scientific reports

Check for updates

OPEN

Decoding bee cleptoparasitism through comparative transcriptomics of *Coelioxoides waltheriae* and its host *Tetrapedia diversipes*

Paulo Cseri Ricardo[1]✉, Maria Cristina Arias[1] & Natalia de Souza Araujo[2]

Cleptoparasitism, also known as brood parasitism, is a widespread strategy among bee species in which the parasite lays eggs into the nests of the host species. Even though this behavior has significant ecological implications for the dynamics of several species, little is known about the molecular pathways associated with cleptoparasitism. To shed some light on this issue, we used gene expression data to perform a comparative analysis between two solitary neotropical bees: *Coelioxoides waltheriae*, an obligate parasite, and their specific host *Tetrapedia diversipes*. We found that ortholog genes involved in signal transduction, sensory perception, learning, and memory formation were differentially expressed between the cleptoparasite and the host. We hypothesize that these genes and their associated molecular pathways are engaged in cleptoparasitism-related processes and, hence, are appealing subjects for further investigation into functional and evolutionary aspects of cleptoparasitism in bees.

Keywords Comparative transcriptomics, Brood parasitism, Cuckoo bee, Solitary bees

Parasitism is an interaction between different species, in which one of them, the parasite, benefits at the expense of another, the host[1,2]. Parasitic species may play an important role in the dynamics of natural populations of host species, for instance, they may affect the susceptibility of their hosts to predation, modify their reproductive patterns, and influence the abundance of endemic and introduced species[3,4]. Cleptoparasitism, alternatively known as brood parasitism, refers to a parasitic behavior where the parasite lays eggs into the nests of the host species. Then during its larval stage, the parasite offspring thrives by consuming the food resources that have been provided by the adult host, ultimately leading to the demise of the host larva or egg, as it is either killed or eaten by the parasite larva. Finally, an adult parasite emerges from the host nest[5].

Cleptoparasitism is widespread in bees (Hymenoptera: Antophila). It is estimated that approximately 13% of the 20,500 bee species in the world[6] are obligate cleptoparasites[7]. Currently, it has been inferred that this behavior has arisen independently 18 times in four out of nine bee families: three distinct times in Apidae[8–10]; probably five times in Megachilidae[11–13]; at least once in Colletidae[14]; and possibly nine in Halictidae[5,15]. In spite of these multiple and independent origins, most cleptoparasitic bees show important convergent adaptations such as the reduction or complete loss of pollen-collecting (e.g., pollen-manipulating brushes and scopa) and nest-building structures (e.g., basitibial and pygidial plates for ground-nesting species)[5]. Compared to non-parasitic species, many cleptoparasites also have a heavily sclerotized cuticle in addition to spines, ridges, crests, carinae or lamellae protecting them from the jaws or sting of host females[5,7]. Moreover, convergent anatomical and physiological changes in the reproductive system of some cleptoparasitic species have been described, such as a greater number of mature oocytes in the ovaries or more ovarioles per ovary[16]. These adaptations allow parasitic females to lay several eggs in a short period of time[5,7].

In addition to their evolutionary relevance, cleptoparasitic species also have an ecological value. Cleptoparasitic bees are considered by far the largest protagonist of solitary bee brood mortality among all natural enemies[17]. In this context, Sheffield et al.[18] suggest that cleptoparasites, especially generalist ones, perform a stabilizing

[1]Departamento de Genética e Biologia Evolutiva – Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil. [2]Unit of Evolutionary Biology & Ecology – Université Libre de Bruxelles, Brussels, Belgium. ✉email: cseri.bio@gmail.com

role in bee communities by attacking dominant and abundant host taxa, which may reduce competition among non-parasitic bee species. These authors also argue that cleptoparasites can serve as indicator taxa for assessing the status of the entire bee community, as their diversity and abundance are closely tied to their host species[18]. However, there is still a deep lack of knowledge regarding the general biology, behavior[7], and particularly the molecular mechanisms underlying the evolution and maintenance of cleptoparasitism. This knowledge gap might be driven by the rarity to find cleptoparasitic bees in nature[7,19].

The abundance of cleptoparasitic bees often exhibits a positive correlation with the density of available host nests[18,20], implying that hosts displaying high levels of gregarious behavior are more susceptible to parasitic attacks. In this context, the species *Coelioxoides waltheriae* Ducke, 1908 and *Tetrapedia diversipes* Klug, 1810 comprise a compelling parasite-host species pair to investigate cleptoparasitism. *T. diversipes*, the host species, has a gregarious behavior, building its nests in naturally pre-existing cavities such as holes in wood[21], and trap-nests. Indeed, *T. diversipes* easy aggregatory behavior in trap-nests allowed the description of the cleptoparasitic behavior of *C. waltheriae*[21], the first recorded for the *Coelioxoides* genus. *Coelioxoides* and *Tetrapedia* are endemic to the Neotropical region[22]. Both genera have been formerly grouped within the same tribe, Tetrapediini (Apidae), as they share a number of similar morphological traits[23]. However, recent molecular-based phylogenies have placed *Coelioxoides* within a large cleptoparasitic clade (Nomadinae: Apidae) not sister to *Tetrapedia*[8–10,24,25]. *C. waltheriae* is considered to be the main natural enemy of *T. diversipes*[26] even though it has been also reported to parasitize nests of other *Tetrapedia* species[27,28].

Recent studies have successfully employed comparative transcriptomic analysis to unveil molecular features related to complex behaviors. For instance, transcriptomics studies in bees have helped to better understand the molecular pathways related to the honeybee (waggle) dance[29], *Varroa* sensitive hygiene behaviour[30], olfactory[31] and visual[32] learning, diapause[33,34], and evolution of task division[35–37]. In this context, herein we use gene expression data to perform a comparative analysis between the correspondent reproductive life stages of *C. waltheriae* and its host, *T. diversipes*. We aimed at identifying diverging gene expression patterns between the solitary pollen-collecting host and the cleptoparasitic bee species. We explored the function of enriched diverging pathways and discussed whether these could be related at some level to their ecological interactions. Instead of assuming any of the differences here observed are causative, we focused on relating our findings to molecular processes for which ecological and/or behavioural relevance has been previously described in the literature to shed some light on the molecular mechanisms potentially related to cleptoparasitic behavior.

Results

Transcriptome assembly and annotation

We sampled reproductive females of *C. waltheriae* and *T. diversipes* from the same trap-nest location and at correspondent life stages. In total, we used five sample replicates per species, consisting of a pool of whole-body extractions from individuals (see "Materials and methods").

After transcriptome assemblies and data treatment (see "Materials and methods") we obtained a set of 18,208 transcripts for *C. waltheriae* and 11,998 for *T. diversipes*. The main quality parameters associated with each assembly are summarized in Table 1.

A total of 10,042 (51.2%) and 7447 (61.9%) transcripts in *C. waltheriae* and *T. diversipes,* respectively, blasted (*e*-value < 1e−5) against protein-coding genes in the UniRef90 database. Moreover, 6253 (31.8%) and 2919 (24.2%) transcripts were estimated to be potential non-coding sequences in *C. waltheriae* and *T. diversipes*. It is worth mentioning that 1397 (7.1%) transcripts were identified as potential contaminants in *C. waltheriae* (Supplementary Material 1). Most were attributed to fungi of the family Tubulinosematidae (Microsporidia): *Tubulinosema ratisbonensis* (1113 transcripts), *Anncaliia algerae* (15 transcripts), and Tubulinosematidae itself (10 transcripts). In addition, 95 (0.48%) of *C. waltheriae* transcripts were identified as very similar to plant protein sequences. Overall, the Fabaceae family was the most representative plants among contaminants (52

Metric	*C. waltheriae*	*T. diversipes*
Total transcripts assembled	87,358	17,866[a]
		40,796[b]
Number of clustered SuperTranscripts[c]	18,208	11,998
SuperTranscripts ≥ 1000 nt	10,195	8051
p fragments mapped[d]	0.93176	0.86625
p bases uncovered[e]	0.0328	0.05047
Transrate optimal score[f]	0.5904	0.50711
Complete BUSCOs (%)	93.4	88.5
Missing BUSCOs (%)	4.1	8.2

Table 1. Main quality parameters of *C. waltheriae* and *T. diversipes* assemblies. [a]Number of transcripts obtained after the reference-based assembly. [b]Number of transcripts obtained after the reference guided de novo assembly. [c]After removing contaminating sequences. [d]Proportion of reads pairs that mapped successfully. [e]Proportion of bases that were not covered by any read. [f]This discrete score represents the accuracy and completeness of the assembly based on the assessment of alignments between contigs and reads. The minimum score is 0, while the maximum is 1.0. For a description of the components used in the calculation of this metric, see Smith-Unna et al.[38].

transcripts), followed by Asteraceae (11 transcripts) and Solanaceae (6 transcripts). Contaminant transcripts of *T. diversipes* have already been described previously[39]. Herein, we identified only 31 (0.25%) transcripts as potential contaminants (Supplementary Material 1), far less than that observed in *C. waltheriae*, which was expected, since the assembly approaches for *T. diversipes* used a reference genome.

Differential orthologs expression

Using Orthofinder[40,41] we identified 4859 orthogroups shared between *C. waltheriae* and *T. diversipes*, of which 3011 were reported as single-copy orthologues. Using blastp[42,43] searches to identify sequences with the highest homology within common orthogroups we additionally selected 1848 one-to-one orthologs, totaling 4154 one-to-one ortholog pairs between the two species. Of these, around half (2096 or 50.45%) were annotated with at least one functional GO term. After correcting for sampling batch effects accounting for species, the number of samples and sampling year (Supplementary Material 2), the significantly differentially expressed orthologs between the species were obtained by overlapping the results from two strategies (1. TMM normalized counts[44] and edgeR 3.34.0[45]; 2. Scale Based Normalization [SCBN] method[46] and NOISeq[47,48]). We identified a total of 646 orthologs as differentially expressed (DE) between the two species (Supplementary Material 3), 335 of which were highly expressed in *C. waltheriae,* and 311 were highly expressed in *T. diversipes.* Among the annotated DE orthologs, two groups caught our attention prior to GO Enrichment Analysis due to their frequency and function. First, a group of eleven transcripts homologous to sequences of PiggyBac transposable element-derived protein genes (*Pgbd4* = 10 transcripts and *Pgbd2* = 1 transcript), most of which (7) were highly expressed in *C. waltheriae* (Fig. 1). Moreover, considering the entire set of transcripts assembled, we found that *C. waltheriae* has more transcripts (n = 68; 0.34% of the transcriptome) of *Pgdb*-like genes than *T. diversipes* (n = 24; 0.19%). Additionally, we found in *T. diversipes* two sequences identified as homologous to Major Royal Jelly Protein/yellow-like (*MRJP/y*) that were highly expressed (Fig. 1).

Through the Gene Ontology term enrichment analysis for all categories (BP, CC, and MF), we identified 54 GO terms enriched among the DE orthologs (Supplementary Material 4). The redundant terms of each GO category were summarized in a two-level hierarchical list for visualization (Fig. 2). Top enriched Cellular Component (CC) terms were related to ion channel complex, transporter complex, and plasma membrane protein complex. Consistently, top enriched Molecular Function (MF) terms were related to channel activity, transmembrane signaling receptor activity, and potassium ion transmembrane activity. The orthologs annotated with these CC and MF terms are shown in Table 2.

