

RESEARCH ARTICLE

Open Access



Host and venom evolution in parasitoid wasps: does independently adapting to the same host shape the evolution of the venom gland transcriptome?

Yi Yang¹ , Shan Xiao¹ , Xianxin Zhao¹ , Yu H. Sun², Qi Fang¹ , Longjiang Fan³, Gongyin Ye¹ and Xinhai Ye^{4,5,6*}

Abstract

Background Venoms have repeatedly evolved over 100 occasions throughout the animal tree of life, making them excellent systems for exploring convergent evolutionary novelty. Growing evidence supports that venom evolution is predominantly driven by prey or host-related selection pressures, and the expression patterns of venom glands reflect adaptive evolution. However, it remains elusive whether the evolution of expression patterns in venom glands is likewise a convergent evolution driven by their prey/host species.

Results We utilized parasitoid wasps that had independently adapted to *Drosophila* hosts as models to investigate the convergent evolution of venom gland transcriptomes in 19 hymenopteran species spanning ~200 million years of evolution. Comparative transcriptome analysis reveals that the global expression patterns among the venom glands of *Drosophila* parasitoid wasps do not achieve higher similarity compared to non-*Drosophila* parasitoid wasps. Further evolutionary analyses of expression patterns at the single gene, orthogroup, and Gene Ontology (GO) term levels indicate that some orthogroups/GO terms show correlation with the *Drosophila* parasitoid wasps. However, these groups rarely include genes highly expressed in venom glands or putative venom genes in the *Drosophila* parasitoid wasps.

Conclusions Our study suggests that convergent evolution may not play a predominant force shaping gene expression levels in the venom gland of the *Drosophila* parasitoid wasps, offering novel insights into the co-evolution between venom and prey/host.

Keywords Convergent evolution, Gene expression, Venom, Comparative transcriptome, Parasitoid wasps

*Correspondence:

Xinhai Ye

yexinhai@zafu.edu.cn

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



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

Background

Convergent evolution offers a valuable opportunity for researchers to study the genetic basis of phenotypic adaptation by observing the independent emergence of similar traits in different evolutionary lineages [1]. Numerous studies have identified instances of convergence at the molecular level, elucidating specific phenotypic adaptations such as vision loss in cave animals, vocal production learning in mammals, and body size reduction in parasitoid wasps [2–4]. One striking example of convergent evolution is animal venom, where the complex trait of venom has evolved multiple times across the animal tree of life [5–7]. Venom consists of a mixture of molecules and plays a crucial role in the ongoing arms race between venomous animals and their prey or hosts [8]. Nevertheless, the extent to which venom evolves under the selection of prey/host remains unclear.

Parasitoid wasps use venom as a conserved weapon to manipulate various physiological pathways in the arthropod hosts, thereby creating a suitable environment for their offspring to develop [9]. In contrast to the predator venom systems, such as those found in snakes, scorpions, and spiders, which primarily rely on venom as toxins for prey capture or defense, the role of venom in parasitoid wasps encompasses a much broader spectrum of effects on their hosts [3, 6]. For instance, venom genes in parasitoid wasps possess the ability to suppress the host's immune system, modulate the metabolism of amino acids and lipids, and regulate the behavior of the host [10–14]. Furthermore, adapting to the specific host puts significant pressure on the repertoire of parasitoid venom, which could lead to host-specific venom properties of parasitoid wasps. Previous studies suggest that changes in their venom repertoires may be largely driven by host shifts [15, 16]. During evolution, several lineages of parasitoid wasps have independently adapted to the same host (or closely related hosts); however, it remains unclear whether these parasitoids often evolve similar venom solutions, i.e., convergent venom evolution. The venom gland of parasitoid wasps is a highly specialized tissue that expresses a small set of genes [17–20] (on average, 4.3% of genes contribute to 80% of total gene expression, based on the data from 18 parasitoid wasps in this study), which mainly reflect the venom components and venom gene regulatory network. Moreover, the venom gene in parasitoid wasps mainly evolves by co-opting existing genes, which alters gene expressions in the venom gland. As a result, the expression patterns of venom gland may reflect the consequences of adaptive evolution [17, 19]. Thus, we hypothesized that independent adaptation to the same or similar hosts could lead to convergent gene expression patterns in the venom glands of parasitoid wasps from different lineages.

Here, we examine this hypothesis by performing a phylogenetic-based comparative transcriptome analysis of venom gland across 19 hymenopteran species, including five parasitoid wasps which represent four independent adaptations to *Drosophila* hosts. Our results reveal that independent adaptation to *Drosophila* hosts does not result in similar venom gland transcriptomes in parasitoid wasps. Phylogenetic analyses on the expression of individual single-copy genes, orthogroups, and Gene Ontology (GO) terms indicate some orthogroups/GO terms exhibiting convergent expression patterns in *Drosophila* parasitoid wasps. However, most of these orthogroups/GO terms do not contain highly expressed genes in the venom gland or putative venom genes across the five *Drosophila* parasitoid wasps or the three larval *Drosophila* parasitoid wasps. These results further support our conclusions that convergent evolution may not be a major force shaping gene expression levels in the *Drosophila* parasitoid wasps. Moreover, our analysis of gene expression dynamics based on single-copy genes indicates that gene expression in the venom gland of parasitoid wasps primarily evolves under stabilizing selection, providing valuable insights into the evolutionary patterns underlying the gene expression in the venom gland of parasitoid wasps.

Results

To test if independent adaptation to the same or similar hosts could lead to convergent gene expression patterns in the venom gland of parasitoid wasps from different lineages, we investigated the venom gland transcriptomes of 19 hymenopteran species spanning six superfamilies, including five *Drosophila* parasitoid wasps that represent four independent adaptations to the *Drosophila* hosts (Additional file 1: Table S1). These five *Drosophila* parasitoid wasps include three larval *Drosophila* parasitoids: *Leptopilina boulardi* (Figitidae), *L. heterotoma* (Figitidae) and *Asobara japonica* (Braconidae), and two pupal *Drosophila* parasitoids: *Trichopria drosophilae* (Diapriidae) and *Pachycrepoideus vindemmiae* (Pteromalidae). Notably, high-quality genomes are available for all species except *T. drosophilae*. We first performed orthology inference analysis using the protein sequences of the 19 species, which allowed us to identify 1479 single-copy orthologous genes shared by all species. Subsequently, we calculated the expression level (TPM) of each gene in each sample and finally generated a normalized expression matrix of the venom gland (TPM10K, see “Methods” for a detailed description) to ensure comparability across species. If independent adaptation to the same host can result in convergent gene expression patterns in the venom gland of parasitoid wasps, it is reasonable to expect that the transcriptome of the venom gland in

Drosophila parasitoid wasps will exhibit similarities. Specifically, we anticipate that the gene expression tree of the venom gland will deviate from the species phylogenetic tree, leading to the clustering of *Drosophila* parasitoid wasps together [2, 21].

Principal component analysis fails to cluster *Drosophila* parasitoid wasps

To explore the overall gene expression patterns of venom glands across parasitoid wasps, we first performed principal component analysis (PCA) on both genome sequence data and expression data based on the 1479 one-to-one orthologues genes. We observed a different pattern between the two PCA plots. The 19

hymenopteran species were strictly separated according to taxon for the genome sequence data (Fig. 1A). However, the transcriptome data of venom glands was contrary to our initial expectation, the *Drosophila* parasitoids data failed to cluster together, nor exhibited clear separation from other species (Fig. 1B, Additional file 2: Fig. S1). We noticed that *T. drosophilae* was separated from other species, which could be an artifact of the lack of reference genome. We further examined the correlation between principal components (PCs) with trait data, including host species and taxonomy (Fig. 1C). Our analysis revealed that PC2 possessed a significant positive correlation with lineages, indicating the dominant role of lineage differences in driving venom gland

