









RESEARCH ARTICLE

REVISED **The first transcriptomes from field-collected individual whiteflies (*Bemisia tabaci*, Hemiptera: Aleyrodidae): a case study of the endosymbiont composition [version 3; referees: 2 approved]**

Previously titled: The first transcriptomes from field-collected individual whiteflies (*Bemisia tabaci*, Hemiptera: Aleyrodidae)

Peter Sseruwagi^{1*}, James Wainaina^{2*}, Joseph Ndunguru¹, Robooni Tumuhimbise ³, Fred Tairo¹, Jian-Yang Guo^{4,5}, Alice Vrielink², Amanda Blythe ², Tonny Kinene ², Bruno De Marchi ^{2,6}, Monica A. Kehoe⁷, Sandra Tanz ², Laura M. Boykin ²

¹Mikocheni Agriculture Research Institute (MARI), Dar es Salaam, P.O. Box 6226, Tanzania

²School of Molecular Sciences and Australian Research Council Centre of Excellence in Plant Energy Biology, University of Western Australia, Perth, WA, 6009, Australia

³National Agricultural Research Laboratories, P.O. Box 7065, Kampala Kawanda - Senge Rd, Kampala, Uganda

⁴Ministry of Agriculture Key Laboratory of Agricultural Entomology, Institute of Insect Sciences, Zhejiang University, Hangzhou, 310058, China

⁵State Key Laboratory for the Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, 100193, China

⁶Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, Brazil

⁷Department of Primary Industries and Regional Development, DPIRD Diagnostic Laboratory Services, South Perth, WA, Australia

* Equal contributors

v3 **First published:** 28 Dec 2017, 1:16 (doi: [10.12688/gatesopenres.12783.1](https://doi.org/10.12688/gatesopenres.12783.1))
Second version: 13 Feb 2018, 1:16 (doi: [10.12688/gatesopenres.12783.2](https://doi.org/10.12688/gatesopenres.12783.2))
Latest published: 08 Mar 2018, 1:16 (doi: [10.12688/gatesopenres.12783.3](https://doi.org/10.12688/gatesopenres.12783.3))

Abstract






Background: *Bemisia tabaci* species (*B. tabaci*), or whiteflies, are the world's most devastating insect pests. They cause billions of dollars (US) of damage each year, and are leaving farmers in the developing world food insecure. Currently, all publically available transcriptome data for *B. tabaci* are generated from pooled samples, which can lead to high heterozygosity and skewed representation of the genetic diversity. The ability to extract enough RNA from a single whitefly has remained elusive due to their small size and technological limitations.

Methods: In this study, we optimised a single whitefly RNA extraction procedure, and sequenced the transcriptome of four individual adult Sub-Saharan Africa 1 (SSA1) *B. tabaci*. Transcriptome sequencing resulted in 39-42 million raw reads. *De novo* assembly of trimmed reads yielded between 65,000-162,000 Contigs across *B. tabaci* transcriptomes.

Results: Bayesian phylogenetic analysis of mitochondrion cytochrome I oxidase (mtCOI) grouped the four whiteflies within the SSA1 clade. BLASTn searches on the four transcriptomes identified five endosymbionts; the primary endosymbiont *Portiera aleyrodidarum* and four secondary endosymbionts:

Open Peer Review

Referee Status:  

	Invited Referees	
	1	2
REVISED version 3 published 08 Mar 2018		 report
REVISED version 2 published 13 Feb 2018	 report	 report
version 1 published 28 Dec 2017	 report	 report

Arsenophonus, *Wolbachia*, *Rickettsia*, and *Cardinium* spp. that were predominant across all four SSA1 *B. tabaci* samples with prevalence levels of between 54.1 to 75%. Amino acid alignments of the *NusG* gene of *P. aleyrodidarum* for the SSA1 *B. tabaci* transcriptomes of samples WF2 and WF2b revealed an eleven amino acid residue deletion that was absent in samples WF1 and WF2a. Comparison of the protein structure of the *NusG* protein from *P. aleyrodidarum* in SSA1 with known *NusG* structures showed the deletion resulted in a shorter D loop.

