Title

A reproducible and sensitive method for generating high-quality transcriptomes from single whitefly salivary glands and other low-input tissues

Journal

Insect Science, 29(5)

ISSN

1672-9609

Authors

Gebiola, Marco

Le, Brandon H

Mauck, Kerry E

Publication Date

2022-10-01

DOI

10.1111/1744-7917.13008

Peer reviewed

Insect science

Insect Science (2022) 0, 1–12, DOI 10.1111/1744-7917.13008

ORIGINAL ARTICLE

A reproducible and sensitive method for generating high-quality transcriptomes from single whitefly salivary glands and other low-input tissues

Marco Gebiola[1], Brandon H. Le[2,3] and Kerry E. Mauck[1]

[1]*Department of Entomology, University of California Riverside, Riverside, California;* [2]*Department of Botany and Plant Sciences, University of California Riverside, Riverside, California and* [3]*Institute of Integrative Genome Biology, University of California Riverside, Riverside, California, USA*

Abstract Transcriptomic studies are an important tool for understanding the molecular pathways underlying host plant use by agricultural pests, including vectors of damaging plant pathogens. Thus far, bulk RNA-Seq has been the main approach for non-model insects. This method relies on pooling large numbers of whole organisms or hundreds of individually dissected organs. The latter approach is logistically challenging, may introduce artifacts of handling and storage, and is not compatible with biological replication. Here, we tested an approach to generate transcriptomes of individual salivary glands and other low-input body tissues from whiteflies (*Bemisia tabaci* MEAM1), which are major vectors of plant viruses. By comparing our outputs to published bulk RNA-Seq datasets for whole whitefly bodies and pools of salivary glands, we demonstrate that this approach recovers similar numbers of transcripts relative to bulk RNA-Seq in a tissue-specific manner, and for some metrics, exceeds performance of bulk tissue RNA-Seq. Libraries generated from individual salivary glands also yielded additional novel transcripts not identified in pooled salivary gland datasets, and had hundreds of enriched transcripts when compared with whole head tissues. Overall, our study demonstrates that it is feasible to produce high quality, replicated transcriptomes of whitefly salivary glands and other low-input tissues. We anticipate that our approach will expand hypothesis-driven research on salivary glands of whiteflies and other Hemiptera, thus enabling novel control strategies to disrupt feeding and virus transmission.

Key words Aleyrodidae; *Bemisia tabaci*; bulk RNA-Seq; insect vectors; low-input RNA-Seq; plant pathogens

Introduction

Bemisia tabaci (Hemiptera: Aleyrodidae) is a complex of a still undetermined number of species of whiteflies (Boykin, 2014), many of which have become globally distributed pests of staple food and fiber crops (Perring *et al*., 2018). Besides direct damage due to removal of phloem sap, which results in production of honeydew and sooty mold, members of the *B. tabaci* species complex damage crops by transmitting at least 300 plant pathogenic viruses (Polston *et al*., 2014). Because they efficiently exploit multiple plant hosts, *B. tabaci* complexes reach high densities at regional scales and move among crops in massive numbers (Naranjo *et al*., 2009; Stansly & Natwick, 2009; Adkins *et al*., 2011). In many affected areas of the world, the chemical controls are beyond reach and whitefly-borne viruses are major drivers of food insecurity. In others, management is primarily via

Correspondence: Marco Gebiola, Department of Entomology, University of California, Riverside, CA, USA. Email: marco.gebiola@gmail.com

multiple prophylactic applications of neonicotinoid insecticides. This approach has deleterious effects on beneficial arthropods (Tsvetkov *et al*., 2017), selects for resistance in whitefly populations (Castle *et al*., 2009), and has unfavorable cost-benefit ratios for growers because most insecticides are only marginally effective in reducing virus acquisition and inoculation (Castle *et al*., 2017). Development of more effective and sustainable whitefly controls requires a fundamental understanding of the molecular mechanisms underlying host plant exploitation and feeding behaviors necessary for virus transmission.

Over the last several years, progress in this area has been facilitated by publication of a draft genome of *B. tabaci* MEAM1 (Chen *et al*., 2016), which has revealed elements that likely contribute to the global pest status of this insect, such as gene families involved in detoxification, virus acquisition and transmission, and insecticide resistance. The draft genome has also facilitated several RNA sequencing (RNA-Seq) studies exploring differential gene expression in viruliferous and non-viruliferous whiteflies at key junctures following virus acquisition (Kaur *et al*., 2017, 2019; Hasegawa *et al*., 2018). These studies derived biological replicates from pools of >100 individual exposed to relevant treatments to provide a whole-body snapshot of similarities and differences in virus effects on whitefly physiology. The feasibility of obtaining RNA-Seq data from a single whitefly has been shown by Sseruwagi *et al*. (2017) and De Marchi *et al*. (2018). While this is an important starting point, obtaining data on gene expression patterns in tissues that are particularly relevant for virus transmission and host defense suppression, such as salivary glands, would represent a major improvement.

Recently, Huang *et al*. (2021) studied salivary proteins of *B. tabaci* MEAM1 using transcriptomic and proteomic approaches, revealing conserved patterns of gene expression between *B. tabaci* and other herbivorous arthropods, as well as unique transcripts and patterns of gene expression specific to *B. tabaci*. Based on this approach, the authors then verified that several of the *B. tabaci* specific genes are differentially expressed in response to changes in host environment. Thus, a more targeted, tissue specific method led directly to insights of potential use for breeding whitefly-resistant crops or developing novel chemical controls. When specific tissues are targeted, such as salivary glands, extraction and pooling of such tissues is normally required to obtain sufficient material for standard library preparation and sequencing protocols. Indeed, Huang *et al*. (2021) pooled 400 individual salivary glands to create a single replicate for sequencing. Extraction of tissues consisting of just a few cells from tiny insects (~1 mm long) is a laborious and time-consuming task. The need to dissect hundreds of salivary glands for RNA-Seq hinders replication and can affect the quality of RNA libraries due to issues arising at the tissue storage step (Shi *et al*., 2021). Additionally, this approach is largely incompatible with manipulative experiments that test hypotheses about connections between environment (e.g., host condition), viruliferous status, and whitefly gene expression in salivary glands.