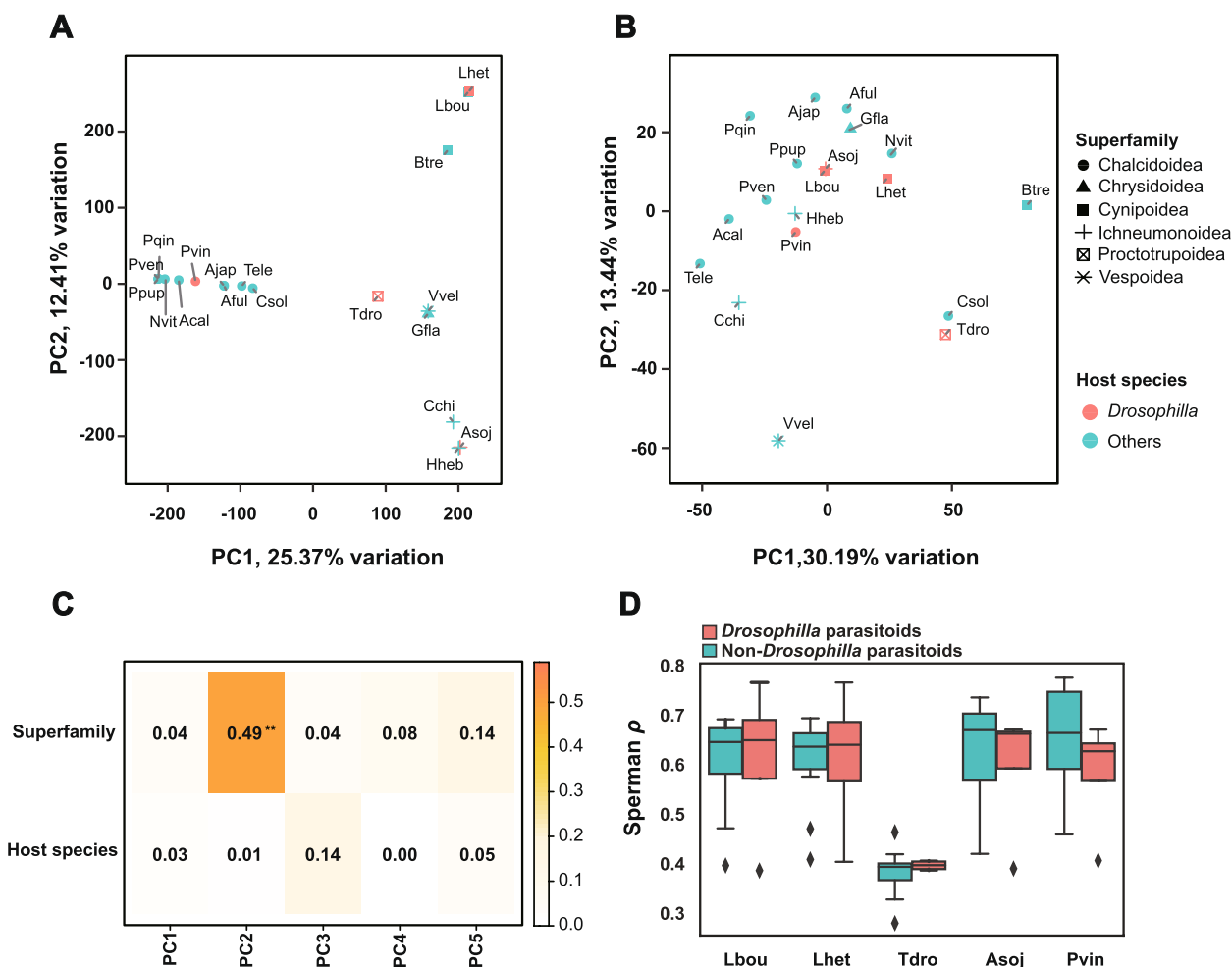


Fig. 1 Principal component analysis (PCA) of venom gland transcriptomes from 19 Hymenoptera species. **A–B** PCA on genomic sequences and on expression data based on the 1479 one-to-one single copy genes. See Additional file 3 for individual data values. **C** Heatmap showing the Pearson's correlation coefficient of the PC1-5 eigenvectors from venom gland gene expression with the phylogeny and host species of parasitoid wasps. *P* values by two-sided Pearson's correlation test (***P* < 0.001). **D** Comparison of the similarity of venom glands between *Drosophila* parasitoids (*n* = 4) and non-*Drosophila* parasitoids (*n* = 14). The Spearman rank correlation coefficient ρ was used to scale the similarity of the venom gland transcriptomes, and no significant difference was observed between *Drosophila* parasitoids and non-*Drosophila* parasitoids (two-sided Wilcoxon rank sum test, *P* > 0.05). See Additional file 3 for individual data values

gene expressions. However, we did not observe any PCs correlated with host species significantly (Fig. 1C). To understand if higher similarity exists within *Drosophila* parasitoids, compared to parasitoid wasps of other host species on the venom gland transcriptomes, we computed the Spearman rank correlation coefficient ρ . As a result, there was no more resemblance observed in venom glands between the *Drosophila* parasitoids and non-*Drosophila* parasitoids (Fig. 1D).

Gene expression phylogeny of venom glands is similar with the species phylogeny

To further dissect the trajectory of venom gland transcriptome evolution, we reconstructed the expression tree using the neighbor-joining (NJ) method based on the Spearman coefficient distance matrix (Additional file 2: Fig. S2). First, we compared the topology of the expression tree with the species phylogenetic tree constructed by the maximum likelihood method based on the protein sequences of the 1479 single-copy orthologous genes (Fig. 2A). We found consistency between the expression tree and the species tree (Fig. 2B). The expression tree correctly clustered the parasitoid wasps from the same superfamily, except for a few outlier species with special traits. Notably, it accurately resolved most of the relationships in Chalcidoidea except for the fig wasp *C. solmsi*, which shifts to adapting on plant hosts. In particular, the five *Drosophila* parasitoids were placed near their close relatives instead of being clustered together themselves. This suggests that the alternations in venom gland gene expression accumulate over evolutionary time, and closely related species tend to exhibit similar venom gland expression atlases.

To quantify the differences in topology between the expression tree and species phylogenetic tree, we employed the PH85 distance (d_T), which is twice the number of internal branches defining different bipartitions of the tips. The calculated distance between the expression tree topology and the species tree topology was determined to be 20 (Fig. 2C). For comparison, we utilized the bootstrap method with 10,000 replications to compare d_T between the transcriptome tree and the species tree with that between the transcriptome tree and the randomly generated tree. We found the d_T between the expression tree and the species tree was significantly smaller than that between the random tree and the species tree (average $d_T=30$, $P<0.001$), and no significant change when using the 50% most highly expressed genes (Additional file 2: Fig. S2). Additionally, none of the obtained topologies of random tree exhibited a d_T value below 20 (Fig. 2C). These results indicated that the gene expression phylogeny of venom glands closely resembled

the species phylogeny and this observation was unlikely to be attributed to chance.

In addition to comparing the topologies, we assessed the correlation between the branch lengths of the expression tree and the species tree. Our results revealed a significant positive correlation between the expression tree of venom glands and the corresponding species tree ($R=0.364$, $P=0.013$, Mantel test; Fig. 2D). This finding further indicates that the overall divergence of gene expression in the venom glands of parasitoid wasps increases with evolutionary time.

Considering that we might miss information from multicopy gene families due to gene loss or gene duplication (i.e., one-to-many and many-to-many orthologs), we also examined another dataset containing 3614 orthogroups (OGs) of orthologous genes shared by all species. Given that OG contains a group of genes derived from a common origin, we used a sum of TPM values in each OG to represent expression abundances at the OG level and generated a normalized (TPM10K) expression matrix to repeat the expression evolutionary analysis. We observed a reduction in the d_T value from the species tree to 18 when using the OG-level expression tree (Fig. 2C, Additional file 2: Fig. S2), and there was a stronger positive correlation of the branch lengths between the species trees and the expression tree (Fig. 2D). This is unsurprising because the expression level of orthogroups is composed of many different genes with variable expression profiles and hence distinguish changes in gene expression patterns less sensitively.

We next ask whether the expression pattern of the functional group in the venom glands of *Drosophila* parasitoid wasps shows convergent shifts. We assigned each gene to specific gene ontology (GO) terms and identified 16,862 GO groups shared by all 19 species. To quantify the expression level of each GO term, we summed the TPM10K values of all genes assigned to that particular GO term. This summation allowed us to generate a GO-level expression matrix. Our subsequent analysis using the GO-level expression data yielded a calculated d_T from the species tree of 24, which is larger compared to both the gene-level and OG-level expression data analyses (Fig. 2C, Additional file 2: Fig. S2). Unexpectedly, we observed that the three larval stage *Drosophila* parasitoid wasps (*L. bouleari*, *L. heterotoma*, and *A. japonica*) formed a monophyletic group in the GO-level expression tree (Additional file 2: Fig. S2). This grouping was supported by a high bootstrap value of 92, indicating strong statistical support for the clustering of these species.

In summary, our expression evolutionary analysis revealed that the expression tree of venom gland was generally consistent with the species phylogeny. This suggests that the convergent adaptation to *Drosophila* may