For the Biological Process (BP) category, the most significant terms in the enrichment analysis were related to camera-type eye morphogenesis, angiogenesis, and regulation of osteoblast differentiation. The main DE orthologs annotated within these terms were identified as homologs of *Fibrillin-2-like*, *Low-density lipoprotein receptor-related protein 6,* and serine/threonine-protein kinases. Moreover, the highest number of DE orthologs

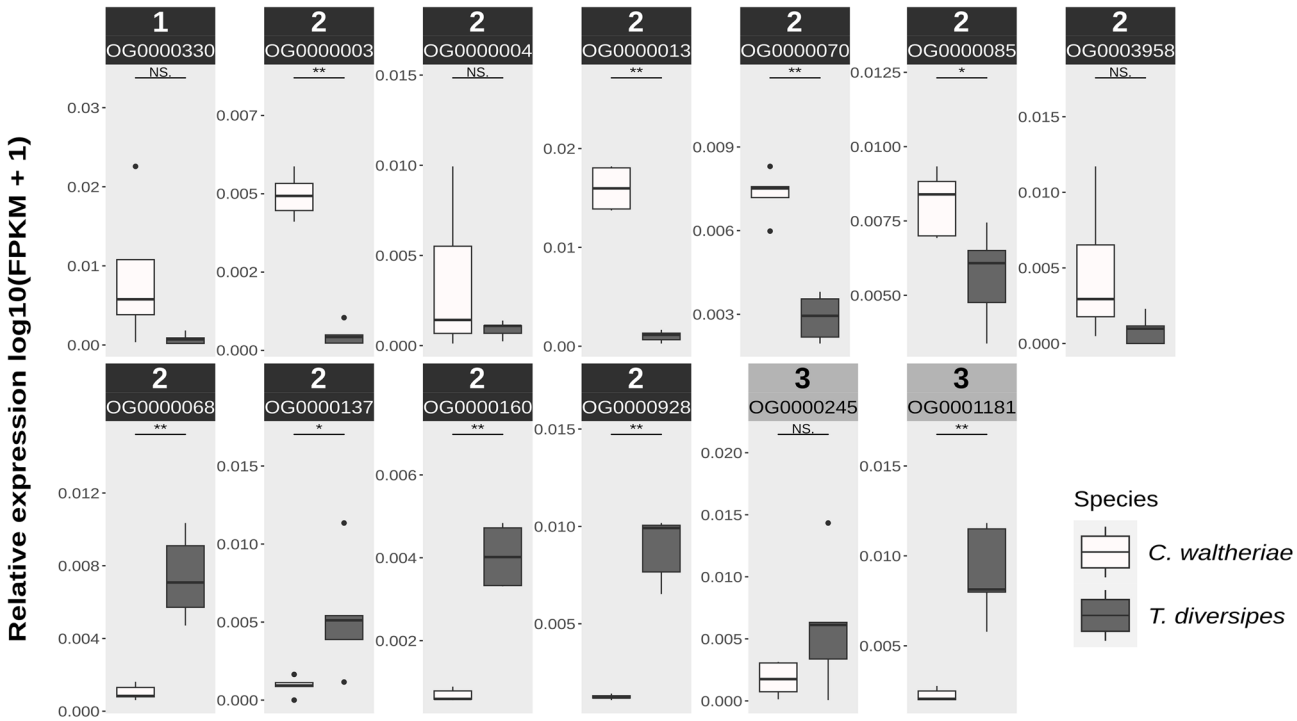

Figure 1. Relative expression of orthologs differentially expressed between the cleptoparasite *Coelioxoides waltheriae* and their host *Tetrapedia diversipes* that were identified as homologs to PiggyBac Transposable Element-derived (1: PGBD2-like; 2: PGBD4-like) and Major Royal Jelly Protein (3) sequences. The respective orthogroups (OGs) are identified according to their IDs in Supplementary Material 3.

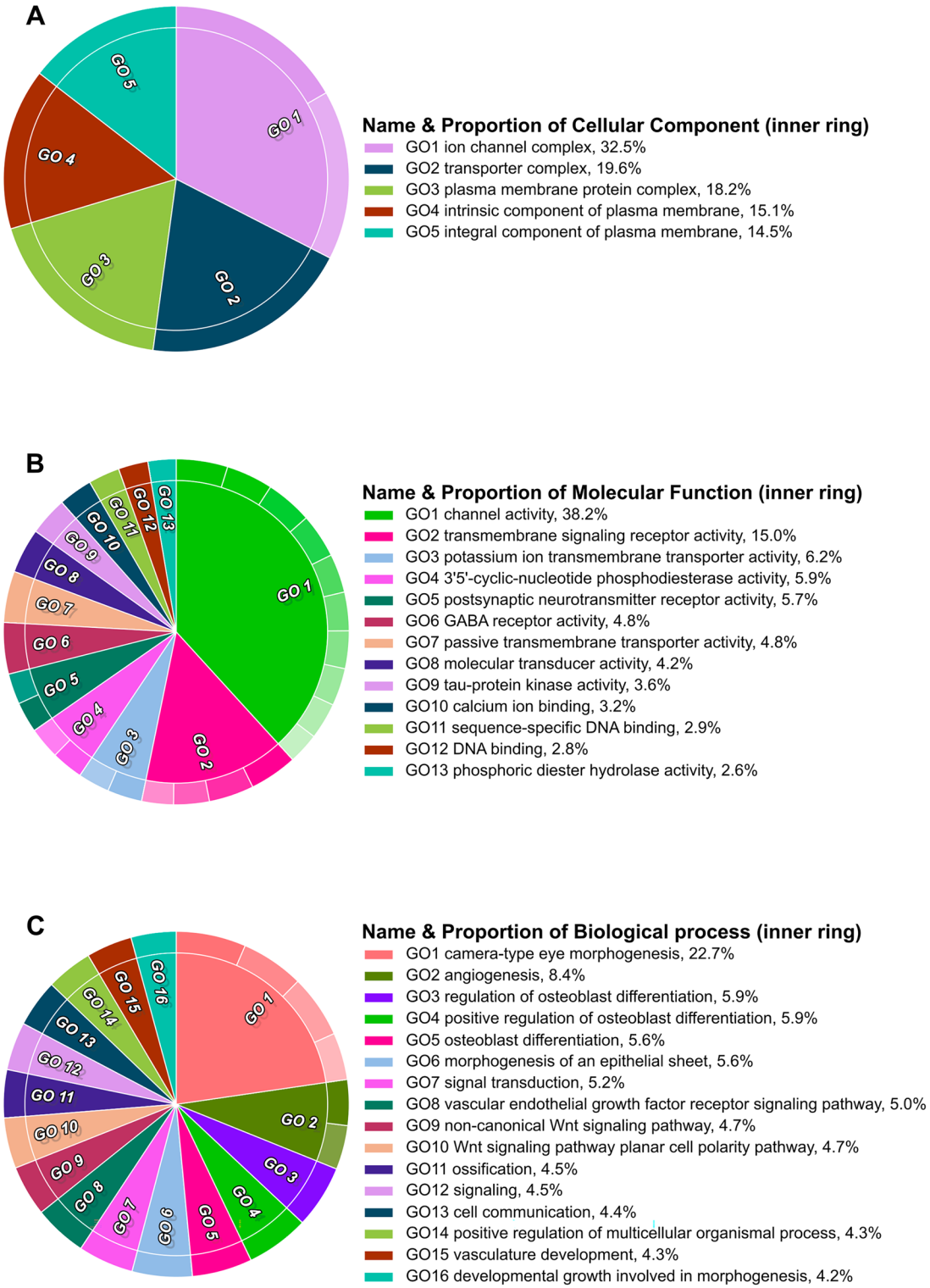

Figure 2. Gene Ontology (GO) terms enriched among the Differentially Expressed (DE) Orthologs between *Coelioxoides waltheriae* and *Tetrapedia diversipes*. Each pie chart represents one of the three main GO categories: (**A**) Cellular Component, (**B**) Molecular Function, and (**C**) Biological Process. Legends next to each chart indicate the representative color, name, and proportion of DE that received an annotation related to the parental GO term. Child terms from the second hierarchical level are represented in the outermost part of the pie chart using shades of the parental representative term color. Parent–child GO term relationships are listed in Supplementary Material 4.

DEO IDs	Expression patterns	Homologs (Swiss-Prot organism)	GO hierarchical groups
Calcium channels			
OG0000284	*C. waltheriae* ↓ *T. diversipes* ↑	*CAC* (*Apis mellifera*)	*Cell component* GO1 ion channel complex GO2 transporter complex GO3 plasma membrane protein complex *Molecular function* GO1 channel activity GO7 passive transmembrane transporter activity GO10 calcium ion binding
OG0000168	*C. waltheriae* ↓ *T. diversipes* ↑	*Trp* (*Drosophila melanogaster*)	*Cell Component* GO1 ion channel complex GO2 transporter complex GO3 plasma membrane protein complex GO4 intrinsic component of plasma membrane GO5 integral component of plasma membrane *Molecular Function* GO1 channel activity GO7 passive transmembrane transporter activity
Sodium channels			
OG0003129	*C. waltheriae* ↑ *T. diversipes* ↓	*NaCP60E* (*Drosophila melanogaster*)	*Cell component* GO1 ion channel complex GO2 transporter complex GO3 plasma membrane protein complex GO4 intrinsic component of plasma membrane GO5 integral component of plasma membrane *Molecular function* GO1 channel activity GO7 passive transmembrane transporter activity
OG0002153	*C. waltheriae* ↑ *T. diversipes* ↓	*unc80* (*Drosophila melanogaster*)	*Cell component* GO1 ion channel complex GO2 transporter complex *Molecular Function* GO1 channel activity GO7 passive transmembrane transporter activity
Potassium channels			
OG0004503	*C. waltheriae* ↓ *T. diversipes* ↑	*KCNQ1* (*Felis catus, Xenopus laevis*)	*Cell component* GO1 ion channel complex GO2 transporter complex GO3 plasma membrane protein complex GO4 intrinsic component of plasma membrane GO5 integral component of plasma membrane *Molecular function* GO1 channel activity GO3 potassium ion transmembrane transporter activity GO7 passive transmembrane transporter activity
OG0000260	*C. waltheriae* ↓ *T. diversipes* ↑	*Sh* (*Drosophila melanogaster*)	*Cell component* GO1 ion channel complex GO2 transporter complex GO3 plasma membrane protein complex GO4 intrinsic component of plasma membrane GO5 integral component of plasma membrane *Molecular function* GO1 channel activity GO3 potassium ion transmembrane transporter activity GO7 passive transmembrane transporter activity
OG0004153	*C. waltheriae* ↑ *T. diversipes* ↓	*KCNT1* (*Homo sapiens*)	*Molecular function* GO1 channel activity GO3 potassium ion transmembrane transporter activity GO7 passive transmembrane transporter activity
OG0002734	*C. waltheriae* ↑ *T. diversipes* ↓	*KCNK9* (*Homo sapiens, Rattus norvegicus*)	*Molecular function* GO1 channel activity GO3 potassium ion transmembrane transporter activity GO7 passive transmembrane transporter activity
Chloride channel			
Continued			

DEO IDs	Expression patterns	Homologs (Swiss-Prot organism)	GO hierarchical groups
OG0000326	*C. waltheriae* ↓ *T. diversipes* ↑	*Grd* (*Drosophila melanogaster*)	*Molecular function* GO1 channel activity GO2 transmembrane signaling receptor activity GO6 GABA receptor activity GO7 passive transmembrane transporter activity GO8 molecular transducer activity
Ligand-gated ion channels			
OG0000029	*C. waltheriae* ↓ *T. diversipes* ↑	*Grik2* (*Mus musculus*, *Xenopus laevis*)	*Cell component* GO1 ion channel complex GO2 transporter complex GO3 plasma membrane protein complex GO4 intrinsic component of plasma membrane GO5 integral component of plasma membrane *Molecular function* GO1 channel activity GO2 transmembrane signaling receptor activity GO3 potassium ion transmembrane transporter activity GO5 postsynaptic neurotransmitter receptor activity GO7 passive transmembrane transporter activity GO8 molecular transducer activity
OG0000962	*C. waltheriae* ↓ *T. diversipes* ↑	*nAChRβ1* (*Drosophila melanogaster*)	*Cell component* GO1 ion channel complex GO2 transporter complex GO3 plasma membrane protein complex GO4 intrinsic component of plasma membrane GO5 integral component of plasma membrane *Molecular function* GO1 channel activity GO2 transmembrane signaling receptor activity GO5 postsynaptic neurotransmitter receptor activity GO7 passive transmembrane transporter activity GO8 molecular transducer activity
G-protein coupled receptors			
OG0002068	*C. waltheriae* ↓ *T. diversipes* ↑	*GABBR1* (*Homo sapiens*, *Mus musculus*)	*Cell component* GO3 plasma membrane protein complex GO4 intrinsic component of plasma membrane GO5 integral component of plasma membrane *Molecular function* GO2 transmembrane signaling receptor activity GO5 postsynaptic neurotransmitter receptor activity GO6 GABA receptor activity GO8 molecular transducer activity
OG0001556	*C. waltheriae* ↑ *T. diversipes* ↓	*Gabbr2* (*Mus musculus*)	*Cell component* GO3 plasma membrane protein complex GO4 intrinsic component of plasma membrane GO5 integral component of plasma membrane *Molecular function* GO2 transmembrane signaling receptor activity GO6 GABA receptor activity GO8 molecular transducer activity
OG0000063	*C. waltheriae* ↑ *T. diversipes* ↓	*Octβ2R* (*Drosophila melanogaster*)	*Molecular function* GO2 transmembrane signaling receptor activity GO8 molecular transducer activity
OG0001943	*C. waltheriae* ↑ *T. diversipes* ↓	*HTR2C* (Mus musculus, *Bos taurus*)	*Molecular function* GO2 transmembrane signaling receptor activity GO8 molecular transducer activity
OG0000814	*C. waltheriae* ↓ *T. diversipes* ↑	*CG31760* (*Drosophila melanogaster*)	*Molecular function* GO2 transmembrane signaling receptor activity GO8 molecular transducer activity
OG0001490	*C. waltheriae* ↓ *T. diversipes* ↑	*Gr64f./Gr5a* (*Drosophila melanogaster*)	*Molecular function* GO2 transmembrane signaling receptor activity GO8 molecular transducer activity
G protein			
OG0004442	*C. waltheriae* ↓ *T. diversipes* ↑	*G-sα60A* (*Anopheles gambiae*)	*Cell component* GO3 plasma membrane protein complex
Translocases			
Continued			