An RNA-Seq approach based on an individual cell or low-input RNA from very few cells has the potential to overcome these issues, thus making the process of getting high-quality RNA libraries from whitefly salivary glands less challenging. This type of approach may also reveal information not discernible when individuals are pooled, such as increased enrichment of genes relevant to specific pathways and improved detection of low abundance transcripts (Aldridge & Teichmann, 2020). Although single cell sequencing technologies and kits have been available for several years, leading to a greater understanding of fundamental biological processes (Aldridge & Teichmann, 2020), among insects use of single cell RNA-Seq has focused on model species such as *Drosophila melanogaster* (Li, 2020) and mosquitoes (Raddi *et al*., 2020). Whitefly salivary glands consist of at least 13 cells each (Ghanim *et al*., 2001) held together tightly in a group that can be dissected, visualized, and removed using a standard stereomicroscope. We hypothesized that single salivary glands would be suitable for an RNA-Seq workflow employing a hybrid single cell sequencing approach that leverages new technologies for working with small quantities of RNA. To test this, we performed a proof-of-concept methodological study to determine the reproducibility and sensitivity of RNA-Seq data obtained from a tiny amount of *B. tabaci* MEAM1 tissues (salivary glands of a single insect, a single insect head, or a pool of 10 heads) as compared with data from bulk RNA-Seq and from a similar low-input approach (SmartSeq2) (He *et al*., 2020). Our results suggest that this method has the potential to make RNA-Seq studies of whiteflies and other non-model insects more insightful and hypothesis driven, specifically by enabling more complex factorial experimental treatment designs.

Materials and methods

Whitefly rearing and feeding

Whiteflies used in this study were obtained from a laboratory culture established from material collected in 2006 in Arizona, USA (Himler *et al*., 2011) and maintained on cowpea (*Vigna unguiculata*) under

climate-controlled conditions of 25 ± 1°C and a 16 : 8 h light : dark photoperiod. A leaf of cowpea with *B. tabaci* pupae was placed onto a 1% agar Petri dish. Adults emerged within 24 h were released on a cowpea plant in a BugDorm-1 insect rearing cage (MegaView Ltd., Taiwan, China) and allowed to feed for 4 days. After this feeding period, adults were aspirated into a 9-dram plastic vial (Bioquip, USA) and placed on ice until use.

RNA isolation, library preparation, and sequencing

The following tissues were dissected from adults onto a 12-well (5 mm) hydrophobic PTFE printed microscope slide (Fisher Scientific, USA) by using custom-made dissecting tools: salivary glands from a single specimen, a single head, a pool of 10 heads. Before dissections, the working bench, microscope, slides and dissecting tools were thoroughly disinfected using 5% bleach followed by 70% ethanol and RNase Away (Molecular BioProducts, USA). A 5-μL drop of Ultrapure distilled water (Invitrogen, USA) was placed in multiple wells, and a chilled whitefly adult was placed in a well. To minimize debris carryover, wings were removed, and the wingless adult was moved to the next well. Here, the abdomen was removed, and the remaining body parts (head + thorax) moved to another well. At this point, depending on the tissue being dissected, either the head was removed and passed onto another clean well, or the thorax and the head were torn apart at the neck juncture, which resulted in salivary glands flowing outside of the body, and removed from the well. The gland dissection was repeated several times until only glands and no appreciable debris were visible inside the well. Cell lysis of salivary glands and single heads ($n = 3$ biological replicates) was performed on the slide using the NEBNext single cell/low-input RNA library prep kit for Illumina (New England Biolabs, USA), using the protocol for low-input RNA.

RNA was extracted also from three pools of 10 heads each. In this case, after dissecting each head as described above, heads were dropped in a 1.7 mL tube containing 30 μL of Tri-Reagent (Molecular Research Center, USA), flash frozen in liquid nitrogen, and ground using a plastic pestle. After grinding, 270 μL of Tri-reagent were added by simultaneously rinsing the pestle. RNA was then purified using the Zymo direct-zol RNA miniprep plus kit (Zymo Research, USA), with the following modifications. The DNase I step was not performed, as preliminary troubleshooting showed a degradation of RNA following this treatment, and RNA was eluted in 15 μL of ultrapure water. RNA quality and quantity of all 10-head pools was checked on a Bioanalyzer using the RNA pico kit (Agilent Technologies, USA).

Hereafter all samples were processed following the NEBNext low-input RNA library prep protocol, irrespective of RNA extraction method, following manufacturer's instructions. cDNA obtained by primer annealing for first strand synthesis followed by reverse transcription, template switching and PCR amplification, and any PCR products hereafter, were cleaned up with Mag-Bind Total Pure NGS kit (Omega Bio-Tek, USA), which uses magnetic beads, following manufacturer's instructions, with final elution in TE buffer. Quality of cleaned cDNA was checked on a Bioanalyzer using the DNA high-sensitivity chip, and concentration was checked on a Qubit. Samples were normalized to the least concentrated sample in 50 μL of final volume. After fragmentation/end prep, the adaptor for the ligation was taken from the NEBNext Multiplex Oligos for Illumina (Index primers set #1), which also provided oligos for the PCR enrichment of adaptor-ligated DNA. The nine libraries thus obtained were cleaned up and checked again on a Bioanalyzer. Here an unexpected extra peak was detected, due to partial purification. Therefore, another round of clean-up was performed after libraries were normalized and pooled in a 25-μL volume of TE buffer. Libraries from single salivary gland, single head and 10-head pools were sequenced on three different lanes of an Illumina Hiseq 4000 at the University of California, Davis (UCD) DNA Technologies Core, with 150-bp paired-end reads mode. Sequenced libraries were demultiplexed by the UCD Bioinformatics Core. Libraries were submitted to GenBank under accession PRJNA755609.

