PLOS ONE

RESEARCH ARTICLE

Transcriptome analysis and comparison reveal divergence between the Mediterranean and the greenhouse whiteflies

Yu-Jun Wang, Hua-Ling Wang, Xiao-Wei Wang, Shu-Sheng Liu*

Ministry of Agriculture Key Laboratory of Agricultural Entomology, Institute of Insect Sciences, Zhejiang University, Hangzhou, China

* shshliu@zju.edu.cn

Citation: Wang Y-J, Wang H-L, Wang X-W, Liu S-S (2020) Transcriptome analysis and comparison reveal divergence between the Mediterranean and the greenhouse whiteflies. PLoS ONE 15(8): e0237744. https://doi.org/10.1371/journal.pone.0237744

Editor: Murad Ghanim, Agricultural Research Organization Volcani Center, ISRAEL

Received: May 8, 2020

Accepted: July 31, 2020

Published: August 25, 2020

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This research was funded by National Natural Science Foundation of China, (Project no.: 31930092). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Both the Mediterranean (MED) species of the *Bemisia tabaci* whitefly complex and the greenhouse whitefly (*Trialeurodes vaporariorum*, TV) are important agricultural pests. The two species of whiteflies differ in many aspects such as morphology, geographical distribution, host plant range, plant virus transmission, and resistance to insecticides. However, the molecular basis underlying their differences remains largely unknown. In this study, we analyzed the genetic divergences between the transcriptomes of MED and TV. In total, 2,944 pairs of orthologous genes were identified. The average identity of amino acid sequences between the two species is 93.6%. The average nonsynonymous (Ka) and synonymous (Ks) substitution rates and the ratio of Ka/Ks of the orthologous genes are 0.0389, 2.23 and 0.0204, respectively. The low average Ka/Ks ratio indicates that orthologous genes tend to be under strong purified selection. The most divergent gene classes are related to the metabolisms of xenobiotics, cofactors, vitamins and amino acids, and this divergence may underlie the different biological characteristics between the two species of whiteflies. Genes of differential expression between the two species are enriched in carbohydrate metabolism and regulation of autophagy. These findings provide molecular clues to uncover the biological and molecular differences between the two species of whiteflies.

Introduction

Many whiteflies (Hemiptera: Aleyrodidae) are important pests of agriculture worldwide, such as some species of the *Bemisia tabaci* whitefly complex and the greenhouse whitefly (*Trialeurodes vaporariorum* Westwood) [1, 2]. While whiteflies of the *B. tabaci* species complex and the greenhouse whitefly are similar in many aspects, they differ in many features such as geographic distribution, range of host plants, virus transmission, and resistance to insecticides [2–8]. Whiteflies of the *B. tabaci* complex distribute in tropic and subtropical regions; some species of this whitefly complex, in particular two species, tentatively named as Middle East-Asia Minor 1 (hereafter MEAM1, formally referred to as the 'B biotype') and Mediterranean (hereafter MED, formally referred to as the 'Q biotype'), have invaded many regions of the world

and caused serious damage to many crops such as cotton and tomato in the last 30 years [1, 9–11]. The greenhouse whitefly *T*. *vaporariorum* (hereafter TV) inhabits the temperate regions and is a major pest of fruit, vegetable and ornamental crops in protected environment [3, 12]. In some regions, MEAM1/MED and TV coexist and show interspecific competition [13, 14]. In these regions of co-existence, they often show apparently different patterns of seasonal abundance: MEAM1/MED become predominant in seasons of relatively high temperatures, while TV becomes predominant in seasons of relatively low temperatures [15]. While all the three species of whiteflies are polyphagous, MEAM1 and MED have a wider range of host plants than TV [3, 4]. In greenhouses where MEAM1/MED and TV co-exist, they differ in patterns of within-plant distribution [16].

Another major difference between MEAM1, MED, and TV lies in their capacity of viral transmission: while whiteflies of the *B*. *tabaci* species complex, including MEAM1 and MED, transmit begomoviruses that include major agents of viral diseases of important crops such as cotton, cassava, and tomato, as well as some other groups of viruses like criniviruses and ipomoviruses, TV is a major vector of criniviruses and torradoviruses but is unable to transmit begomoviruses [2, 17]. Some criniviruses, for example *Tomato chlorosis virus*, are transmitted by both TV and MED [18, 19]. MED has developed much higher levels of resistance to major classes of insecticides than TV [6, 20, 21]. For example, MED had developed up to 1900-fold resistance to imidacloprid, and 1200-fold resistance to thiamethoxam, while TV had developed only 23.8- and 20.4-fold resistance to these two insecticides [6, 7]. However, the molecular basis underneath the differences between the whiteflies of the *B*. *tabaci* species complex and TV remains largely unknown.

RNA-seq provides an efficient approach to analyze the transcriptome of an organism and also an efficient method to discover new genes of interest [22, 23]. Pairwise comparisons between MED, MEAM1, and Asia II 3 (a native species of whitefly) have been conducted at the sequence and gene expression levels, indicating that sequence divergence of gene clusters include cytochrome P450, glutathione metabolism, and oxidative phosphorylation, and highly expression divergent genes are mainly related to basic metabolism and detoxification [24–26]. So far, several RNA-seq studies have been analyzed on whiteflies in relation to host adaption [27–30], insecticides resistance [31–33] and virus transmission [34–38]. Some detoxification genes such as cytochrome 450 monooxygenases (P450s), glutathione S-transferases (GSTs) and UDP-glucosyltransferases (UGTs) were found related with both host adaption and insecticide resistance [30–32, 39, 40]. Moreover, the detoxification gene expression patterns can shape the ability of *Bemisia* species to utilize multiple plant hosts [27].

In this study, first, we reassembled the previous version of transcriptome of MED [41] using Trinity to obtain the unigenes of similar length to those of the published TV transcriptome [42]; next, we compared the orthologous genes derived from transcriptomes of the two species; and finally, we analyzed the differential gene expression between MED and TV. Our major purpose was to find important genes that may contribute to the divergence of the two whitefly species at both genetic and expression level.

Materials and methods

Reads assembly, functional annotation and coding sequence prediction for MED

The raw reads of MED were downloaded from NCBI Short Read Archive (SRA), accession number: SRX018661. Before assembly, the raw reads were preprocessed by cutting off low quality base pairs within 20 bps (reads with unknown sequences 'N' or average quality score less than 20) of each read in the 3'ends by costumed Perl script to ensure no loss of information

from paired ends. Then the preprocessed reads were assembled using the Trinity software (trinityrnaseq-r20110519) with default parameters [43].

Sequences were annotated by searching against the NCBI nr database with a cut-off E-value of $1.0E^{-5}$using Blastx [44]. Gene Ontology (GO) annotation was analyzed using Blast2GO software [45]. The GO terms were retrieved from Blastx hits with an e-value threshold of $1.0E^{-5}$. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation was performed using Blastx software against the KEGG database. The best potential coding sequences (CDS) from each of the reconstructed transcripts were predicted using the software BestORF (http://www.softberry.com/berry.phtml?topic=bestorf&group=programs&subgroup=gfind) with parameters trained with *Drosophila* ESTs. Predicted CDS that start with "ATG" start codon and end with "TAA"/"TGA"/"TAG" stop codon were assumed as complete CDS.

