Title

Mechanisms of Novel Host Use by Bactrocera tau (Tephritid: Diptera) Revealed by RNA Transcriptomes

Journal

Journal of Insect Science, 20(5)

ISSN

1536-2442

Authors

Shi, Wei
Roderick, George
Zhang, Gen-Song

Publication Date

2020-09-01

DOI

10.1093/jisesa/ieaa102

Peer reviewed

Journal of Insect Science, (2020) 20(5): 22; 1–10
doi: 10.1093/jisesa/ieaa102
Research

Mechanisms of Novel Host Use by *Bactrocera tau* (Tephritid: Diptera) Revealed by RNA Transcriptomes

Wei Shi,[1,3] George Roderick,[2] and Gen-Song Zhang[1]

[1]School of Ecology and Environment Science, Yunnan University, 650091 Kunming, China, [2]Department of Environmental Science Policy and Management, University of California Berkeley, Berkeley, CA 94720, and [3]Corresponding author, e-mail: shiwei55@126.com

Subject Editor: Juan Rull

Received 11 May 2020; Editorial decision 21 August 2020

Abstract

Use of novel plant hosts can facilitate the establishment and range expansion of herbivorous invasive species. However, the inherent mechanisms of novel host use are still unclear in many herbivorous species. Here, we examine mechanisms of novel host use in the invasive tephritid fruit fly *Bactrocera tau* (Walker)(Diptera:Tephritidae) by documenting changes in the RNA transcriptomes associated with a novel host. RNA transcripts of *B. tau* were obtained with high-throughput sequencing from samples continuously reared on two traditional Cucurbitaceae hosts and a novel host (banana). We found transcriptome variation was strongly associated with feeding on banana. Moreover, *B. tau* feeding on banana contained more differentially expressed genes (DEGs) and more annotated categories of DEGs in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database with 1,595 DEGs and 21 major annotated pathways. The annotated categories of DEGs in individuals reared on banana differed with from those individuals feeding on other hosts and were enriched in oxidative phosphorylation, citrate cycle pathway, and four other carbohydrate pathways. For *B. tau* feeding on banana, the predominant numbers of upregulated genes in the mitochondrial *NADH* (56 on average) and a relatively higher numbers of upregulated genes (13 on average) were found in oxidative phosphorylation and the TCA pathway, respectively. Changes in RNA transcriptomes associated with novel host use, especially for genes related to energy and carbohydrate metabolism, help to explain how *B. tau* can be successful in use of novel hosts and may be useful in developing novel strategies for control of tephritid flies.

Key words: *Bactrocera tau*, novel host use, oxidative phosphorylation, RNA transcriptome, TCA pathway

Use of novel hosts allows herbivorous invasive insects to expand their range and potentially diverge genetically from source populations (Charlery de la Massellère et al. 2017). When herbivorous insects enter new environments, exploitation of novel hosts can facilitate colonization, establishment, and maintenance of stable populations (Cronin and Abrahamson 2001). Therefore, understanding how insect invaders use novel hosts is critical to understanding the inherent mechanisms of biological invasion for many invasive species.

Tephritid fruit flies include numerous well-known invasive species and are found in many regions worldwide (Clarke 2017). Invasion of tephritid flies to new regions appears to be closely associated with the ability to adapt to novel hosts and there are numerous cases that have demonstrated that novel hosts can influence colonization and distribution. For example, the feeding on novel host honeysuckle (Caprifoliaceae: *Lonicera*) allowed the successful expansion of *Rhagoletis pomonella* (Walsh) in the United States (Mattsson et al. 2015). Similarity, using papaya (*Carica papaya*) facilitated the successful introduction of *Bactrocera cucurbitae* (Coquillett) into Hawaii (Piñero et al. 2017).

The use and/or eventual adaptation to novel hosts involves several processes, including identification of the hosts by the flies (Tallamy 2000), physiological and chemical adaptation to nutritional composition and secondary metabolites of hosts (Charlery de la Massellère et al. 2017), as well as host-finding behavior mediated by neurophysiology (Tallamy 2000). Each of these processes leading to host adaptation has a genetic basis. For example, in the cotton bollworm, *Helicoverpa*, genes encoding serine proteases are fundamental in adaptation to different hosts (Chikate et al. 2013).

In studies of tephritid fruit flies, previous research has focused on single processes or single genes in host adaptation. For instance, the *Rh6* gene was found to play an important role in host locating in *Bactrocera minax* (Enderiein) (Wang et al. 2019). *Trypsin* genes are frequently associated with host digestion in *Bactrocera dorsalis* (Hendel) (Li et al. 2017), and olfactory receptor genes and *OBPs* genes (odorant-binding proteins genes) are regarded as exerting key roles in host odor identification in *B. dorsalis* (Li et al. 2017, Miyazaki et al. 2018). However, studies of many species show that the use of novel hosts is a complex process regulated by multiple genes. For example, in novel host shifts of *Polygonia c-album*

(Lepidoptera: Nymphalidae), ribosome, digestion, and detoxification genes jointly affected host use (Celorio-Mancera et al. 2013). Still, there are very few comprehensive, multi-locus studies of the mechanisms of novel host use for tephritid flies.

High-throughput DNA sequencing is a highly efficient technology to analyze genes and related metabolic pathways based on RNA transcriptomes (Guo et al. 2018). This technology has been applied extensively in elucidating the mechanisms of environmental adaption in insects (Barth et al. 2018). For tephritid flies, high-throughput sequencing has been used to reveal genetic mechanisms associated with pupariation (Chen et al. 2018), flight capability (Guo et al. 2018), and mating (Nagalingam et al. 2018). However, research using this technology to reveal the underlying causes of novel host use is still rare.

Bactrocera tau is an economically important tephritid species of and is distributed throughout tropical and subtropical Asia and the South Pacific (Singh et al. 2010). The species has been listed as an important quarantine pest by many countries (Ooi and Wee 2016). *Bactrocera tau* mainly attacks hosts of the family Cucurbitaceae (Christenson and Foote 1960). In recent years, the species began to infest novel hosts from other plant families including Leguminosae (e.g., *Phaseolus vulgaris* L.) (Sumrandee et al. 2011), Myrtaceae (e.g., *Psidium guajava* L.) (Hasyim et al. 2008), and Rutaceae (e.g., *Citrus*) (Zhang and Chen 2012). The host range of *B. tau* has now expanded to more than 80 plants species (Huang et al. 2005).

In China, *B. tau* was geographically restricted to southern regions (Shi et al. 2014), but recently has expanded to central and northern China, including Shanxi and Shaanxi provinces (Huang et al. 2005). During its expansion, *B. tau* began to use other hosts in addition to traditional Cucurbitaceae hosts. For example, in central China, the fly attacked carambola (*Averrhoa carambola L.*) (Huang et al. 2005) and in the north of China, orange (*Citrus*) (Zhang and Chen 2012).