Conclusions: The use of field-collected specimens means time and money will be saved in future studies using single whitefly transcriptomes in monitoring vector and viral interactions. Our method is applicable to any small organism where RNA quantity has limited transcriptome studies.

Keywords

Bacterial endosymbionts, sub-Saharan Africa, cassava, smallholder farmers, NusG, next generation sequencing

- 1 **Kai-Shu Ling**, USDA-ARS (United States Department of Agriculture - Agricultural Research Service), USA
- 2 **Henryk H. Czosnek**, The Hebrew University of Jerusalem, Israel

Discuss this article

Comments (0)

Corresponding author: Laura M. Boykin (laura.boykin@uwa.edu.au)

Author roles: **Sseruwagi P:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Wainaina J:** Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Ndunguru J:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Tumuhimise R:** Data Curation, Investigation, Methodology, Resources, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Tairo F:** Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Guo JY:** Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Vrielink A:** Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Blythe A:** Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kinene T:** Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **De Marchi B:** Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kehoe MA:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Tanz S:** Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Boykin LM:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

How to cite this article: Sseruwagi P, Wainaina J, Ndunguru J *et al.* **The first transcriptomes from field-collected individual whiteflies (*Bemisia tabaci*, Hemiptera: Aleyrodidae): a case study of the endosymbiont composition [version 3; referees: 2 approved]** Gates Open Research 2018, 1:16 (doi: [10.12688/gatesopenres.12783.3](https://doi.org/10.12688/gatesopenres.12783.3))

Copyright: © 2018 Sseruwagi P *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: This work was supported by Mikocheni Agricultural Research Institute (MARI), Tanzania through the “Disease Diagnostics for Sustainable Cassava Productivity in Africa” project, Grant no. OPP1052391 that is jointly funded by the Bill and Melinda Gates Foundation and The Department for International Development (DFID) of the United Kingdom (UK). The Pawsey Supercomputing Centre provided computational resources with funding from the Australian Government and the Government of Western Australia supported this work. J.M.W is supported by an Australian Award scholarship by the Department of Foreign Affairs and Trade (DFAT).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

First published: 28 Dec 2017, 1:16 (doi: [10.12688/gatesopenres.12783.1](https://doi.org/10.12688/gatesopenres.12783.1))

REVISED Amendments from Version 2

We have updated the title to "The first transcriptomes from field-collected individual whiteflies (*Bemisia tabaci*, Hemiptera: Aleyrodidae): a case study of the endosymbiont composition", as per the two reviewers suggestions.

See referee reports

Introduction

Members of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex are classified as the world's most devastating insect pests. There are 34 species globally¹ and the various species in the complex are morphologically identical. They transmit over 100 plant viruses^{2,3}, become insecticide resistant⁴, and ultimately cause billions of dollars in damage annually for farmers⁵. The adult whiteflies are promiscuous feeders, and will move between viral infected crops and native weeds that act as viral inoculum 'sources', and deposit viruses to alternative crops that act as viral 'sinks' while feeding.

The crop of importance for this study was cassava (*Manihot esculenta*). Cassava supports approximately 800 million people in over 105 countries as a source of food and nutritional security, especially within rural smallholder farming communities⁶. Cassava production in Sub-Saharan Africa (SSA), especially the East African region, is hampered by whitefly-transmitted DNA and RNA viruses.