DEO IDs	Expression patterns	Homologs (Swiss-Prot organism)	GO hierarchical groups
OG0001155	*C. waltheriae* ↑ *T. diversipes* ↓	*Atpα* (*Drosophila melanogaster*)	*Cell component* GO2 transporter complex GO3 plasma membrane protein complex GO4 intrinsic component of plasma membrane GO5 integral component of plasma membrane *Molecular function* GO1 channel activity GO3 potassium ion transmembrane transporter activity
OG0001472	*C. waltheriae* ↑ *T. diversipes* ↓	*nrv2/nrv1* (*Drosophila melanogaster*)	*Cell component* GO2 transporter complex GO3 plasma membrane protein complex GO4 intrinsic component of plasma membrane GO5 integral component of plasma membrane
OG0000864	*C. waltheriae* ↓ *T. diversipes* ↑	*UQCRFS1* (*Lagothrix lagotricha*, *Colobus polykomos*)	*Molecular function* GO1 channel activity
DNA-binding protein			
OG0003803	*C. waltheriae* ↑ *T. diversipes* ↓	*Hr38* (*Drosophila melanogaster*)	*Molecular function* GO2 transmembrane signaling receptor activity GO8 molecular transducer activity GO11 sequence-specific DNA binding GO12 DNA binding
Low-density lipoprotein-receptor-related			
OG0004128	*C. waltheriae* ↑ *T. diversipes* ↓	*LRP6* (*Homo sapiens*, *Mus musculus*)	*Cell component* GO3 plasma membrane protein complex *Molecular function* GO2 transmembrane signaling receptor activity GO8 molecular transducer activity
Cell adhesion molecules			
OG0000727	*C. waltheriae* ↑ *T. diversipes* ↓	*NRXN1* (*Homo sapiens*, *Mus musculus*)	*Cell component* GO4 intrinsic component of plasma membrane GO5 integral component of plasma membrane *Molecular function* GO2 transmembrane signaling receptor activity GO8 molecular transducer activity GO10 calcium ion binding
OG0000165	*C. waltheriae* ↑ *T. diversipes* ↓	*Stan* (*Drosophila melanogaster*)	*Molecular function* GO2 transmembrane signaling receptor activity GO8 molecular transducer activity GO10 calcium ion binding
Protein phosphatase			
OG0000836	*C. waltheriae* ↓ *T. diversipes* ↑	*Ptp69D* (*Drosophila melanogaster*)	*Molecular function* GO2 transmembrane signaling receptor activity GO8 molecular transducer activity
Tyrosine-protein kinase			
OG0002698	*C. waltheriae* ↓ *T. diversipes* ↑	*Abl* (*Drosophila melanogaster*)	*Molecular function* GO2 transmembrane signaling receptor activity GO8 molecular transducer activity

Table 2. Differentially expressed (DE) orthologs between *Coelioxoides waltheriae* and *Tetrapedia diversipes* grouped according to their association to the top enriched GO terms for Cell Component and Molecular Function categories. The orthologs are represented by the ID of their respective orthogroups. Expression patterns of orthologs comparatively between *C. waltheriae* and *T. diversipes* are indicated by arrows: ↑ (upregulated) and ↓ (downregulated). GO hierarchical groups refer to CirGo terms as shown in Fig. 2.

were associated with cell communication, signaling, and signal transduction BP terms (Supplementary Material 5).

Discussion

We employed comparative transcriptomics to investigate molecular distinctions during the reproductive stage of *Coelioxoides waltheriae* (a cleptoparasite) and its host species, *Tetrapedia diversipes*. The present study represents an initial effort to obtain some insights about the molecular mechanisms involved in bee cleptoparasitism by comparing orthologs expression data between a cleptoparasitic species and their host, both inhabiting the same location and at equivalent developmental stages. Considering the constraints of cross-species transcriptomic comparisons, our findings shed light on broad molecular variations between the cleptoparasite and its host, suggesting possible ecological implications.

We overlapped two statistical approaches to perform a cross-species comparison and to reduce the number of false positives in our analyses. The first of these methods was the SCBN, a recently proposed method for count normalization optimized to deal with cross-species comparisons[46]. Secondly, we used a traditional differential expression workflow with the edgeR Bioconductor package[45]. By overlapping the results from these two approaches we consistently retrieved 646 differentially expressed orthologs between the two species, with PiggyBac Transposable Element-derived like (*Pgbd-like*) genes being the most frequent (Fig. 1). Overall, we found

many orthologs annotated as *Pgbd-like,* and several of them were found to be highly expressed in *C. waltheriae*. The PiggyBac transposon superfamily is widespread among eukaryotes. The first PiggyBac element was described from a cell culture of *Trichoplusia ni* (Lepidoptera: Noctuidae), the cabbage looper moth[49,50]. Since then, several PiggyBac-like sequences have been described in a variety of organisms. Some of these transposable elements lost transposase activity and are called domesticated[51]. Molecular domestication is a process that occurs due to a transposition inactivating mutation, resulting in the loss of mobility of the transposable element[52]. These insertions may lead to the emergence of new cellular activities, either by altering the coding and/or regulatory regions in which these elements are inserted in the genome or by the evolution of the former TE genes into new genes[52,53]. Thus, these findings lead us to hypothesize whether the domestication of PiggyBac-like transcripts and its regulatory consequences could be one of the molecular mechanisms involved in the evolution and adaptation of *C. waltheriae*, justifying their differential expression between the studied host-parasite species. Indeed, in a recent study, it was observed through comparative genomic analyses that retroviral or transposable elements have undergone a recent or ongoing spread in the genome of a Nomadinae cleptoparasite[54]. This indicates a possible involvement of TEs in the evolution of cleptoparasitism. Future research should further explore this hypothesis, investigating the evolution of transposable and transposable-derived elements in host-parasites genomes and their expression pattern in other cleptoparasitic species of Apidae.

We also identified two MRJP/Yellow-like sequences as highly expressed in *T. diversipes* while none were upregulated in *C. waltheriae*. The MRJP/Yellow gene family encodes multifunctional yellow-like proteins identified in arthropods and in several bacteria[55,56]. Nonetheless, the MRJP-like genes, as part of this family, seem to be restricted to Hymenoptera[56]. Even though these genes have been associated with olfactory learning—particularly regarding *mrjp1* expression in the mushroom body (Kenyon cells)[31,57,58]—and functional studies have shown that MRJPs may have immunoregulatory and antibacterial effects[56], in bees, MRJPs are mostly associated with larvae feeding, development and with the regulation of phenotypic plasticity and age-polyethism in workers[59]. In honeybees, MRJPs are known to be essential components of the larvae diet, with MRJP1–3 and 5 being the most abundant proteins of larval food[56]. These proteins are synthesized by honeybee nurse workers in the specialized hypopharyngeal glands and offered to the immature offspring through a special food jelly (royal jelly)[60]. In contrast to honeybees, bumblebees do not produce royal jelly. However, their hypopharyngeal glands express one MRJP ortholog[61]. The production of MRJP is not dependent on brood-feeding activity in bumblebees, and protein signals were identified in the abdominal parts of the digestive tract in queens and workers. This suggests that bumblebee MRJP does not have a nutritional function but is instead involved in food digestion and/or modification, consistent with the putatively ancient function of bee hypopharyngeal glands in food digestion[61]. In solitary bees the role of MRJP-like genes is currently poorly understood. One hypothesis for the high expression of MRJPs in *T. diversipes* could be the association of these proteins with larval food. As a solitary bee species, founder females of *T. diversipes* are also responsible for larvae food provisioning and an increased expression of transcripts related to MRJP-like genes could suggest that these proteins may be component of larval food. Yet, the role of MRJPs-like may not be necessarily nutritional/developmental, as they could alternatively play an antibacterial role and/or aid in the processing of bee products (e.g., formation of pollen-pellets and pollen-bread)[62–64].

Several Biological Process (BP) GO terms enriched among DE orthologs were found to be typical or exclusively related to vertebrate species, such as camera-type eye morphogenesis, angiogenesis, and osteoblast differentiation. It is important to note that molecular data banks, including UniRef, predominantly rely on scientific evidence derived from a limited number of model species[65]. As a result, the nature and extent of GO annotations reflect the aspects investigated in these organisms being only extrapolated to other non-model species under the assumption that putative orthologs are functionally equivalent[66]. However, when dealing with species that are evolutionarily distant, orthologs may have completely different biological functions[67]. Therefore, extrapolating the biological significance of these top enriched BP terms to the invertebrate species investigated in this study is challenging. On the other hand, the enriched Molecular Functions (MF) and Cellular Components (CC) GO terms results were more relatable. The top enriched GO terms for the CC category were related to "ion channel complex", "transporter complex", and "plasma membrane protein complex". Consistently, top enriched MF terms were related to "channel activity", "transmembrane signaling receptor activity", and "potassium ion transmembrane activity". Most of the DE orthologs annotated along with these terms (Table 2) express proteins with a variety of functional roles that are putatively relevant for cleptoparasitic adaptations. In summary, in *C. waltheriae* upregulated orthologs annotated to these terms were related to olfaction, learning, neuronal excitability, synaptic organization, and other neuronal processes. Adaptations in these genetic pathways could be related to the successful exploitation of host resources and the development of specialized behaviors in cleptoparasitic bees. Meanwhile, orthologs highly expressed in *T. diversipes* were involved in chemosensory and sensory perception, particularly gustatory receptor genes involved in sweet taste response, motor, and mitochondrial activity, as well as neurotransmission, genetic mechanisms that could be associated with foraging and nest provisioning performed by *T. diversipes* foundresses. In boxes 1 and 2, we detailed the differentially expressed orthologs associated with the top MF and CC GO terms enriched by addressing their putative functions.

Cleptoparasitic bees are thought to rely on olfactory signals to locate host nests[68,69]. Indeed, morphological differences in antennae sensillae composition have been observed between cleptoparasites and their host species, with cleptoparasites having a greater prevalence of olfactory structures[70]. Previous studies have also shown that chemosensory-related genes (CRGs) are highly expressed in hymenopteran parasitoids[71–74]. Chemosensory cues, along with other external stimuli, are processed in sensory systems, triggering the formation of both long-term and short-term memories[75]. These adaptations may contribute to the localization and memorization of host nest aggregations[76]. In addition to the sensory-related genes described in box 1, we identified other potential chemosensory-related genes (CRGs) upregulated in the cleptoparasite, including Ionotropic Receptors (IRs), a gene family involved in olfaction, gustation, and other sensory perceptions[77,78]. Unlike many other chemosensory gene families, IRs exhibit a remarkable degree of conservation across species[79]. The *IR25a* homolog was found

highly expressed in *C. waltheriae*. This gene, along with *IR8a*, functions as a co-receptor (IRco) for the formation of functional IR complexes in conjunction with a ligand-binding receptor protein (IR X)[80]. Furthermore, we detected an upregulated putative odorant receptor (OR) homologous to the *Polistes dominula Or85c-like* sequence in *C. waltheriae*. Although this specific OR's role in odour perception remains unknown, computational analysis in *P. dominula* suggests its involvement in parasitoid sensory perception. It is important to note that our current approach may have missed lineage-specific CRGs, and further investigation into these lineage-specific CRGs in cleptoparasitic lineages is crucial to gain insights into the underlying processes of parasitism in bees.