We hypothesize that novel host use of *B. tau* is key for its success as an invader and subsequent geographical spread, yet mechanisms of adaptation are not yet clear. Here, we address this problem using high-throughput sequencing technology. We manipulated changes in hosts use from traditional to novel hosts and documented variation in transcriptomes. For invasive herbivorous species, such as *B. tau*, understanding the metabolism of host use will be critical in understanding processes of range expansion and subsequent adaption but also for developing new methods of control.

Materials and Methods

Bactrocera tau Collection and Rearing

Summer squash (*Cucurbita pepo* var. *fastigata* L.) infested by *B. tau* in Yuanjiang river valley were brought into the laboratory at Yunnan University, China. After 2 wk of rearing, more than 300 *B. tau* emerging from summer squash were collected and used in the following studies.

The flies were divided into four experimental units or treatments. Each unit was fed one of following host fruits: cucumber (*Cucumis sativus* L.), pumpkin (*Cucurbita pepo* var. *pepo* L.), orange (*Citrus*), or banana (*Musa paradisiaca* Colla), following methods described by Shi et al. (2017). Cucumber and pumpkin are traditional Cucurbitaceae hosts for *B. tau*, while orange and banana are novel hosts for the fly. About 60 adult flies were used in each treatment. The flies were reared continuously on each of the four host plants. Rearing was conducted in insect-rearing cages (60 × 40 × 45 cm), which were placed in an environmental chamber maintained at 27 ± 1°C, 70 ± 5% relative humidity, and a photoperiod of L12:D12 h (lights 0700–1900). After 10 d of adult emergence in each generation, we kept 60 fly adults to initiate the next generation. Each treatment was replicated three times. *Bactrocera tau* feeding on cucumber, pumpkin, and banana were reared continuously for 15 generations, while the flies reared on orange died within the first 3 generations and were excluded from the study.

Over 60 larvae of *B. tau* were collected from each of the 10th and 15th generations, respectively, for each host unit treatment. The larvae collected were snap-frozen in liquid nitrogen and immediately stored at −80°C until RNA isolation. This resulted in six sample groups: C10 (the 10th generation feeding on cucumber), B10 (the 10th generation feeding on banana), and P10 (the 10th generation feeding on pumpkin), as well as the corresponding C15, B15, and P15, generations, respectively.

De Novo Transcriptome Sequencing

Each of six groups mentioned above was sampled by pooling of 30 larvae, which were homogenized under liquid nitrogen in a mortar. Whole RNA was isolated from each of six larval homogenizations with the RNeasy plus Micro Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. A second replicate was completed for each group, resulting in 12 RNA pools. Sequencing of RNA purification, quality control, and quantification were performed according to the description by Guo et al. (2018) for a study of *B. dorsalis*. Following transcriptome sequencing of 12 larvae mRNA using the Illumina HiSeq2500 platform, cDNA library construction and raw data analysis were conducted by BIOMARKER Technologies.

Data Filtering and Assembly

A large number of raw sequencing reads were produced in fastq format and processed through and in-house PERL scripts. From these raw reads, clean reads were obtained by removing adapters, poly-N, and low reads (i.e., reads with unknown nucleotides >5%, reads with more than 20% of base quality scores <10). To estimate the quality of the reads, GC content, Q value (quality score, or error probability of base calling), and sequence duplication level of the clean data were calculated. In this analysis, Q30 was set as a criterion with the error probability of base calling 0.1%. Low-quality reads were filtered out. The higher-quality reads were assembled into unigenes by TRINITY (version r20140413p1) (Grabherr et al. 2011) and mapped to the reference genome sequence of *Ceratitis capitata* (NCBI Assembly: GCF_000347755.3) and treated as the basis for analysis. All sequencing data were deposited in the NCBI and can be accessed in the Short Read Archive (SRA) under accession number PRJAN655252.

Analysis of Differentially Expressed Genes

Searching for differentially expressed genes (DEGs) between two sample stages was performed using DESeq (Wang et al. 2019). The resulting *P* values were adjusted using the Benjamini and Hochberg's (1995) approach for controlling the false discovery rate (FDR). The genes expressed differentially were those had an FDR less than 0.01 and twofold differences (the absolute value of log2 ratio ≥2) between two different samples. We examined status of DEGs of two types, namely upregulated and downregulated. The upregulated genes were those for which expression was promoted, while downregulated genes were those for which expression was restrained.

Gene Expression

The fragments per kilobase of transcript per million mapped reads (or FPKM) was used to estimate the gene expression levels of the unigenes, and calculated by the following formula (Chen et al. 2018):

$$\text{FPKM} = \text{cDNA Fragments} / \text{Mapped Fragments} \times \text{Transcript Lengthe (kb)}$$

Gene Function Annotations

The potential gene functions of all DEGs were annotated by comparing unigenes against different databases including Nr (NCBI non-redundant protein sequences, NCBI blast 2.2.28), Swiss-Prot (http://www.ebi.ac.uk/uniprot), COG (Clusters of Orthologous Groups, http://www.ncbi.nlm.nih.gov/COG/), KOG (eukaryotic Orthologous Groups, http://www.ncbi.nlm.nih.gov/KOG/), Pfam (Protein family, http://pfam.xfam.org/, HMMER 3.0 package), GO (Gene Ontology, http://www.geneontology.org/, Blast2Gov2.5 (Götz et al. 2008), and KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.jp/kegg/).

KOBAS (version v2.0.12) (Mao et al. 2005) was used to test whether the enrichments of genes expressed differentially for those genes in the KEGG database. KEGG is a database represented by large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies, which mainly are used to analyze high-level functions and relationships among biological systems. The topGO software (Alexa and Rahnenfuhrer 2016) was used to estimate the enrichments of DEGs in the GO database. *P*-value <0.05 was set as a threshold to estimate if the enrichments of the unigenes were significant.

Sample Comparisons

We compared C10, C15, B10, B15, P10, and P15 resulting in 15 pairwise sample transcript comparisons (see Table 1). These sample comparisons were divided into three groups: A, B, and C (Table 1). The A group consisted of three sample comparisons rearing respectively on the same host at different generations. The B group included four sample comparisons feeding on two Cucurbitaceae hosts. The C group was composed of eight comparisons, each of which included one sample feeding on the Cucurbitaceae host and one sample feeding on banana.

Results

Transcriptome Data

Transcriptome data were obtained from six *B. tau* larvae sample from experimental units, feeding on three different hosts at two developmental stages (see Table 2). An average of approximately 24M clean reads was produced for each *B. tau* sample with an average GC content of 52.68% and a Q30 of 90%. The ratio of reads was mapped onto the genome and reached 84% on average, reflecting high-quality data. Detailed information for each sample transcript has been listed in Table 2.