Among the DNA viruses, nine cassava-infecting cassava mosaic begomoviruses (CMBs) including: *African cassava mosaic virus* (ACMV), *African cassava mosaic Burkina faso virus* (ACMBV), *African cassava mosaic Madagascar virus* (ACMMV), *East African cassava mosaic virus* (EACMV), *East African cassava mosaic Cameroon virus* (EACMCV), *East African cassava mosaic Kenya virus* (EACMKV), *East African cassava mosaic Malawi virus* (EACMMV), *East African cassava mosaic Zanzibar virus* (EACMZV) and *South African cassava mosaic virus* (SACMV) have been reported in SSA⁷. The DNA viruses cause cassava mosaic disease (CMD) leading to 28–40% crop losses with estimated economic losses of up to \$2.7 billion dollars per year in SSA⁸. The CMD pandemics in East Africa, and across other cassava producing areas in SSA, were correlated with *B. tabaci* outbreaks⁹. Relevant to this study are two RNA ipomoviruses, in the family *Potyviridae*: *Cassava brown streak virus* (CBSV) and the *Ugandan cassava brown streak virus* (UCBSV), both devastating cassava in East Africa. *Bemisia tabaci* species have been hypothesized to transmit these RNA viruses with limited transmission efficiency^{10,11}. Recent studies have shown that there are multiple species of these viruses¹², which further strengthens the need to obtain data from individual whiteflies as pooled samples could contain different species with different virus composition and transmission efficiency. In addition, CBSV has been shown to have a higher rate of evolution than UCBSV¹³ increasing the urgency of understanding the role played by the different whitefly species in the system.

Endosymbionts and their role in *B. tabaci*

Viral-vector interactions within *B. tabaci* are further influenced by bacterial endosymbionts forming a tripartite interaction. *B. tabaci* has one of the highest numbers of endosymbiont bacterial infections with eight different vertically transmitted bacteria reported^{14–17}. They are classified into two categories; primary (P) and secondary (S) endosymbionts, many of which are in specialised cells called bacteriocytes, while a few are also found scattered throughout the whitefly body. A single obligate *P-endosymbiont* *P. aleyrodidarum* is systematically found in all *B. tabaci* individuals. *Portiera* has a long co-evolutionary history with all members of the *Aleyrodinae* subfamily¹⁶. In this study, we further explore genes within the *P. aleyrodidarum* retrieved from individual whitefly transcriptomes, including the transcription termination/antitermination protein *NusG*. *NusG* is a highly conserved protein regulator that suppresses RNA polymerase, pausing and increasing the elongation rate^{18,19}. However, its importance within gene regulation is species specific; in *Staphylococcus aureus* it is dispensable^{20,21}.

The S-endosymbionts are not systematically associated with hosts, and their contribution is not essential to the survival and reproduction. Seven facultative S-endosymbionts, *Wolbachia*, *Cardinium*, *Rickettsia*, *Arsenophonus*, *Hamiltonella defensa* and *Fritschea bemisiae* and *Orientia-like organism* have been detected in various *B. tabaci* populations^{14,22–25}. The presence of S-endosymbionts can influence key biological parameters of the host. *Hamiltonella* and *Rickettsia* facilitate plant virus transmission with increased acquisition and retention by whiteflies²⁵. This is done by protection and safe transit of virions in the haemolymph of insects through chaperonins (*GroEL*) and protein complexes that aid in protein folding and repair mechanisms²⁰.

Application of next generation sequencing in pest management of *B. tabaci*

The advent of next generation sequencing (NGS) and specifically transcriptome sequencing has allowed the unmasking of this tripartite relationship of vector-viral-microbiota within insects^{24,26–28}. Furthermore, NGS provides an opportunity to better understand the co-evolution of *B. tabaci* and its bacterial endosymbionts²⁶. The endosymbionts have been implicated in influencing species complex formation in *B. tabaci* through conducting sweeps on the mitochondrial genome²⁷. Applying transcriptome sequencing is essential to reveal the endosymbionts and their effects on the mitogenome of *B. tabaci*, and predict potential hot spots for changes that are endosymbiont induced.

Several studies have explored the interaction between whitefly and endosymbionts^{29,30} and have resulted in the identification of candidate genes that maintain the relationship^{31,32}. This has been explored as a source of potential RNAi pesticide control targets^{32–34}. RNAi-based pest control measures also provide opportunities to identify species-specific genes for target gene sequences for knock-down. However, to date all transcriptome sequencing has involved pooled samples, obtained through rearing several generations of isolines of a single species to ensure high quantities of RNA for subsequent sequencing. This remains a major bottle neck in particular within arthropoda, where

collected samples are limited due to small morphological sizes³². In addition, the development of isolines is time consuming and often has colonies dying off mainly due to inbreeding depression³⁴.