C. waltheriae and *T. diversipes* lineages, both belonging to the Apidae family, diverged over 77 million years ago[11]. Thus, compared to several cleptoparasites, that prefer hosts from other bee families (e.g., cleptoparasites of the Nomadinae clade)[10], the phylogenetic proximity between *C. waltheriae* and *T. diversipes* could minimize the "phylogenetic noise" in transcriptomic comparisons. Still, cross-species transcriptomic comparisons present challenges, and certain caveats should be considered in the light of the present results. First, our analytical strategy focused on conserved orthologs, thus neglecting non-orthologous or rapidly evolving genes that may play a crucial role in cleptoparasitism evolution. Additionally, due to the divergent life history trajectories of these two species, the differentially expressed orthologs identified could be associated with non-behavioral or random species-specific adaptations and not directly related to cleptoparasitic adaptations. Thus, we argue that applying the methodology used in this study to other bee host-parasite species pairs should allow the differentiation of species-specific adaptations from shared molecular mechanisms. Lastly, we conducted a broad-scale transcriptome analysis, meaning that the RNA-Seq data used were obtained from whole-body extractions. As an exploratory analysis, this approach allowed some initial broad insights, but it is limited in capturing genes with tissue-specific expression or complex expression patterns across the body[81]. We also used sample replicates containing a variable number of pooled individuals for *C. waltheriae*, to account for this sampling strategy we normalized and treated the gene expression counts for batch effects across replicates, and this treatment might have reduced even further our power to detect genes with subtle expression differences. Further research should therefore delve into the obtained results by investigating the expression of these genes in a tissue-specific context, in order to elucidate their functions more accurately. Finally, we analyzed a specific life stage of each species by comparing females performing reproductive activities. We considered that this stage comprises some of the most distinguishing behavioural differences between the host and cleptoparasite species, yet comparative analyses of multiple life stages could provide a more detailed profile of cleptoparasitism across developmental stages.

Box 1: Orthologs upregulated in *C. waltheriae* and associated to top enriched GO terms

In *C. waltheriae*, we found upregulated orthologs that may have roles in the olfactory and learning pathways, potentially representing adaptations to cleptoparasitic behavior[68,69].

Among the upregulated sequences in *C. waltheriae*, we discovered homologs to the invertebrate *Octβ2R* and vertebrate *HTR2C*, receptors for biogenic amines octopamine (OA) and serotonin (5-HT), respectively. Biogenic amines play important physiological roles in organisms, modulating neuronal, metabolic, and physiological processes[82]. In hymenopteran species, OA has been linked to locomotor activity, sensory (gustatory, olfactory, and visual) sensitivity, aggressive behavior, and (associative and non-associative) learning[76,83,84]. *Octβ2R* has been associated with rewarding reinforcement signaling in *Drosophila*, indicating its potential role in behavioral responses[85]. In honeybees, octopaminergic signaling is implicated in appetitive learning[86]. Overexpression of octopamine receptors in honeybees has also been associated with oxidative stress, neuroinflammation, and olfactory dysfunction[87]. Similarly, 5-HT acts as a neurotransmitter, neuromodulator, and hormone in insects[88], regulating various behaviors such as aggression, mating, feeding, locomotion, and olfaction[89–95]. Invertebrate 5-HT receptors are found in the central and peripheral nervous systems (CNS and PNS, respectively), mediating both excitatory and inhibitory actions[96,97]. The signaling of 5-HT receptors in *Drosophila* is linked to aggression, sleep, circadian behavior, feeding, mating, learning, and memory[98–103].

We also observed upregulation of the subunit of a neurotransmitter receptor in *C. waltheriae*, a homolog of the mammalian *Gabbr2* (Gamma-aminobutyric acid type B receptor subunit 2). Meanwhile the *Gabbr1* (Gamma-aminobutyric acid type B receptor subunit 1) was upregulated in *T. diversipes*. Gamma-aminobutyric acid (GABA) is the major neurotransmitter for inhibitory synaptic transmissions in the CNS of both vertebrates and invertebrates[104,105]. GABA signaling influences insect behaviors such as learning and memory, locomotor activity, and odor processing[106–110]. $GABA_B$ receptors, coupled with ionotropic $GABA_A$ receptors, mediate the action of GABA[111]. These receptors regulate complex behaviors and nervous system functions by inhibiting GABA release and reducing the release of other neurotransmitters[112]. For instance, in *Drosophila* and *Heliothis virescens*, $GABA_B$ receptors are involved in the olfactory pathway[89,113]. In *Apis mellifera*, $GABA_B$ is associated with locomotor behavior[106].

In *C. waltheriae*, we also observed the upregulation of *NaCP60E*, a voltage-gated sodium channel. This gene has been mainly associated with olfactory function, but it is also expressed in other tissues[114,115] and in invertebrates without olfaction[116]. In *Drosophila*, decreased expression of *NaCP60E* has been shown to decrease the olfactory response to benzaldehyde[117]. Also, a dense concentration of NaCP60E proteins in neuron axons of chemosensory organs suggests that this gene is involved in the olfactory system[118]. Thus, the upregulation of *NaCP60E* suggests its involvement in olfaction-related processes in the cleptoparasitic bee.

C. waltheriae also showed increased expression of potassium channel orthologs, specifically homologs of human *KCNT1* and *KCNK9* genes. *KCNT1* encodes a sodium-gated potassium channel that is activated by neuromodulators[119]. Therefore, it is suggested that it regulates neuronal excitability and may play a role in several behaviors[120]. *KCNK9* encodes a two-pore domain potassium channel that is involved in resting membrane potentials[121]. It can also be a target for modulatory molecules, such as volatile anesthetics and

neurotransmitters[122]. Additionally, in *Drosophila*, *KCNK9* orthologs (*Task6* and *Task7*) are preferentially expressed in the antennal lobes and may play a role in olfactory memory formation[123].

Furthermore, we identified in *C. waltheriae* the upregulation of genes associated with neuronal excitability, including circadian rhythm, memory formation, and development. These include *Unc80*, *Hr38*, and *Lrp6*. Unc80 is a regulatory component of the sodium leak channel NALCN complex[124], which play a role in modulating resting membrane potential, neuronal excitability, firing rates, and pacemaker activity[125,126]. In *Drosophila*, for instance, *Unc80* is necessary for circadian rhythmicity[124]. It is thought that Unc80 may also be involved in bee learning and memory[31]. The gene *Hr38, a nuclear receptor*, is involved in the transcriptional control of the dopamine synthesis pathway[127], the cuticle gene expression[128], and the ecdysteroid signaling pathways[129]. In addition, it has been suggested that *Hr38* may play an important role in high neuronal functions such as memory formation, courtship behavior, and circadian rhythm[130–132]. *Lrp6* belongs to the low-density lipoprotein receptor family of cell surface receptors and is an essential component of the Wnt signaling pathway, which controls several biological processes throughout the development and adult life of metazoans[133]. Expression of the *Drosophila* ortholog of *Lrp6* (*arr*) has been associated with CNS morphogenesis and organization[134], as well as long-term memory formation[135].

We also found upregulated orthologs of cell adhesion molecules (CAMs) genes, including *NRXN1* and *Fmi/Stan*. In *Drosophila*, *NRXN1* gene is required for proper organization and growth at neuromuscular junctions[136], and it also regulates sleep and synaptic plasticity[137]. In honeybees, the expression of *NRXN1* is increased under sensory stimulation, suggesting a link between sensory processing and associative learning[138]. *Drosophila Fmi*/*Stan* encodes a cadherin that regulates planar cell polarity[139], axon guidance in photoreceptor neurons[140], dendrite morphogenesis of sensory neurons[141], and neuronal morphogenesis of the mushroom body[142]. The upregulation of CAMs genes can be involved with their role in synaptic organization and plasticity, which may be relevant to the cleptoparasitic lifestyle of *C. waltheriae*.

Lastly, we found upregulated orthologs of *ATPα* and nrv2/nrv1 genes in *C. waltheriae*, which encode components of the Na + /K + ATPase[143,144]. Na + /K + ATPase is important for maintaining the balance of sodium and potassium ions across the cell membrane[145]. Defects in the sodium–potassium pump can lead to neuronal dysfunctions, such as circadian rhythm disturbances, locomotor problems, and auditory mechanosensation[146–148].

Box 2: Orthologs upregulated in *T. diversipes* and associated to top enriched GO terms

T. diversipes exhibited upregulated sequences related to chemosensory perception, particularly in the context of gustatory receptor (GR) genes. One such sequence was identified in *T. diversipes* as the *Gr5a* gene while in *C. waltheriae*, the corresponding sequence was identified as the *Gr64f.* gene. Despite the difference in annotation, both genes belong to the same conserved subfamily that is involved in the sweet taste response observed in *Drosophila*[149]. Nevertheless, they are located on different chromosomes in *Drosophila* (*Gr5a*: X chromosome; *Gr64f.*: third chromosome)[150]. These receptors are considered the primary basis for sugar reception in *Drosophila*, and their co-expression in gustatory receptor neurons is necessary for the response to certain sugars[149,151]. Since bees primarily rely on flower nectar as an energy source, sugar detection is crucial for their survival[152]. Thus, a higher expression of these receptors in *T. diversipes* may be attributed to their intense foraging activity while provisioning food in their nests, an activity not performed by the parasite. Another upregulated sequence in *T. diversipes* was also related to GR, the transcript homologous to the *G-sα60A* gene of *Anopheles gambiae*. Studies on female antennae of *A. gambiae* have suggested that *G-sα60A* isoforms participate in olfactory signal transduction[153]. It has also been proposed that *Gsα* in *Drosophila* is responsible for Gr5a-mediated sweet taste perception[154].

Apart from GR genes, several other upregulated genes in *T. diversipes* were associated with sensory perception but were notably involved in distinct pathways when compared to those upregulated in *C. waltheriae*. For example, the *Drosophila Sh* (*Shaker*) gene encodes a voltage-gated potassium channel (Kv)[155] that is involved in shaping and firing the action potential[156]. It is also expressed in the retina and different regions of the CNS, including the mushroom body neuropil[157,158]. Furthermore, evidence suggests that *Sh* regulation affects olfactory learning and memory[159]. Another putative Kv upregulated in *T. diversipes* was a homolog to the vertebrate *KCNQ1* gene, which regulates essential physiological processes all over the body[160]. In *Drosophila*, *KCNQ1* ortholog (*dKCNQ*) is expressed in the brain cortical neurons, cardia, and in the nurse cells and oocytes in the ovary[161,162]. Apart from its role in the fly's normal heartbeat[162] and early embryonic development[161], *dKCNQ* has been implicated in age-dependent memory impairment and is required in α/β Mushroom Body Neurons for setting short-term memories[163].

Another upregulated sequence in *T. diversipes*, homologous to the *Drosophila trp* (Transient Receptor Potential) gene, may also be related to sensory perception. TRP proteins have been associated with phototransduction and olfactory response to CO_2 in *Drosophila*[164,165].

An upregulated homolog of the *Drosophila abl* (Abelson) gene, which encodes a non-receptor tyrosine kinase, was also identified upregulated in *T. diversipes*. The *Drosophila abl* gene is expressed in the axons of the CNS and plays a very important role in establishing axonal connections during CNS development[166]. Also, it has been suggested that Abl-mediated phosphorylation is an important mechanism for the fly visual development[167].

Motor activity-related genes also exhibited higher expression levels in *T. diversipes*. One such gene is the *cacophony* (*Cac*), which encodes a voltage-gated calcium channel that plays a role in regulating neurotransmitter release at neuromuscular synapses in *Drosophila*[168]. Similarly, a homolog of the *Drosophila Ptp69D* gene, which encodes a receptor of tyrosine phosphatase that is associated with mechanosensory neuron development[169], was also upregulated gene in *T. diversipes*. *Ptp69D* might influence some elements of motor function in adult *T. diversipes* females.

Additionally, in *T. diversipes,* there was a higher expression of a homolog to the *UQCRFS1* gene. This gene encodes an iron-sulfur protein (Rieske Fe-S protein—RISP)[170] and is involved in electron transfer in bc1 complex[171]. Furthermore, *T. diversipes* displayed higher expression of homologs of other genes associated to electron transport activity, including genes with mitochondrial activity, indicating an increased energy demand during the founder life stage. These upregulated genes included, *ETFA*, *MRPS18C*, *mdh2*, *Cox6al*, and *Ndufs2*. The *ETFA* gene encodes an electron transfer flavoprotein subunit alpha, which is involved in mitochondrial fatty acid beta-oxidation and amino acid catabolism[172]. *MRPS18C* encodes a mitochondrial ribosomal protein that plays a role in protein synthesis within mitochondria[173]. The *mdh2* gene encodes a malate dehydrogenase, an enzyme involved in the citric acid cycle[174]. *Cox6al* encodes a subunit of cytochrome c oxidase[175], and *Ndufs2* encodes a core subunit of NADH dehydrogenase (Complex I)[176], another two of the critical enzymes in the electron transport chain.