Identification of orthologous genes

Transcriptome sequences of TV were obtained from InsectaCentral database (http://www.insectacentral.org/). Identification of orthologous genes was performed according to the previous descriptions [24]. Briefly, transcriptome sequences of MED and TV were reciprocally blasted to obtain pairs of sequences with best hit to each other with a minimum match length of 200 bp. Then each pair of sequences were searched against the Swiss-prot database using Blastx, and only the pairs of sequences that were unambiguously mapped to the same protein (E value of $1.0E^{-5}$) were retained. The 5'UTR and 3'UTR regions were designated based on predicted CDS. The sequence pairs that contain predicted CDS longer than 150 bp were defined as orthologous genes.

Analysis of sequence divergence and estimation of substitution rates

The divergence of orthologous genes at nondegenerate (nd), four-fold degenerate (4d), CpG and non-CpG regions were calculated according to the previous description [24]. The nonsynonymous sites (Ka), synonymous sites (Ks), and the Ka/Ks ratios were calculated using the YN method with the KaKs Calculator [46].

Analysis of differential gene expression

Stock cultures of MED whitefly (*mtCO1* GenBank accession: GQ371165) and TV were maintained on cotton *Gossypium hirsutum* (Malvaceae) cv. Zhe-Mian 1793 in a climate room of 27 ± 1˚C, 14 h light:10 h darkness, and 70 ± 10% relative humidity. Several hundred female adults from MED and TV were collected for further tests, with two biological replicates for each of the species. Total RNA of each sample was isolated using SV total RNA isolation system (Promega) according to the manufacturer's protocol, respectively. Sequencing libraries were generated using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, USA). Each library was sequenced on an Illumina Hiseq 2000/2500 platform in Novogene Bioinformatics Technology Co., Ltd. (Beijing, China). Clean reads of MED and TV were obtained from NCBI BioProject PRJNA545218 (MED: SRR9141092, SRR9141088; TV: SRR141082, SRR9141090). After removing reads containing adapter or ploy-N, RSEM [47] was used to map the processed RNA-seq reads of each sample to the orthologous region of the two whiteflies [48, 49]. Differential expression analysis between the two species was conducted using edgeR [50]. Differential expression genes were selected with thresholds based on FDR P-value 0.05 and fold change 2. Goseq [51] was used for GO and KEGG enrichment analysis. 'BH' method was used for adjusted p-value [52].

Results and discussions

Reassembly the transcriptome of MED

To improve quality of the MED transcriptome previously reported, Trinity software was used to *de novo* reassemble the sequencing. The reads were assembled into 95,441 sequences (N50 = 725bp) with the length cut off of 200bp. Sequence analysis showed that 12,050 of the Trinity assembled sequences are longer than 1000 bp and 2,761 sequences longer than 2,000 bp, compared to those obtained using the SOAP method of which 4,591 sequences are longer than 1,000 and 662 sequences longer than 2,000, indicating that the new assembled transcriptome has been substantially improved.

Annotation of predicted proteins

For functional annotation, the Trinity assembly results were searched against the NCBI non-redundant (nr) protein database using BLASTx. A total of 27,728 sequences returned significant BLAST hits (e-value$<1.0E^{-5}$). Of them, 9,673 sequences are annotated with GO terms (E-value $<1.0E^{-5}$), 5,398 match in "biological process", 8,582 in "molecular function", and 3,080 in "cellular component". In addition, 8,469 sequences could be assigned to 293 KEGG pathways.

Identification of orthologous genes between MED and TV

To compare the sequence divergence between MED and TV, bidirectional best hit approach, which had been widely used to identify orthologous genes [24, 53, 54], was used to find orthologous genes between the transcriptomes of the two species of whitefly. To remove potential paralogs, these putative orthologous genes were further screened against the Swiss-prot database. Only pairs of sequences that mapped unambiguously to the same protein in Swiss-prot database with an e-value $< 1.0E^{-5}$ were selected as orthologous genes. Totally, 4,850 pairs of orthologs were kept with an average length of 591 bp and 82.02% identity (ranging from 76.6% to 100%). The untranslated region (UTR) of each sequence pair was identified based on the predicted coding region. Among the 4,850 pairs of orthologs, 57 pairs contain 5'UTR, and 54 pairs contain 3'UTR. After removing the UTRs, the CDS of all the orthologs were obtained. The CDS sequences containing unexpected stop codon were further filtered, resulting in 2,944 pairs of orthologous CDS sequences. The average length of the orthologous genes is 555 bp with an average similarity of 81.9%, which is much lower than that between the MEAM1 and MED species of the *B. tabaci* species complex (mean = 99.2%). The average GC content of the orthologous CDS is 42.3%, a value slightly lower than those of MEAM1 and MED species.

Sequence divergence between the orthologous genes

Among the 2,944 orthologous gene pairs, the overall divergence in CDS is 18.1%. In non-CpG sites, the divergence is lower (16.1%); whereas in the CpG sites, the divergence (37.2%) is 2.3 times as high as that of non-CpG sites (Table 1). Nucleotides in coding regions were further classified as non-degenerative (nd) sites (any nucleotide substitutions produce amino acid change) and four-fold degenerate (4d) sites (no changes cause amino acid replacement). From a total of 1,634.08 kb of coding region sequences, 954.77 kb are nd sites, and 223.91 kb are 4d sites. At nd sites, the overall divergence is 3.7%, whereas the overall divergence at 4d sites (56.4%) is 15.4 times of that at the nd sites (Table 1). These results indicate that the nd sites evolve under extensive functional constraints because any nucleotide substitutions at nd sites will produce amino acid changes.

Table 1. Sequence divergence between MED and TV transcriptomes.

	%CpG	%GC	*Loci*	% differences		Compared kb
				Mean	SE	
CDS	10.02	42.34	2944			
All				18.12	0.06	1634.08
No CpG				16.06	0.06	1470.31
CpG				37.20	0.14	163.77
nd sites*	8.6	43.89	2944			
All				3.67	0.04	954.77
No CpG				3.50	0.04	872.66
CpG				5.45	0.12	82.12
4d sites#	17.94	35.98	2944			
All				56.38	0.18	223.91
No CpG				50.38	0.19	183.73
CpG				85.11	0.28	40.18

*nd sites: non-degenerative sites where any nucleotide substitutions produce amino acid change.

#4d sites: four-fold degenerate sites where no changes cause amino acid replacement.

https://doi.org/10.1371/journal.pone.0237744.t001

Synonymous and non-synonymous sites between the orthologous genes

To identify genes undergoing purifying or positive selections, rates of nonsynonymous (Ka) and synonymous (Ks) substitutions, a measure widely used to measure the intensity and mode of selection, between MED and TV ortholog pairs were estimated [55]. Among the 2,944 pairs of CDS, both a Ka and a Ks rate could be calculated for 2,742 orthologs. The mean of Ks is 2.23 (median value = 1.98), indicating that synonymous sites had substituted more than 2 times on average. The median Ks value is higher than that of the comparison between human and chicken (1.66) [56]. The Ka/Ks ratio between MED and TV (average ratio = 0.0204) is much lower than those between the three species, i.e., Asia II 3, MEAM1 and MED of the *B. Tabaci* whitefly complex; the average ratios are 0.198 between Asia II 3 and MEAM1, 0.201 between Asia II 3 and MED, and 0.225 between MED and MEAM 1 [24, 25]. The Ka/Ks ratio between MED and TV is even much lower than those of rodent-human (0.170) [57], chicken-human (0.052) [56], and 12 *Drosophila* species (0.06 to 0.11) [58]. The low Ka/Ks ratio is consistent with the high 4d /nd ratio, suggesting that the orthologues genes have been under high purified selection.