It is against this background that we sought to develop a method for single whitefly transcriptomes to understand the virus diversity within different whitefly species. We did not detect viral reads, probably an indication that the sampled whitefly was not carrying any viruses, but as proof of concept of the method, we validated the utility of the data generated by retrieving the microbiota *P-endosymbionts* and *S-endosymbionts* that have previously been characterised within *B. tabaci*. In this study we report the endosymbionts present within field-collected individual African whiteflies, as well as characterisation and evolution of the *NusG* genes present within the *P-endosymbionts*.

Methods

Whitefly sample collection and study design

In this study, we sampled whiteflies in Uganda and Tanzania from cassava (*Manihot esculenta*) fields. In Uganda, fresh adult whiteflies were collected from cassava fields at the National Crops Resources Research Institute (NaCRRI), Namulonge, Wakiso district, which is located in central Uganda at 32°36'E and 0°31'N, and 1134 meters above sea level. The whiteflies obtained from Tanzania were collected on cassava in a countrywide survey conducted in 2013. The samples: WF2 (Uganda) and WF1, WF2a, and WF2b (Tanzania) used in this study were collected on CBSD-symptomatic cassava plants. In all the cases, the whitefly samples were kept in 70% ethanol in Eppendorf tubes until laboratory analysis. The whitefly samples were used for a two-fold function; firstly, to optimise a single whitefly RNA extraction protocol and secondly, to unmask RNA viruses and endosymbionts within *B. tabaci* as a proof of concept. In addition, we obtained a *NusG* sequence from a Brazilian NW2 isolate (De Marchi, unpublished) and other downloaded and published *NusG* sequences from GeneBank) to ensure phylogenetic representation across whitefly species.

Extraction of total RNA from single whitefly

RNA extraction was carried out using the ARCTURUS® PicoPure® kit (Arcturus, CA, USA), which is designed for fixed paraffin-embedded (FFPE) tissue samples. Briefly, 30 µl of extraction buffer were added to an RNase-free micro centrifuge tube containing a single whitefly and ground using a sterile plastic pestle. To the cell extract an equal volume of 70% ethanol was added. To bind the RNA onto the column, the RNA purification columns were spun for two minutes at 100 x g and immediately followed by centrifugation at 16,000 x g for 30 seconds. The purification columns were then subjected to two washing steps using wash buffer 1 and 2 (ethyl alcohol). The purification column was transferred to a fresh RNase-free 0.5 ml micro centrifuge tube, with 30 µl of RNase-free water added to elute the RNA. The column was incubated at room temperature for five minutes, and subsequently spun for one minute at 1,000 x g,

followed by 16,000 x g for one minute. The eluted RNA was returned into the column and re-extracted to increase the concentration. Extracted RNA was treated with DNase using the TURBO DNA free kit, as described by the manufacturer (Ambion, Life Technologies, CA, USA). Concentration of RNA was done in a vacuum centrifuge (Eppendorf, Germany) at room temperature for 1 hour, the pellet was suspended in 15 µl of RNase-free water and stored at -80°C awaiting analysis. RNA was quantified, and the quality and integrity assessed using the 2100 Bioanalyzer (Agilent Technologies, CA, USA). Dilutions of up to x10 were made for each sample prior to analysis in the bioanalyzer.

cDNA and Illumina library preparation

Total RNA from each individual whitefly sample was used for cDNA library preparation using the Illumina TruSeq Stranded Total RNA Preparation kit as described by the manufacturer (Illumina, CA, USA). Subsequently, sequencing was carried out using the HiSeq2000 (Illumina) on the rapid run mode generating 2 x 50 bp paired-end reads. Base calling, quality assessment and image analysis were conducted using the HiSeq control software v1.4.8 and Real Time Analysis v1.18.61 at the Australian Genome Research Facility (Perth, Australia).