A total of 111,695 transcripts and 59,538 unigenes with N50 of 2,257 and 1,617 bp were obtained from *B. tau* assembled transcriptomes. The length range of transcripts and unigenes were provided in Supp Fig. S1 (online only), revealing that 5,615 unigenes and 17,704 transcripts were longer than 2,000 bp, demonstrating that transcriptome sequencing was effective.

DEG Analysis

Of all unigenes, 50% of DEGs were associated with host use. There was a smaller fraction of genes (18%) that were differentially expressed between different generations (10th generations and 15th generations) (see Fig. 1 and Table 3).

The numbers of genes expressed differentially in each sample comparison are shown in Fig. 1. The DEGs distributions and numbers in A, B, and C groups (see above) are provided in Table 1.

In A group (Table 3), the mean number of DEGs was 985 with 351 upregulated and 633 downregulated. The proportion of the average number of DEGs found in A group of the total DEGs detected was 18% (Fig. 1 and Table 3). In comparison, in B group, an average of 1,010 genes were expressed differentially, and 425 were upregulated and 585 were downregulated. A similar proportion of the average DEGs found in B group versus total DEGs detected was observed with 19% (Table 3).

In C group, an average of 1,595 genes were expressed differentially of which 1,001 were upregulated and 594 were downregulated (Fig. 1 and Table 3). C was the group consisting of samples reared on both traditional hosts and novel hosts at the same time. Compared with A and B, a relatively higher proportion of the average numbers of DEGs in C group of the total number of DEGs detected was recorded (31%). These results suggest that *B. tau* using the novel

Table 1. Numbers of unigenes annotated in different databases for each sample comparison of *B. tau*

Sample comparison	Group	Annotated	COG	GO	KEGG	KOG	Pfam	Swiss-Prot	Nr
P10_vs_P15	A	835	239	150	148	455	559	596	830
B10_vs_B15	A	771	237	96	149	376	464	492	749
C10_vs_C15	A	782	209	165	170	401	513	554	776
P10_vs_C10	B	649	172	133	120	306	420	425	630
P10_vs_C15	B	808	252	150	171	441	560	572	792
P15_vs_C10	B	1,083	288	198	210	553	688	761	1,066
P15_vs_C15	B	627	139	104	94	279	362	445	612
P10_vs_B10	C	1,521	455	269	330	814	992	1,062	1,514
P10_vs_B15	C	1,288	404	250	307	747	899	952	1,275
P15_vs_B10	C	1,345	406	269	292	751	930	968	1,341
P15_vs_B15	C	1,057	292	195	234	573	703	775	1,040
C10_vs_B10	C	1,656	446	309	354	830	1,036	1,144	1,628
C10_vs_B15	C	1,247	319	208	240	586	738	849	1,223
C15_vs_B10	C	1,367	380	274	318	763	950	1,000	1,357
C15_vs_B15	C	826	236	162	193	442	570	605	821

Table 2. Assembled transcriptome data for different sample treatments

	Pumpkin (P)		Cucumber (C)		Banana (B)	
	P10	P15	C10	C15	B10	B15
Clean read number	23,768,653	22,310,067	23,578,847	21,853,456	28,937,971	20,572,148
% ≥ Q30	91.09%	90.68%	90.24%	90.51%	85.23%	90.29%
Mapped reads number	20,479,103	18,793,620	20,040,917	18,394,928	24,004,095	16,932,613
Mapped ratio	86.16%	84.24%	85.00%	84.17%	82.95%	82.31%

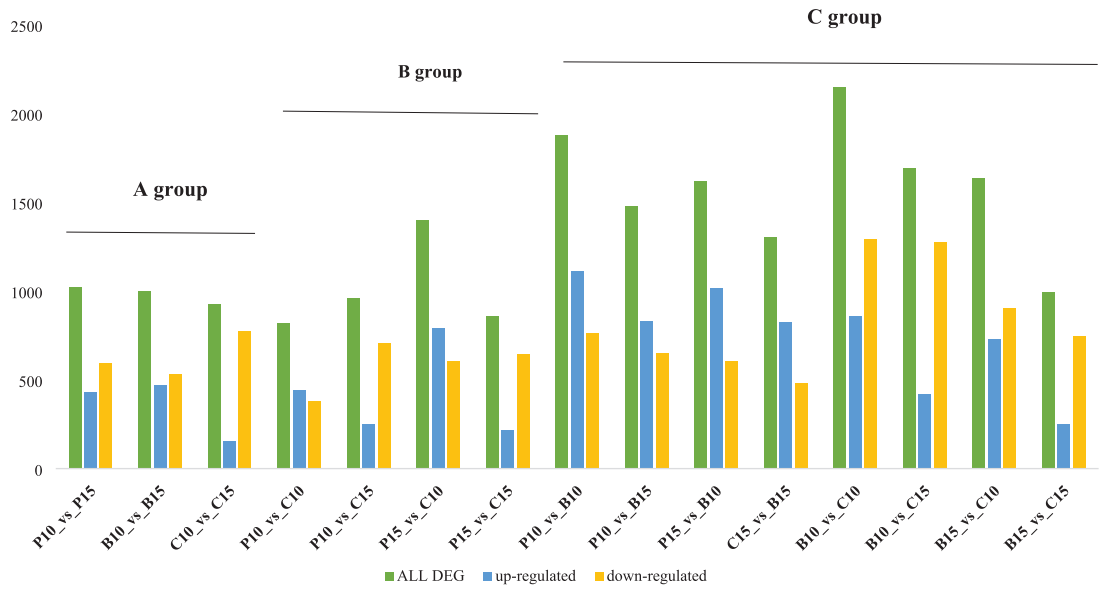

Fig. 1. Numbers of differentially expressed genes (DEGs) of each sample combination, including all DEGs, upregulated genes and downregulated genes. A, B, and C are the same groups as noted in Table 2.

Table 3. The average numbers of differentially expressed genes (DEGs) for each sample comparison group (A, B, and C) of *B. tau* and corresponding proportions (the average numbers of DEGs in each group that account for the total numbers of founded DEGs)

Group	All DEGs	Proportion	Upregulated	Downregulated
A	985	18%	352	633
B	1,011	19%	425	585
C	1,595	31%	1,001	594
Generation related	985	18%	352	633
Host related	1,303	50%	1,426	1,179

Group A consisted of sample comparisons feeding one host respectively and from different generations; group B included sample comparisons feeding on two traditional hosts and from different generations; and group C included sample comparisons feeding on two traditional and novel hosts. DEGs were genes between two sample comparison groups with FDR <0.01 and log2 ≥2.

host (banana) triggered more genes to be expressed differentially (see heatmap, Fig. 2 and Supp Fig. S2 [online only])

Gene Function Annotations

Homolog Annotation

The total numbers of DEGs that have been previously annotated in different databases are shown in Table 1. In the Nr protein database, approx. 67% of the annotated genes showed the best homolog from the *Ceratitis capitata*, another tephritid fly, while 20% genes were found as best homologs in other species. Approximately 2% of genes were most similar to *Drosophila melanogaster* (vinegar fly) and to *Clonorchis sinensis* (liver fluke), and also 0.9% annotated genes corresponded to *Camponotus floridanus* (ant) (see Fig. 3). These numbers reflect in part the amount of data currently available for other organisms.