Similarity of orthologous sequences

The 2,944 pairs of orthologous CDS sequences show a mean homology of 81.9%, ranging from 70.2% to 100%. And the average homology is much lower than those shown by pairwise comparisons between species within the *Bemisia tabaci* whitefly complex (MED-MEAM1: 99.2%, MEAM1-Asia II 3: 98.3%, and MED-Asia II 3: 98.2%). Among the 2,944 orthologous gene pairs, only 18 genes show 100% homology, which is much fewer than pairwise comparisons between species within the *Bemisia tabaci* whitefly complex (MED-MEAM1: 604 and MEAM1-Asia II 3: 94) [24, 25]. This result is in line with the wider genetic distance between MED and TV compared with that between MED and MEAM1 or between MEAM1 and Asia II 3. The average identity of amino acid sequences is 93.6%, ranging from 71.7% to 100% (Fig 1a), much lower than that among species within the *Bemisia tabaci* whitefly complex (within MEAM1, MED and Asia II 3, higher than 99%). The average identity is higher than that of chicken-human (~75%) [56], rodent-human (~88%) [57], and the majority of pairwise

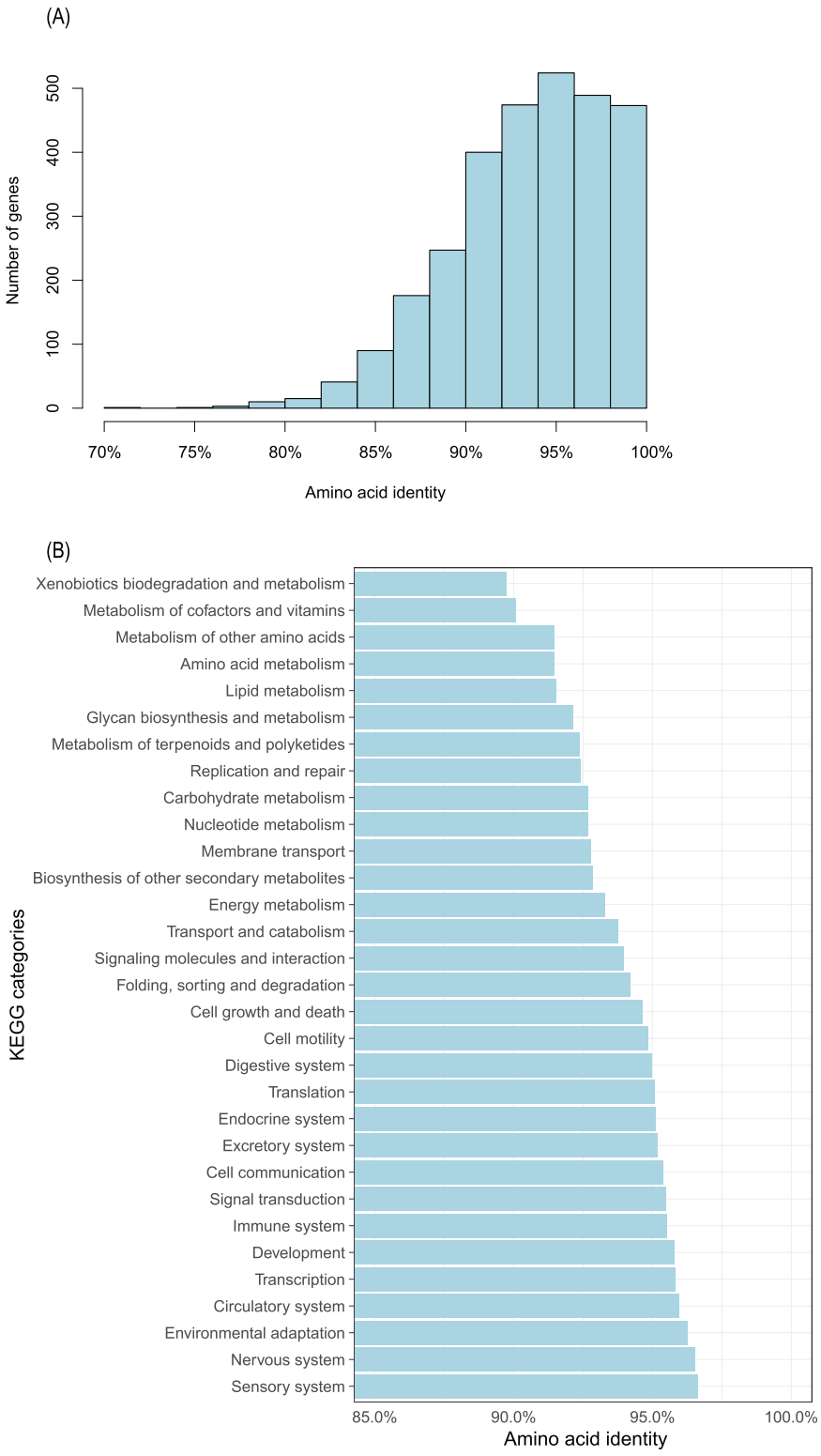

Fig 1. Sequence identity of orthologous. A, the distribution of amino acid identity; and B, the distribution of amino acid identity in KEGG categories.

https://doi.org/10.1371/journal.pone.0237744.g001

comparisons within 7 drosophilids (*D. melanogaster*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. mojavensis*, *D. virilis* and *D. grimshawi* except for one comparision: *D. melanogaster* -*D. erecta*) [58]. Among these orthologous CDS sequences, the most divergence gene pair is *Rho GTPase-activating protein 190* that is related to olfactory learning and memory in *Drosophila* [59].

Next, the orthologous genes were matched to the KEGG pathways to see the distribution of these divergent genes within each pathway (S1 Table). The most highly divergent category was xenobiotics biodegradation and metabolism, followed by categories of metabolism of cofactors, vitamins, amino acids and lipids (Fig 1b, S2 Table); these categories are also highly divergent between species of the *B. tabaci* whitefly complex [24, 25]. Some genes of MEAM1 and TV related to signaling pathways such as junction, spliceosome, synapse and secretion, show relatively low Ka/Ks ratio and probably have been under more purified selection (S1 Table), a pattern similar to those shown by comparison of transcriptome and genome sequences of other species of the *B. tabaci* whitefly complex [24, 25, 60]. This pattern of similarity might be a common feature among most species of whiteflies, and the key pathways play important roles in the divergence of different whiteflies.