Finally, several upregulated genes in *T. diversipes* were homologous to neurotransmitter receptors, including *nAChRβ1*, *Grik2*, and *Grd*. The *nAChRβ1* gene encodes a subunit of the nicotinic acetylcholine receptor (nAChR), which is the primary excitatory neurotransmitter in insects. It is involved in olfactory learning and memory formation, as well as modulation of aggressive behavior[177,178]. *Grik2* encodes the vertebrate glutamate receptor ionotropic kainate 2, which contribute to rapid synaptic transmission[179]. However, the role of kainate receptors in CNS glutamatergic circuits of insects is not well understood[180]. The *Grd* gene encodes a *Drosophila* chloride-channel homolog of the mammalian GABA receptor delta subunit that may respond to different neurotransmitters[181]. This gene has been linked to mediating the glycine response[181] and forming functional ionotropic GABA receptors[182]. Additionally, an upregulated homolog of the *Drosophila CG31760* gene was found in *T. diversipes*. *CG31760* is a probable G-protein coupled receptor (GPCR), whose expression was highest in the adult fly brain however its exact function is unknown[183].

Conclusion

In conclusion, our study provides global insights into the molecular mechanisms of cleptoparasitism in bees. We observed differential gene expression between the cleptoparasite *C. waltheriae* and its host *T. diversipes* particularly involving genes related to signal transduction, sensory perception, learning, and memory formation. These findings suggest the importance of sensory adaptations and learning in host-parasite interactions. We also identified a higher abundance of transcripts derived from transposable elements in *C. waltheriae* transcriptome, indicating these could be involved in gene neofunctionalization for parasite-specific adaptations. Additionally, the host species exhibited highly expressed genes from the Major Royal Jelly Proteins family. In bees, these genes are associated with various functions, including nutritional, immunity, and developmental regulation. We hypothesize that the differential expression of these genes could be related to nest provisioning in *T. diversipes*, a task not performed by the parasite. Further investigation is required to fully understand the role of all these molecular mechanisms in cleptoparasitism and their potential tissue-specific functions. Moreover, we propose that the methodology employed in the present study should be extended to other bee host-parasite species pairs, allowing for better differentiation of species-specific adaptations and shared molecular mechanisms underlying cleptoparasitism and its convergent traits in bees.

Material and methods

Sample collection and RNA sequencing

T. diversipes samples were collected and sequenced in a previous study[33]. Briefly, females were collected during their reproductive period (November to December of 2012, here called G1, and March to July of 2013, here called G2), at Universidade de São Paulo (23° 33′ 53.2″ S 46° 43′ 51.7″ W), while provisioning their nests. G1 and G2 represent the foundresses of different reproductive generations. Individuals were always sampled between 10:00 A.M. and 12:30 P.M., immediately frozen in liquid nitrogen and stored at − 80 °C. RNA extractions were performed individually using the whole body. Each sample replicated contained pooled RNA from three individuals. Here we included in the analyses RNA-Seq data from five replicates: three from G1 and two from G2. For more details see[33].

Adult females of *C. waltheriae* were collected at the same location as *T. diversipes* samples while hovering at the entry of *T. diversipes* trap nesting spots. This behaviour is typical of *C. waltheriae* females looking for locations to lay their eggs[21]. Two individuals were collected in March 2013 (along with *T. diversipes* samples), and six more were collected later between November 2015 and March 2016. These individuals were also sampled between 10:00 A.M. and 12:30 P.M., immediately frozen in liquid nitrogen, and stored at − 80 °C. Total RNA extractions were performed for each individual separately using the RNeasy® Mini Kit (Qiagen, Austin, Texas, USA), following the manufacturer's instructions. RNA was extracted from the whole body of the samples to enable a broad-scale transcriptomic analysis. Quantification and quality assessment of RNA was performed initially using NanoDrop™ Lite (Thermo Fisher Scientific, Wilmington, Delaware, USA) and later with the Agilent 2100 Bioanalyzer system (Agilent Technologies, Palo Alto, California, USA) before the library preparations. *C. waltheriae* samples were divided into five replicates: two containing only one individual in each (samples collected in 2013); and three containing the pooled RNA of two individuals (collected between 2015 and 2016). Library preparation and sequencing of the first two replicates (2013 samples) were performed by Macrogen (Macrogen Inc., Seoul, South Korea), along with *T. diversipes* samples from[33], using the Illumina® HiSeq 2000 platform. The other three replicates of *C. waltheriae* (sampled from 2015–2016) were sequenced at LaCTAD (Unicamp, Campinas, Brazil), also using an Illumina® HiSeq 2000 sequencer.

C. waltheriae RNA sequencing resulted in 285,806,598 raw reads. After cleaning, read number decreased to 169,961,662. For *T. diversipes*, 343,059,264 raw reads were used, resulting in 190,339,784 cleaned reads after trimming.

Transcripts assembly

Reads quality assessment was performed using the FastQC 0.11.5[184]. Trimmomatic 0.38[185] was used to quality trimming the raw reads (options: SLIDINGWINDOW:4:30 TRAILING:3 MINLEN:80 AVGQUAL:30 HEADCROP:14). Cleaned reads were then digitally normalized (20 × coverage) using a script (*insilico read normalization*) implemented within the Trinity toolbox[186,187]. Transcriptome assembly was performed differently for each species. *C. waltheriae* transcriptome was assembled following Trinity 2.10.0[186,187] de novo assembly protocol with default parameters. For *T. diversipes*, a draft genome was available (Santos et al., unpublished), so two different approaches were used: (1) a genome-guided transcriptome assembly, using STAR 2.7.3[188] and Cufflinks 2.2.1[189]; and (2) a reference guided de novo assembly with Trinity, using the bam file generated by the STAR software. The transcripts resulting from these two assemblies were clustered into SuperTranscripts[190] by using CD-Hit[191], Corset[192], and Lace[190]. This clustering approach was also applied to the de novo assembly of *C. waltheriae*, to make the datasets more comparable between the two species and to reduce transcript redundancy in *C. waltheriae*. Downstream differential expression analyses were performed using the SuperTranscripts non-redundant datasets. Quality parameters of the transcriptomes were analyzed with Transrate 1.0.3[38] and BUSCO 5.2.2[193] (using hymenoptera_odb10 database).

Annotation and expression analysis of orthologs

Non-redundant transcripts were annotated with Annocript 2.0.1[194] pipeline using the UniProt Reference Clusters (UniRef90)[195] and UniProtKB/Swiss-Prot[196] databases from February 2021. Transcripts with significant blast hits (*e*-value < 1e−5) against possible contaminants (i.e., Acari, Alveolata, Archaea, Bacteria, Fungi, Rhizaria, Rhodophyta, Viridiplantae, and Viruses) in the UniRef90 database were removed from the final data set using custom scripts (https://github.com/PauloCseri/Annotation.git).

To identify orthologs between the two species the amino acid sequences of the resulting ORFs predicted in the Annocript pipeline were used as input to Orthofinder 2.1.1[40,41]. Then, we selected the best matching transcripts per species in each orthogroup to get one-to-one ortholog data. This selection was performed by filtering the transcripts of the highest bitscore value (better alignment) between sequences of the two species in the same orthogroup. These alignments were performed using the blastp algorithm[42,43] with default parameters. Differential expression (DE) analyses were then performed using only the orthologs identified in both species.

To estimate gene expression levels, we used Bowtie 2.3.5.1[197] to align the cleaned reads to their respective ortholog transcripts set, i.e., *C. waltheriae* sample reads were aligned to *C. waltheriae* ortholog transcripts and *T. diversipes* reads to its corresponding set of ortholog transcripts. Then, we used RSEM 1.3.3[198] for read counting. The significant DE orthologs between the two species were identified by combining the following approaches: (1) using the edgeR 3.34.0[45] with TMM normalized counts[44]; and (2) using NOISeq 2.36.0[47,48] with the normalization factor calculated through the SCBN method to normalize read counts[46]. This normalization strategy is optimized for cross-species DE analyses[46] and is implemented in the Bioconductor package SCBN 1.10.0. Only significant DE orthologs (|Log2FC| ≥ 2, adjusted *p-value* < 0.05) commonly identified as so in these two analyses were selected for the resulting set of DEs. It is worth mentioning that as the samples were not sequenced in the same batch (different years), we additionally adjusted these counts for batch effects using the ComBat-seq method[199] available on the Bioconductor package sva 3.38.0[200]. In detail, for the batch parameter, we used the different sequencing times as factors (0: initial sequencing of *C. waltheriae* and *T. diversipes* samples; 1: later sequencing of *C. waltheriae* samples only) and for the *group* parameter (biological condition), we used the respective species (0: *C. waltheriae*; 1: *T. diversipes*).

GO functional analysis

Assuming that sequences from the same orthogroup descend from a single gene in the last common ancestor[40,41] and hence they likely have similar functions, the ensemble of functional GO annotations from all SuperTranscripts belonging to the same orthogroup were used for functional analyses. The Bioconductor package PloGO2 1.4.0[201] was used to plot and visualize the GO annotations of DE orthologs. To test whether any GO term was enriched among the DE orthologs in comparison to all other orthologs identified we used the Bioconductor package TopGO 2.20.0[202]. Redundant GO enriched terms were summarized in a two-level hierarchical GO set using the REVIGO web server[203] for simplification, and these hierarchical sets were represented in charts generated by the CirGO 2.0 software[204].

Data availability

The raw sequence reads have been deposited in the NCBI Sequence Read Archive (SRA) under the following accession numbers: SRR24187037, SRR24187038, SRR24187039, SRR24187040, SRR24187041, SRR24187042, SRR24187043, SRR24187044, SRR24187045, SRR24187046. The corresponding BioProject accession number is PRJNA955762. The transcriptome assemblies, as well as the sets of SuperTranscripts, are available on FigShare (https://doi.org/10.6084/m9.figshare.23264771.v1).

Received: 23 August 2023; Accepted: 4 March 2024
Published online: 29 May 2024

References

1. Schmidt, G. D. & Roberts, L. S. *Foundations of Parasitology* (McGraw-Hill, 2009).
2. Phillips, R. S. *Parasitism: The Variety of Parasites* (Wiley, 2012).
3. Lefèvre, T. *et al.* The ecological significance of manipulative parasites. *Trends Ecol. Evol.* **24**, 41–48 (2009).
4. Dobson, A. P. & Hudson, P. J. Parasites, disease and the structure of ecological communities. *Trends Ecol. Evol.* **1**, 11–15 (1986).