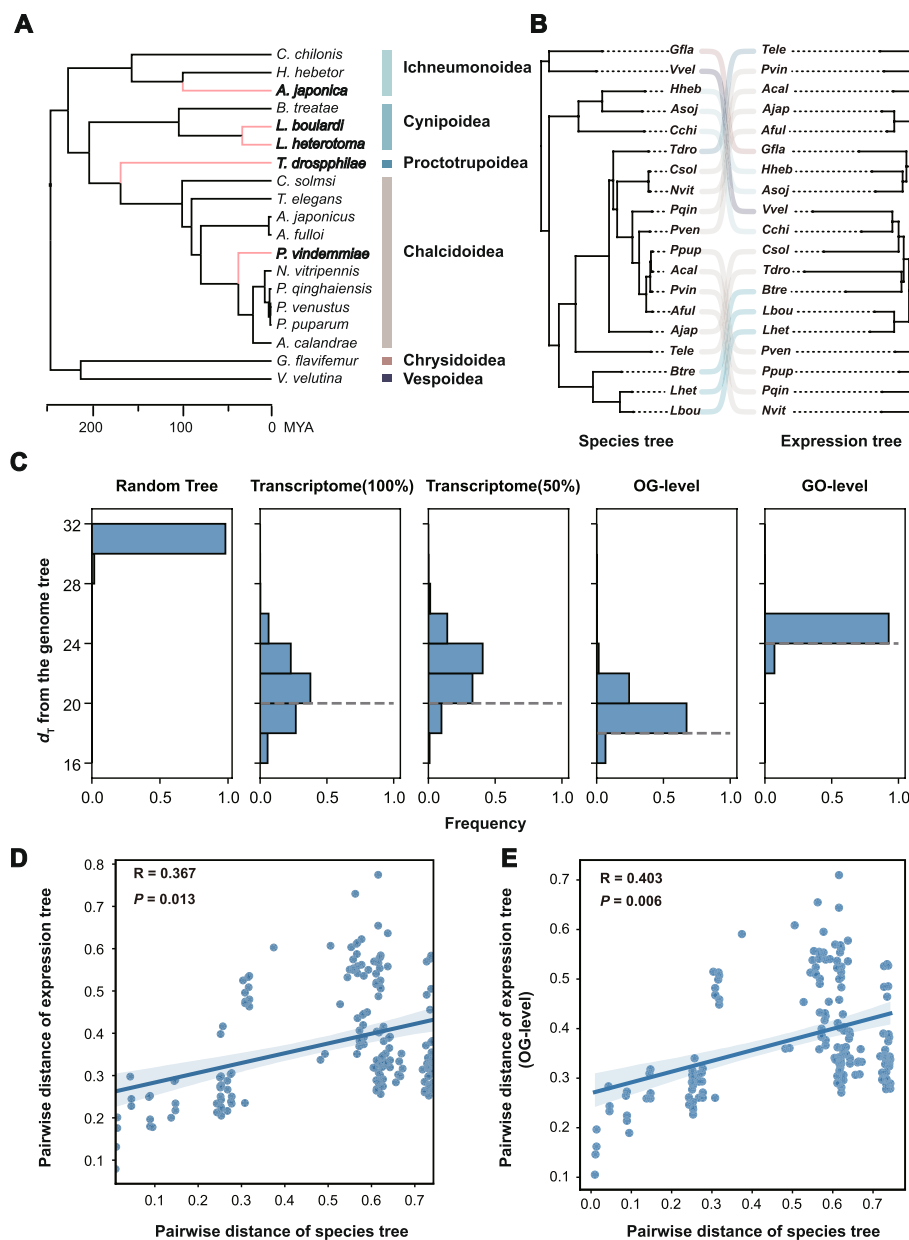


Fig. 2 Comparative transcriptome analysis of venom glands from 19 Hymenoptera species. **A** Species phylogeny of 19 Hymenoptera species. The maximum-likelihood tree was reconstructed by IQ-TREE based on the multisequence alignments of 1479 one-to-one orthologous genes, and branches in red represent four independent origins of adapting to *Drosophila* hosts. All nodes received 100% bootstrap support. **B** Topology comparison between the species tree and the venom gland expression tree. Left is the species phylogeny obtained from **A**, and right is the expression tree reconstructed by the NJ method based on the Spearman distance matrix of the venom gland. **C** Frequency distributions of topology distance (PH85 distance, d_T) between the genome tree and the random tree, transcriptome tree, transcriptome tree with the 50% most highly expressed genes, transcriptome tree based on OG-level expression matrix, and transcriptome tree based on GO-level expression matrix. Each distribution was obtained based on 10,000 random trees or bootstrapped trees ($n = 10,000$). Dotted line represented the observed d_T between genome tree and other trees based on the original data. See Additional file 3 for individual data values. **D–E** Pairwise comparison of the distance matrix between the species tree and the expression tree based on gene-level expression matrix ($n = 171$) (**D**) and based on the OG-level expression matrix ($n = 171$) (**E**). The two-sided Mantel test was used to test for significant correlation between the distance matrices. See Additional file 3 for individual data values

not contribute to the evolution of venom gland expression landscape of parasitoid wasps.

Phylogenetic analysis based on the expression of putative venom genes does not support clustering in *Drosophila* parasitoid wasps

Parasitoid wasps typically possess a relatively modest array of venom proteins, with reported numbers ranging from 59 to 210 across ten species analyzed in this study [17, 19, 20, 22–28]. Our above analysis, which included 1479 single-copy genes, 3614 orthogroups, and 16,862 GO terms likely exceeds the number of venom genes. This potential overrepresentation could obscure the detection of convergent signals in the venom gland transcriptomes of *Drosophila* parasitoid wasps. To address this issue, we first identified the highly expressed genes in the venom glands of each species. Orthogroups/GO terms containing these genes were assigned as venom highly expressed gene-containing groups (VHEOGs/VHEGOs). Among the 1479 single-copy orthogroups, 3614 orthogroups, and 16,862 GO terms shared by the 19 species, 560 single-copy orthogroups, 1202 shared orthogroups, and 13,880 shared GO terms were classified as VHEOGs/VHEGOs, respectively. We then performed phylogenetic analyses of the expression patterns focusing solely on these VHEOGs/VHEGOs. The results were consistent with previous findings. The five *Drosophila* parasitoid wasps did not cluster in the expression trees based on either dataset. The three larval *Drosophila* parasitoid wasps still formed a monophyletic group, but the bootstrap value was reduced to 48 when using the expression matrix based on 13,880 VHEGOs (Additional file 2: Fig. S3).

We further identified putative venom genes in each species and assigned relevant orthogroups and GO terms as venom gene-containing groups (VCOGs/VCGOs). Specifically, there were 106 single-copy orthogroups, 284 shared orthogroups, and 7563 shared GO terms classified as VCOGs/VCGOs, respectively. Similar phylogenetic analyses were carried out using the expression matrices based on these VCOGs/VCGOs. Again, the expression trees failed to cluster the five *Drosophila* parasitoid wasps from either dataset. Notably, we did not observe the clustering of the three larval *Drosophila* parasitoids when using the expression matrix based on the 7563 VCGOs (Additional file 2: Fig. S4). We then filtered our dataset to include only highly expressed genes in the venom gland and putative venom genes of the five *Drosophila* parasitoid wasps. This refinement resulted in 134 single-copy VHEOGs, 326 shared VHEOGs, 8235 shared VHEGOs, 26 single-copy VCOGs, 74 shared VCOGs, and 3914 shared VCGOs, respectively. Subsequent analysis of the expression trees based on these datasets yielded

similar results. We did not observe the clustering of the five *Drosophila* parasitoid wasps or the three larval *Drosophila* parasitoid wasps in either dataset (Additional file 2: Fig. S5, Additional file 2: Fig. S6).

Limited evidence for convergent evolution in the expression profiles of individual orthogroups/GO terms among *Drosophila* parasitoid wasps

The results that lack a strong convergent signal of global expression patterns in the venom gland of *Drosophila* parasitoid wasps caused us to examine if individual genes, orthogroups, and functional groups showed similar patterns which may correspond to the convergent adaptation to *Drosophila* hosts. To this end, we first constructed the expression tree of each 1479 one-to-one single copy orthologues genes by the NJ method based on the Euclidian distance matrix, respectively. We next screened each of the topologies and tended to identify expression trees in which the five *Drosophila* parasitoid wasps formed a monophyletic group, which was particularly effective in detecting adaptive evolution in gene expression data [2, 21]. We did not observe any genes that meet this criterion. Only one gene (*Frizzled-2*) where the expression tree showed a monophyletic group of four *Drosophila* parasitoid wasps (*L. boulandi*, *L. heterotoma*, *A. japonica*, and *T. drosophilae*) and three genes (*ADP-ribose pyrophosphatase*, *nuclear pore membrane glycoprotein*, and *aconitate hydratase*) where the expression tree showed a monophyletic group of three *Drosophila* parasitoid wasps (*L. boulandi*, *L. heterotoma*, and *A. japonica/P. vindemmiae*) (Fig. 3A, Additional file 1: Table S2). This observation falls significantly below the expected values ($P=0.00016$, binomial test), indicating that there is a lack of genes displaying a similar expression pattern in the venom gland of *Drosophila* parasitoid wasps. We next performed the same analysis on each of the 3124 orthogroups shared by the 19 species. We identified one additional orthogroup (OG0004615: serine/threonine protein kinase) where the expression tree clustered the four *Drosophila* parasitoid wasps (*L. boulandi*, *L. heterotoma*, *A. japonica*, and *T. drosophilae*) and nine orthogroups where the expression trees exhibited a monophyletic grouping of three *Drosophila* parasitoid wasps (Additional file 1: Table S3), significantly below the expectation values ($P=6.5e-07$, binomial test). Collectively, these findings suggest that, consistent with the overall expression patterns in the venom gland of *Drosophila* parasitoid wasps, convergent signal also could not be detected at both single-gene and orthogroup levels.

For the GO-level analysis, we extended the same topology examination to all the 16,862 GO terms shared by the 19 species. The results were consistent with the individual gene level analysis. We did not find any GO terms where