Analysis of differential expression of orthologous genes

First, RNA-seq single-end reads in each sample were mapped to *de novo* transcriptome sequences of MED or TV, respectively. The mapping rate of MED ranges from 87.5% to 89.9%, while that of TV ranges from 68.0% to 70.9% (S3 Table). Biological replication showed high reproducibility (Pearson correction was 0.98 in MED and 0.94 in TV). In order to compare differential expression in orthologous genes cross species, RNA-seq reads in each sample were mapped to orthologous regions in 2,944 pair of orthologs. Approximately 8%-10% of RNA-seq reads were mapped to the orthologous regions in each of samples (S3 Table). Genes that had absolute of $\log_2$ MED:TV expression ratio >1 and FDR adjusted $P< 0.05$ were considered as differential expression genes (DEG). In the 2,944 orthologs, 394 genes show over-expression in MED, and 409 genes show over-expression in TV (Fig 2). Some KEGG pathways show a trend to enrich (single p-value < 0.05, but adjusted p-value >0.05). Interestingly, four out of five genes of MED show over-expression in 'regulation of autophagy' pathway. Autophagy is a cellular degradation system, which plays an important role in homeostatic process, development, and pathology [61]. Some carbohydrate metabolism ('galactose metabolism', 'pentose phosphate pathway' and 'fructose and mannose metabolism') tend to be DEG enriched (Table 2; S4 Table), a feature that may be related to the strong capacity in utilizing a wide range of host plants by MED and TV. Likewise, these pathways were also DEGs enriched in the comparisons of expression divergence between MEAM1, MED and Asia II 3 [26]. However, the 'Oxidative phosphorylation' pathway was not DEG enriched (6 DEG out of 20, p-value = 0.39), suggesting that DEG in carbohydrate may not participate in energy metabolism but in other physiological activities. Carbohydrates not only provide energy and structural component, but also participate in various physiological activities such as protection against exposure to unfavorable temperatures, reproduction and embryonic development [62–64].

Genes related with amino acids, vitamins and cofactors

Phloem sap is deficient in several essential amino acids [65] and phloem sap-sucking insects harbor microbial endosymbionts to complement the requirement of these amino acids. As phloem sap-sucking insects, whiteflies harbor the primary endosymbiont *Portiera aleyrodidarum* and one to several secondary symbionts for provision of some essential amino acids, vitamins and cofactors [66–68]. For instance, in MED, seven essential amino acids (Arg, His,

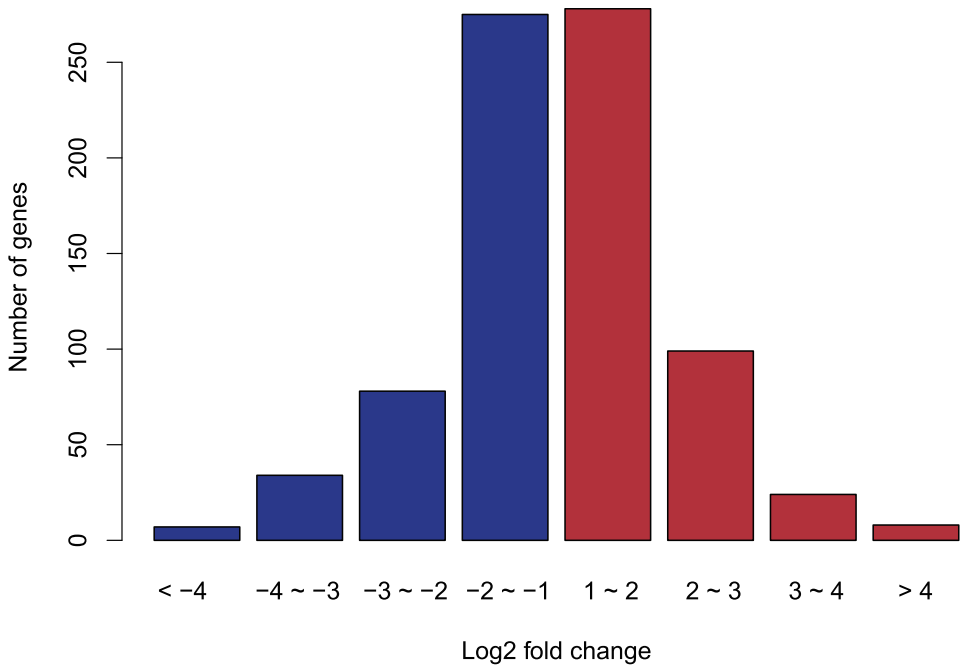

Fig 2. The log 2 ratio distribution of differentially expressed orthologous genes. Blue, MED over-expressed, and red, TV over-expressed.

https://doi.org/10.1371/journal.pone.0237744.g002

Lys, Val, Met, Ile and Leu) are complemented by both whitefly and symbionts, while the other three (Phe, Thr, Trp) are provided by *Portiera* alone [66]. MED and TV have high sequence divergence in categories of metabolism of cofactors, vitamins, amino acids. In order to recognize genes of the whitefly genomes rather than those of symbionts, the orthologous genes were blasted against genome sequences of *Portiera*, *Hamiltonella defensa*, MED and MEAM1. The blast results illustrate that all these genes show high identity with those of MED or MEAM1 genomes, and all top NRs hit from animals, showing that all these orthologous genes come from genomes of the whiteflies and not from the symbionts. High divergent genes are over-represented in both amino acid metabolism (26 out of 65, hyper-test $p<0.05$) and vitamin metabolism (21 out of 37, hyper-test $p<0.05$) (S5 Table). At the expression level, a few genes related to amino acids, vitamins and cofactors show significantly differential expression between MED and TV (24 out of 99) and none of the KEGG pathways are enriched by differential expression genes (S5 Table). Among these genes, some have both high sequences divergence and high expression difference such as *aminoacylase*, *biotin synthase*, and *FAD synthetase*, and may be important in amino acid and vitamin biosynthesis [60, 66] (Table 3). These divergent sequences and/or expression between MED and TV may play an important role in determining their capacity to utilize different host plants.

Table 2. Enriched KEGG pathway among the DEGs.

KEGG pathway	No. of DEGs	No. of genes	p-value	Adjusted p-value
Regulation of autophagy [PATH:ko04140]	4	5	0.016	0.941
Galactose metabolism [PATH:ko00052]	7	13	0.025	0.941
Pentose phosphate pathway [PATH:ko00030]	5	8	0.028	0.941
Fructose and mannose metabolism [PATH:ko00051]	7	14	0.034	0.941

https://doi.org/10.1371/journal.pone.0237744.t002

Table 3. Differential expression genes in metabolism of amino acids, vitamins and cofactors.