5. Michener, C. D. *The Bees of the World* (The Johns Hopkins University Press, 2007).
6. Ascher, J. S. & Pickering, J. *Discover Life Bee Species Guide and World Checklist (Hymenoptera: Apoidea: Anthophila)*. http://www.discoverlife.org/mp/20q?guide=Apoidea_species (2020).
7. Danforth, B. N., Minckley, R. L. & Neff, J. L. *The Solitary Bees* (Princeton University Press, 2019).
8. Cardinal, S., Straka, J. & Danforth, B. N. Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *Proc. Natl. Acad. Sci. USA* **107**, 16207–16211 (2010).
9. Bossert, S. *et al.* Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. *Mol. Phylogenet. Evol.* **130**, 121–131 (2019).
10. Sless, T. J. L. *et al.* Phylogenetic relationships and the evolution of host preferences in the largest clade of brood parasitic bees (Apidae: Nomadinae). *Mol. Phylogenet. Evol.* **166**, 107326 (2022).
11. Litman, J. R., Praz, C. J., Danforth, B. N., Griswold, T. L. & Cardinal, S. Origins, evolution, and diversfication of cleptoparasitic lineages in long-tongued bees. *Evolution* **67**, 2982–2998 (2013).
12. Trunz, V., Packer, L., Vieu, J., Arrigo, N. & Praz, C. J. Comprehensive phylogeny, biogeography and new classification of the diverse bee tribe Megachilini: Can we use DNA barcodes in phylogenies of large genera?. *Mol. Phylogenet. Evol.* **103**, 245–259 (2016).
13. Gonzalez, V. H., Gustafson, G. T. & Engel, M. S. Morphological phylogeny of Megachilini and the evolution of leaf-cutter behavior in bees (Hymenoptera: Megachilidae). *J. Melittol.* **85**, 1–123. https://doi.org/10.17161/jom.v0i85.11541 (2019).
14. Magnacca, K. N. & Danforth, B. N. Evolution and biogeography of native Hawaiian *Hylaeus* bees (Hymenoptera: Colletidae). *Cladistics* **22**, 393–411 (2006).
15. Danforth, B. N. *et al.* Phylogeny of Halictidae with an emphasis on endemic African Halictinae. *Apidologie* **39**, 86–101 (2008).
16. Rozen, J. G. Eggs, ovariole numbers, and modes of parasitism of cleptoparasitic bees, with emphasis on neotropical species (Hymenoptera: Apoidea). *Am. Mus. Novit.* **3413**, 1–36 (2003).
17. Minckley, R. L. & Danforth, B. N. Sources and frequency of brood loss in solitary bees. *Apidologie* **50**, 515–525 (2019).
18. Sheffield, C. S., Pindar, A., Packer, L. & Kevan, P. G. The potential of cleptoparasitic bees as indicator taxa for assessing bee communities. *Apidologie* **44**, 501–510 (2013).
19. Oertli, S., Müller, A. & Dorn, S. Ecological and seasonal patterns in the diversity of a species-rich bee assemblage (Hymenoptera: Apoidea: Apiformes). *Eur. J. Entomol.* **102**, 53–63 (2005).
20. Polidori, C., Borruso, L., Boesi, R. & Andrietti, F. Segregation of temporal and spatial distribution between kleptoparasites and parasitoids of the eusocial sweat bee, *Lasioglossum malachurum* (Hymenoptera: Halictidae, Mutillidae). *Entomol. Sci.* **12**, 116–129 (2009).
21. Alves-dos-Santos, I., Melo, G. A. R. & Rozen, J. G. Biology and immature stages of the Bee Tribe Tetrapediini (Hymenoptera: Apidae). *Am. Mus. Novit.* **3377**, 1–45 (2002).
22. Moure, J. S. Tetrapediini Michener & Moure, 1957. In *Catalogue of Bees (Hymenoptera, Apoidea) in the Neotropical Region—online version* (2012).
23. Roig-Alsina, A. *Coelioxoides* Cresson, a parasitic genus of Tetrapediini (Hymenoptera: Apoidea). *J. Kans. Entomol. Soc.* **63**, 279–287 (1990).
24. Danforth, B. N., Cardinal, S., Praz, C., Almeida, E. A. B. & Michez, D. The impact of molecular data on our understanding of bee phylogeny and evolution. *Annu. Rev. Entomol.* **58**, 57–78 (2013).
25. Martins, A. C., Luz, D. R. & Melo, G. A. R. Palaeocene origin of the Neotropical lineage of cleptoparasitic bees Ericrocidini-Rhathymini (Hymenoptera, Apidae). *Syst. Entomol.* **43**, 510–521 (2018).
26. Rocha-Filho, L. C. & Garófalo, C. A. Natural history of *Tetrapedia diversipes* (Hymenoptera: Apidae) in an Atlantic semideciduous forest remnant surrounded by coffee crops, *Coffea arabica* (Rubiaceae). *Ann. Entomol. Soc. Am.* **109**, 183–197 (2016).
27. Araújo, P. C. S., Lourenço, A. P. & Raw, A. Trap-nesting bees in montane grassland (Campo Rupestre) and Cerrado in Brazil: Collecting generalist or specialist nesters. *Neotrop. Entomol.* **45**, 482–489 (2016).
28. Lima, R., Oliveira, D. M. & Garófalo, C. A. Interaction network and niche analysis of natural enemy communities and their host bees (Hymenoptera: Apoidea) in fragments of Cerrado and Atlantic forest. *Sociobiology* **65**, 591 (2018).
29. Feng, W. *et al.* Understanding of waggle dance in the honey bee (*Apis mellifera*) from the perspective of long non-coding RNA. *Insects* **13**, 111 (2022).
30. Mondet, F. *et al.* Antennae hold a key to *Varroa*-sensitive hygiene behaviour in honey bees. *Sci. Rep.* **5**, 10454 (2015).
31. Fahad Raza, M. *et al.* Differential gene expression analysis following olfactory learning in honeybee (*Apis mellifera* L.). *PLoS ONE* **17**, e0262441 (2022).
32. Li, L. *et al.* Large-scale transcriptome changes in the process of long-term visual memory formation in the bumblebee, *Bombus terrestris*. *Sci. Rep.* **8**, 534 (2018).
33. Araujo, N. S., Santos, P. K. F. & Arias, M. C. RNA-Seq reveals that mitochondrial genes and long non-coding RNAs may play important roles in the bivoltine generations of the non-social Neotropical bee *Tetrapedia diversipes*. *Apidologie* **49**, 3–12 (2018).
34. Santos, P. K. F., de Souza Araujo, N., Françoso, E., Zuntini, A. R. & Arias, M. C. Diapause in a tropical oil-collecting bee: Molecular basis unveiled by RNA-Seq. *BMC Genom.* **19**, 305 (2018).
35. Araujo, N. S. & Arias, M. C. Gene expression and epigenetics reveal species-specific mechanisms acting upon common molecular pathways in the evolution of task division in bees. *Sci. Rep.* **11**, 3654 (2021).
36. Berens, A. J., Hunt, J. H. & Toth, A. L. Comparative transcriptomics of convergent evolution: Different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Mol. Biol. Evol.* **32**, 690–703 (2015).
37. Saleh, N. W. & Ramírez, S. R. Sociality emerges from solitary behaviours and reproductive plasticity in the orchid bee *Euglossa dilemma*. *Proc. R. Soc. B* **286**, 20190588 (2019).
38. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M. & Kelly, S. TransRate: Reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* **26**, 1134–1144 (2016).
39. Araujo, N. S., Zuntini, A. R. & Arias, M. C. Getting useful information from RNA-Seq contaminants: A case of study in the oil-collecting bee *Tetrapedia diversipes* transcriptome. *OMICS* **20**, 491–492 (2016).
40. Emms, D. M. & Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
41. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
42. Altschul, S. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
43. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
44. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
45. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
46. Zhou, Y. *et al.* A statistical normalization method and differential expression analysis for RNA-seq data between different species. *BMC Bioinform.* **20**, 163 (2019).
47. Tarazona, S. *et al.* Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* **43**, gkv711 (2015).

48. Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. Differential expression in RNA-seq: A matter of depth. *Genome Res.* **21**, 2213–2223 (2011).
49. Fraser, M. J., Smith, G. E. & Summers, M. D. Acquisition of host cell DNA sequences by baculoviruses: Relationship between Host DNA insertions and FP mutants of *Autographa californica* and *Galleria mellonella* nuclear polyhedrosis viruses. *J. Virol.* **47**, 287–300 (1983).
50. Cary, L. C. *et al.* Transposon mutagenesis of baculoviruses: Analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* **172**, 156–169 (1989).
51. Bouallègue, M., Rouault, J. D., Hua-Van, A., Makni, M. & Capy, P. Molecular evolution of piggyBac superfamily: From selfishness to domestication. *Genome Biol. Evol.* **9**, 323–339 (2017).
52. Volff, J.-N. Turning junk into gold: Domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays* **28**, 913–922 (2006).
53. Jangam, D., Feschotte, C. & Betrán, E. Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet* **33**, 817–831 (2017).
54. Sless, T. J. L., Searle, J. B. & Danforth, B. N. Genome of the bee *Holcopasites calliopsidis:* A species showing the common apid trait of brood parasitism. *G3* **12**, 160 (2022).
55. Ferguson, L. C., Green, J., Surridge, A. & Jiggins, C. D. Evolution of the insect *Yellow* gene family. *Mol. Biol. Evol.* **28**, 257–272 (2011).
56. Buttstedt, A., Moritz, R. F. A. & Erler, S. Origin and function of the major royal jelly proteins of the honeybee (*Apis mellifera*) as members of the *yellow* gene family. *Biol. Rev.* **89**, 255–269 (2014).
57. Hojo, M., Kagami, T., Sasaki, T., Nakamura, J. & Sasaki, M. Reduced expression of *major royal jelly protein 1* gene in the mushroom bodies of worker honeybees with reduced learning ability. *Apidologie* **41**, 194–202 (2010).
58. Kucharski, R., Maleszka, R., Hayward, D. C. & Ball, E. E. A royal jelly protein is expressed in a subset of kenyon cells in the mushroom bodies of the honey bee brain. *Naturwissenschaften* **85**, 343–346 (1998).
59. Dobritzsch, D., Aumer, D., Fuszard, M., Erler, S. & Buttstedt, A. The rise and fall of major royal jelly proteins during a honeybee (*Apis mellifera*) workers' life. *Ecol. Evol.* **9**, 8771–8782 (2019).
60. Winkler, P., Sieg, F. & Buttstedt, A. Transcriptional control of honey bee (*Apis mellifera*) major royal jelly proteins by 20-hydroxyecdysone. *Insects* **9**, 122 (2018).
61. Kupke, J., Spaethe, J., Mueller, M. J., Rössler, W. & Albert, Š. Molecular and biochemical characterization of the major royal jelly protein in bumblebees suggest a non-nutritive function. *Insect Biochem. Mol. Biol.* **42**, 647–654 (2012).
62. Fratini, F., Cilia, G., Mancini, S. & Felicioli, A. Royal Jelly: An ancient remedy with remarkable antibacterial properties. *Microbiol. Res.* **192**, 130–141 (2016).
63. Park, H. G. *et al.* Antibacterial activity of major royal jelly proteins of the honeybee (*Apis mellifera*) royal jelly. *J. Asia Pac. Entomol.* **22**, 737–741 (2019).
64. Šimúth, J. Some properties of the main protein of honeybee (*Apis mellifera*) royal jelly. *Apidologie* **32**, 69–80 (2001).
65. Hassani-Pak, K. & Rawlings, C. Knowledge discovery in biological databases for revealing candidate genes linked to complex phenotypes. *J. Integr. Bioinform.* **14**, 20160002 (2017).
66. Gaudet, P. & Dessimoz, C. Gene ontology: Pitfalls, biases, and remedies. *Methods Mol. Biol.* **1446**, 189–205 (2017).
67. Koonin, E. V. Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* **39**, 309–338 (2005).
68. Cane, J. H. Olfactory evaluation of *Andrena* host nest suitability by kleptoparasitic *Nomada* bees (Hymenoptera: Apoidea). *Anim. Behav.* **31**, 138–144 (1983).
69. Dotterl, S. Antennal responses of an oligolectic bee and its cleptoparasite to plant volatiles. *Plant Signal. Behav.* **3**, 296–297 (2008).
70. Galvani, G. L., González-Vaquero, R. A., Guerra-Navarro, C. & Settembrini, B. P. Antennal sensilla of cleptoparasitic and non-parasitic bees in two subfamilies of Apidae. *Apidologie* **48**, 437–449 (2017).
71. Qi, Y. *et al.* Transcriptome analysis of an endoparasitoid wasp *Cotesia chilonis* (Hymenoptera: Braconidae) reveals genes involved in successful parasitism. *Arch Insect Biochem. Physiol.* **88**, 203–221 (2015).
72. Zhou, C.-X., Min, S.-F., Yan-Long, T. & Wang, M.-Q. Analysis of antennal transcriptome and odorant binding protein expression profiles of the recently identified parasitoid wasp, *Sclerodermus* sp. *Comp. Biochem. Physiol. Part D* **16**, 10–19 (2015).
73. Zhao, Y. *et al.* Transcriptome and expression patterns of chemosensory genes in antennae of the parasitoid wasp *Chouioia cunea*. *PLoS ONE* **11**, e0148159 (2016).
74. Nie, X. P. *et al.* Antennal transcriptome and odorant binding protein expression profiles of an invasive mealybug and its parasitoid. *J. Appl. Entomol.* **142**, 149–161 (2018).
75. Leavell, B. C. & Bernal, X. E. The cognitive ecology of stimulus ambiguity: A predator-prey perspective. *Trends Ecol. Evol.* **34**, 1048–1060 (2019).
76. Haverkamp, A. & Smid, H. M. A neuronal arms race: The role of learning in parasitoid–host interactions. *Curr. Opin. Insect Sci.* **42**, 47–54 (2020).
77. Rimal, S. & Lee, Y. The multidimensional ionotropic receptors of *Drosophila melanogaster*. *Insect. Mol. Biol.* **27**, 1–7 (2018).
78. van Giesen, L. & Garrity, P. A. More than meets the IR: The expanding roles of variant Ionotropic Glutamate Receptors in sensing odor, taste, temperature and moisture. *F1000Res* **6**, 1753 (2017).
79. Eyun, S. *et al.* Evolutionary history of chemosensory-related gene families across the arthropoda. *Mol. Biol. Evol.* **34**, 1838–1862 (2017).
80. Breer, H., Fleischer, J., Pregitzer, P. & Krieger, J. Molecular mechanism of insect olfaction: Olfactory receptors. In *Olfactory Concepts of Insect Control: Alternative to insecticides* 93–114 (Springer International Publishing, 2019).
81. Johnson, B. R., Atallah, J. & Plachetzki, D. C. The importance of tissue specificity for RNA-seq: Highlighting the errors of composite structure extractions. *BMC Genom.* **14**, 586 (2013).
82. Sinakevitch, I. T., Wolff, G. H., Pflüger, H.-J. & Smith, B. H. Editorial: Biogenic amines and neuromodulation of animal behavior. *Front. Syst. Neurosci.* **12**, 31 (2018).
83. Blenau, W. & Baumann, A. Octopaminergic and tyraminergic signaling in the honeybee (*Apis mellifera*) Brain. In *Trace Amines and Neurological Disorders* 203–219 (Elsevier, 2016).
84. Manfredini, F., Brown, M. J. F. & Toth, A. L. Candidate genes for cooperation and aggression in the social wasp *Polistes dominula*. *J. Comp. Physiol. A* **204**, 449–463 (2018).
85. Burke, C. J. *et al.* Layered reward signalling through octopamine and dopamine in *Drosophila*. *Nature* **492**, 433–437 (2012).
86. Agarwal, M. *et al.* Dopamine and octopamine influence avoidance learning of honey bees in a place preference assay. *PLoS ONE* **6**, e25371 (2011).
87. Farooqui, T. A potential link among biogenic amines-based pesticides, learning and memory, and colony collapse disorder: A unique hypothesis. *Neurochem. Int.* **62**, 122–136 (2013).
88. Monastirioti, M. Biogenic amine systems in the fruit fly *Drosophila melanogaster*. *Microsc. Res. Tech.* **45**, 106–121 (1999).
89. Dacks, A. M., Green, D. S., Root, C. M., Nighorn, A. J. & Wang, J. W. Serotonin modulates olfactory processing in the antennal lobe of *Drosophila*. *J. Neurogenet.* **23**, 366–377 (2009).
90. Zhang, X. & Gaudry, Q. Functional integration of a serotonergic neuron in the *Drosophila* antennal lobe. *Elife* **5**, 16836 (2016).
91. Ro, J. *et al.* Serotonin signaling mediates protein valuation and aging. *Elife* **5**, 16843 (2016).