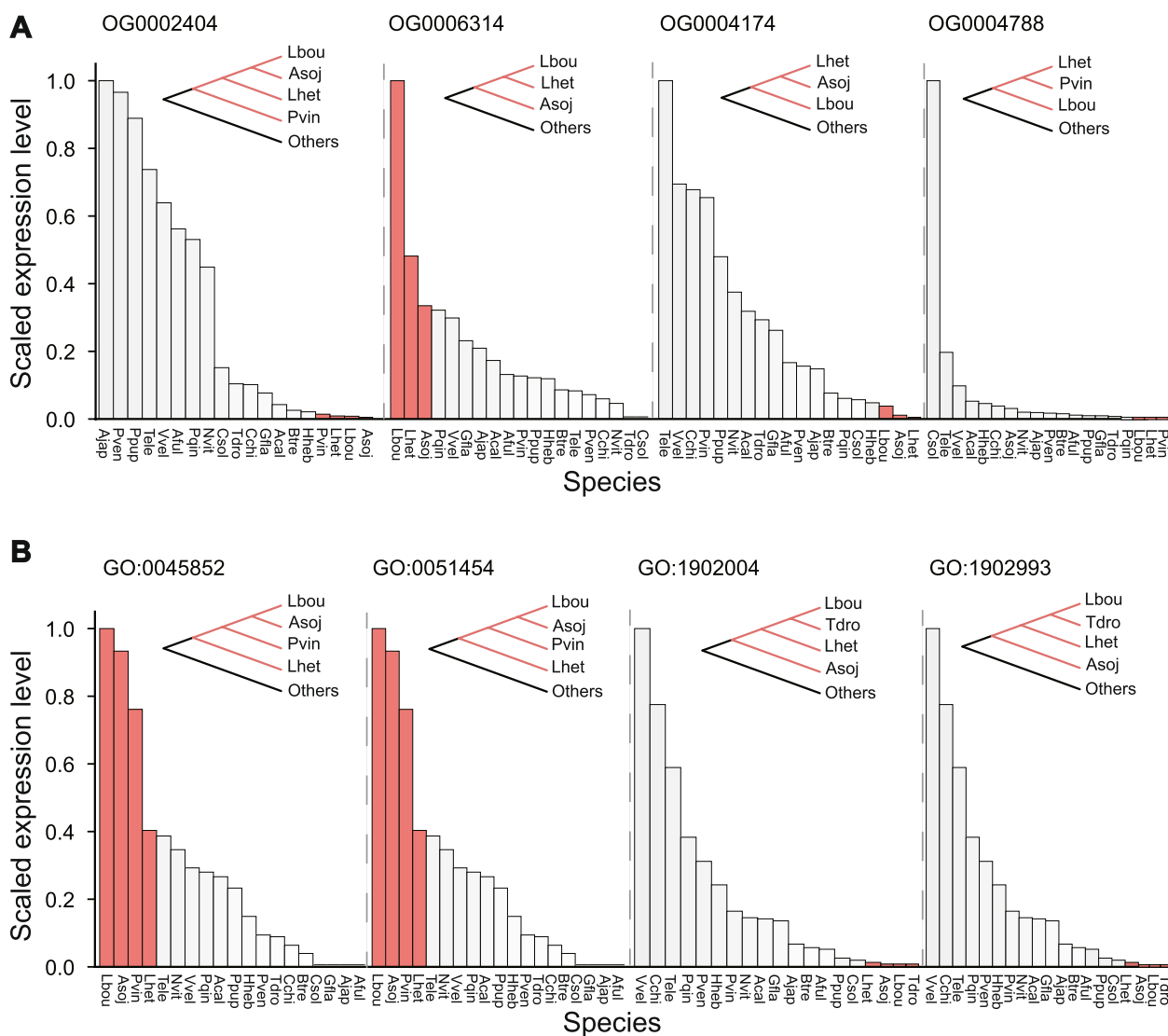


Fig. 3 Representative cases of expression trees showing monophyletic grouping in *Drosophila* parasitoid wasps. **A** Genes exhibiting a monophyletic grouping of four or three *Drosophila* parasitoid wasps in the expression tree. **B** GO terms exhibiting a monophyletic grouping of four *Drosophila* parasitoid wasps in the expression tree. The red branches in the tree indicate clustering of *Drosophila* parasitoid wasps. The bar plot illustrates the scaled expression levels of each gene or GO term across various species. These genes or GO terms demonstrate differential expression, either higher or lower, within the *Drosophila* monophyletic group compared to other species (Red). See Additional file 3 for individual data values

the expression tree showed a monophyletic group of all five *Drosophila* parasitoid wasps, and only four GO terms (pH elevation, intracellular pH elevation, positive regulation of amyloid-beta formation, and positive regulation of amyloid precursor protein catabolic process) where the expression tree showed a monophyletic group of the remaining four *Drosophila* parasitoid wasps, excluding *T. drosophilae* (Fig. 3B, Additional file 1: Table S4). Interestingly, we found 74 GO terms where the expression tree showed a monophyletic group of three of the five *Drosophila* parasitoid wasps, in which the clustering of three

larval stage *Drosophila* parasitoid wasps (*L. boulandi*, *L. heterotoma*, and *A. japonica*) was significantly higher than the expectation of 7.4 (16 out of 74, $P=0.002$, binomial test, Additional file 1: Table S4). To determine if the observation of 16 GO terms which clustering the three larval stage *Drosophila* parasitoid wasps was caused by chance, we enumerated all the 969 possible combinations of selecting three parasitoid wasps from the 19 hymenopteran species and counted the number of GO terms that supported each combination. We ranked each combinations based on the total number of supported GO terms

and found that 698 out of 969 combinations (72%) had a supported GO number ≤ 16 (Additional file 2: Fig. S7). This suggested that the three larval stage *Drosophila* parasitoid wasps exhibited a higher degree of clustering in the GO-level expression tree compared to other species. We next meticulously examined the expression patterns of each of the 16 GO terms and found that the three larval stage *Drosophila* parasitoid wasps showed either higher (14 GO terms) or lower (2 GO terms) expressions than other species (Fig. 3B). These GO categories primarily encompassed transporter, central nervous system myelination, and peptide hormone processing (Additional file 2: Fig. S8). Although we found a small number of GO terms where the expression tree showed a monophyletic grouping of the three larval stage parasitoid wasps, the majority of GO terms did not exhibit convergent signals in the *Drosophila* parasitoid wasps. Further analysis revealed that 4 of the 16 GO terms (GO:0016486: peptide hormone processing, GO:0022010: central nervous system myelination, GO:0032286: central nervous system myelin maintenance, and GO:0080132: fatty acid alpha-hydroxylase activity) contained genes previously identified as highly expressed in the venom gland of the three larval *Drosophila* parasitoid wasps. Only one GO term (GO:0016486: peptide hormone processing) contained a putative venom gene (*neprilysin*) identified in *A. japonica* (Additional file 1: Table S5). These findings suggest that the observed convergence in the expression of these 16 GO terms in the three larval *Drosophila* parasitoid wasps may be attributed to other conserved genes rather than actual venom compositions.

In addition to directly comparing the topology of the expression tree to the species tree for each orthogroup/GO term, we used Bayesian phylogenetic generalized linear mixed models (MCMCglmm models) to identify orthogroups/GO terms whose expression levels correlate with host species. Phylogenetic relationships and different genes within the same orthogroups or GO terms were treated as random effects in each model. Among the 1479 single copy genes, we identified 19 genes whose expression levels showed significant correlation with *Drosophila* parasitoid wasps (Additional file 1: Table S6). One of these genes (*Frizzled-2*) was previously identified as forming a monophyletic grouping of four *Drosophila* parasitoid wasps. However, only two genes (Pvin00505 and Pvin06853, see details in Additional file 1: Table S6) were found to be highly expressed in the venom gland of *P. vindeminae*, and none of these genes were identified as putative venom genes across the five *Drosophila* parasitoid wasps. For the 3124 shared orthogroups, we identified an additional 52 orthogroups where the expression

patterns significantly correlated with *Drosophila* parasitoid wasps (Additional file 1: Table S7). Among these, 14 orthogroups contained genes previously identified as highly expressed genes in the venom gland of *Drosophila* parasitoid wasps. Five orthogroups contained genes previously defined as putative venom genes in the *Drosophila* parasitoid wasps. However, none of these orthogroups contained genes highly expressed in the venom gland or putative venom genes across the five *Drosophila* parasitoid wasps. In the GO-level analysis, we identified 65 GO terms with expression levels significantly correlated with *Drosophila* parasitoid wasps (Additional file 1: Table S8). Among these, 23 GO terms contained genes previously identified as highly expressed in the venom gland of *Drosophila* parasitoid wasps, with 12 GO terms harboring putative venom genes in *Drosophila* parasitoid wasps. Nevertheless, only two GO terms (GO:0005471: ATP:ADP antiporter activity; GO:1990737: response to manganese-induced endoplasmic reticulum stress) contained highly expressed genes in the venom gland across all five *Drosophila* parasitoid wasps, and none of these GO terms contained putative venom genes across all five *Drosophila* parasitoid wasps.

Furthermore, we ran an additional MCMCglmm model for each GO term to detect convergent signals among the three larval *Drosophila* parasitoid wasps. We identified 37 GO terms whose expression levels were significantly associated with the three larval *Drosophila* parasitoid wasps, with six of these GO terms overlapping with the 16 GO terms identified using the topology comparison method (Additional file 1: Table S9, Additional file 1: Table S10). We observed that nine GO terms contained genes previously identified as highly expressed in the venom gland across the three larval *Drosophila* parasitoid wasps. Four GO terms contained putative venom genes specific to one or two larval *Drosophila* parasitoid wasps, but not across all three larval *Drosophila* parasitoid wasps.

Altogether, our analyses of expression levels, employing both topology comparison and MCMCglmm models across individual genes, orthogroups, and GO terms, reveal that a small number of orthogroups or GO terms exhibit similar expression patterns in the five *Drosophila* parasitoid wasps or the three larval *Drosophila* parasitoid wasps. However, most of these identified orthogroups or GO terms do not contain highly expressed genes or putative venom genes across the five *Drosophila* parasitoid wasps or the three larval *Drosophila* parasitoid wasps. This suggests that convergent evolution is not a major force shaping gene expression levels in the venom gland of *Drosophila* parasitoid wasps.