KO ID	KO_description	log2FC	FDR	Identity of amino acid sequences
K01939	purA, ADSS; adenylosuccinate synthase [EC:6.3.4.4]	1.24	5.92E-05	93.36%
K01580	E4.1.1.15, gadB, gadA, GAD; glutamate decarboxylase [EC:4.1.1.15]	-1.20	4.16E-04	92.52%
K14677	ACY1; aminoacylase [EC:3.5.1.14]	2.68	3.57E-06	82.71% *
K00318	PRODH; proline dehydrogenase [EC:1.5.-.-]	1.01	1.04E-04	93.22%
K00819	E2.6.1.13, rocD; ornithine—oxo-acid transaminase [EC:2.6.1.13]	-1.86	1.82E-04	95.92%
K13253	DNOS; nitric-oxide synthase, invertebrate [EC:1.14.13.39]	-1.56	3.74E-02	94.12%
K00456	CDO1; cysteine dioxygenase [EC:1.13.11.20]	-1.72	6.85E-06	90.36% *
K00108	E1.1.99.1, betA, CHDH; choline dehydrogenase [EC:1.1.99.1]	2.50	1.86E-02	92.59%
K06101	ASH1L; histone-lysine N-methyltransferase ASH1L [EC:2.1.1.43]	-2.39	1.23E-03	90.27% *
K00453	E1.13.11.11, TDO2; tryptophan 2,3-dioxygenase [EC:1.13.11.11]	4.57	3.00E-02	94.12%
K00451	HGD, hmgA; homogentisate 1,2-dioxygenase [EC:1.13.11.5]	1.60	3.57E-05	85.32% *
K09478	ACADSB; short/branched chain acyl-CoA dehydrogenase [EC:1.3.99.12]	1.83	3.01E-06	90.53% *
K01012	bioB; biotin synthase [EC:2.8.1.6]	2.99	3.98E-06	85.96% *
K01737	queD, ptpS, PTS; 6-pyruvoyltetrahydropterin/6-carboxytetrahydropterin synthase [EC:4.2.3.12 4.1.2.50]	-1.14	8.72E-04	92.22%
K03783	punA; purine-nucleoside phosphorylase [EC:2.4.2.1]	2.72	4.12E-06	95.00%
K06133	LYS5, acpT; 4'-phosphopantetheinyl transferase [EC:2.7.8.-]	1.88	1.78E-03	88.42% *
K00430	E1.11.1.7; peroxidase [EC:1.11.1.7]	-1.34	8.93E-06	93.07%
K00699	UGT; glucuronosyltransferase [EC:2.4.1.17]	1.94	1.73E-06	90.74% *
K15734	SDR16C5; all-trans-retinol dehydrogenase (NAD+) [EC:1.1.1.105]	3.41	1.18E-05	96.25%
K00953	FLAD1; FAD synthetase [EC:2.7.7.2]	-1.08	2.84E-02	78.21% *
K00861	RFK, FMN1; riboflavin kinase [EC:2.7.1.26]	-1.28	7.53E-03	86.07% *
K06125	COQ2; 4-hydroxybenzoate hexaprenyltransferase [EC:2.5.1.-]	2.53	3.49E-06	84.93% *
K06126	COQ6; ubiquinone biosynthesis monooxygenase Coq6 [EC:1.14.13.-]	1.89	3.62E-04	86.59% *
K01800	maiA, GSTZ1; maleylacetoacetate isomerase [EC:5.2.1.2]	-1.77	1.08E-03	96.97%

* Identity of amino acid sequence was lower than lower quantile of all orthologous genes.

https://doi.org/10.1371/journal.pone.0237744.t003

Table 4. Fold change of expression and identity of amino acid sequences in P450s.

Gene name of MED	Gene name of TV	Log_2FC	FDR of DGE	Identity of protein sequences
CYP6DB3	*CYP6DB2*	0.63	0.71	85.07% *
CYP4C64	*CYP4C63*	0.30	0.34	89.94% *
CYP301A1	*CYP301A1*	1.98	0.10	92.93%
CYP18A1	*CYP18A1*	0.44	0.55	93.69%
CYP4CR2	*CYP4CR1*	0.64	0.08	84.52% *
CYP301B1	*CYP301B1*	-0.30	0.76	96.58%
CYP380C14	*CYP380D1*	-1.37	0.44	86.36% *

* Identity of amino acid sequence was lower than lower quantile of all orthologous genes (90.8%).

https://doi.org/10.1371/journal.pone.0237744.t004

Genes related to metabolism of xenobiotics

The most highly divergent KEGG category is xenobiotic metabolism, which may contribute to the differences in host plant range and insecticide resistance between MED and TV. Detoxification of plant toxic compounds and resistance to insecticides can be enhanced by over-expression of GSTs, UGTs, P450s [39, 40, 69–72]. Since gene families of cytochrome P450s and UGTs are expanded in whiteflies [42, 60, 73], the number of orthologous genes in these categories are limited. Gene duplication is a mechanism of adaption to the environment [74], and the duplication of P450 genes is associated with insecticide resistance [75]. In the 2,944 orthologous genes, nine of the 13 genes related to xenobiotic metabolism show relatively high divergence of protein sequences, including one GST (79.17%), and one UGT (90.74%) (S6 Table). Four out of the seven P450 genes show relatively high divergence (Table 4), and among them *CYP4C64* was shown to be associated with imidacloprid resistance in MED [76]. At the expression level, only one GST (*GSTZ1*) is MED over-expressed, and one UGT is TV over-expressed, while all the seven P450 genes do not differ in expression between MED and TV (Table 4, S6 Table). In previous studies on the transcriptomes of MEAM1, MED and Asia II 3, numerous genes related to xenobiotic metabolism were shown to have high divergence [24, 26], while the majority of genes related to drug metabolic pathway were shown to be similarly expressed in the two invasive whiteflies MED and MEAM1 but the expression of these genes in MEAM1 and MED is higher than that in the indigenous whitefly Asia II 3 [26]. RNA-seq analysis across different species and host-plants show that the similar expression patterns of detoxification related genes associated with wide host range of whiteflies [27]. Thus, MED and TV that do not differ in expression of detoxification related genes may share a similar pattern: high detoxification gene expression, wide host range. On the other hand, a wide host range is probably associated with high insecticide resistance in whitefly [30]. Therefore, high sequence divergence and non-differential expression of detoxification related genes between MED and TV may associate with the difference in performance on plants of a wide range and insecticide resistance.

Conclusion

In this study, we reassembled the transcriptome of MED and analyzed the divergence of sequences and expression level between MED and TV. Analysis of sequences divergence of 2,944 orthologous genes showed that these genes have been under strong purified selection. Some genes related to metabolism of xenobiotics, cofactors, vitamins, and amino acids show high protein sequence divergence between the two species of whiteflies. Genes showing differential expression were found to be enriched in carbohydrate metabolism and regulation of

autophagy. These analyses provide valuable molecular references to investigate and understand the biological and molecular differences between MED and TV, and potentially differences between other species of whiteflies.

Supporting information

S1 Table. Identity of amino acid sequences and Ka/Ks in KEGG pathways.
(XLSX)

S2 Table. Identity of amino acid sequences and Ka/Ks in KEGG categories.
(XLSX)

S3 Table. Statistics of reads mapped to orthologous genes.
(XLSX)

S4 Table. Differential expressed genes in carbohydrate metabolism.
(XLSX)

S5 Table. Fold change of expression and identity of amino acid sequences in metabolism of amino acids, vitamins and cofactors.
(XLSX)

S6 Table. Fold change of expression and identity of amino acid sequences in metabolism of xenobiotics.
(XLSX)

Author Contributions

Data curation: Yu-Jun Wang.

Formal analysis: Yu-Jun Wang.

Funding acquisition: Shu-Sheng Liu.

Supervision: Xiao-Wei Wang, Shu-Sheng Liu.