GO Annotation

The functions of a total 4,605 unigenes were predicted in the GO database. The functional annotations of DEGs in all unigenes can be classified into three main categories (Supp Fig. S3 [online only]): biological process, cellular components, and molecular function, with an average of 14 annotations for each category. Within the biological process category, the DEGs were enriched in cell, membrane, cell part, and organelle functions. Genes encoding for binding and catalytic activity were enriched in the molecular function category. In cellular components, genes were mainly annotated in metabolic process, developmental process, biological regulation, and cellular process (Supp Fig. S3 [online only]).

COG Annotation

The functions of 5,225 unigenes were further predicted using the COG database. Genes expressed differentially for all unigenes were

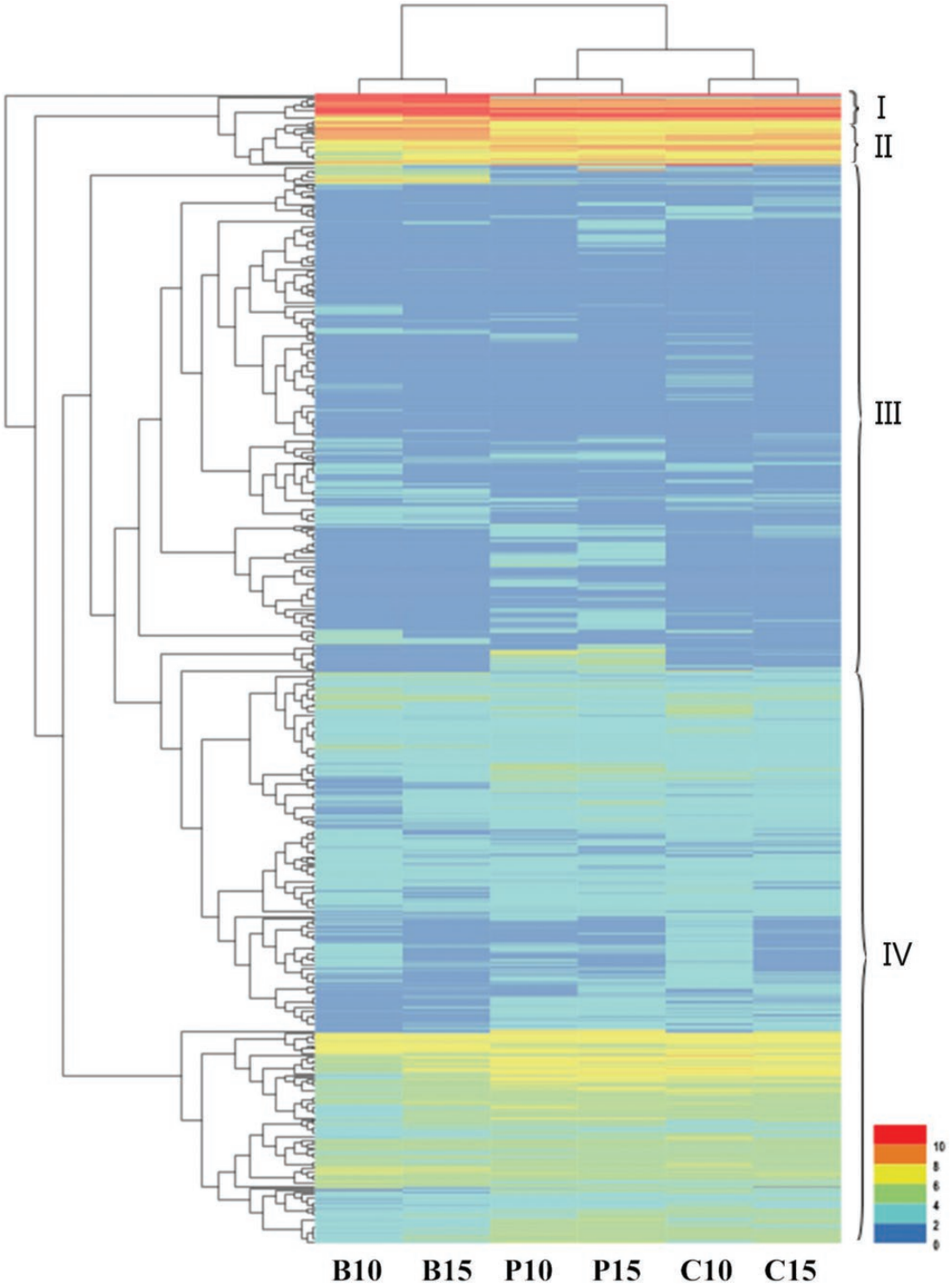

Fig. 2. The hierarchical clustering heatmap of differentially expressed genes (DEGs). Different columns in the map represent different samples, while different lines represent the log2 values of FPKM from different genes (see text). Red color indicates upregulated expression, whereas dark blue indicated downregulated expression. The roman numerais of right the heatmap listed the major DEGs annotated categories referencing Table 4 including I: oxidative phosphorylation, citrate cycle (TCA cycle), and starch and sucrose metabolism; II: amino sugar and nucleotide sugar, glutathione metabolism, and drug metabolism cytochrome P450; III: folate biosynthesis, and lysosome and pyruvate metabolism; IV: valine, leucine, and aminoacyl-Trna biosynthesis, and isoleucine biosynthesis.

mapped into 24 COG categories (Supp Fig. S4 [online only]). The largest category was 'general function prediction only' followed by 'transcription', 'cytoskeleton', 'signal transduction mechanisms', 'energy production and conversion', and 'amino acid transport and metabolism'. Certainly, the functions of many genes are unknown.

KEGG Annotation

Genes expressed differentially among different sample combinations were also searched by blast in KEGG pathways. There were 3,330 unigenes that were assigned in 435 KEGG pathways. The top 21 pathways were listed in Table 4 according to the enrichment analysis

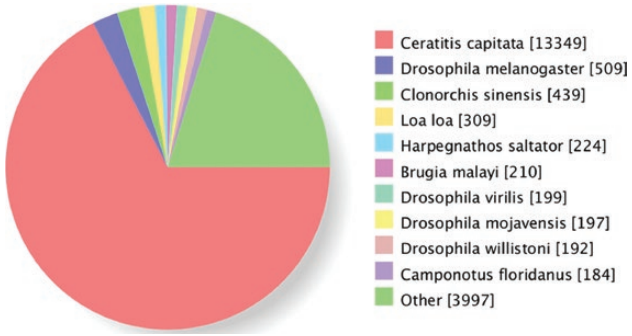

Fig. 3. Distribution of species identified with homologies to *Bactrocera tau* based on unigene blast in Nr database.