92. Neckameyer, W. S., Coleman, C. M., Eadie, S. & Goodwin, S. F. Compartmentalization of neuronal and peripheral serotonin synthesis in *Drosophila melanogaster*. *Genes Brain Behav.* **6**, 756–769 (2007).
93. Kamyshev, N. G., Smirnova, G. P., Savvateeva, E. V., Medvedeva, A. V. & Ponomarenko, V. V. The influence of serotonin and p-chlorophenylalanine on locomotor activity of *Drosophila melanogaster*. *Pharmacol. Biochem. Behav.* **18**, 677–681 (1983).
94. Alekseyenko, O. V. *et al.* Single serotonergic neurons that modulate aggression in *Drosophila*. *Curr. Biol.* **24**, 2700–2707 (2014).
95. Pooryasin, A. & Fiala, A. Identified serotonin-releasing neurons induce behavioral quiescence and suppress mating in *Drosophila*. *J. Neurosci.* **35**, 12792–12812 (2015).
96. Tierney, A. J. Structure and function of invertebrate 5-HT receptors: A review. *Comp. Biochem. Physiol. A* **128**, 791–804 (2001).
97. Barnes, N. M. & Sharp, T. A review of central 5-HT receptors and their function. *Neuropharmacology* **38**, 1083–1152 (1999).
98. Nichols, C. D. 5-HT_2 receptors in *Drosophila* are expressed in the brain and modulate aspects of circadian behaviors. *Dev. Neurobiol.* **67**, 752–763 (2007).
99. Yuan, Q., Joiner, W. J. & Sehgal, A. A sleep-promoting role for the *Drosophila* serotonin receptor 1A. *Curr. Biol.* **16**, 1051–1062 (2006).
100. Johnson, O., Becnel, J. & Nichols, C. D. Serotonin 5-HT_2 and 5-HT_{1A}-like receptors differentially modulate aggressive behaviors in *Drosophila melanogaster*. *Neuroscience* **158**, 1292–1300 (2009).
101. Lyu, Y. *et al. Drosophila* serotonin 2A receptor signaling coordinates central metabolic processes to modulate aging in response to nutrient choice. *Elife* **10**, 1–67 (2021).
102. Johnson, O., Becnel, J. & Nichols, C. D. Serotonin receptor activity is necessary for olfactory learning and memory in *Drosophila melanogaster*. *Neuroscience* **192**, 372–381 (2011).
103. Becnel, J., Johnson, O., Luo, J., Nässel, D. R. & Nichols, C. D. The serotonin 5-HT_7Dro receptor is expressed in the brain of *Drosophila*, and is essential for normal courtship and mating. *PLoS ONE* **6**, e20800 (2011).
104. Goetz, T., Arslan, A., Wisden, W. & Wulff, P. $GABA_A$ receptors: Structure and function in the basal ganglia. *Prog. Brain Res.* **160**, 21–41 (2007).
105. Lummis, S. C. R. GABA receptors in insects. *Comp. Biochem. Physiol. C* **95**, 1–8 (1990).
106. Mustard, J. A., Jones, L. & Wright, G. A. GABA signaling affects motor function in the honey bee. *J. Insect Physiol.* **120**, 103989 (2020).
107. Leal, S. M. & Neckameyer, W. S. Pharmacological evidence for GABAergic regulation of specific behaviors in *Drosophila melanogaster*. *J. Neurobiol.* **50**, 245–261 (2002).
108. Raccuglia, D. & Mueller, U. Temporal integration of cholinergic and GABAergic inputs in isolated insect mushroom body neurons exposes pairing-specific signal processing. *J. Neurosci.* **34**, 16086–16092 (2014).
109. Choudhary, A. F., Laycock, I. & Wright, G. A. γ-Aminobutyric acid receptor A-mediated inhibition in the honeybee's antennal lobe is necessary for the formation of configural olfactory percepts. *Eur. J. Neurosci.* **35**, 1718–1724 (2012).
110. Dupuis, J. P. *et al.* Homomeric RDL and heteromeric RDL/LCCH3 GABA receptors in the honeybee antennal lobes: Two candidates for inhibitory transmission in olfactory processing. *J. Neurophysiol.* **103**, 458–468 (2010).
111. Bettler, B., Kaupmann, K., Mosbacher, J. & Gassmann, M. Molecular structure and physiological functions of $GABA_B$ receptors. *Physiol. Rev.* **84**, 835–867 (2004).
112. Pinard, A., Seddik, R. & Bettler, B. GABAB receptors: Physiological functions and mechanisms of diversity. *Adv. Pharmacol.* **58**, 231–255 (2010).
113. Pregitzer, P., Schultze, A., Raming, K., Breer, H. & Krieger, J. Expression of a $GABA_B$: Receptor in olfactory sensory neurons of sensilla trichodea on the male antenna of the Moth *Heliothis virescens*. *Int. J. Biol. Sci.* **9**, 707–715 (2013).
114. Gosselin-Badaroudine, P. *et al.* Biophysical characterization of the honeybee DSC1 orthologue reveals a novel voltage-dependent Ca^{2+} channel subfamily: Ca_v4. *J. Gen. Physiol.* **148**, 133–145 (2016).
115. Hong, C. & Ganetzky, B. Spatial and temporal expression patterns of two sodium channel genes in *Drosophila*. *J. Neurosci.* **14**, 5160–5169 (1994).
116. Liebeskind, B. J., Hillis, D. M. & Zakon, H. H. Evolution of sodium channels predates the origin of nervous systems in animals. *Proc. Natl. Acad. Sci. USA* **108**, 9154–9159 (2011).
117. Kulkarni, N. H., Yamamoto, A. H., Robinson, K. O., Mackay, T. F. C. & Anholt, R. R. H. The DSC1 channel, encoded by the smi60E locus, contributes to odor-guided behavior in *Drosophila melanogaster*. *Genetics* **161**, 1507–1516 (2002).
118. Castella, C., Amichot, M., Bergé, J.-B. & Pauron, D. DSC1 channels are expressed in both the central and the peripheral nervous system of adult *Drosophila melanogaster*. *Invertebr. Neurosci.* **4**, 85–94 (2001).
119. Santi, C. M. Opposite regulation of slick and slack K^+ channels by neuromodulators. *J. Neurosci.* **26**, 5059–5068 (2006).
120. Budelli, G. *et al.* SLO2 channels are inhibited by all divalent cations that activate SLO1 K^+ channels. *J. Biol. Chem.* **291**, 7347–7356 (2016).
121. Goldstein, S. A. N., Bockenhauer, D., O'Kelly, I. & Zilberberg, N. Potassium leak channels and the KCNK family of two-p-domain subunits. *Nat. Rev. Neurosci.* **2**, 175–184 (2001).
122. Talley, E. M. & Bayliss, D. A. Modulation of TASK-1 (Kcnk3) and TASK-3 (Kcnk9) potassium channels. *J. Biol. Chem.* **277**, 17733–17742 (2002).
123. Döring, F., Scholz, H., Kühnlein, R. P., Karschin, A. & Wischmeyer, E. Novel *Drosophila* two-pore domain K^+ channels: Rescue of channel function by heteromeric assembly. *Eur. J. Neurosci.* **24**, 2264–2274 (2006).
124. Lear, B. C. *et al.* UNC79 and UNC80, putative auxiliary subunits of the NARROW ABDOMEN ion channel, are indispensable for robust circadian locomotor rhythms in *Drosophila*. *PLoS ONE* **8**, e78147 (2013).
125. Flourakis, M. *et al.* A conserved bicycle model for circadian clock control of membrane excitability. *Cell* **162**, 836–848 (2015).
126. Shi, Y. *et al.* Nalcn is a 'leak' sodium channel that regulates excitability of brainstem chemosensory neurons and breathing. *J. Neurosci.* **36**, 8174–8187 (2016).
127. Sekine, Y. *et al.* p38 MAPKs regulate the expression of genes in the dopamine synthesis pathway through phosphorylation of NR4A nuclear receptors. *J. Cell Sci.* **124**, 3006–3016 (2011).
128. Kozlova, T., Lam, G. & Thummel, C. S. *Drosophila* DHR38 nuclear receptor is required for adult cuticle integrity at eclosion. *Dev. Dyn.* **238**, 701–707 (2009).
129. Baker, K. D. *et al.* The *Drosophila* orphan nuclear receptor DHR38 mediates an atypical ecdysteroid signaling pathway. *Cell* **113**, 731–742 (2003).
130. Fujita, N. *et al.* Visualization of neural activity in insect brains using a conserved immediate early gene, *Hr38*. *Curr. Biol.* **23**, 2063–2070 (2013).
131. Takayanagi-Kiya, S. & Kiya, T. Activity-dependent visualization and control of neural circuits for courtship behavior in the fly *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **116**, 5715–5720 (2019).
132. Liu, W. *et al.* Dibutyl phthalate disrupts conserved circadian rhythm in *Drosophila* and human cells. *Sci. Total Environ.* **783**, 147038 (2021).
133. Clevers, H. & Nusse, R. Wnt/β-catenin signaling and disease. *Cell* **149**, 1192–1205 (2012).
134. Koizumi, K. *et al.* RNA interference screen to identify genes required for *Drosophila* embryonic nervous system development. *Proc. Natl. Acad. Sci. USA* **104**, 5626–5631 (2007).
135. Tan, Y., Yu, D., Busto, G. U., Wilson, C. & Davis, R. L. Wnt signaling is required for long-term memory formation. *Cell Rep.* **4**, 1082–1089 (2013).