Writing – original draft: Yu-Jun Wang.

Writing – review & editing: Hua-Ling Wang, Shu-Sheng Liu.

References

1. Jones D. Plant viruses transmitted by whiteflies. Eur J Plant Pathol. 2003; 109(3):195–219. https://doi.org/10.1023/A:1022846630513
2. Navas-Castillo J, Fiallo-Olive E, Sanchez-Campos S. Emerging virus diseases transmitted by whiteflies. Annu Rev Phytopathol. 2011; 49:219–48. https://doi.org/10.1146/annurev-phyto-072910-095235 PMID: 21568700.
3. Byrne DN, Bellows TS Jr. Whitefly biology. Annu Rev Entomol. 1991; 36(1):431–57.
4. Roditakis NE. Host plants of greenhouse whitefly *Trialeurodes vaporariorum* westwood (Homoptera: Aleyrodidae) in crete. Attractiveness and impact on whitefly life stages. Agr Ecosyst Environ. 1990; 31 (3):217–24. https://doi.org/10.1016/0167-8809(90)90221-X
5. Iida H, Kitamura T, Honda K. Comparison of egg-hatching rate, survival rate and development time of the immature stage between B- and Q-biotypes of *Bemisia tabaci* (Gennadius) (Homoptera: Aleyrodidae) on various agricultural crops. Appl Entomol Zool. 2009; 44(2):267–73. https://doi.org/10.1303/aez.2009.267
6. Longhurst C, Babcock JM, Denholm I, Gorman K, Thomas JD, Sparks TC. Cross-resistance relationships of the sulfoximine insecticide sulfoxaflor with neonicotinoids and other insecticides in the whiteflies *Bemisia tabaci* and *Trialeurodes vaporariorum*. Pest Manag Sci. 2013; 69(7):809–13. https://doi.org/10.1002/ps.3439 PMID: 23203347.

7. Karatolos N, Denholm I, Williamson M, Nauen R, Gorman K. Incidence and characterisation of resistance to neonicotinoid insecticides and pymetrozine in the greenhouse whitefly, *Trialeurodes vaporariorum* Westwood (Hemiptera: Aleyrodidae). Pest Manag Sci. 2010; 66(12):1304–7. https://doi.org/10.1002/ps.2014 PMID: 20799247.

8. Wang Z, Yan H, Yang Y, Wu Y. Biotype and insecticide resistance status of the whitefly *Bemisia tabaci* from China. Pest Manag Sci. 2010; 66(12):1360–6. https://doi.org/10.1002/ps.2023 PMID: 20824685.

9. De Barro PJ, Liu SS, Boykin LM, Dinsdale AB. *Bemisia tabaci*: a statement of species status. Annu Rev Entomol. 2011; 56:1–19. https://doi.org/10.1146/annurev-ento-112408-085504 PMID: 20690829.

10. Liu SS, Colvin J, De Barro PJ. Species concepts as applied to the whitefly *Bemisia tabaci* systematics: how many species are there? J Integr Agri. 2012; 11(2):176–86. https://doi.org/10.1016/s2095-3119(12)60002-1

11. Boykin LM, De Barro PJ. A practical guide to identifying members of the *Bemisia tabaci* species complex: and other morphologically identical species. Front Ecol Evol. 2014; 2:45. https://doi.org/10.3389/fevo.2014.00045

12. Wainaina JM, De Barro P, Kubatko L, Kehoe MA, Harvey J, Karanja D, et al. Global phylogenetic relationships, population structure and gene flow estimation of *Trialeurodes vaporariorum* (Greenhouse whitefly). Bull Entomol Res. 2017:1–9. https://doi.org/10.1017/S0007485317000360 PMID: 28532532

13. Tsueda H, Tsuchida K. Differences in spatial distribution and life history parameters of two sympatric whiteflies, the greenhouse whitefly (*Trialeurodes vaporariorum* Westwood) and the silverleaf whitefly (*Bemisia argentifolii* Bellows & Perring), under greenhouse and laborator conditions. Appl Entomol Zool. 1998; 33(3):379–83.

14. Zhang GF, Li DC, Liu TX, Wan FH, Wang JJ. Interspecific interactions between *Bemisia tabaci* biotype B and *Trialeurodes vaporariorum* (Hemiptera: Aleyrodidae). Environ Entomol. 2011; 40(1):140–50. https://doi.org/10.1603/EN10135 PMID: 22182623.

15. Ramos N, Neto A, Arsénio S, Mangerico E, Stigter L, Fortunato E, et al. Situation of the whiteflies *Bemisia tabaci* and *Trialeurodes vaporariorum* in protected tomato crops in Algarve (Portugal). EPPO Bulletin. 2002; 32(1):11–5.

16. Arnó J, Albajes R, Gabarra R. Within-plant distribution and sampling of single and mixed infestations of *Bemisia tabaci* and *Trialeurodes vaporariorum* (Homoptera: Aleyrodidae) in winter tomato crops. J Econ Entomol. 2006; 99(2):331–40. https://doi.org/10.1603/0022-0493-99.2.331 PMID: 16686130

17. Czosnek H, Hariton-Shalev A, Sobol I, Gorovits R, Ghanim M. The incredible journey of begomoviruses in their whitefly vector. Viruses. 2017; 9(10). https://doi.org/10.3390/v9100273 PMID: 28946649

18. Wisler GC, Li RH, Liu HY, Lowry DS, Duffus JE. *Tomato chlorosis virus*: a new whitefly-transmitted, phloem-limited, bipartite closterovirus of tomato. Phytopathology. 1998; 88(5):402–9. https://doi.org/10.1094/PHYTO.1998.88.5.402 PMID: 18944918

19. Navas-Castillo J, Camero R, Bueno M, Moriones E. Severe yellowing outbreaks in tomato in spain associated with infections of *Tomato chlorosis virus*. Plant Dis. 2000; 84(8):835–7. https://doi.org/10.1094/PDIS.2000.84.8.835 PMID: 30832134

20. Horowitz AR, Kontsedalov S, Khasdan V, Ishaaya I. Biotypes B and Q of *Bemisia tabaci* and their relevance to neonicotinoid and pyriproxyfen resistance. Arch Insect Biochem Physiol. 2005; 58(4):216–25. https://doi.org/10.1002/arch.20044 PMID: 15756703

21. Nauen R, Stumpf N, Elbert A. Toxicological and mechanistic studies on neonicotinoid cross resistance in Q-type *Bemisia tabaci* (Hemiptera: Aleyrodidae). Pest Manag Sci. 2002; 58(9):868–75. https://doi.org/10.1002/ps.557 PMID: 12233176

22. Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet. 2010; 11(1):31–46. https://doi.org/10.1038/nrg2626 PMID: 19997069.

23. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2010; 12(2):87–98. https://doi.org/10.1038/nrg2934 PMID: 21191423

24. Wang XW, Luan JB, Li JM, Su YL, Xia J, Liu SS. Transcriptome analysis and comparison reveal divergence between two invasive whitefly cryptic species. BMC Genomics. 2011; 12:458. https://doi.org/10.1186/1471-2164-12-458 PMID: 21939539

25. Wang XW, Zhao QY, Luan JB, Wang YJ, Yan GH, Liu SS. Analysis of a native whitefly transcriptome and its sequence divergence with two invasive whitefly species. BMC Genomics. 2012; 13(1):529. https://doi.org/10.1186/1471-2164-13-529 PMID: 23036081

26. Wang YL, Wang YJ, Luan JB, Yan GH, Liu SS, Wang XW. Analysis of the transcriptional differences between indigenous and invasive whiteflies reveals possible mechanisms of whitefly invasion. PLoS ONE. 2013; 8(5):e62176. https://doi.org/10.1371/journal.pone.0062176 PMID: 23667457

27. Malka O, Santos-Garcia D, Feldmesser E, Sharon E, Krause-Sakate R, Delatte H, et al. Species-complex diversification and host-plant associations in *Bemisia tabaci*: A plant-defence, detoxification

perspective revealed by RNA-Seq analyses. Mol Ecol. 2018; 27(21):4241–56. Epub 2018/09/18. https://doi.org/10.1111/mec.14865 PMID: 30222226

28. Xia WQ, Wang XR, Liang Y, Liu SS, Wang XW. Transcriptome analyses suggest a novel hypothesis for whitefly adaptation to tobacco. Scientific reports. 2017; 7(1):12102. Epub 2017/09/25. https://doi.org/10.1038/s41598-017-12387-3 PMID: 28935950

29. Xie W, Wu Q, Wang S, Jiao X, Guo L, Zhou X, et al. Transcriptome analysis of host-associated differentiation in *Bemisia tabaci* (Hemiptera: Aleyrodidae). Frontiers in physiology. 2014; 5:487. Epub 2014/12/30. https://doi.org/10.3389/fphys.2014.00487 PMID: 25540625

30. Pym A, Singh KS, Nordgren Å, Davies TGE, Zimmer CT, Elias J, et al. Host plant adaptation in the polyphagous whitefly, *Trialeurodes vaporariorum*, is associated with transcriptional plasticity and altered sensitivity to insecticides. BMC Genomics. 2019; 20(1):996. Epub 2019/12/21. https://doi.org/10.1186/s12864-019-6397-3 PMID: 31856729

31. Ilias A, Lagnel J, Kapantaidaki DE, Roditakis E, Tsigenopoulos CS, Vontas J, et al. Transcription analysis of neonicotinoid resistance in Mediterranean (MED) populations of *B. tabaci* reveal novel cytochrome P450s, but no nAChR mutations associated with the phenotype. BMC Genomics. 2015; 16:939. Epub 2015/11/18. https://doi.org/10.1186/s12864-015-2161-5 PMID: 26573457

32. Tian L, Song T, He R, Zeng Y, Xie W, Wu Q, et al. Genome-wide analysis of ATP-binding cassette (ABC) transporters in the sweetpotato whitefly, *Bemisia tabaci*. BMC Genomics. 2017; 18(1):330. Epub 2017/04/28. https://doi.org/10.1186/s12864-017-3706-6 PMID: 28446145

33. Yang N, Xie W, Yang X, Wang S, Wu Q, Li R, et al. Transcriptomic and proteomic responses of sweetpotato whitefly, *Bemisia tabaci*, to thiamethoxam. PLoS ONE. 2013; 8(5):e61820. Epub 2013/05/15. https://doi.org/10.1371/journal.pone.0061820 PMID: 23671574

34. Ding TB, Li J, Chen EH, Niu JZ, Chu D. Transcriptome profiling of the whitefly *Bemisia tabaci* MED in response to single infection of *Tomato yellow leaf curl virus*, *Tomato chlorosis virus*, and their co-infection. Frontiers in physiology. 2019; 10:302. Epub 2019/04/20. https://doi.org/10.3389/fphys.2019.00302 PMID: 31001125

35. Geng L, Qian LX, Shao RX, Liu YQ, Liu SS, Wang XW. Transcriptome profiling of whitefly guts in response to *Tomato yellow leaf curl virus* infection. Virol J. 2018; 15(1):14. Epub 2018/01/18. https://doi.org/10.1186/s12985-018-0926-6 PMID: 29338737

36. Hasegawa DK, Chen W, Zheng Y, Kaur N, Wintermantel WM, Simmons AM, et al. Comparative transcriptome analysis reveals networks of genes activated in the whitefly, Bemisia tabaci when fed on tomato plants infected with tomato yellow leaf curl virus. Virology. 2018; 513:52–64. Epub 2017/10/17. https://doi.org/10.1016/j.virol.2017.10.008 PMID: 29035786.

37. Kaur N, Chen W, Fei Z, Wintermantel WM. Differences in gene expression in whitefly associated with CYSDV-infected and virus-free melon, and comparison with expression in whiteflies fed on ToCV- and TYLCV-infected tomato. BMC Genomics. 2019; 20(1):654. Epub 2019/08/17. https://doi.org/10.1186/s12864-019-5999-0 PMID: 31416422

38. Kaur N, Chen W, Zheng Y, Hasegawa DK, Ling KS, Fei Z, et al. Transcriptome analysis of the whitefly, *Bemisia tabaci* MEAM1 during feeding on tomato infected with the crinivirus, *Tomato chlorosis virus*, identifies a temporal shift in gene expression and differential regulation of novel orphan genes. BMC Genomics. 2017; 18(1):370. Epub 2017/05/13. https://doi.org/10.1186/s12864-017-3751-1 PMID: 28494755

39. Yang N, Xie W, Jones CM, Bass C, Jiao X, Yang X, et al. Transcriptome profiling of the whitefly *Bemisia tabaci* reveals stage-specific gene expression signatures for thiamethoxam resistance. Insect Mol Biol. 2013; 22(5):485–96. https://doi.org/10.1111/imb.12038 PMID: 23889345.

40. He C, Xie W, Yang X, Wang SL, Wu QJ, Zhang YJ. Identification of glutathione S-transferases in *Bemisia tabaci* (Hemiptera: Aleyrodidae) and evidence that GSTd7 helps explain the difference in insecticide susceptibility between *B. tabaci* Middle East-Minor Asia 1 and Mediterranean. Insect Mol Biol. 2018; 27(1):22–35. Epub 2017/08/03. https://doi.org/10.1111/imb.12337 PMID: 28767183.

41. Wang XW, Luan JB, Li JM, Bao YY, Zhang CX, Liu SS. *De novo* characterization of a whitefly transcriptome and analysis of its gene expression during development. BMC Genomics. 2010; 11:400. https://doi.org/10.1186/1471-2164-11-400 PMID: 20573269

42. Karatolos N, Pauchet Y, Wilkinson P, Chauhan R, Denholm I, Gorman K, et al. Pyrosequencing the transcriptome of the greenhouse whitefly, *Trialeurodes vaporariorum* reveals multiple transcripts encoding insecticide targets and detoxifying enzymes. BMC Genomics. 2011; 12:56. https://doi.org/10.1186/1471-2164-12-56 PMID: 21261962

43. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011; 29(7):644–52. https://doi.org/10.1038/nbt.1883 PMID: 21572440

44. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25 (17):3389–402. https://doi.org/10.1093/nar/25.17.3389 PMID: 9254694

45. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005; 21(18):3674–6. https://doi.org/10.1093/bioinformatics/bti610 PMID: 16081474.