Stabilizing selection is the primary evolutionary force driving gene expression divergence in the venom gland of parasitoid wasps

It is widely accepted that venom is an adaptive trait, and the diversity in venom composition and function is subject to strong natural selection as a result of adaptation towards specific diets [29, 30]. However, our transcriptome evolutionary analyses did not detect convergent signals in the evolution of gene expression in the venom gland of parasitoid wasps that have independently adapting to *Drosophila* hosts. In contrast, our findings indicated a phylogenetic signal in the evolution of global gene expression in the venom gland of parasitoid wasps. This has prompted us to investigate the underlying evolutionary forces that driving gene expression variation in the venom gland of parasitoid wasps. More specifically, we aim to identify if expression variation in individual genes in the venom gland of parasitoid wasps is evolving through genetic drift, stabilizing selection or directional selection. To achieve this, we tested, for each 1479 one-to-one single copy gene separately, the model fit to three common models of trait evolution along the time-calibrated species tree including Brownian motion (BM), a single-optimum Ornstein–Uhlenbeck (OU) process or an OU model with branch shift (OUM). The BM model assumes that the expression variation between species accumulated over time due to divergence and evolutionary rate, often likened to genetic drift in evolutionary modeling. OU model introduces a selective regime that

pulls expression values towards an optimum, reflecting stabilizing selection. The OU model can be expanded to incorporate branch-specific events (OUM), resembling directional selection in specific lineages [31]. We found that 1115 out of the 1479 genes (75.3%) exhibited a better fit to the OU model compared to the other two models (Fig. 4A). This suggests that the majority of genes have evolved in accordance with the OU model, providing evidence that stabilizing selection has played a primary role in shaping the gene expression patterns in the venom gland of parasitoid wasps. However, we found that only 65 genes were identified to be putative venom protein-encoding genes, indicating the observation is likely driven by most genes not encoding venoms. Additionally, we conducted further analysis of our expression data by comparing the expected gene expression divergence under an OU model to the observed expression data. We employed a simulation approach where we generated expression levels for 10,000 genes along the known time-calibrated species phylogeny with a range of different α values under an OU model, respectively. Subsequently, we computed the mean squared expression distance for each pairwise species based on the simulated expression levels and compared them to the observed data. We found that our observed gene expression data had a close fit to the simulated data (Fig. 4B), further suggesting that the stabilizing selection is the primary evolutionary force driving gene expression divergence in the venom gland of parasitoid wasps.

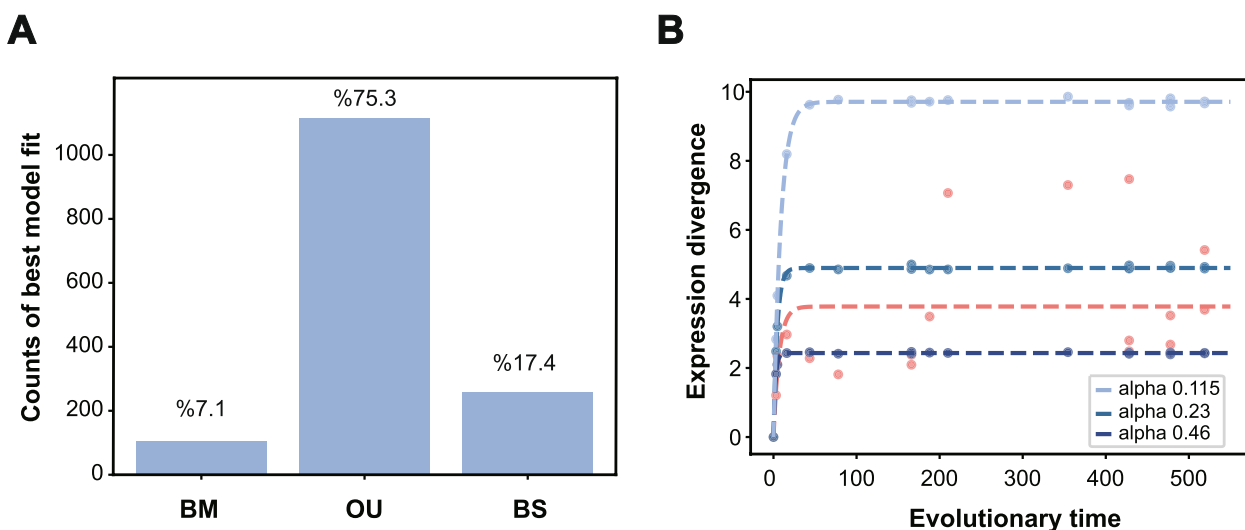


Fig. 4 Gene expression dynamics in the venom gland across hymenopteran species. **A** The number of per-gene expression patterns that have evolved under a BM, OU, or OUM model of trait evolution. See Additional file 3 for individual data values. **B** Pairwise mean squared distance between *P. puparum* and other species across evolutionary time. Red: observed gene expression divergences. Blue: expression divergence from 10,000 simulated trajectories under OU model with different values of α ($n = 10,000$). See Additional file 3 for individual data values

Discussion

Herein, we examined if convergent signal could be detected in the evolution of gene expression profiles of venom gland among the five *Drosophila* parasitoid wasps by analyzing the transcriptomes of venom gland from 19 hymenopteran species. Our results demonstrate that the venom transcriptomes of *Drosophila* parasitoid wasps do not exhibit higher similarity compared to other parasitoid wasps. We observed a correlation between the expression tree of venom glands and the phylogenetic tree of the species. Interestingly, we found that the three larval *Drosophila* parasitoid wasps formed a monophyletic group in the GO-level expression tree, indicating a potential convergent signal in the overall expression patterns of these three species. However, this convergence was not observed when focusing solely on the expression patterns of putative venom genes, suggesting that the observed convergence may be attributed to other genes rather than actual venom compositions. Further analysis at the individual gene, orthogroup, and GO-level expression data revealed that a small number of orthogroups/GO terms displayed convergent signals in the *Drosophila* parasitoid wasps. However, most of these identified orthogroups or GO terms do not encompass genes that are highly expressed in the venom gland or presumed to be venom genes in *Drosophila* parasitoid wasps. Overall, our findings provide evidence that independent adaptation to *Drosophila* hosts does not lead to convergence in the transcriptomes of venom gland in parasitoid wasps. In addition, our analysis of expression dynamics on each of 1479 single copy genes highlights that most of genes are evolved under an OU model, indicating that stabilizing selection pressures drive the evolution of gene expression in venom gland of parasitoid wasps. This is consistent with previous researches showing that stabilizing selection plays a major role in the evolution of gene expression across a diverse range of species, including *Drosophila*, African cichlid fishes, and mammals [32–34]. Stabilizing selection on gene expression plays a crucial role in maintaining the stability and functionality of gene expression networks, strengthening the expression levels of genes around the optimal range. However, it is worth noting that our detection of stabilizing selection was limited to the 1479 conserved one-to-one single copy genes across the 19 hymenopteran species, most of which do not encode venoms. Given the rapid evolution of venom genes in parasitoid wasps, further datasets and analyses are necessary to understand the evolution of venom gene expression patterns.

There are still several caveats that should be considered in our study. Firstly, venom genes in parasitoid wasps exhibit a rapid rate of evolution, accompanied by frequent gene gain and loss events [9, 17]. In the case of the

two *Anastatus* wasps that diverged approximately 3 million years ago, more than 50% of the venom genes have undergone significant changes [19]. Although our comparative transcriptome analysis considered the impact of gene gain and loss, i.e., analysis of orthologous gene families, we may still have missed some information within the context of rapid venom evolution, such as new venom genes originating from de novo genes or lateral gene transfer events [17, 19, 35]. Secondly, given the extremely rich biodiversity and highly variable host choice of parasitoid wasps, our study only represents the tip of the iceberg, and further in-depth studies and genome/transcriptome sequencing are needed to explore the full view of venom evolution in parasitoid wasps.

Conclusions

In this study, we employ parasitoid wasps that have independently adapted to *Drosophila* hosts as models to explore if the evolution of venom gland expression patterns is likewise a convergent evolution driven by their prey/host species. Through a phylogenetic-based comparative transcriptome study of the venom gland of 19 hymenopteran species including five *Drosophila* parasitoid wasps that represent four independent adaptations to the *Drosophila* hosts, we observed that the expression patterns among the venom gland of *Drosophila* parasitoid wasps do not achieve higher similarity compared to non-*Drosophila* parasitoid wasps. We detected limited convergent signals in the expression levels of individual genes, orthogroups, and GO terms. Our findings show that convergent evolution is not a major force shaping gene expression levels in the venom gland of *Drosophila* parasitoid wasps. Further analysis of gene expression dynamics based on single-copy genes highlights that stabilizing selection is the predominant force driving the evolution of gene expression in the venom gland of parasitoid wasps.

Methods

Data collection

We selected species for convergent transcriptome evolutionary analysis in *Drosophila* parasitoid wasps considering the following two criteria: (1) representing the main lineages of Hymenoptera (12) having a high-quality reference genome and at least one RNA-seq data of the venom gland. There are four *Drosophila* parasitoid wasps that meet the criteria including three larval parasitoid wasps (*L. bouleari* and *L. heterotoma* from Cynipoidea, *A. japonica* from Ichneumonoidea) and one pupal parasitoid wasp (*P. vindemmiae* from Chalcidoidea). For comparison, we sequenced the transcriptomes of venom gland of another pupal parasitoid wasp of *Drosophila* species (*T. drosophilae* from Proctotrupeoidea) and

included it in our selection of species though it lacks a reference genome. Finally, the datasets comprised 18 parasitoid wasps representing five main superfamilies of Hymenoptera and one killer wasp as outgroup species (Additional file 1: Table S1). These species contained five *Drosophila* parasitoid wasps representing four independent adaptations to the *Drosophila* hosts.

RNA-seq of venom gland of *T. drosophilae*

The parasitoid wasps *T. drosophilae* were reared on the pupa of the fruit flies (*Drosophila melanogaster*) under laboratory conditions at 25 ± 1 °C, $65 \pm 5\%$ relative humidity (RH), and a photoperiod of 14: 10 h (light:dark), as described in Yang et al. (2020) [27]. The venom glands (100 for each replicates) and carcass (i.e., female adult tissues excluding the venom gland) (20 for each replicates) of *T. drosophilae* (3 days old female adult) were dissected in cold PBS buffer and then washed three times. The total RNA was isolated using the TRIzol-based method (Invitrogen). Paired-end RNA-seq libraries (150 bp) for venom glands and carcasses were prepared using the TruSeq RNA Library Prep Kits (Illumina), then sequenced on the Illumina HiSeq X Ten platform. Three biological replicates were prepared for each sample. To get comprehensive reference sequences of *T. drosophilae* for downstream analysis, we assembled the generated RNA-seq data and all the publicly available RNA-seq data of *T. drosophilae* into unigenes by Trinity [36] v2.13.2 with default parameters (Additional file 1: Table S1). The open reading frames of each transcript were predicted by TransDecoder v5.7.0 (<https://transdecoder.github.io/>) with default parameters.