RNA-Seq data analyses

We tested reproducibility and sensitivity of our limited tissue approach. Quality metrics of libraries were examined using FastQC (Babraham Bioinformatics). Libraries were then cleaned by removing erroneous *k*-mers, discarding read pairs for which one of the reads was deemed unfixable, trimming adapters and low-quality bases, removing unwanted rRNA reads and remaining overrepresented sequences following the protocol by Freedman & Weeks (2021). Reproducibility was assessed between replicates of each tissue sample (salivary gland, 1 head, 10 heads), between treatments and between our datasets and the relevant datasets of Huang *et al*. (2021), who obtained 150-bp paired-end RNA-Seq data from a pool of 400 salivary glands (SRR10527109, hereafter Huang20SG) and from a pool of 40 whole bodies of MEAM1 adults (SRR10527110, hereafter Huang20WB), and of He et al. (2020), who obtained 100-bp single-end read RNA-Seq data from a single salivary gland of MEAM1 using a Smart-seq2 approach (SRR10780448,

SRR10780449, SRR10780450, hereafter He20SG). All these libraries were mapped against the MEAM1 reference genome version 1.2 (Chen *et al*., 2016, www.whiteflygenomics.org) using STAR v2.7.5 (Dobin *et al*., 2013). Sensitivity, here defined as the degree to which gene enrichment from limited tissues compares with bulk RNA-Seq or low-input RNA-Seq data from the same tissue, was tested quantitatively (set analysis) and qualitatively (hierarchical clustering).

Correlation analysis

For the correlation plot, BAM files were initially processed with the multiBamSummary tool in the deeptools2 suite version 3.4.0 (Ramírez *et al*., 2016) with parameters "–binsize 10000" and the plot was generated using the plotCorrelation tool with parameters "-c pearson –removeOutliers –plotNumbers –skipZeros."

Set analysis

Gene counts of mapped reads were obtained by featurecounts (Liao *et al*., 2014) as implemented in the Subread v2.0.2 package (Liao *et al*., 2013) and manually normalized by reads per million (RPM). Biological replicates were pooled and genes with RPM < 1 were filtered out. Set analysis and visualization was generated in R 4.0.3 (R Core Team, 2021) using the UpSetR package version 1.4.0 (Conway & Gehlenborg, 2019).

Hierarchical clustering

Genes were filtered to select the top 2000 most varying genes based on RPM (i.e., largest standard deviation) and the hierarchical clustering heatmap was generated in R 4.0.3 (R Core Team, 2021) using the packages: tidyverse version 1.3.1 (Wickham *et al*., 2019) and pheatmap version 1.0.12 (Kolde, 2019).

Results

To assess the feasibility of studying gene expression activity within a single salivary gland, we generated RNA-Seq libraries, in triplicate, from a single salivary gland isolated from whitefly feeding on cowpea plants (CP_G). As controls, we also generated libraries, in triplicate, from a single whitefly head (CP_H) and 10-head pool (CP_10H). We sequenced, in 150-bp paired-end mode, ∼336 million read pairs from the nine libraries, with ∼17-72 million read pairs per library. After quality control processing of the raw reads, ∼2-55 million cleaned read pairs were generated per library, with an average mapping rate of 90% against the MEAM1 genome (Table 1).

Data reproducibility

Pearson's correlation coefficients for the biological replicates ranged from 0.73 to 0.93 (single gland CP_G), 0.69 to 0.82 (single head CP_H), and 0.93 to 0.97 (10 heads CP_10H) (Fig. 1). This indicates that the data are highly reproducible among the replicates. Compared to a pool of 400 isolated glands (Huang *et al*., 2021), our single gland samples were also strongly correlated (r from 0.51 to 0.71) indicating that our low-input RNA approach captured a good representation of the gland transcriptome. In contrast, 100-bp single-end read RNA-Seq data obtained from individual salivary glands processed using a Smart-seq2 approach (He20SG) had low correlation with the pooled gland sample ($0.19 < r < 0.51$) and our single gland samples ($0.24 < r < 0.62$) despite having good reproducibility among each other (Pearson's correlation coefficients greater than 0.83) (Fig. 1).

Qualitative detection of gland transcripts

To determine if our single gland dataset can qualitatively detect the salivary gland transcriptome, we carried out a set intersection analysis of our dataset against published salivary gland transcriptomes. We first filtered for genes with RPM ≥ 1 in the CP_G [our data], He20SG, Huang20SG, and Huang20WB datasets and compared the detected genes in a set analysis. We detected 10525, 9637, 11935, and 10417 genes, respectively, in the CP_G, He20SG, Huang20SG, and Huang20WB datasets (Fig. 1, blue bars). Approximately 83% of the genes detected in the bulk gland RNA-seq dataset (Huang20SG) were detected in our single gland dataset (Fig. 1, orange bars). Additionally, there were 309 genes identified in our single gland dataset that were not detected in the bulk RNA-seq dataset or He20SG dataset (Fig. 2, red bar), and 46 genes that were not detected in our other two datasets (Fig. S1). The proportion of genes unique to our gland dataset (309) versus genes unique to the bulk gland datasets (1038) is also remarkable, given the ratio of 3 : 400 glands used. The same pattern is apparent when comparing our single and pooled head datasets against the bulk datasets (Figs. S2 and S3).

Table 1 Rates of mapping against the *Bemisia tabaci* MEAM1 reference genome.