Analysis of NGS data using the supercomputer

Assembly of RNA transcripts: Raw RNA-Seq reads were trimmed using *Trimmomatic*. The trimmed reads were used for *de novo* assembly using *Trinity*³⁵ with the following parameters: time -p srun --export=all -n 1 -c \${NUM_THREADS} Trinity --seqType fq --max_memory 30G --left 2_1.fastq --right 2_2.fastq --SS_lib_type RF --CPU \${NUM_THREADS} --trimmomatic --cleanup --min_contig_length 1000 -output _trinity min_glue = 1, V = 10, edge_thr = 0.05, min_kmer_cov = 2, path_reinforcement_distance = 150, and group pairs distance = 500.

BLAST analysis of transcripts and annotation: BLAST searches of the transcripts under study were carried out on the *NCBI* non-redundant nucleotide database using the default cut-off on the Magnus Supercomputer at the Pawsey Supercomputer Centre Western Australia. Transcripts identical to known bacterial endosymbionts were identified and the number of genes from each identified endosymbiont bacteria determined.

Phylogenetic analysis of whitefly mitochondrial cytochrome oxidase I (COI)

Phylogenetic analysis of whitefly mitochondrial cytochrome oxidase I (COI): The phylogenetic relationship of mitochondrial cytochrome oxidase I (mtCOI) of the whitefly samples in this study were inferred using a Bayesian phylogenetic method implemented in *MrBayes* (version 3.2.2)³⁶. The optimal substitution model was selected using Akaike Information Criteria (AIC) implemented in the *Jmodel test*³⁷.

Sequence alignment and phylogenetic analysis of *NusG* gene

Sequence alignment and phylogenetic analysis of *NusG* gene in *P. aleyrodidarum* across *B. tabaci* species: Sequence alignment of the *NusG* gene from the P-endosymbiont *P. aleyrodidarum* from the SSA1 *B. tabaci* in this study was compared with

another *B. tabaci* species, *Trialeurodes vaporariorum* and *Alerodius dispersus* using MAFFT (version 7.017)³⁸. The Jmodel version 2³⁷ was used to search for phylogenetic models with the Akaike information criterion selecting the optimal that was to be implemented in MrBayes 3.2.2. MrBayes run was carried out using the command: “lset nst=6 rates=gamma” for 50 million generations, with trees sampled every 1000 generations. In each of the runs, the first 25% (2,500) trees were discarded as burn in.

Analysis and modelling the structure of the NusG gene

The structures for *Portiera aleyrodidarum* BT and *B. tabaci* SSA1 whitefly were predicted using Phyre2³⁹ with 100% confidence and compared to known structures of NusG from other bacterial species. All models were prepared using Pymol (The PyMOL Molecular Graphics System, Version 1.5.0.4).

Results

RNA extraction and NGS optimised for individual *B. tabaci* samples

In this study, we sampled four individual adult *B. tabaci* from cassava fields in Uganda (WF2) and Tanzania (WF1, WF2a, WF2b). Total RNA from single whitefly yielded high quality

RNA with concentrations ranging from 69 ng to 244 ng that were used for library preparation and subsequent sequencing with Illumina Hiseq 2000 on a rapid run mode. The number of raw reads generated from each single whitefly ranged between 39,343,141 and 42,928,131 (Table 1). After trimming, the reads were assembled using Trinity resulting in 65,550 to 162,487 transcripts across the four SSA1 *B. tabaci* individuals (Table 1).