Orthology inference

A total of 395,994 genes from the 19 hymenopteran species were assigned into 26,787 Orthogroups using OrthoFinder [37] v2.5.4. Among these, 1208 orthogroups were single-copy orthologs across all 19 species except *T. drosophilae*. We further recovered 271 single-copy orthologs nested within large gene families with multiple homologs in one or more species by Orthosnap [38]. Because of the lacking genome information of *T. drosophilae*, we identified the best-hit genes of *T. drosophilae* in the single-copy orthogroups shared by other 18 species and obtained a total of 1479 one-to-one single copy orthogroups used for further analysis.

Phylogenetic tree

To reconstruct the species phylogenetic tree, we aligned the protein sequences of 1479 single-copy orthogroups from all 19 species by MAFFT [39] v7.48776 with the L-INS-I model, respectively. After filtering by trimAl [40] v1.2, we generated a super-gene sequence by

concatenating these sequences and constructed a maximum likelihood phylogenetic tree using IQ-Tree [41] v2.0 with 1000 replicates for ultrafast bootstrap analysis. The best-fit substitution model (JTT+F+I+G4) estimated by ModelFinder [42] was used. We utilized r8s [43] v1.81 to estimate the divergent times between the species. Four time points based on a previous study were used to calibrate the tree: Ichneumonoidea: 151–218 mya, Chalcidoidea: 105–159 mya, Aculeata: 160–224 mya, and Apocrita: 203–276 mya.

Expression matrix

The raw RNA-seq reads were processed by fastp [44] v0.23.2 for quality control and filtering, then aligned to each species' reference genome using STAR [45] v2.7.10a, respectively. For *T. drosophilae*, which lacks the reference genome, we mapped the RNA-seq reads to the unigenes assembled by Trinity [36] v2.13.2. The gene-level read counts for each gene were summarized by featureCounts [46] v2.0.3, then normalized into TPM values. Since the RNA-seq data were sampled from different species, and the TPM values are not directly comparable across species when genes have been gained and lost in the course of evolution, we used TPM10K to account for different sequencing depths among species, which has proven to be efficient for gene expression comparison across species [47]. The TPM10K was calculated following the formula: $TPM10K_i = TPM_i \times N / 10^4$, where N is the number of reference genes, and the mean of TPM10K values in each library is normalized to 10^2 . The TPM10K values were further log₂-transformed, and the median TPM10K values from biological replicates were calculated. Three expression matrices based on the TPM10K values were generated for transcriptome evolution analysis: (1) 1479 one-to-one orthologous genes identified previously, (2) OG-level expression matrix generated by summing the TPM10K values of genes in the same orthogroups, and (3) GO-level expression matrix generated by summing the TPM10K values of genes in the same GO terms. To ensure the robustness of our analysis on OG-level and GO-level expression datasets, phylogenetic analyses of venom gland transcriptome overall expression patterns were conducted using both mean and sum methods for each orthogroup. The results demonstrated consistent findings across methods, with the same topology observed regardless of the approach used (Additional file 2: Fig. S9).

Identification of highly expressed genes in the venom gland and putative venom genes

Given that the venom gland of parasitoid wasps is a highly specialized tissue which usually highly expressed a small number of genes [17, 19, 20]. In our analyzed 18

parasitoid wasps, on average, 4.3% of genes accounted for 80% of the total expressions of the venom gland. Consequently, we defined highly expressed genes as those comprising the top 80% of total gene expression in each venom gland transcriptome. We identified 13,470 highly expressed genes in the venom gland across the 19 species (on average, 708 highly expressed genes in the venom gland of each species). Notably, this dataset of highly expressed genes in the venom gland covered more than 80% of genes encoding venom proteins in 10 out of the 19 species, where venom proteins had been previously identified using a combination of transcriptomic and proteomic approaches [17, 19, 20, 22–28]. To further identify the putative venom genes for the remaining nine species, we considered both transcriptomic evidence and the presence of a signal peptide. Specifically, a putative venom gene was defined as one previously identified as highly expressed in the venom gland and containing a signal peptide predicted by SignalP v5.0 [48]. Notably, for *T. drosophilae*, where RNA-seq data were generated both for the venom gland and carcass (i.e., female adult tissues excluding the venom gland), we applied additional criteria: genes highly expressed in the venom gland compared to the carcass (\log_2 -fold change ≥ 2 and adjusted $P \leq 0.05$). In total, we identified 2591 putative venom genes across the 19 species. We then assigned orthogroups/GO terms as venom highly expressed gene-containing orthogroups/GO terms (VHEOGs/VHEGOs) or venom gene-containing orthogroups/GO terms (VCOGs/VCGOs) that contained highly expressed genes or putative venom genes in at least one species. To focus on the expression patterns of these venom-associated genes, we filtered our previously identified 1479 single-copy genes, 3614 shared orthogroups, and 16,862 shared GO terms to include only VHEOGs/VHEGOs or VCOGs/VCGOs. Subsequently, we generated six expression matrices based on the TPM10K values: (1) 560 single-copy VHEOGs, (2) 1202 shared VHEOGs, (3) 13,880 VHEGOs, (4) 106 single-copy VCOGs, (5) 284 shared VCOGs, and (6) 7563 VCGOs. We further narrowed down our datasets to include only venom highly expressed genes and putative venom genes of the five *Drosophila* parasitoid wasps. This resulted in 134 single-copy VHEOGs, 326 VHEOGs, 8,235 VHEGOs, 26 single-copy VCOGs, 74 VCOGs, and 3914 VCGOs.

Transcriptome similarity analysis

To explore the overall expression patterns of the venom glands among the 19 hymenopteran species, we performed PCA using R package PCAtools (<https://github.com/kevinblighe/PCAtools>) based on the expression matrix obtained previously. For the PCA on the genome sequences, we first employed the one hot encoding to

convert the protein sequence data of the 1479 one-to-one single copy genes used for species phylogeny inference into numerical values using the OneHotEncoder function from the sklearn module (<https://scikit-learn.org/stable/about.html#citing-scikit-learn>) in Python. Subsequently, we performed PCA using the PCAtools package in R. The correlation between each PCs with the trait data were calculated with eigencorplot function in PCAtools package. The Spearman rank correlation coefficients between venom gland transcriptomes were calculated by cor function in R and were used to quantify the transcriptome similarity. The distance between *Drosophila* parasitoids and non-*Drosophila* parasitoids were compared with the Wilcoxon rank sum test.

Expression tree

The neighbor-joining expression tree of the venom gland was constructed using the ape [49] package in R based on the 1-Spearman rank correlation coefficient distance matrix. We utilized the boot.phylo function from the ape package to perform 10,000 replications for bootstrap analysis. The random tree of the 19 hymenopteran species was generated by the rtree function from ape package in R. To compare the topology of the species tree and the expression tree, we used the cophylo function in the phytools [50] to generate the co-phylogenetic plot. The PH85 distance between the topologies of two phylogenetic trees was calculated using the dist.topo function from the ape package in R. The pairwise distance matrix of the species phylogenetic tree and the expression tree was obtained by ETE Toolkit [51] and tested for correlation between phylogenetic and expression matrices with Mantel tests using the mantel library in Python.

Phylogenetic generalized linear mixed models of gene expression and *Drosophila* parasitoid wasps

For each of the 1479 single-copy genes, 3664 shared Gene Ontology (GO) terms, and 16,862 shared orthogroups, we constructed MCMCglmm models to assess the correlation of gene expression in the venom gland with *Drosophila* parasitoid wasps. We categorized all species into *Drosophila* parasitoid wasps and non-*Drosophila* parasitoid wasps. Phylogenetic relationships and other factors, such as different genes within the same OG/GO terms, were treated as random effects. In all models, the expression values of genes from the 19 species were fitted as a Gaussian response variable, with the categorization of *Drosophila* parasitoid wasps or non-*Drosophila* parasitoid wasps as predictor variables. A prior of covariance V of 1.00 and a degree of belief parameter (ν) of 0.002 were applied in each model. Each model ran for 1,000,000 iterations with a burn-in of 100,000 iterations and a thinning of

500 iterations using the MCMCglmm package [52]. Model convergence was assessed using the geweke.diag function in coda package, considering convergence achieved when the absolute value of the returned Z-score was less than 1.96 [53]. For the analysis of each GO term, where we previously identified whole-gene expression convergence signals in the transcriptomes of venom gland among the three larval *Drosophila* parasitoid wasps, we ran an additional MCMCmodel for each GO terms to divide species into larval *Drosophila* parasitoid wasps and non-larval *Drosophila* parasitoid wasps using the same parameters. We determined the cutoff score of the posterior mean for each model using previously reported methods [54]. Briefly, we employed fitdistrplus [55] to fit the posterior mean to a high-probability distribution (e.g., normal or logistic), then calculated statistics such as Z-scores for the best-fit distribution as the cutoff. Genes were considered significantly correlated with *Drosophila* parasitoid wasps if pMCMC was less than 0.05 and the absolute value of the posterior mean exceeded the cut score [54].