RNA-Seq libraries	# Input reads	% Uniquely mapped reads	% Reads mapped to multiple loci	% Unmapped reads
CP_G1 (salivary glands)	13 515 339	88.43	2.50	9.07
CP_G2 (salivary glands)	30 259 531	88.94	2.61	8.45
CP_G3 (salivary glands)	55 375 806	89.34	2.01	8.65
CP_H1 (1 head)	2 419 138	86.95	3.47	9.58
CP_H2 (1 head)	29 532 073	87.87	3.43	8.70
CP_H3 (1 head)	21 213 597	89.81	2.79	7.40
CP_10H_1 (10 heads)	28 991 903	91.47	2.36	6.17
CP_10H_2 (10 heads)	25 314 206	91.73	2.21	6.06
CP_10H_3 (10 heads)	35 913 560	91.56	2.07	6.37
SRR10527109 (Huang20SG)	23 667 382	89.04	3.21	7.75
SRR10527110 (Huang20WB)	24 794 274	86.02	4.61	9.37
SRR10780448 (He20SG)	20 212 661	95.78	2.32	1.90
SRR10780449 (He20SG)	18 261 265	96.05	2.40	1.50
SRR10780450 (He20SG)	20 528 340	95.66	2.47	1.87

Detection of gland-enriched transcripts

To determine whether our single gland transcriptome can detect gland-enriched transcripts, we filtered and selected the top 2000 most varying genes among the single gland, single head, and 10 head samples. Hierarchical clustering of these 2000 genes identified transcripts enriched in each sample (i.e., gland, single head, 10 heads). In particular, the large number of transcripts enriched in the single gland samples compared to the single- or 10-head samples suggests that the transcriptome data was able to detect gland-enriched transcripts that were likely diluted in the single-head or 10-head samples (Fig. 3, third clade from the top).

Discussion

RNA-Seq is a powerful tool for quantifying insect gene expression in response to external stimuli and treatment conditions. Approaches based on pooled populations of many minute individuals can reveal differential expression patterns, coexpression networks, alternative splicing patterns, and the presence of variants. While this can work well for whole-insect transcriptomics studies, it is not well-suited to tissue-specific research questions. Pooling of dissected tissues is logistically challenging and may increase variation due to handling time and degradation. RNA-Seq of small insect tissues is therefore often not compatible with replication or simple factorial experimental designs. To address this, we tested a low-input RNA-Seq approach by comparing expression patterns in single *B. tabaci* MEAM1 salivary gland and other limited tissues to those generated through pooling of 400 *B. tabaci* MEAM1 salivary glands (Huang *et al*., 2021) and a Smart-Seq2 salivary gland dataset (He *et al*., 2020). Our results suggest that this hybrid approach is reproducible and sensitive, with the ability to reveal enrichment patterns not evident in pooled tissue populations.

The low-input RNA method described here relies on processing RNA from salivary glands of a single whitefly, a single whitefly head, and pools of 10 whitefly heads using a kit marketed also for single cell RNA library preparation (NEBNext Single Cell/Low Input RNA Library Prep Kit for Illumina). RNA-Seq data derived from libraries prepared with the low-input RNA method and sequenced using 150-bp paired-end reads were comparable to data obtained from pooling many salivary glands (bulk RNA-Seq) or of whole adult bodies (Huang *et al*., 2021). The gland dataset is of particular interest, given the importance of this tissue for virus retention and inoculation, host plant defense manipulation, detoxification, and resistance to insecticides (Chen *et al*., 2016). Somewhat unexpectedly, there was poor correlation between our gland datasets and the Smart-Seq2 datasets (He *et al*., 2020), despite both having been obtained from a single MEAM1 whitefly. However, from the little details available on the Smart-Seq2 datasets (supplementary information provided for He *et al*., 2020), we can speculate that experimental differences may partially explain this low correlation. One possible reason for low correlation is differences in handling. In our dissections, we were careful to leave glands intact while removing contaminating head material. Details on the dissections are not available