(Supp Fig. S5 [online only]). Few DEGs (six on average) in A group were mapped into about 12 KEEP annotations, which included 'folate biosynthesis', 'lysosome', 'galactose metabolism', 'starch and sucrose metabolism', 'caffeine metabolism', 'metabolism of xenobiotics by cytochrome *P450*', and 'DNA replication'.

Most of the enriched DEGs in A group, consisting of three sample comparisons reared on the same host at different generations, were downregulated. A few upregulated genes were found associated with 'valine, leucine and isoleucine biosynthesis', 'aminoacyl-tRNA biosynthesis', and 'lysosome', corresponding to comparisons of pumpkin at different generations. In A group, 16 genes showed upregulation in 'metabolism of xenobiotics by cytochrome P450' in response to samples of different generations feeding on banana (Table 4). The four upregulated genes of these banana-reared samples scattered in other pathways including 'folate biosynthesis', 'caffeine metabolism', 'TCA cycle', 'drug metabolism cytochrome P450', 'glycolysis/gluconeogenesis', and 'butanoate metabolism' (Table 4).

In B group including four sample comparisons feeding on two traditional hosts, the DEGs were annotated into about 15 KEGG pathways and similar enrichments of DEGs as A were found in B group including 'folate biosynthesis', 'galactose metabolism', 'starch and sucrose metabolism', 'lysosome', 'drug metabolism cytochrome P450', and 'metabolism of xenobiotics by cytochrome P450' (Supp Fig. S5 [online only]). Most DEGs of four sample comparisons enriched in 'folate biosynthesis' and 'lysosome' were downregulated, while most DEGs enriched in 'drug metabolism cytochrome P450' and 'metabolism of xenobiotics by cytochrome P450' were upregulated (Table 4).

In the C group composed of individuals reared on traditional hosts and novel hosts at the same time, both the number of DEGs annotated and the categories enriched KEGG annotations (21 annotations) among eight sample comparisons were higher than in A and B groups. Moreover, the classifications of annotations enriched in KEGG in C group were rather different than A and B. The DEGs of eight sample comparisons in C group were mainly centralized in 'oxidative phosphorylation' and 'citrate cycle' with predominant numbers of upregulated genes enriching 59 and 13 on average, respectively. Other processes with high enrichments in C group involved 'pyruvate metabolism', 'valine, leucine and isoleucine biosynthesis', 'glycolysis/gluconeogenesis', and 'butanoate metabolism' (Supp Fig. S5 [online only] and Table 4).

Classification and Expressional Status of DEGs in Main Metabolism Pathways

Oxidative Phosphorylation

A large number of DEGs of sample comparisons of C group were annotated in the oxidative phosphorylation pathway (Supp Fig. S5 [online only] and Supp Table S1 [online only]). Most genes of *NADH* dehydrogenase, succinate dehydrogease/furnarate reductase, cytochrome c reductase, cytochrome c oxidase, and F-type ATPase (Eukaryotes) in oxidative phosphorylation pathway were upregulated. The highest numbers of upregulated genes were found for *NADH* dehydrogenase, with 59 on average in each sample comparison of C group (Supp Table S1 [online only]). Among the *NADH* dehydrogenase genes, the expression levels (FPKM) of ND5 and *Ndufs7* gene were highest with an average of 977 and 1,058, respectively. The expression levels of *ND1* and *NDufs5* gene were also very high with an average of 1,683 and 952, respectively.

Citrate

Most sample comparisons enriched in the citrate pathway were from A and B groups. The annotations of DEGs in this pathway mainly were in three categories including proteases, glycosidases, sulfatases, and cytosol (see Supp Fig. S5 [online only] and Supp Table S2 [online only]). Moreover, most of these DEGs were downregulated. The expression levels of malate dehydrogenase, hyaluronate lyase, and succinate-CoA ligase gene were also relatively high with averages of 458, 323, and 587, respectively

Lysosome

Most sample comparisons enriched in the lysosome pathway were from A and B groups. The annotations of DEGs in this pathway mainly were in three categories including proteases, glycosidases, sulfatases, and cytosol (see Supp Fig. S5 [online only] and Supp Table S3 [online only]). Moreover, most of these DEGs were downregulated.

Folate Biosynthesis

The folate pathway was enriched mainly in A and B groups (Supp Fig. S5 [online only] and Supp Table S4 [online only]). The DEGs of this pathway were found in genes encoding sepiapterin reductase, dihydrotolate reductase, alkaline phosphatase, and guanosine triphosphate cyclohydrolase enzymes. Most of these folate genes were downregulated.

Discussion

In this study, we present the first RNA transcriptomes for *B. tau* using high-throughput sequencing, and document variation associated with host use. The transcripts obtained were of high quality with an average of approximately 23M clean reads of Q30 >90% and 83% reads mapping ratio for each sample (Table 2), which are similar to those available for *B. dorsalis* (Chen et al. 2018). The transcriptomes of *B. tau* provide a reference for comparison to other tephritid fly species and also provide baseline data for additional research, including investigations of functional genes.

As expected, the effect of number of rearing generations on transcript variation of *B. tau* was very weak (only 18% of genes were differentially expressed between different generations). The A group containing sample comparisons rearing on the same host but different generations had both the lowest numbers and categories of DEGs with only 18% expressing differently genes accounting for all unigenes (Table 3). A similar pattern was found in the transcriptome data of *Polygonia c-album* (Lepidoptera: Nymphalidae) (Celorio-Mancera et al. 2013). By contrast, we found hosts can exert a major effect on transcript variation of *B. tau* with 50% of DEGs in all unigenes associated with the novel host, banana feeding comparisons (Table 3).