Gene expression dynamics

To investigate the underlying forces acting on gene expression evolution, we employed three evolutionary models to analyze the expression data of 1479 single-copy genes. These models included the Brownian motion model (BM), a single-optimum Ornstein–Uhlenbeck (OU) process, and an OU model with branch shift (OUM), implemented in the OUwie package (<http://thej022214.github.io/OUwie>). The models were compared by calculating the difference in the Akaike information criterion, and the best-fit model was determined for each gene. Additionally, we conducted simulations by generating 10,000 sets of simulated gene expression data under the OU model with varying α values (0.23, which was the mean α estimated by the single-optimum OU process, as well as 0.46 and 0.115, representing twice and half of the mean α value, respectively) using the OUwie package. We next compared the transcriptome divergence by calculating the mean squared distance between the observed data and simulated data.

Abbreviations

TPM	Transcripts per million
PCA	Principal component analysis
NJ	Neighbor-joining
GO	Gene Ontology
VHEOGs	Venom highly expressed gene-containing orthogroups
VHEGOs	Venom highly expressed gene-containing GO terms
VCOGs	Venom gene-containing orthogroups
VCGOs	Venom gene-containing GO terms

BM	Brownian motion
OU	Ornstein–Uhlenbeck
OUM	OU model with branch shift
RH	Relative humidity

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-01974-2>.

Additional file 1: Tables S1–S10. TableS1 - List of species and data resources selected for venom gland transcriptome analysis used in this study. TableS2 - Gene information for which the expression tree showed a monophyly of four or three *Drosophila* parasitoid wasps. TableS3 - OG information for which the expression tree showed a monophyly of four or three *Drosophila* parasitoid wasps. TableS4 - GO information for which the expression tree showed a monophyly of four or three *Drosophila* parasitoid wasps. TableS5 - GO terms contained genes highly expressed in the venom gland or putative venom genes in *Drosophila* parasitoid wasps. TableS6- MCMCglmm models identified genes whose expression levels were significantly correlated with *Drosophila* parasitoid wasps. TableS7 - MCMCglmm models identified orthogroups whose expression levels were significantly correlated with *Drosophila* parasitoid wasps. TableS8 - MCM-Cglmm models identified GO terms whose expression levels were significantly correlated with *Drosophila* parasitoid wasps. TableS9- MCMCglmm models identified GO terms whose expression levels were significantly correlated with the three larval *Drosophila* parasitoid wasps. TableS10 - GO information whose expression levels showed convergent signals in the three larval *Drosophila* parasitoid wasps identified using both topology comparison method and MCMCglmm models.

Additional file 2: Figures S1–S9. FigS1 - Principal component analysis on expression data based on the 1,479 one-to-one single copy genes. FigS2 - Neighbor-joiningtrees constructed using different expression data from the venom gland. FigS3 - Neighbor-joiningtrees constructed using expression data from the VHEOGs across the 19 species. FigS4 - Neighbor-joiningtrees constructed using expression data from the VCOGs across the 19 species. FigS5 - Neighbor-joiningtrees constructed using expression data from the VHEGOs across the five *Drosophila* parasitoid wasps. FigS6 - Neighbor-joiningtrees constructed using expression data from the VCOGs across the five *Drosophila* parasitoid wasps. FigS7 - Number of GO terms for which the expression tree supports the monophyly of each of the 969 possible combinations of selecting three parasitoid wasps from the 19 hymenopteran species. FigS8 - Gene Ontologycategories where the expression tree showed a monophyly group of the three larval stage *Drosophila* parasitoid wasps. FigS9 - Neighbor-Joiningtrees constructed using three different methods for generating OG-level expression matrices.

Additional file 3. Individual data point values for Figures

Acknowledgements

We thank Dr. Ziwen Teng at Qingdao Agricultural University for providing *Trichopria drosophilae* samples. We thank Dr. Ryusuke Niwa at University of Tsukuba for sharing the genome and transcriptome data of *Asobara japonica*.

Authors' contributions

X.H.Y., G.Y.Y., and Y.Y. conceived and designed the project. Y.Y., X.H.Y., and S.X. performed the experiments and analyzed the data. Y.Y. wrote the original draft of the paper and led the revisions, and X.H.Y., G.Y.Y., X.X.Z., L.J.F., Q.F., and Y.H.S. critically reviewed several versions of the manuscript. All authors read and approved the manuscript.

Funding

This work was supported by the Key Program of National Natural Science Foundation of China (NSFC) (grant no. 32330085 to G.Y.Y.), the Program of NSFC (Grant no. 32202376 to X.H.Y., no. 32302408 to Y.Y.), the Young Elite Scientists Sponsorship Program by China Association for Science and Technology (grant no. 2022QNRC001 to X.H.Y.), and the China Postdoctoral Science Foundation (grant no. 2021M700125 to X.H.Y., no. 2022M722803 to Y.Y.).

Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files and publicly available repositories. The raw sequencing data generated in this study have been deposited to the Genome Sequence Archive (GSA: CRA016166) [56] in National Genomics Data Center (<https://ngdc.cncb.ac.cn/>). The related datasets and codes are available on GitHub (<https://github.com/yangyi09/Comparative-transcriptome-analysis>) and archived at Zenodo (<https://doi.org/10.5281/zenodo.13073502>) [57]. All other genomic resources and transcriptomic datasets analyzed in this study were obtained from previous studies [17–20, 22–28, 58–64] or downloaded from NCBI (<https://www.ncbi.nlm.nih.gov>) [65–72], InsectBase (<http://v2.insect-genome.com/>) [73] and iVenomDB (<http://www.waspomics.com/venom/index.php>) [74]. The individual data point values are listed in Additional file 3.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹State Key Laboratory of Rice Biology and Breeding & Ministry of Agricultural and Rural Affairs Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Institute of Insect Sciences, Zhejiang University, Hangzhou, China. ²Department of Biology, University of Rochester, Rochester, NY, USA. ³Institute of Bioinformatics & Institute of Crop Science, Zhejiang University, Hangzhou, China. ⁴College of Advanced Agriculture Science, Zhejiang A&F University, Hangzhou, China. ⁵Shanghai Institute for Advanced Study, Zhejiang University, Shanghai, China. ⁶College of Computer Science and Technology, Zhejiang University, Hangzhou, China.

Received: 19 February 2024 Accepted: 5 August 2024

Published online: 15 August 2024

References

- Losos JB. *Improbable destinies: how predictable is evolution?* 1st ed. London: Riverhead Books; 2017.
- Stern DB, Crandall KA. The Evolution of gene expression underlying vision loss in cave animals. *Mol Biol Evol.* 2018;35:2005–14.
- Xu H, Ye X, Yang Y, Yang Y, Sun YH, Mei Y, et al. Comparative genomics sheds light on the convergent evolution of miniaturized wasps. *Mol Biol Evol.* 2021;38:5539–54.
- Wirthlin ME, Schmid TA, Elie JE, Zhang X, Kowalczyk A, Redlich R, et al. Vocal learning-associated convergent evolution in mammalian proteins and regulatory elements. *Science.* 2024;383:eabn3263.
- Schendel V, Rash LD, Jenner RA, Undheim EAB. The diversity of venom: the importance of behavior and venom system morphology in understanding its ecology and evolution. *Toxins.* 2019;11:666.
- Zancollì G, Casewell NR. Venom systems as models for studying the origin and regulation of evolutionary novelties. *Mol Biol Evol.* 2020;37:2777–90.
- Zancollì G, Reijnders M, Waterhouse RM, Robinson-Rechavi M. Convergent evolution of venom gland transcriptomes across Metazoa. *Proc Natl Acad Sci.* 2022;119:e2111392119.
- Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol Evol.* 2013;28:219–29.
- Mrinalini, Werren JH. Parasitoid wasps and their venoms. In: Gopalakrishnakone P, Malhotra A, editors. *Evolution of venomous animals and their toxins.* Dordrecht: Springer Netherlands; 2015. p. 1–26.
- Fang Q, Wang L, Zhu Y, Stanley DW, Chen X, Hu C, et al. *Pteromalus puparum* venom impairs host cellular immune responses by decreasing expression of its scavenger receptor gene. *Insect Biochem Mol Biol.* 2011;41:852–62.
- Moreau SJM, Asgari S. Venom proteins from parasitoid wasps and their biological functions. *Toxins.* 2015;7:2385–412.
- Yan Z, Fang Q, Liu Y, Xiao S, Yang L, Wang F, et al. A venom serpin splicing isoform of the endoparasitoid wasp *Pteromalus puparum* suppresses host prophenoloxidase cascade by forming complexes with host hemolymph proteinases. *J Biol Chem.* 2017;292:1038–51.
- Arvidson R, Kaiser M, Lee SS, Urenda J-P, Dail C, Mohammed H, et al. Parasitoid jewel wasp mounts multipronged neurochemical attack to hijack a host brain. *Mol Cell Proteomics.* 2019;18:99–114.
- Wang J, Jin H, Schlenke T, Yang Y, Wang F, Yao H, et al. Lipidomics reveals how the endoparasitoid wasp *Pteromalus puparum* manipulates host energy stores for its young. *Biochim Biophys Acta Mol Cell Biol Lipids.* 2020;1865:158736.
- Cavigliasso F, Mathé-Hubert H, Kremmer L, Rebuf C, Gatti J-L, Malausa T, et al. Rapid and differential evolution of the venom composition of a parasitoid wasp depending on the host strain. *Toxins.* 2019;11:629.
- Cavigliasso F, Mathé-Hubert H, Gatti J-L, Colinet D, Poirié M. Parasitic success and venom composition evolve upon specialization of parasitoid wasps to different host species. *Peer Community J.* 2021;1:e51. <https://doi.org/10.24072/pcjournal.28>.
- Martinson EO, Mrinalini, Kelkar YD, Chang C-H, Werren JH. The evolution of venom by co-option of single-copy genes. *Curr Biol.* 2017;27:2007–2013.e8.
- Yang Y, Ye X, Dang C, Cao Y, Hong R, Sun YH, et al. Genome of the pincer wasp *Gonatopus flavifemur* reveals unique venom evolution and a dual adaptation to parasitism and predation. *BMC Biol.* 2021;19:145.
- Ye X, Yang Y, Zhao C, Xiao S, Sun YH, He C, et al. Genomic signatures associated with maintenance of genome stability and venom turnover in two parasitoid wasps. *Nat Commun.* 2022;13:6417.
- Ye X, He C, Yang Y, Sun YH, Xiong S, Chan KC, et al. Comprehensive isoform-level analysis reveals the contribution of alternative isoforms to venom evolution and repertoire diversity. *Genome Res.* 2023;33:1554–67.
- Yang J-R, Maclean CJ, Park C, Zhao H, Zhang J. Intra and interspecific variations of gene expression levels in yeast are largely neutral: (Nei Lecture, SMBE 2016, Gold Coast). *Mol Biol Evol.* 2017;34:2125–39.
- Xiao S, Ye X, Wang S, Yang Y, Fang Q, Wang F, et al. Genome assembly of the ectoparasitoid wasp *Theocolax elegans*. *Sci Data.* 2023;10:159.
- Yu K, Chen J, Bai X, Xiong S, Ye X, Yang Y, et al. Multi-omic identification of venom proteins collected from artificial hosts of a parasitoid wasp. *Toxins.* 2023;15:377.
- Teng Z-W, Xiong S-J, Xu G, Gan S-Y, Chen X, Stanley D, et al. Protein discovery: combined transcriptomic and proteomic analyses of venom from the endoparasitoid *Cotesia chilonis* (Hymenoptera: Braconidae). *Toxins.* 2017;9:E135.
- Yan Z, Fang Q, Wang L, Liu J, Zhu Y, Wang F, et al. Insights into the venom composition and evolution of an endoparasitoid wasp by combining proteomic and transcriptomic analyses. *Sci Rep.* 2016;6:19604.
- de Graaf DC, Aerts M, Brunain M, Desjardins CA, Jacobs FJ, Werren JH, et al. Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies. *Insect Mol Biol.* 2010;19(Suppl 1):11–26.
- Yang L, Yang Y, Liu M-M, Yan Z-C, Qiu L-M, Fang Q, et al. Identification and comparative analysis of venom proteins in a pupal ectoparasitoid *Pachycrepoideus vindemmia*. *Front in Physiol.* 2020;11:9.
- Goecks J, Mortimer NT, Mobley JA, Bowersock GJ, Taylor J, Schlenke TA. Integrative approach reveals composition of endoparasitoid wasp venoms. *PLoS One.* 2013;8:e64125.
- Dalry JC, Wüster W, Thorpe RS. Diet and snake venom evolution. *Nature.* 1996;379:537–40.
- Barlow A, Pook CE, Harrison RA, Wüster W. Coevolution of diet and prey-specific venom activity supports the role of selection in snake venom evolution. *Proc Royal Soc B.* 2009;276:2443–9.
- Price PD, Palmer Drogue DH, Taylor JA, Kim DW, Place ES, Rogers TF, et al. Detecting signatures of selection on gene expression. *Nat Ecol Evol.* 2022;6:1035–45.
- Lemos B, Meiklejohn CD, Cáceres M, Hartl DL. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution.* 2005;59:126–37.