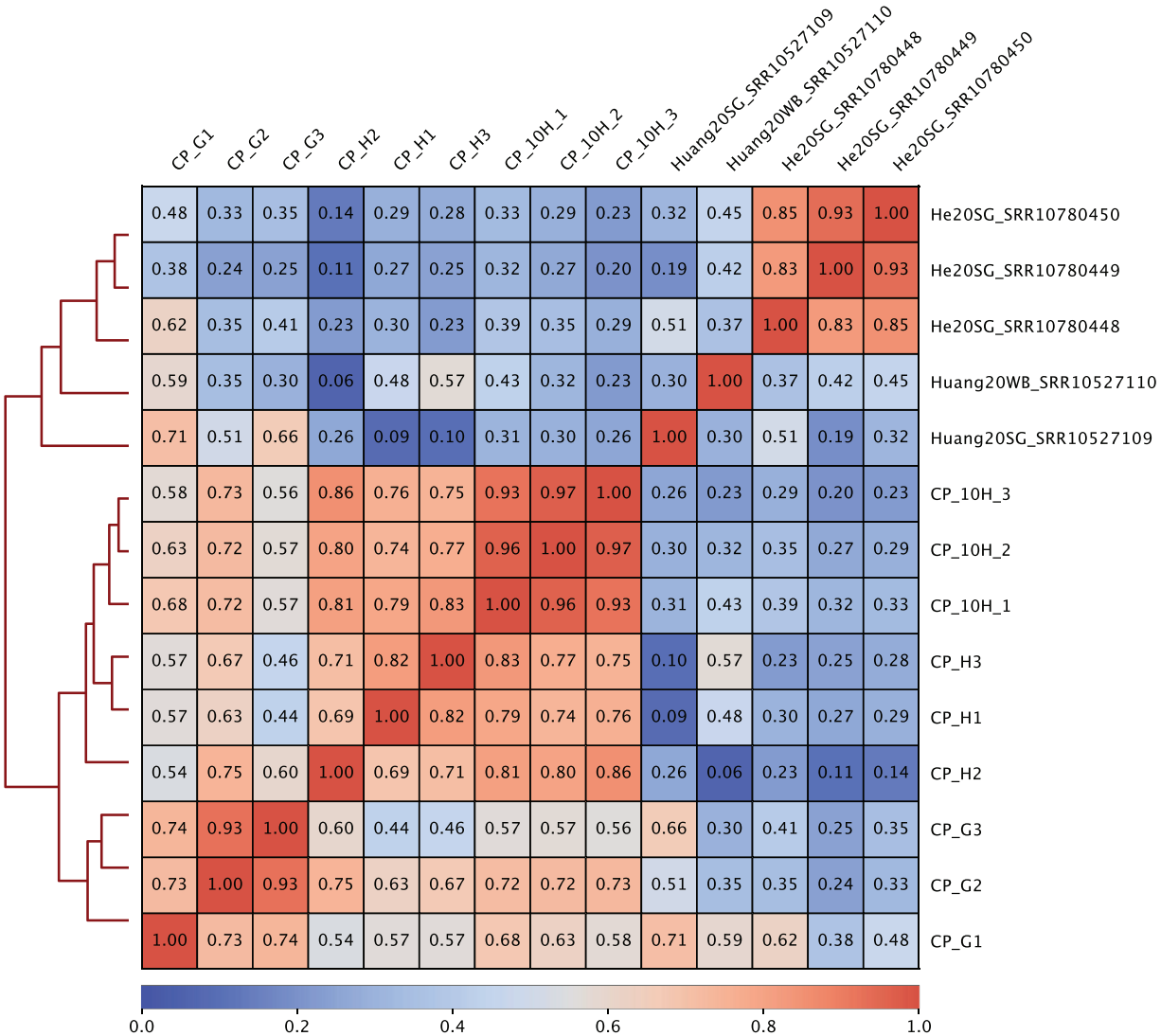

Fig. 1 Heatmap showing Pearson's correlation coefficients between biological replicates, treatments, and external datasets. The color bar represents the correlation coefficients from 0 (no correlation, blue) to 1 (perfect correlation, red). The branches represent grouping of the individual samples, respectively.

in He *et al*. (2020), so we cannot assess whether handling specifics differed substantially between our approaches. Rearing environment may also have influenced dataset congruence. Both the bulk RNA seq dataset (400 pooled glands, Huang *et al*., 2021) and the Smart-Seq2 datasets (He *et al*., 2020) were derived from whiteflies feeding on tomato, while our datasets were derived from whiteflies feeding on cowpea. It is possible that host plant-based differences in salivary gland gene expression are magnified using low-tissue input approaches (our approach or the Smart-Seq-2) relative to pooled tissue approaches. Some of the datasets provided by He *et al*. (2020) also include exposure to geminiviruses, which are transmitted by whiteflies in a circulative persistent manner. This may have also contributed to reduced correlations between the Smart-Seq2 datasets and both the bulk RNA-seq and low-input RNA datasets. Overall, the correlational analyses suggest that low-input approaches may be advantageous for exploring subtle changes in transcript presence/absence or abundance during host switching or in response to virus infection, which can be validated in future studies.

Besides being able to capture a large proportion of the salivary gland transcriptome present in the bulk RNA-seq

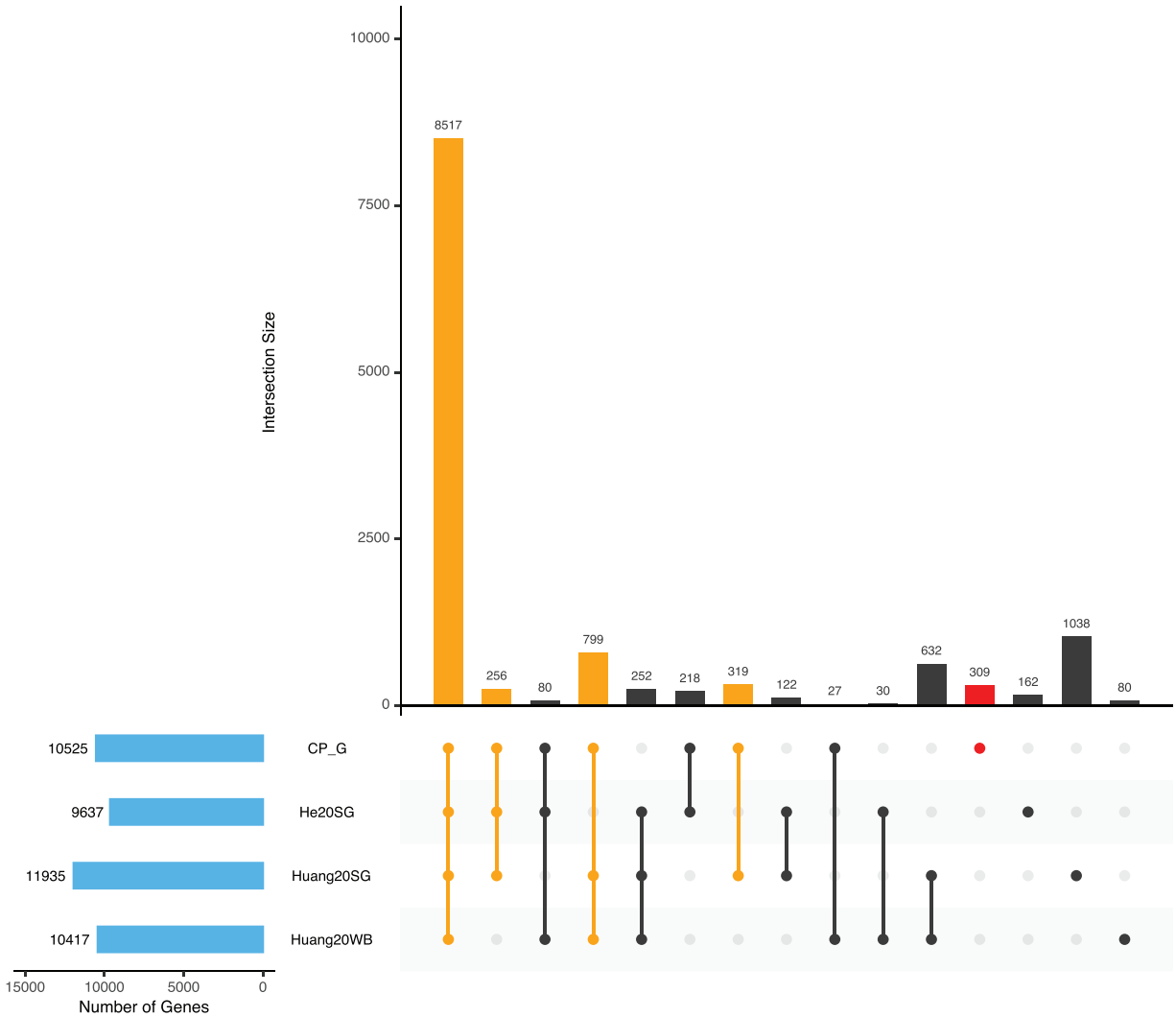

Fig. 2 UpSet plot showing overlaps in expressed genes between datasets. Lower left horizontal bar plot (in blue): total number of genes detected (RPM > 1) in at least one biological replicate, from each sample type. Top vertical bar plot: number of genes represented in the overlap diagram (represented with dots). Samples with solid color dots (orange, black, and red) indicate an overlap and grey dots indicate no overlap. Genes detected in the single gland sample and the external datasets are highlighted in orange. Genes detected only in our single gland dataset are highlighted in red.