Comparison of endosymbionts within the SSA1 *B. tabaci* samples

Comparison of the diversity of bacterial endosymbionts across individual whitefly transcripts was conducted with BLASTn searches on the non-redundant nucleotide database and by identifying the number of genes from each bacterial endosymbiont (Supplementary Table 1). We identified five main endosymbionts including: *P. aleyrodidarum*, the primary endosymbionts and four secondary endosymbionts: *Arsenophonus*, *Wolbachia*, *Rickettsia* sp, and *Cardinium* spp (Table 2). *P. aleyrodidarum* predominate across all four SSA1 *B. tabaci* study samples based on number of core gene families identified for WF1 (74.82%), WF2 (72.18%), WF2a (53.17%) and WF2b (71.70%). This

Table 1. Summary statistics from De novo Trinity assemble of Illumina paired end individual whitefly transcriptomes.

	WF1	WF2	WF2a	WF2b
Total Number of reads	39,343,141	42,587,057	42,513,188	42,928,131
Number of reads after trimming for quality	34,470,311 (87.61%)	39,898,821 (93.69%)	40,121,377 (94.37%)	40,781,932 (95.00%)
Transcripts	65,550	73,107	162,487	104,539
Number of endosymbiont contigs matching core genes	417	446	568	569
All transcript Contigs (N50)	505	525	1,084	1,018
Only longest Contigs (N50)	468	484	707	746

Table 2. Number of Contigs matching the core genes of bacteria endosymbionts across the four SSA1 whitefly transcriptome.

	Endosymbionts	Number of Contigs matching core genes of respective endosymbionts using BLAST			
		WF1	WF2	WF2a	WF2b
Primary	<i>Portiera</i>	312 (74.82%)	322 (72.18%)	302 (53.17%)	408 (71.70%)
	<i>Hamiltonella</i>	NA	NA	1 (0.002%)	NA
	<i>Rickettsia</i>	11 (2.64%)	8 (1.79%)	147 (25.88%)	NA
	<i>Wolbachia</i>	32 (7.67%)	25	71 (12.5%)	46 (8.08%)
	<i>Cardinium</i>	NA	NA	9 (1.58%)	6 (1.05%)
	<i>Fritschea</i>	NA	NA	NA	NA
Secondary	<i>Arsenophonus</i>	62 (14.87%)	91 (20.40%)	38 (6.69%)	109 (19.16%)
	Total	417	446	568	569

was followed by *Arsenophonus*, *Wolbachia*, *Rickettsia* sp, and *Cardinium* spp, which occurred at an average of 18.0%, 5.9%, 1.6% and <1%, respectively across all four study samples (Table 2). The secondary endosymbiont *Hamiltonella* had the least number of core genes (n=1) and was detected in only one of the SSA1 *B. tabaci* (WF2a) samples.

Phylogenetic analysis of single whitefly mitochondrial cytochrome oxidase I (COI)

B. tabaci is recognized as a species complex of 34 species based on the mitochondrion cytochrome oxidase I^{1,38,39}. We therefore used cytochrome oxidase I (COI) transcripts of the four individual whitefly to ascertain *B. tabaci* species status and their phylogenetic relation using reference *B. tabaci* COI GenBank sequences found at <http://www.whiteflybase.org>. All four COI sequences clustered within Sub-Saharan Africa 1 clade (SSA1) species with greater than 99% identity (www.whiteflybase.org).

Sequence alignment and Bayesian phylogenetic analysis of NusG gene

Nucleotide and amino acid sequence alignments of the NusG in *P. aleyrodidarum* were conducted for several whitefly species including: *B. tabaci* (SSA1), Mediterranean (MED) and Middle East Asia Minor 1 (MEAM1), New World 2 (NW2), *T. vaporariorum* (Greenhouse whitefly) and *Aleurodicus dispersus*. The

alignment identified 11 missing amino acids in the NusG sequences for the SSA1 *B. tabaci* samples: WF2 and WF2b, *T. vaporariorum* (Greenhouse whitefly) and *Aleurodicus dispersus*. However, all 11 amino acids were present in samples WF1 and WF2a, MED, MEAM1 and NW2 (Figure 1). Bayesian phylogenetic relationships of the NusG sequences of *P. aleyrodidarum* for the different whitefly species clustered all four SSA1 *B. tabaci* (WF1, WF2, WF2a and WF2b) within a single clade together with ancestral *B. tabaci* from GenBank (Figure 2). The SSA1 clade was supported by posterior probabilities of 1 with *T. vaporariorum* and *Aleurodicus*, which formed clades at the base of the phylogenetic tree (Figure 2).