Table 4. Major pathway classification and numbers of differentially expressed genes (DEGs) in the KEGG database for each sample comparison for *B. tau*

Functional annotations	Categories	A group			B group				C group							
		P10/P15	B10/B15	C10/C15	P10/C10	P10/P15	P15/C10	P15/C15	P15/C15	P10/B15	P15/B10	P15/B15	C10/B10	C10/B15	C15/B10	C15/B15
Folate biosynthesis	M		1D/5U	3D	4D	3D	4D/1U	6D	1D					4U	4D/1U	2D/1U
Insect hormone biosynthesis	M	4D	2U			5D	1D/2U		4D	2D		2U		2U		
Starch and sucrose metabolism	M	4D	1D/2U	5D	1D/1U	5D/1U	3U		3D/2U	3D/3U	3D/1U		2D/5U	1D/3U	3D/5U	
Caffeine metabolism	M	1D	4U	1D		1D	1U	1U	1D	1D				4U	1U	
Oxidative phosphorylation			4D					3U	1D /63U	1D /63U	60U	63U	1D/58U	1D /45U	1D /65U	1D /55U
Citrate cycle (TCA cycle)	M		3U						1D /16U	15U	1D/14U	15U	15U	13U	16U	
Drug metabolism cytochrome P450	M		3D/4U	4D	4U	3U	3U	1D/3U	3D		3D		3U			
Metabolism of xenobiotics by cytochrome P450	M		16U/1D	14D	16U	1U	1U	15D/1U	1D		1D		1U			
Pyruvate metabolism	M		2U/1D		1D/8U				3D/8U	1D/8U	2D/8U	2D/9U	8D/1U	7D	9D	7D
Valine, leucine, and isoleucine biosynthesis	M	2U						2D		1D/2U	4D/1U	2D/1U		2D	2D	2D
Aminoacyl-Trna biosynthesis	M	4U	2U					4D						4D/1U		
Amino sugar and nucleotide sugar metabolism	M							1U		1D/5U			4D/6U		7D/6U	4D/1U
Glycolysis/ gluconeogenesis	M		1D/4U						3D/11U	10U	1D/11U	1D/11U	9D/4U	7D/1U	13D/2U	9D
Butanoate metabolism	M									2U		2U		2D	2D	2D
Glutathione metabolism	M		5D/4U			7D			8D/2U							
Lysosome	C	4D/2U		4D/2U	6D/2U	6D/1U	4D/7U	3D		7D/2U						
DNA replication	G			10D											10U	
Progesterone-mediated oocyte maturation	O			8D	3U										8U	

U = upregulated genes, D = downregulated genes; M = metabolism, C = cellular processes, O = organismal system.

Transcript patterns of *B. tau* varied greatly among different hosts. The transcript pattern of flies feeding on orange was not tested in the end, because flies feeding on oranges died within the first three generations, perhaps related to ripeness of fruit. In the future, we hope to adjust rearing methods in the lab, so that we can add the transcript variation of flied reared continuously on orange. In contrast, the transcript of *B. tau* feeding on banana (Musaceae family), which has been reported as a potential novel host for *B. tau* (Lin et al. 2005), showed a clear difference from other traditional hosts in gene expressional patterns (Fig. 2 and Supp Fig. S2 [online only]). The C group including comparisons of samples feeding on banana with those feeding on other hosts contained both greater numbers and annotated categories of DEGs than those in B group, which compared *B. tau* feeding on two cucurbit hosts. Moreover, the annotated categories of DEGs in C group differ sharply from those in B group. The results suggest that both genes and metabolism pathways (Table 4) have been triggered in the use of a novel hosts for *B. tau*. These differences are consistent with findings from previous research concerning survival rate and genetic structure of *B. tau* feeding on the three same hosts as in this study (Shi et al. 2017).

Because much variation of transcriptomes of *B. tau* in this study is associated with use of a novel host, one can address the question, 'what triggers the novel host usage of *B. tau*?' We suggest the utilization of the novel host activated the related genes of various metabolism pathways to be expressed differently that regulate this process. Annotated DEGs of sample comparisons in C group are mainly enriched in oxidative phosphorylation and citrate cycle pathway (Table 4). In contrast, the similar annotations of DEGs in A and B groups are enriched in folate biosynthesis and lysosome pathways (Table 4). Other studies, including one for *B. dorsalis* (Chen et al. 2017), have speculated about different genes and biochemical pathways may be expressed differentially in response to different diets.

The highest DEGs enrichments of sample comparisons in the C group were observed in the oxidative phosphorylation pathway with a large number of upregulated genes (Table 4 and Supp Table S1 [online only]). The oxidative phosphorylation pathway is an energy metabolism pathway, which mainly provides energy with the form of ATP for most organisms (Wilson and Vinogradov 2014). One hypothesis is that the flies using the novel host of banana in this study expended more energy and the genes of energy production were activated to regulate this process. These results suggest that the oxidative phosphorylation pathway may play a core role in the process of novel host use of banana.

In our study, we found the predominant numbers of most upregulated genes (59 on average) in the oxidative phosphorylation pathway belonged *NADH* genes (Table 4 and Supp Table S1 [online only]) and the expression levels of some *NADH* genes are very high, including *ND5* and *Ndufs7*, with average 977 and 1,058 expression levels respectively. *NADH* genes are complex I, which are the largest component in oxidative phosphorylation and mainly involved in ATP synthesis (Wirth et al. 2016). Studies of activities of many species, including honeybees (Strachecka et al. 2019), *Drosophila* (Hur et al. 2013), and *Acyrthosiphon pisum* (pea aphid, Valmalette et al. 2012), show that *NADH* is associated with bioenergetic functions. However, the related functions of *NADH* genes in *B. tau* require verification, such as by using gene knock-outs or other approaches.

The citrate cycle also likely plays an important role for the novel host use in *B. tau*, as suggested by relatively high enrichments of DEGs of C group in the citrate cycle with relatively more upregulated genes with 13 on average (Table 4 and Supp Table S2 [online only]). The citrate cycle is one pathway associated with the metabolism of sugars, lipids, and amino acids (Zhang et al. 2019) and also producing energy (Raimundo et al. 2011). Genes encoding for key enzymes in this cycle, such as malic dehydrogenase, hyaluronate lyase, and succinate-CoA ligase (Supp Table S2 [online only]) (Zhang et al. 2019), were upregulated among most sample comparisons of C group. In addition to TCA, carbohydrate metabolism includes starch and sucrose, amino sugar and nucleotide sugar, and glycolysis/gluconeogenesis metabolism, which also showed high enrichments in samples of C group (Supp Fig. S5 [online only] and Table 4). Together, these results suggest use of the novel banana host for *B. tau* is related to exploiting more sugars in addition to using more energy. Indeed, oxidative phosphorylation and carbohydrate metabolism were both associated with novel host utilization for *B. tau* in this study. Thus, both genes of energy and sugar metabolism were involved in regulating banana use of *B. tau*.