(83%), our method detected expression of over 300 genes in our gland libraries (CP_G) that were not detected in the bulk RNA-seq dataset, and 46 genes that were not detected in our other two datasets (Fig. S1). This suggests that we captured both abundant and rare salivary gland-enriched transcripts, which is particularly useful for identifying tissue-specific pathways involved in virus transmission or host exploitation. Examination of quantitative expression patterns revealed that many gene transcripts are enriched specifically in salivary gland tissues (Fig. 3). This confirms that focusing on specific tissues can dramatically improve detection of transcripts that could be otherwise diluted and wrongly considered of low abundance if multiple tissues were used together.

Beyond providing tissue-specific information through a logistically feasible protocol, a key advantage of a low-input RNA approach is the capacity to sequence several replicates for each treatment to estimate the underlying expression distribution for any given gene. Single-cell sequencing workflows must strike a cost-effective balance between replication (10s to 1000s of cells) and sequencing depth, which is a fixed number of achievable reads (Zhang *et al*., 2020). Sequencing fewer cells will provide deeper coverage for each cell but may not be

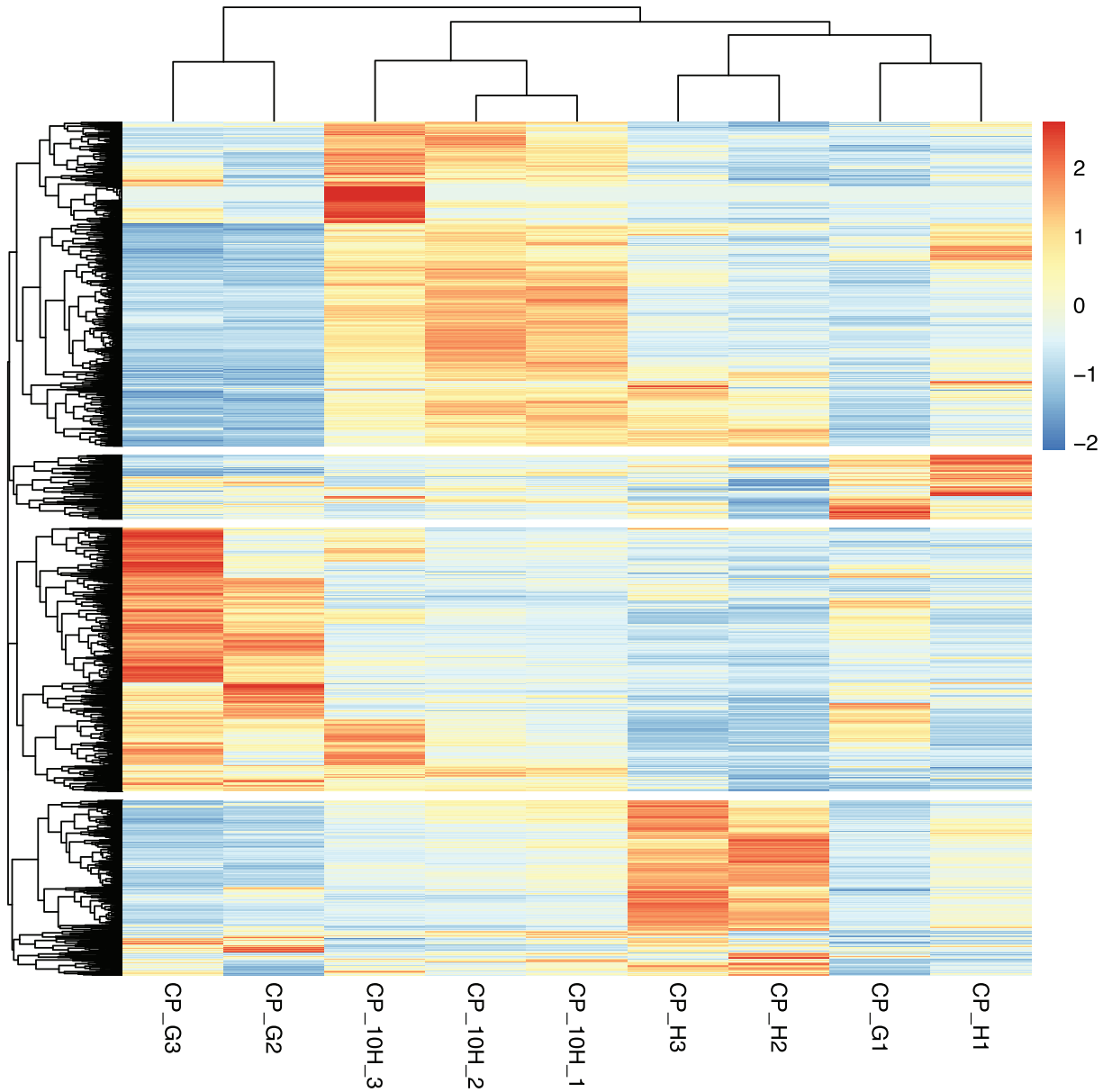

Fig. 3 Hierarchical clustering heatmap of the 2000 most varying genes across tissue samples. Color scale bar indicates *z*-score of each gene across all tissue samples.

useful for understanding the distribution of expression, while sequencing more cells provides information about the distribution but comes with a lot of noise (Zhang *et al*., 2020). In our study, after filtering to remove rRNA reads and the most overexpressed transcripts, two of the three salivary gland libraries (CP_G2 and CP_G3) produced more reads (∼30.2–55.3 million) than the single pooled salivary gland library (∼23.6 million) published by Huang *et al*. (2021), with the third single gland library (CP_G1) still producing a little more than half of the number of reads present in the pooled gland library (∼13.5 million reads) (Table 1).