136. Banerjee, S., Venkatesan, A. & Bhat, M. A. Neurexin, Neuroligin and Wishful Thinking coordinate synaptic cytoarchitecture and growth at neuromuscular junctions. *Mol. Cell. Neurosci.* **78**, 9–24 (2017).
137. Larkin, A. *et al.* Neurexin-1 regulates sleep and synaptic plasticity in *Drosophila melanogaster*. *Eur. J. Neurosci.* **42**, 2455–2466 (2015).
138. Biswas, S. *et al.* Sensory regulation of *Neuroligins* and *Neurexin I* in the honeybee brain. *PLoS ONE* **5**, e9133 (2010).
139. Usui, T. *et al.* Flamingo, a seven-pass transmembrane cadherin, regulates planar cell polarity under the control of frizzled. *Cell* **98**, 585–595 (1999).
140. Hakeda-Suzuki, S. *et al.* Golden Goal collaborates with Flamingo in conferring synaptic-layer specificity in the visual system. *Nat. Neurosci.* **14**, 314–323 (2011).
141. Kimura, H., Usui, T., Tsubouchi, A. & Uemura, T. Potential dual molecular interaction of the *Drosophila* 7-pass transmembrane cadherin Flamingo in dendritic morphogenesis. *J. Cell Sci.* **119**, 1118–1129 (2006).
142. Reuter, J. E. *et al.* A mosaic genetic screen for genes necessary for *Drosophila* mushroom body neuronal morphogenesis. *Development* **130**, 1203–1213 (2003).
143. Lebovitz, R. M., Takeyasu, K. & Fambrough, D. M. Molecular characterization and expression of the ($Na^+ + K^+$)-ATPase alpha-subunit in *Drosophila melanogaster*. *EMBO J.* **8**, 193–202 (1989).
144. Sun, B. & Salvaterra, P. M. Two *Drosophila* nervous system antigens, Nervana 1 and 2, are homologous to the beta subunit of Na^+, K^+-ATPase. *Proc. Natl. Acad. Sci. USA* **92**, 5396–5400 (1995).
145. Palladino, M. J., Bower, J. E., Kreber, R. & Ganetzky, B. Neural dysfunction and neurodegeneration in *Drosophila* Na^+ /K^+ ATPase alpha subunit mutants. *J. Neurosci.* **23**, 1276–1286 (2003).
146. Roy, M., Sivan-Loukianova, E. & Eberl, D. F. Cell-type–specific roles of Na^+ /K^+ ATPase subunits in *Drosophila* auditory mechanosensation. *Proc. Natl. Acad. Sci. USA* **110**, 181–186 (2013).
147. Damulewicz, M., Rosato, E. & Pyza, E. Circadian regulation of the Na^+/K^+-Atpase alpha subunit in the visual system is mediated by the pacemaker and by retina photoreceptors in *Drosophila melanogaster*. *PLoS ONE* **8**, e73690 (2013).
148. Trotta, N., Rodesch, C. K., Fergestad, T. & Broadie, K. Cellular bases of activity-dependent paralysis in *Drosophila* stress-sensitive mutants. *J. Neurobiol.* **60**, 328–347 (2004).
149. Dahanukar, A., Lei, Y.-T., Kwon, J. Y. & Carlson, J. R. Two Gr genes underlie sugar reception in *Drosophila*. *Neuron* **56**, 503–516 (2007).
150. Ueno, K. *et al.* Trehalose sensitivity in *Drosophila* correlates with mutations in and expression of the gustatory receptor gene *Gr5a*. *Curr. Biol.* **11**, 1451–1455 (2001).
151. Jiao, Y., Moon, S. J., Wang, X., Ren, Q. & Montell, C. Gr64f is required in combination with other gustatory receptors for sugar detection in *Drosophila*. *Curr. Biol.* **18**, 1797–1801 (2008).
152. Jung, J. W., Park, K. W., Ahn, Y.-J. & Kwon, H. W. Functional characterization of sugar receptors in the western honeybee, *Apis mellifera*. *J. Asia Pac. Entomol.* **18**, 19–26 (2015).
153. Rützler, M., Lu, T. & Zwiebel, L. J. G_α encoding gene family of the malaria vector mosquito *Anopheles gambiae*: Expression analysis and immunolocalization of $AG_{\alpha q}$ and $AG_{\alpha o}$ in female antennae. *J. Comp. Neurol.* **499**, 533–545 (2006).
154. Ueno, K. *et al. Gsα* is involved in sugar perception in *Drosophila melanogaster*. *J. Neurosci.* **26**, 6143–6152 (2006).
155. Papazian, D. M., Schwarz, T. L., Tempel, B. L., Jan, Y. N. & Jan, L. Y. Cloning of genomic and complementary DNA from *Shaker*, a putative potassium channel gene from *Drosophila*. *Science* **1979**(237), 749–753 (1987).
156. Wicher, D., Walther, C. & Wicher, C. Non-synaptic ion channels in insects: Basic properties of currents and their modulation in neurons and skeletal muscles. *Prog. Neurobiol.* **64**, 431–525 (2001).
157. Mottes, J. R. & Iverson, L. E. Tissue-specific alternative splicing of hybrid *Shaker*/*lacZ* genes correlates with kinetic differences in Shaker K^+ currents in vivo. *Neuron* **14**, 613–623 (1995).
158. Rogero, O., Hämmerle, B. & Tejedor, F. J. Diverse expression and distribution of *Shaker* potassium channels during the development of the *Drosophila* nervous system. *J. Neurosci.* **17**, 5108–5118 (1997).
159. Cowan, T. M. & Siegel, R. W. *Drosophila* mutations that alter ionic conduction disrupt acquisition and retention of a conditioned odor avoidance response. *J. Neurogenet.* **3**, 187–201 (1986).
160. Jespersen, T., Grunnet, M. & Olesen, S.-P. The KCNQ1 potassium channel: From gene to physiological function. *Physiology* **20**, 408–416 (2005).
161. Wen, H. *et al.* A *Drosophila* KCNQ channel essential for early embryonic development. *J. Neurosci.* **25**, 10147–10156 (2005).
162. Ocorr, K. *et al.* KCNQ potassium channel mutations cause cardiac arrhythmias in *Drosophila* that mimic the effects of aging. *Proc. Natl. Acad. Sci. USA* **104**, 3943–3948 (2007).
163. Cavaliere, S., Malik, B. R. & Hodge, J. J. L. KCNQ channels regulate age-related memory impairment. *PLoS ONE* **8**, e62445 (2013).
164. Agam, K. *et al.* Metabolic stress reversibly activates the *Drosophila* light-sensitive channels TRP and TRPL *in vivo*. *J. Neurosci.* **20**, 5748–5755 (2000).
165. Badsha, F. *et al.* Mutants in *Drosophila* TRPC channels reduce olfactory sensitivity to carbon dioxide. *PLoS ONE* **7**, e49848 (2012).
166. Gertler, F. B., Bennett, R. L., Clark, M. J. & Hoffmann, F. M. *Drosophila* abl tyrosine kinase in embryonic CNS axons: A role in axonogenesis is revealed through dosage-sensitive interactions with disabled. *Cell* **58**, 103–113 (1989).
167. Xiong, W., Dabbouseh, N. M. & Rebay, I. Interactions with the abelson tyrosine kinase reveal compartmentalization of eyes absent function between nucleus and cytoplasm. *Dev. Cell* **16**, 271–279 (2009).
168. Kawasaki, F. Active zone localization of presynaptic calcium channels encoded by the cacophony locus of *Drosophila*. *J. Neurosci.* **24**, 282–285 (2004).
169. Sun, Q., Schindelholz, B., Knirr, M., Schmid, A. & Zinn, K. Complex genetic interactions among four receptor tyrosine phosphatases regulate axon guidance in *Drosophila*. *Mol. Cell. Neurosci.* **17**, 274–291 (2001).
170. Pennacchio, L. A. *et al.* Structure, sequence and location of the *UQCRFS1* gene for the human Rieske Fe-S protein. *Gene* **155**, 207–211 (1995).
171. Saraste, M. Oxidative phosphorylation at the *fin de siècle*. *Science* **1979**(283), 1488–1493 (1999).
172. Henriques, B. J., Katrine Jentoft Olsen, R., Gomes, C. M. & Bross, P. Electron transfer flavoprotein and its role in mitochondrial energy metabolism in health and disease. *Gene* **776**, 145407 (2021).
173. Marygold, S. J. *et al.* The ribosomal protein genes and *Minute* loci of *Drosophila melanogaster*. *Genome Biol.* **8**, R216 (2007).
174. Wang, L., Lam, G. & Thummel, C. S. *Med24* and *Mdh2* are required for *Drosophila* larval salivary gland cell death. *Dev. Dyn.* **239**, 954–964 (2010).
175. Grossman, L. I., Rosenthal, N. H., Akamatsu, M. & Erickson, R. P. Cloning, sequence analysis, and expression of a mouse cDNA encoding cytochrome c oxidase subunit VIa liver isoform. *Biochim. Biophys. Acta* **1260**, 361–364 (1995).
176. Dunham-Snary, K. J. *et al.* Ndufs2, a core subunit of mitochondrial complex I, is essential for acute oxygen-sensing and hypoxic pulmonary vasoconstriction. *Circ Res* **124**, 1727–1746 (2019).
177. Gauthier, M. & Grünewald, B. Neurotransmitter systems in the honey bee brain: Functions in learning and memory. In *Honeybee Neurobiology and Behavior* 155–169 (Springer, 2012).
178. Alekseyenko, O. V. *et al.* Serotonergic modulation of aggression in *Drosophila* involves GABAergic and cholinergic opposing pathways. *Curr. Biol.* **29**, 2145-2156.e5 (2019).

179. Tikhonov, D. B. & Magazanik, L. G. Origin and molecular evolution of ionotropic glutamate receptors. *Neurosci. Behav. Physiol.* **39**, 763–773 (2009).
180. Li, Y. *et al.* Novel functional properties of *Drosophila* CNS glutamate receptors. *Neuron* **92**, 1036–1048 (2016).
181. Frenkel, L. *et al.* Organization of circadian behavior relies on glycinergic transmission. *Cell Rep.* **19**, 72–85 (2017).
182. Henry, C. *et al.* Heterogeneous expression of GABA receptor-like subunits LCCH3 and GRD reveals functional diversity of GABA receptors in the honeybee *Apis mellifera*. *Br. J. Pharmacol.* **177**, 3924–3940 (2020).
183. Bastian, F. B. *et al.* The Bgee suite: Integrated curated expression atlas and comparative transcriptomics in animals. *Nucleic Acids Res.* **49**, D831–D847 (2021).
184. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2018).
185. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
186. Grabherr, M. G. *et al.* Trinity: Reconstructing a full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2013).
187. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
188. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
189. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
190. Davidson, N. M., Hawkins, A. D. K. & Oshlack, A. SuperTranscripts: A data driven reference for analysis and visualisation of transcriptomes. *Genome Biol.* **18**, 148 (2017).
191. Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–682 (2010).
192. Davidson, N. M. & Oshlack, A. Corset: Enabling differential gene expression analysis for de novoassembled transcriptomes. *Genome Biol.* **15**, 410 (2014).
193. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
194. Musacchia, F., Basu, S., Petrosino, G., Salvemini, M. & Sanges, R. Annocript: A flexible pipeline for the annotation of transcriptomes able to identify putative long noncoding RNAs. *Bioinformatics* **31**, 2199–2201 (2015).
195. Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B. & Wu, C. H. UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).
196. Bairoch, A. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Res.* **24**, 21–25 (1996).
197. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
198. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323–333 (2011).
199. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: Batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform.* **2**, 1–10 (2020).
200. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
201. Pascovici, D. & Wu, J. *PloGO2: Plot Gene Ontology and KEGG Pathway Annotation and Abundance. R Package Version 1.4.0* (2021).
202. Alexa, A. & Rahnenfuhrer, J. *topGO: Enrichment Analysis for Gene Ontology. R Package Version 2.20.0* (2020).
203. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).
204. Kuznetsova, I., Lugmayr, A., Siira, S. J., Rackham, O. & Filipovska, A. CirGO: An alternative circular way of visualising gene ontology terms. *BMC Bioinform.* **20**, 84 (2019).

Acknowledgements

The authors acknowledge Susy Coelho Oliveira and Thiago Geronimo Pires Alegria for technical assistance, Isabel Alves dos-Santos and Guaraci Duran Cordeiro for support in specimen sampling, and Priscila Karla Ferreira dos Santos for initial assistance in data analysis.

Author contributions

P.C.R.: Methodology, investigation, data analyses and interpretation. N.S.A.: Conceptualization, methodology, sampling, assistance in data analyses and interpretation. M.C.A.: Conceptualization, supervision, fund acquisition. All authors contributed to the preparation of the manuscript.

Funding

This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—CAPES, Brasil (scholarship to PCR, Proc. 1783961 and 88882.377416/2019-01, Finance Code 001), CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico (research sponsorship to MCA—306932/2016-4) and Fundação de Amparo à Pesquisa do Estado de São Paulo—FAPESP (Proc. 2013/12530-4, 2016/24669-5 and 2019/23186-9). NSA is supported by the Fonds de la Recherche Scientifique—FNRS (Grant ID 40005980).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-56261-5.

Correspondence and requests for materials should be addressed to P.C.R.