Compared with the use of banana, DEGs associating with use of two traditional Cucurbitaceae hosts were mainly involved in folate biosynthesis and lysosome pathways (Supp Tables S3 and S4 [online only]), which include the metabolism of cofactors and vitamins and transport and catabolism, respectively. Alkaline phosphatase genes in folate biosynthesis and lysosome genes with disintoxication functions (Pascua-Maestro et al. 2017, Terra et al. 2018) showed downregulation in samples feeding on Cucurbitaceae hosts (Supp Tables S3 and S4 [online only]). Moreover, genes for sepiapterin reductase and guanosine triphosphate cyclohydrolase in folate biosynthesis (Supp Table S4 [online only]), which are reported to regulate nervous activity of *Drosophila* (Jacobson and Manos 1989, Kim et al. 1994), were also downregulated in the Cucurbitaceae host samples in this study. These results suggest that genes involved in regulating the host use of *B. tau* feeding traditional Cucurbitaceae hosts of pumpkin and cucumbers are not so different.

Unlike the findings for *B. tau* in this study, research on the host shifting of *Polygonia c-album* (Lepidoptera: Nymphalidae) from *Urtica dioica* (nettle) to a novel host, *Ribes* (gooseberries), showed changes in ribosomal, digestion-, and detoxification-related genes (Celorio-Mancera et al. 2013). Although digestion- and detoxification-related genes (such as genes of 'lysosome', 'drug metabolism cytochrome *P450*', and 'metabolism of xenobiotics by cytochrome *P450*') were also enriched in the sample comparisons for those feeding on banana in this study (Table 4), these genes did not show the largest effects associated with banana use by *B. tau*. In addition, results of studies related novel host use in a cicada revealed that digestion-, detoxification-, oxidation-reduction-, water-deprivation-related genes regulated host adaptation (Hou and Wei 2019). Here, we show that *B. tau* feeding on different hosts will face various nutrients and chemical constituents produced by different host plants (Vogel et al. 2014). So different genes and metabolism pathways were stimulated as part of the process of novel host use.

This study provided some clues to help understand the mechanisms associated with novel host use in an invasive herbivorous insect. This study suggests that genes of oxidative phosphorylation involving energy metabolism and carbohydrate metabolism are associated with use of the novel host banana for *B. tau*. In particular, mitochondrial *NADH* genes and genes of citrate cycle pathway are key genes regulating processes associated with use of bananas by *B. tau* in comparison to expression of genes associated with use of different traditional cucurbit hosts. The specific functions of *NADH* genes are still unclear and will require other approaches to resolve.

Supplementary Data

Supplementary data are available at *Journal of Insect Science* online.

Acknowledgments

We are grateful to two anonymous reviewers, who gave revised suggestions to improve the manuscript. This research was supported by the National Basic Research Program of China (grant no. 2014, 31460163) and State Scholarship Fund issued by the China Scholarship Council 2019 (CSC, 201907035005).

References Cited

Alexa, A., and J. Rahnenfuhrer. 2016. TopGO: enrichment analysis for gene ontology. R package version 2.24.0. Available online: https://bioconductor.org/packages/release/bioc/html/topGO.html. Accessed on June 2, 2020.

Barth, M. B., K. Buchwalder, A. Y. Kawahara, X. Zhou, S. Liu, N. Krezdorn, B. Rotter, R. Horres, and A. K. Hundsdoerfer. 2018. Functional characterization of the *Hyles euphorbiae* hawkmoth transcriptome reveals strong expression of phorbol ester detoxification and seasonal cold hardiness genes. Front. Zool. 1: 15–20.

Benjamini, Y. H., and Y. Hochberg. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. Sta. 57: 289–300.

Celorio-Mancera, M., C. W. Wheat, H. Vogel, L. Söderlind, N. Janz, and S. Nylin. 2013. Mechanisms of macroevolution: polyphagous plasticity in butterfly larvae revealed by RNA-Seq. Mol. Ecol. 22: 4884–4895.

Charlery de la Masselière, M., V. Ravigné, B. Facon, P. Lefeuvre, F. Massol, S. Quilici, and P. F. Duyck. 2017. Changes in phytophagous insect host ranges following the invasion of their community: long-term data for fruit flies. Ecol. Evol. 14: 5181–5190.

Chen, E. H., Q. L. Hou, D. D. Wei, H. B. Jiang, and J. J. Wang. 2017. Phenotypes, antioxidant responses, and gene expression changes accompanying a sugar-only diet in *Bactrocera dorsalis* (Hendel) (Diptera: Tephritidae). BMC Evol. Biol. 17: 194.

Chen, E. H., Q. L. Hou, W. Dou, D. D. Wei, Y. Yue, R. L. Yang, S. F. Yu, K. De Schutter, G. Smagghe, and J. J. Wang. 2018. RNA-seq analysis of gene expression changes during pupariation in *Bactrocera dorsalis* (Hendel) (Diptera: Tephritidae). BMC Genomics 21: 693.

Chikate, Y. R., V. A. Tamhane, R. S. Joshi, V. S. Gupta, and A. P. Giri. 2013. Differential protease activity augments polyphagy in *Helicoverpa armigera*. Insect Mol. Biol. 22: 258–272.

Christenson, L. D., and R. H. Foote. 1960. Biology of fruit flies. Annu. Rev. Entomol. 5: 171–192.

Clarke, A. R. 2017. Why so many polyphagous fruit flies (Diptera: Tephritidae)? A further contribution to the 'generalism' debate. Biol. J. Linn. Soc. Lond. 120: 245–257.

Cronin, J. T., and W. G. Abrahamson. 2001. Goldenrod stem galler preference and performance effects of multiple herbivores and plant genotypes. Oecologia 127: 87–96.

Götz, S., J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talón, J. Dopazo, and A. Conesa. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36: 3420–35.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, et al. **2011**. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29: 644–652.

Guo, S., Z. Zhao, L. Liu, Z. Li, and J. Shen. 2018. Comparative transcriptome analyses uncover key candidate genes mediating flight capacity in *Bactrocera dorsalis* (Hendel) and *Bactrocera correcta* (Bezzi) (Diptera: Tephritidae). Int. J. Mol. Sci. 19: 396.

Hasyim, A., M. Muryati, and W. J. de Kogel. 2008. Population fluctuation of adult males of the fruit fly, *Bactrocera tau* Walker (Diptera: Tephritidae) in passion fruit orchards in relation to abiotic factors and sanitation. Indonesian J. Agr. Sci. 9: 29–33.

Hou, Z., and C. Wei. 2019. De novo comparative transcriptome analysis of a rare cicada, with identification of candidate genes related to adaptation to a novel host plant and drier habitats. BMC Genomics 20: 182.

Huang, K., Q. X. Guo, Y. Yu, and Z. H. Huang. 2005. Risk analysis of *Bactrocera* (Zeugodacus) *tau* (Walker). Wuyi Sci. J. 21: 77–80.

Hur, J. K., M. K. Zinchenko, S. Djuranovic, and R. Green. 2013. Regulation of Argonaute slicer activity by guide RNA 3' end interactions with the N-terminal lobe. J. Biol. Chem. 288: 7829–7840.