The better performing CP_G2 and CP_G3 libraries are each derived from 0.25% of the material used to generate the pooled gland library (Huang *et al*., 2021). This could be due to differences in depth of sequencing or, possibly, issues with library normalization, which can occur for both low-input and bulk tissue protocols.

Hierarchical clustering analysis that included non-salivary gland tissues (Fig. 3), along with the nearly halved number of reads, suggest that library CP_G1 could be removed as an outlier, which is a typical practice when generating tens to hundreds of transcriptomes. For example, in a single-cell RNA-Seq study that generated between 100 and 200 single-celled malaria parasite transcriptomes, about 20%–25% were removed as outliers before data analysis (Reid *et al*., 2018). When pooling hundreds of salivary glands, outlier removal is not an option. With our method, we can establish thresholds for inclusion based on a combination of (1) previous studies and available genomic resources, (2) transcript enrichment relative to nontarget adjacent tissues (as we did here), or (3) the relative abundance of a suite of marker genes and the probability of their recovery (as in Zhang *et al*., 2020). Based on the present results, we recommend that future studies using our method for salivary gland transcriptomics in whiteflies, and possibly other Hemiptera, should aim for replication of at least 10 single gland libraries per treatment and at least 20 million reads per library. This recommendation is a starting point and should be validated in future studies that include an experimental treatment, such as host-switching.

Inclusion of single head and pooled head datasets in our study was essential to study consistency in expression patterns among single gland library replicates and establish that CP_G1 is likely an outlier. Beyond this, the head datasets can help us to answer the question of whether heads can be used as more logistically feasible proxy for salivary gland dissection, as has been done in some prior studies with other Hemiptera (Thorpe *et al*., 2016). According to hierarchical clustering analysis of the 2000 most varying genes, several hundred gene transcripts are enriched in salivary gland libraries relative to head libraries (Fig. 3). Additionally, our individual salivary gland dataset had at least 46 unique transcripts relative to both whole head databases (which could have been missed by sampling only heads), and the 10-head database had over 1000 transcripts not represented in the salivary gland database (which may give a false indication of importance if extrapolated for salivary gland specific processes) (Fig. S1). These metrics indicate that different conclusions may be drawn from sampling whole heads in lieu of salivary gland dissections. On the positive side, both whole head databases (CP_10H and CP_H) share many more unique transcripts with each other (1239) relative to shared transcripts between CP_10H and CP_G (262) and CP_H and CP_G (42) (Fig. S1), providing indirect evidence that our dissecting method results in minimal contamination of glands with head RNA. Unique transcripts were also identified in both whole head datasets (851 for CP_H [Fig. S2] and 994 for CP_10H [Fig. S3]) relative to pooled whole-body or salivary gland RNA-Seq datasets (Huang *et al*., 2021), and read counts were also comparable despite reduced tissue inputs, providing evidence that our hybrid workflow is adaptable to producing transcriptomes for other individual body regions.

Overall, our study demonstrates that a low-input, paired-end RNA-Seq workflow is a promising approach for hypothesis-driven transcriptome studies on minute tissues in whiteflies, and possibly other pathogen-transmitting hemipteran insects such as aphids and psyllids. Single salivary gland libraries required less handling time to generate, yielded equivalent or higher-quality data relative to a library of 400 pooled glands, and showed expression patterns that are distinct from adjacent head tissues. Additionally, our hybrid workflow produced better results than a comparable low-input RNA method (Smart-Seq2), which used 100 bp single-end reads instead of 150 bp paired-end reads. It is also adaptable for other low-input tissues, such as individual body regions. Using the workflow described here, researchers can now pursue hypothesis-driven manipulative studies to better understand cause and effect relationships among whiteflies, other organisms, and the environment.

Acknowledgments

This study was funded by USDA NIFA grant #2019-67014-29359 to K.M. and the National Institute of Health Training Grant in Environmental Toxicology T32 ES018827 to B.H.L. We thank Jaimie Kenney and Perla Achi for assistance with whitefly rearing, Penglin Sun for help with library preparation, Clay and Holly Clark at the UCR Genomics Core for BioAnalyzer runs, and Siranoosh Ashtari at the UC Davis DNA Technologies Core for assistance with sequencing. Figure in the graphical abstract was created with BioRender.com.

Disclosure

The authors declare that they have no conflict of interest.

References

Adkins, S., Webster, C.G., Kousik, C.S., Webb, S.E., Roberts, P.D., Stansly, P.A. *et al*. (2011) Ecology and management of whitefly-transmitted viruses of vegetable crops in Florida. *Virus Research*, 159, 110–114.

Aldridge, S. and Teichmann, S.A. (2020) Single cell transcriptomics comes of age. *Nature Communications*, 11, 4307.

Boykin, L.M. (2014) *Bemisia tabaci* nomenclature: lessons learned. *Pest Management Science*, 70, 1454–1459.

Castle, S.J., Palumbo, J. and Prabhaker, N. (2009) Newer insecticides for plant virus disease management. *Virus Research*, 141, 131–139.

Castle, S.J., Palumbo, J., Merten, P., Cowden, C. and Prabhaker, N. (2017) Effects of foliar andsystemic insecticides on whitefly transmission and incidence of *Cucurbit yellow stuntingdisorder virus*. *Pest Management Science*, 73, 1462–1472.

Chen, W., Hasegawa, D.K., Kaur, N., Kliot, A., Valle Pinheiro, P., Luan, J. *et al*. (2016) The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biology*, 14, 110.