Structure analysis of Portiera NusG genes

Structures of the NusG protein sequence of the primary endosymbiont *P. aleyrodidarum* in the four SSA1 *B. tabaci* samples were predicated using Phyre2 with 100% confidence, and compared to known structures of NusG from other bacterial species including (*Shigella flexneri*, *Thermus thermophilus* and *Aquifex aeolicus*; (PDB entries 2KO6, 1NZ8 and 1MIH, respectively) and Spt4/5 from yeast (*Saccharomyces cerevisiae*; PDB entry 2EXU)^{18,40,41}. The 11-residue deletion was found in a loop region that is variable in length and structure across bacterial species, but is absent from archaeal and eukaryotic species (Figure 3 and Figure 4A). The effect of the deletion appears to shorten the loop

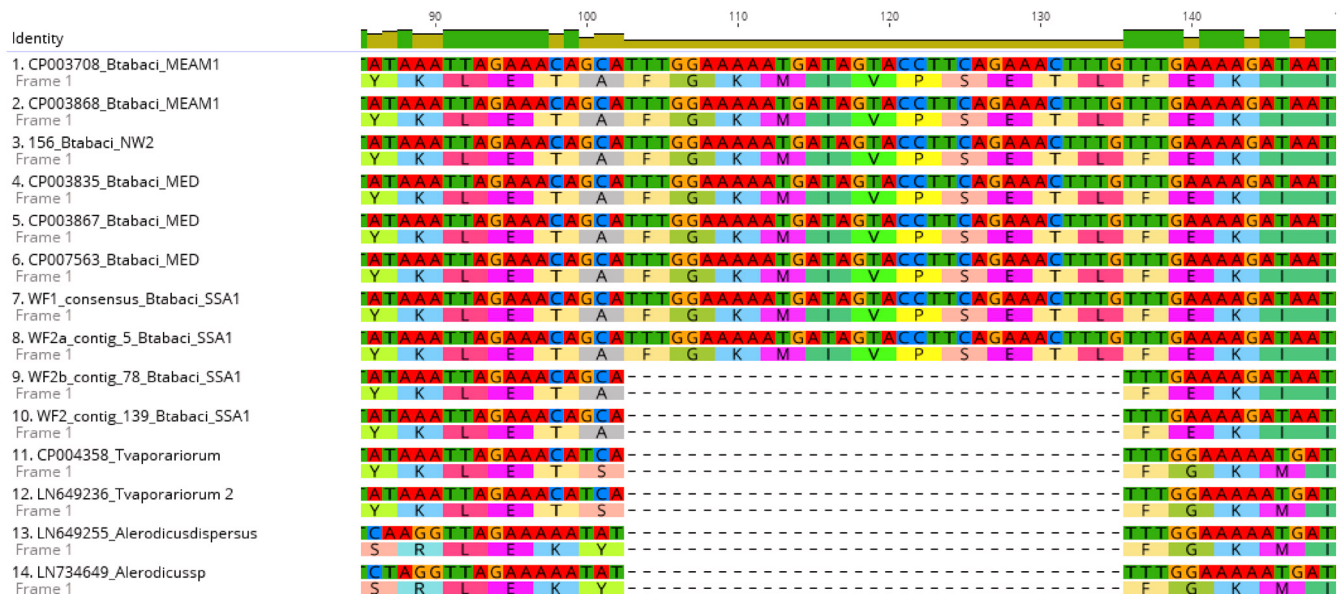


Figure 1. Sequence alignment of nucleotide sequences of NusG gene in *P. aleyrodidarum* across whitefly species sequences using MAFFT v 7.017.

