–2489.
- Whiteman, N.K., Gloss, A.D., Sackton, T.B., Groen, S.C., Humphrey, P.T., Lapoint, R.T. *et al.* 2012. Genes involved in the evolution of herbivory by a leaf-mining, *Drosophilid* fly. *Genome Biol. Evol.* **4**: 900–916.
- Wickham, H. 2016. ggplot2: elegant graphics for data analysis.
- Wigglesworth, V.B. 1957. The physiology of insect cuticle. *Annu. Rev. Entomol.* **2**: 37–54.
- Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B. *et al.* 2004. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* **306**: 1937–1940.
- Yassin, A., Debat, V., Bastide, H., Gidaszewski, N., David, J.R. & Pool, J.E. 2016. Recurrent specialization on a toxic fruit in an island *Drosophila* population. *Proc. Natl. Acad. Sci. USA* **113**: 4771–4776.
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T.L. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**: 134.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1 Phylogeny of insect Osiris orthologs.

Figure S2 Synteny of Osiris genes within the Holometabola.

Figure S3 Correlation matrix of developmental Osiris expression, *B. terrestris*.

Figure S4 Correlation matrix of developmental Osiris expression, *P. barbatus*.

Figure S5 Correlation matrix of developmental Osiris expression, *F. exsecta*.

Figure S6 Plot of correlation statistical significance by strength for all four species.

Figure S7 Comparative NMDS plots of Osiris developmental expression.

Figure S8 Guide phylogeny of *D. melanogaster* Osiris genes.

Figure S9 Phylo-heat-map of expression: 16 h embryo vs. 48 h pupa.

Figure S10 Phylo-heat-map of expression: 18 h embryo vs. 2nd instar larva.

Figure S11 Phylo-heat-map of expression: 48 h vs. 72 h pupa.

Figure S12 Co-expression module development correlation: *D. melanogaster*.

Figure S13 Co-expression module development correlation: *B. terrestris*.

Figure S14 Co-expression module development correlation: *P. barbatus*.

Figure S15 Co-expression module development correlation: *F. exsecta*.

Table S1 Between caste differences in Osiris gene expression.

Data S1 Relative gene expression tables (RPKM, reads per kilobase of transcript per million mapped reads) for species (except fruit fly) used in gene expression analyses.

Data S2 Mapping of genes to co-expression modules for all species used in co-expression analyses.

Data S3 Osiris gene name and co-expression module grouping for all species used in co-expression analyses.

Received 3 October 2017; revised 18 December 2017; accepted 20 December 2017