Jacobson, K. B., and R. E. Manos. 1989. Effects of sepiapterin and 6-acetyldihydrohomopterin on the guanosine triphosphate cyclohydrolase I of mouse, rat and the fruit-fly *Drosophila*. Biochem. J. 260: 135–141.

Kim, D. H., H. J. Kang, S. H. Park, and K. Kobashi. 1994. Characterization of beta-glucosidase and beta-glucuronidase of alkalotolerant intestinal bacteria. Biol. Pharm. Bull. 17: 423–426.

Li, Y. L., M. Z. Hou, G. M. Shen, X. P. Lu, Z. Wang, F. X. Jia, J. J. Wang, and W. Dou. 2017. Functional analysis of five trypsin-like protease genes in the oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae). Pestic. Biochem. Physiol. 136: 52–57.

Lin, M. Y., S. K. Chen, and Y. C. Liu. 2005. The host plants of *Bactrocera tau* in Taiwan. Research Report on Agricultural Improvement Farm in Tainan District. 45: 39–52.

Mao, X., T. Cai, J. G. Olyarchuk, and L. Wei. 2005. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics 21: 3787–3793.

Mattsson, M., G. R. Hood, J. L. Feder, and L. A. Ruedas. 2015. Rapid and repeatable shifts in life-history timing of *Rhagoletis pomonela* (Tephritidae) following colonization of novel host plants in the Pacific Northwestern United States. Ecol. Evol. 5: 5823–5827.

Miyazaki, H., J. Otake, H. Mitsuno, K. Ozaki, R. Kanzaki, A. Chui-Ting Chieng, A. Kah-Wei Hee, R. Nishida, and H. Ono. 2018. Functional characterization of olfactory receptors in the oriental fruit fly *Bactrocera dorsalis* that respond to plant volatiles. Insect Biochem. Mol. Biol. 101: 32–46.

Nagalingam, K., B. A. V. der Chloé, Q. Yujia, L. C. Stephen, R. C. Anthony, and J. P. Peter. 2018. Plant-mediated female transcriptomic changes post-mating in a tephritid fruit fly, *Bactrocera tryoni*. Genome Bio. Evol. 10: 94–107.

Ooi, Y. T., and S. L. Wee. 2016. Sexual maturation, mating propensity and remating incidence of *Zeugodacus tau* (Walker) (Diptera:Tephritidae). J. Asia-Pac. Entomol. 19: 451–457.

Pascua-Maestro, R., S. Diez-Hermano, C. Lillo, M. D. Ganfornina, and D. Sanchez. 2017. Protecting cells by protecting their vulnerable lysosomes: identification of a new mechanism for preserving lysosomal functional integrity upon oxidative stress. PLoS Genet. 13: e1006603.

Piñero, J. C., S. K. Souder, and R. I. Vargas. 2017. Vision-mediated exploitation of a novel host plant by a tephritid fruit fly. PLoS ONE 12: e0174636.

Raimundo, N., B. E. Baysal, and G. S. Shadel. 2011. Revisiting the TCA cycle: signaling to tumor formation. Trends Mol. Med. 17: 641–649.

Shi, W., C. Kerdelhué, and H. Ye. 2014. Genetic structure and colonization history of the fruit fly *Bactrocera tau* (Diptera: Tephritidae) in China and Southeast Asia. J. Econ. Entomol. 107: 1256–1265.

Shi, W., T. Y. Yang, H. Ye, and J. Cao. 2017. Impact of host plants on genetic variation in the *Bactrocera tau* (Diptera: Tephritidae) based on molecular markers. J. Entomol. Sci. 52: 411–426.

Singh, S. K., D. Kumar, and V. V. Ramamurthy. 2010. Biology of *Bactrocera (Zeugodacus) tau* (Walker). Entomol. Res. 40: 259–263.

Strachecka, A., M. Grzybek, A. A. Ptaszynska, A. Los, J. Chobotow, and R. Rowinski. 2019. Lactate dehydrogenase activity in hive and forager honeybees may indicate delayed onset muscle soreness-preliminary studies. Biochemistry 84: 435–440.

Sumrandee, C., J. R. Milne, and V. Baimai. 2011. Ovipositor morphology and host relations of the *Bactrocera tau* complex (Diptera: Tephritidae) in Thailand. Songklanakarin J. Sci. Technol. 33: 247–254.

Tallamy, D. W. 2000. Physiological issues in host range expansion. pp. 11 ± 26. *In* R. Van Driesche, T. Heard, A. McClay, and R. Reardon (eds.), Proceedings of Session Host Specificity Testing of Exotic Arthropod Biological Control Agents: The Biological Basis for Improvement in Safety. International Symposium on Biological Control of Weeds, Bozeman, Montana.

Terra, W. R., R. O. Dias, P. L. Oliveira, C. Ferreira, and T. M. Venancio. 2018. Transcriptomic analyses uncover emerging roles of mucins, lysosome/secretory addressing and detoxification pathways in insect midguts. Curr. Opin. Insect Sci. 29: 34–40.

Valmalette, J. C., A. Dombrovsky, P. Brat, C. Mertz, M. Capovilla, and A. Robichon. 2012. Light-induced electron transfer and ATP synthesis in a carotene synthesizing insect. Sci. Rep. 2: 579.

Vogel, H., C. Badapanda, E. Knorr, and A. Vilcinskas. 2014. RNA-sequencing analysis reveals abundant developmental stage-specific and immunity-related genes in the pollen beetle *Meligethes aeneus*. Insect Mol. Biol. 23: 98–112.

Wang, Y., A. A. Andongma, Y. Dong, Z. Chen, P. Xu, X. Ren, M. N. Krosch, A. R. Clarke, and C. Niu. 2019. Rh6 gene modulates the visual mechanism of host utilization in fruit fly *Bactrocera mi*nax. Pest Manag. Sci. 75: 1621–1629.

Wilson, D. F., and S. A. Vinogradov. 2014. Mitochondrial cytochrome c oxidase: mechanism of action and role in regulating oxidative phosphorylation. J. Appl. Physiol. (1985). 117: 1431–1439.

Wirth, C., U. Brandt, C. Hunte, and V. Zickermann. 2016. Structure and function of mitochondrial complex I. Biochim. Biophys. Acta 1857: 902–914.

Zhang, X. Y., and G. Q. Chen. 2012. Observation of *Bactrocera tau* infesting orange. Zhejiang Agric. Sci. 9: 1274–1275.

Zhang, W. C., J. M. Wells, K. H. Chow, H. Huang, M. Yuan, T. Saxena, M. A. Melnick, K. Politi, J. M. Asara, D. B. Costa, et al. **2019**. miR-147b-mediated TCA cycle dysfunction and pseudohypoxia initiate drug tolerance to EGFR inhibitors in lung adenocarcinoma. Nat. Metab. 1: 460–474.