Conway, J.R. and Gehlenborg, N. (2019) UpSetR: a more scalable alternative to Venn and Euler diagrams for visualizing intersecting sets. R package version 1.4.0. https://CRAN.R-project.org/package=UpSetR

De Marchi, R.B., Kinene, T., Mbora Wanaina, J., Krause-Sakate, R. and Boykin, L. (2018) Comparative transcriptome analysis reveals genetic diversity in the endosymbiont *Hamiltonella* between native and exotic populations of *Bemisia tabaci* from Brazil. *PLoS ONE*, 13, e0201411.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S. *et al*. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15–21.

Freedman, A. and Weeks, N. (2021) *Best practices for de novo transcriptome assembly with trinity*. Available at: https://informatics.fas.harvard.edu/best-practices-for-de-novo-transcriptome-assembly-with-trinity.html (Accessed: 7 June 2021)

Ghanim, M., Rosell, R.C., Campbell, L.R., Czosnek, H., Brown, J.K. and Ullman, D.E. (2001) Digestive, salivary, and reproductive organs of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) B type. *Journal of Morphology*, 248, 22–40.

Hasegawa, D.K., Chen, W., Zheng, Y., Kaur, N., Wintermantel, W.M., Simmons, A.M. *et al*. (2018) 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*, 513, 52–64.

He, Y.Z., Wang, Y.M., Yin, T.Y., Fiallo-Olivé, E., Liu, Y.Q., Hanley Bowdoin, L. *et al*. (2020) A plant DNA virus replicates in the salivary glands of its insect vector via recruitment of host DNA synthesis machinery. *Proceedings of the National Academy of Sciences USA*, 117, 16928–16937.

Himler, A.G., Adachi-Hagimori, T., Bergen, J.E., Kozuch, A., Kelly, S.E., Tabashnick, B.E. *et al*. (2011) Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science*, 332, 254–256.

Huang, H.J., Ye, Z.X., Lu, G., Zhang, C.X., Chen, J.P. and Li, J.M. (2021) Identification of salivary proteins in the whitefly *Bemisia tabaci* by transcriptomic and LC-MS/MS analyses. *Insect Science*, 28, 1369–1381.

Kaur, N., Chen, W., Zheng, Y., Hasegawa, D.K., Ling, K.S., Fei, Z. *et al*. (2017) 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*, 18, 370.

Kaur, N., Chen, W., Fei, Z. and Wintermantel, W.M. (2019) 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*, 20, 654.

Kolde, R. (2019) pheatmap: Pretty Heatmaps. R package version 1.0.12. https://CRAN.R-project.org/package=pheatmap

Liao, Y., Smyth, G.K. and Shi, W. (2013) The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*, 41, e108.

Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30, 923–930.

Li, H. (2020) Single-cell RNA sequencing in *Drosophila*: technologies and applications. *WIREs Developmental Biology*, 10(5), e396.

Naranjo, S.E., Castle, S.J., De Barro, P.J. and Liu, S.S. (2009) Population dynamics, demography, dispersal and spread of *Bemisia tabaci*. *Bemisia: Bionomics and Management of a Global Pest* (eds. P.A. Stansly & S.E. Naranjo), pp. 185–226. Springer, Dordrecht.

Perring, T.M., Stansly, P.A., Liu, T.-X., Smith, H.A. and Andreason, S.A. (2018) Chapter 4 - Whiteflies: biology, ecology, and management. *Sustainable Management of Arthropod Pests of Tomato* (eds. W. Wakil, G.E. Brust & T.M. Perring), pp. 73–110. Academic Press, San Diego.

Polston, J.E., De Barro, P. and Boykin, L.M. (2014) Transmission specificities of plant viruses with the newly identified species of the *Bemisia tabaci* species complex. *Pest Management Science*, 70, 1547–1552.

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Raddi, G., Barletta, A.B.F., Efremova, M., Ramirez, J.L., Cantera, R., Teichmann, S.A. *et al*. (2020) Mosquito cellular immunity at single-cell resolution. *Science*, 369, 1128–1132.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S. *et al*. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44, W160–W165.

Reid, A.J., Talman, A.M., Bennett, H.M., Gomes, A.R., Sanders, M.J., Illingworth, C.J.R. *et al*. (2018) Single-cell

RNA-seq reveals hidden transcriptional variation in malaria parasites. *eLife*, 7, e33105

Shi, H., Zhou, Y., Jia, E., Pan, M., Bai, Y. and Ge, Q. (2021) Bias in RNA-seq library preparation: current challenges and solutions. *BioMed Research International*, 2021, 6647597.

Sseruwagi, P., Mbora Wainaina, J., Ndunguru, J., Tumuhimbise, R., Tairo, F., Guo, J.Y. *et al*. (2017) The first transcriptomes from field-collected individual whiteflies (*Bemisia tabaci*, Hemiptera: Aleyrodidae): a case study of the endosymbiont composition. *Gates Open Research*, 1, 16.

Stansly, P.A. and Natwick, E.T. (2009) Integrated systems for managing *Bemisia tabaci* in protected and open field agriculture. *Bemisia: Bionomics and Management of a Global Pest* (eds. P.A. Stansly & S.E. Naranjo), pp. 467–497. Springer, Dordrecht.

Thorpe, P., Cock, P.J.A. and Bos, J. (2016) Comparative transcriptomics and proteomics of three different aphid species identifies core and diverse effector sets. *BMC Genomics*, 17, 172.

Tsvetkov, N., Samson-Robert, O., Sood, K., Patel, H.S., Malena, D.A., Gajiwala, P.H., *et al*. (2017) Chronic exposure to neonicotinoids reduces honey bee health near corn crops. *Science*, 356, 1395–1397.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., Francois, R. *et al*. (2019) Welcome to the Tidyverse. *Journal of Open Source Software*, 4, 1686.

Zhang, M.J., Ntranos, V. and Tse, D. (2020) Determining sequencing depth in a single-cell RNA-seq experiment. *Nature Communications*, 11, 774.

Manuscript received August 18, 2021
Final version received December 15, 2021
Accepted January 6, 2022

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1 UpSet plot showing overlaps in expressed genes between our datasets.

Fig. S2 UpSet plot showing overlaps in expressed genes between our one-head and all other datasets.

Fig. S3 UpSet plot showing overlaps in expressed genes between our 10-head and all other datasets.