33. Bedford T, Hartl DL. Optimization of gene expression by natural selection. *Proc Natl Acad Sci*. 2009;106:1133–8.
34. Chen J, Swofford R, Johnson J, Cummings BB, Rogel N, Lindblad-Toh K, et al. A quantitative framework for characterizing the evolutionary history of mammalian gene expression. *Genome Res*. 2019;29:53–63.
35. Martinson EO, Martinson VG, Edwards R, Mrinalini, Werren JH. Laterally transferred gene recruited as a venom in parasitoid wasps. *Mol Biol Evol*. 2016;33:1042–52.
36. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52.
37. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:238.
38. Steenwyk JL, Goltz DC, Iii TJB, Li Y, Shen X-X, Rokas A. OrthoSNAP: a tree splitting and pruning algorithm for retrieving single-copy orthologs from gene family trees. *PLoS Biol*. 2022;20:e3001827.
39. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
40. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3.
41. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37:1530–4.
42. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587–9.
43. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*. 2003;19:301–2.
44. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–90.
45. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
46. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30.
47. Munro C, Zapata F, Howison M, Siebert S, Dunn CW. Evolution of gene expression across species and specialized zooids in Siphonophora. *Mol Biol Evol*. 2022;39:msac027.
48. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019;37:420–3.
49. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019;35:526–8.
50. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 2012;3:217–23.
51. Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol*. 2016;33:1635–8.
52. Hadfield JD. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J Stat Softw*. 2010;33:1–22.
53. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. *R News*. 2006;6:7–11.
54. Zhu P, Liu W, Zhang X, Li M, Liu G, Yu Y, et al. Correlated evolution of social organization and lifespan in mammals. *Nat Commun*. 2023;14:372.
55. Delignette-Muller ML, Dutang C. fitdistrplus: an R package for fitting distributions. *J Stat Softw*. 2015;64:1–34.
56. Yang Y. RNA-seq of the venom gland and carcass of *Trichopria drosophilae*. *Genome Sequence Archive*. 2024. <https://ngdc.cnca.ac.cn/gsa/browse/CRA016166>.
57. Yang Y. Codes and datasets for the comparative transcriptome analysis of venom glands in parasitoid wasps. *Zenodo*. 2024. <https://doi.org/10.5281/zenodo.13073502>
58. Ye X, Yang Y, Tian Z, Xu L, Yu K, Xiao S, et al. A high-quality de novo genome assembly from a single parasitoid wasp. 2020. Preprint at <https://doi.org/10.1101/2020.07.13.200725>.
59. Ye X, Xiong S, Teng Z, Yang Y, Wang J, Yu K, et al. Genome of the parasitoid wasp *Cotesia chilonis* sheds light on amino acid resource exploitation. *BMC Biol*. 2022;20:118.
60. Kamiyama T, Shimada-Niwa Y, Tanaka H, Katayama M, Kuwabara T, Mori H, et al. Whole-genome sequencing analysis and protocol for RNA interference of the endoparasitoid wasp *Asobara japonica*. *DNA Res*. 2022;29:dsac019.
61. Chen J, Fang G, Pang L, Sheng Y, Zhang Q, Zhou Y, et al. Neofunctionalization of an ancient domain allows parasites to avoid intraspecific competition by manipulating host behaviour. *Nat Commun*. 2021;12:5489.
62. Huang J, Chen J, Fang G, Pang L, Zhou S, Zhou Y, et al. Two novel venom proteins underlie divergent parasitic strategies between a generalist and a specialist parasite. *Nat Commun*. 2021;12:234.
63. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, et al. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science*. 2010;327:343–8.
64. Dalla Benetta E, Antoshechkin I, Yang T, Nguyen HQM, Ferree PM, Akbari OS. Genome elimination mediated by gene expression from a selfish chromosome. *Sci Adv*. 2020;6:eaaz9808.
65. Shimada-Niwa Y, Kamiyama T, Tanaka H, Toyoda A, Ito T. *Asobara japonica* strain Tokyo isolate Genome #3, whole genome shotgun sequencing project. *GenBank*. 2022. <https://identifiers.org/ncbi/insdc:BQM100000000>.
66. Ye XH, Fang Q, Ye GY, Li F. *Cotesia chilonis* isolate zju, whole genome shotgun sequencing project. *GenBank*. 2021. <https://identifiers.org/ncbi/insdc:RJVT010000000>.
67. Yang Y. *Gonatopus flavifemur* isolate zju, whole genome shotgun sequencing project. *GenBank*. 2021. <https://identifiers.org/ncbi/insdc:JAFJZ010000000>.
68. Fang G, Chen J. *Leptopilina bouvardi* isolate dan_hultmark-lab, whole genome shotgun sequencing project. *GenBank*. 2021. <https://identifiers.org/ncbi/insdc:JADEYJ010000000>.
69. Dong Z. *Leptopilina heterotoma* isolate ZJUHH_Lhet, whole genome shotgun sequencing project. *GenBank*. 2023. <https://identifiers.org/ncbi/insdc:JAVJHQ000000000>.
70. Antoshechkin I, Benetta ED, Yang T, Nguyen HQM, Ferree PM, Akbari OS. *Nasonia vitripennis* strain AsymCx, whole genome shotgun sequencing project. *GenBank*. 2020. <https://identifiers.org/ncbi/insdc:WELF000000000>.
71. Xiao S. *Theocolax elegans* breed Tele_zju isolate zju2018, whole genome shotgun sequencing project. *GenBank*. 2022. <https://identifiers.org/ncbi/insdc:JANSJX000000000>.
72. Ye XH, Yang Y, He C, Ye GY. *Pteromalus puparum* isolate zju_2012, whole genome shotgun sequencing project. *GenBank*. 2023. <https://identifiers.org/ncbi/insdc:VCDM000000000>.
73. Mei Y, Jing D, Tang S, Chen X, Chen H, Duanmu H, et al. InsectBase 2.0: a comprehensive gene resource for insects. *Nucleic Acids Res*. 2022;50:D1040–5.
74. Chen L, Lang K, Zhang B, Shi J, Ye X, Stanley DW, et al. iVenomDB: a manually curated database for insect venom proteins. *Insect Sci*. 2023;30:264–6.

Publisher's Note

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