46. Zhang Z, Li J, Zhao XQ, Wang J, Wong GKS, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genom Proteom Bioinf. 2006; 4(4):259–63. https://doi.org/10.1016/S1672-0229(07)60007-2 PMID: 17531802

47. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. https://doi.org/10.1186/1471-2105-12-323 PMID: 21816040

48. Zhu Y, Li M, Sousa AM, Šestan N. XSAnno: a framework for building ortholog models in cross-species transcriptome comparisons. BMC Genomics. 2014; 15(1):343. https://doi.org/10.1186/1471-2164-15-343 PMID: 24884593

49. LoVerso PR, Cui F. A Computational Pipeline for Cross-Species Analysis of RNA-seq Data Using R and Bioconductor. Bioinform Biol Insights. 2015; 9:165–74. https://doi.org/10.4137/BBI.S30884 PMID: 26692761

50. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26(1):139–40. https://doi.org/10.1093/bioinformatics/btp616 PMID: 19910308

51. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010; 11(2):R14. https://doi.org/10.1186/gb-2010-11-2-r14 PMID: 20132535

52. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological). 1995; 57(1):289–300.

53. Osada N, Hashimoto K, Kameoka Y, Hirata M, Tanuma R, Uno Y, et al. Large-scale analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between *M. fascicularis* and *M. mulatta*. BMC Genomics. 2008; 9:90. https://doi.org/10.1186/1471-2164-9-90 PMID: 18294402

54. Liu T, Tang S, Zhu S, Tang Q, Zheng X. Transcriptome comparison reveals the patterns of selection in domesticated and wild ramie (*Boehmeria nivea* L. Gaud). Plant Mol Biol. 2014; 86(1–2):85–92. https://doi.org/10.1007/s11103-014-0214-9 PMID: 24934879

55. Koonin EV, Wolf YI. Constraints and plasticity in genome and molecular-phenome evolution. Nat Rev Genet. 2010; 11(7):487–98. https://doi.org/10.1038/nrg2810 PMID: 20548290

56. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004; 432 (7018):695–716. https://doi.org/10.1038/nature03154 PMID: 15592404

57. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature. 2004; 428(6982):493–521. https://doi.org/10.1038/nature02426 PMID: 15057822

58. Heger A, Ponting CP. Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. Genome Res. 2007; 17(12):1837–49. https://doi.org/10.1101/gr.6249707 PMID: 17989258

59. Billuart P, Winter CG, Maresh A, Zhao X, Luo L. Regulating axon branch stability: the role of p190 RhoGAP in repressing a retraction signaling pathway. Cell. 2001; 107(2):195–207. https://doi.org/10.1016/s0092-8674(01)00522-0 PMID: 11672527

60. Xie W, Yang X, Chen C, Yang Z, Guo L, Wang D, et al. The invasive MED/Q *Bemisia tabaci* genome: a tale of gene loss and gene gain. BMC Genomics. 2018; 19(1):68. https://doi.org/10.1186/s12864-018-4448-9 PMID: 29357812

61. Klionsky DJ, Emr SD. Autophagy as a regulated pathway of cellular degradation. Science. 2000; 290 (5497):1717–21. https://doi.org/10.1126/science.290.5497.1717 PMID: 11099404

62. Wolfe GR, Hendrix DL, Salvucci ME. A thermoprotective role for sorbitol in the silverleaf whitefly, *Bemisia argentifolii*. J Insect Physiol. 1998; 44(7–8):597–603. https://doi.org/10.1016/s0022-1910(98)00035-3 PMID: 12769942

63. Salvucci ME, Stecher DS, Henneberry TJ. Heat shock proteins in whiteflies, an insect that accumulates sorbitol in response to heat stress. J Therm Biol. 2000; 25(5):363–71. https://doi.org/10.1016/s0306-4565(99)00108-4 PMID: 10838175

64. Chippendale GM. The functions of carbohydrates in insect life processes. In: M R, editor. Biochemistry of insects. New York: Academic Press; 1978. p. 1–54.

65. Douglas AE. Phloem-sap feeding by animals: problems and solutions. J Exp Bot. 2006; 57(4):747–54. https://doi.org/10.1093/jxb/erj067 PMID: 16449374.

66. Luan JB, Chen W, Hasegawa DK, Simmons AM, Wintermantel WM, Ling KS, et al. Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. Genome Biol Evol. 2015; 7(9):2635–47. https://doi.org/10.1093/gbe/evv170 PMID: 26377567

67. Zchori-Fein E, Lahav T, Freilich S. Variations in the identity and complexity of endosymbiont combinations in whitefly hosts. Frontiers in microbiology. 2014; 5:310. https://doi.org/10.3389/fmicb.2014.00310 PMID: 25071729

68. Skaljac M, Zanic K, Ban SG, Kontsedalov S, Ghanim M. Co-infection and localization of secondary symbionts in two whitefly species. BMC Microbiol. 2010; 10(1):142. https://doi.org/10.1186/1471-2180-10-142 PMID: 20462452

69. Tijet N, Helvig C, Feyereisen R. The cytochrome P450 gene superfamily in *Drosophila melanogaster*: Annotation, intron-exon organization and phylogeny. Gene. 2001; 262(1–2):189–98. https://doi.org/10.1016/s0378-1119(00)00533-3 PMID: 11179683

70. Low WY, Ng HL, Morton CJ, Parker MW, Batterham P, Robin C. Molecular evolution of glutathione S-transferases in the genus *Drosophila*. Genetics. 2007; 177(3):1363–75. https://doi.org/10.1534/genetics.107.075838 PMID: 18039872

71. Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, et al. Evolution of supergene families associated with insecticide resistance. Science. 2002; 298(5591):179–81. https://doi.org/10.1126/science.1076781 PMID: 12364796

72. Guo L, Xie W, Wang S, Wu Q, Li R, Yang N, et al. Detoxification enzymes of *Bemisia tabaci* B and Q: biochemical characteristics and gene expression profiles. Pest Manag Sci. 2014; 70(10):1588–94. Epub 2014/02/04. https://doi.org/10.1002/ps.3751 PMID: 24488614.

73. Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, et al. The draft genome of whitefly *Bemisia* tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. BMC Biol. 2016; 14(1):110. https://doi.org/10.1186/s12915-016-0321-y PMID: 27974049

74. Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc Biol Sci. 2012; 279(1749):5048–57. https://doi.org/10.1098/rspb.2012.1108 PMID: 22977152

75. Wondji CS, Irving H, Morgan J, Lobo NF, Collins FH, Hunt RH, et al. Two duplicated P450 genes are associated with pyrethroid resistance in *Anopheles funestus*, a major malaria vector. Genome Res. 2009; 19(3):452–9. https://doi.org/10.1101/gr.087916.108 PMID: 19196725

76. Yang X, Xie W, Wang SL, Wu QJ, Pan HP, Li RM, et al. Two cytochrome P450 genes are involved in imidacloprid resistance in field populations of the whitefly, *Bemisia tabaci*, in China. Pestic Biochem Physiol. 2013; 107(3):343–50. https://doi.org/10.1016/j.pestbp.2013.10.002 PMID: 24267696