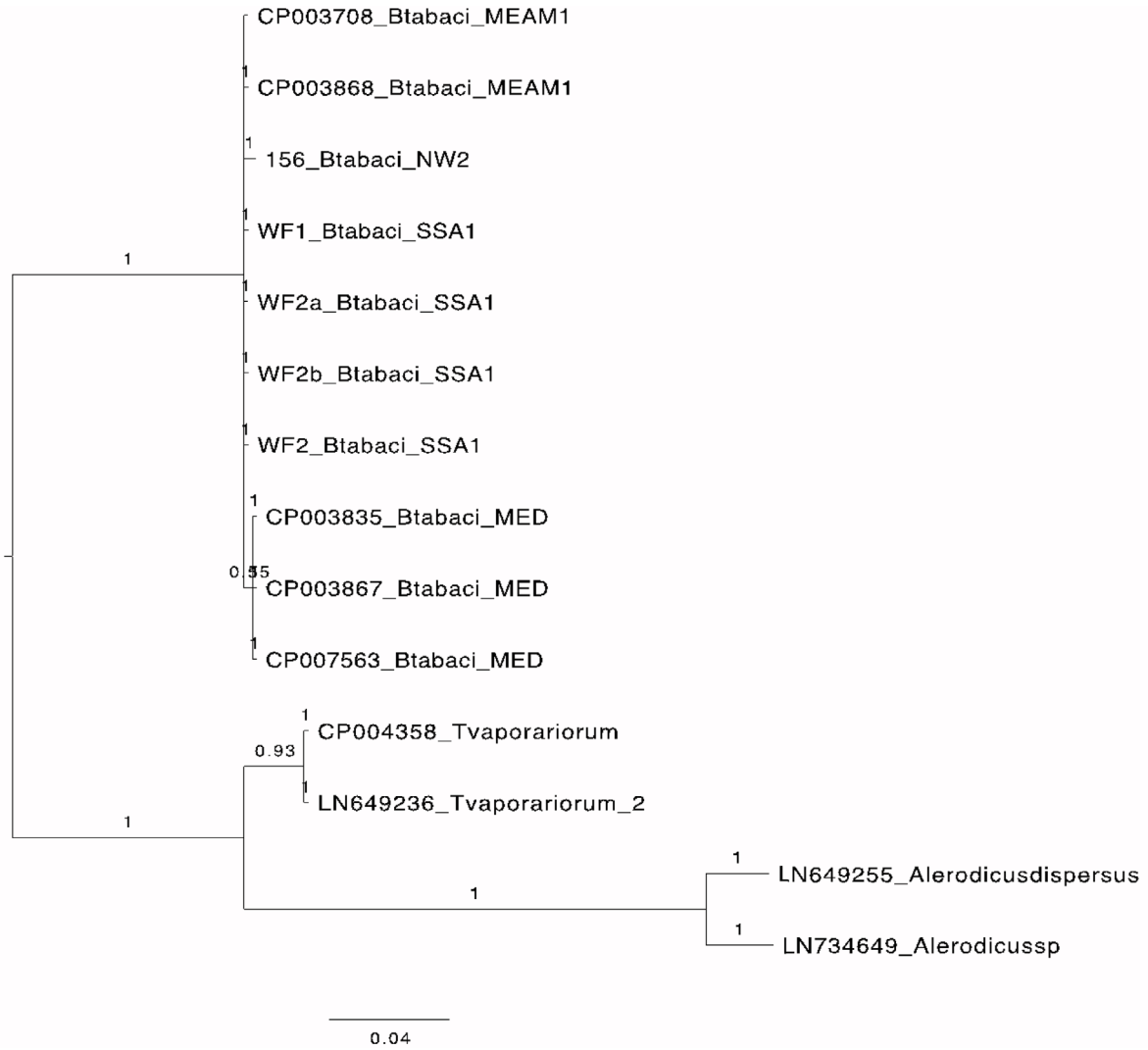


Figure 2. Bayesian phylogenetic tree of *NusG* gene of *P. aleyrodidarum* across whitefly species using MrBayes -3.2.2.

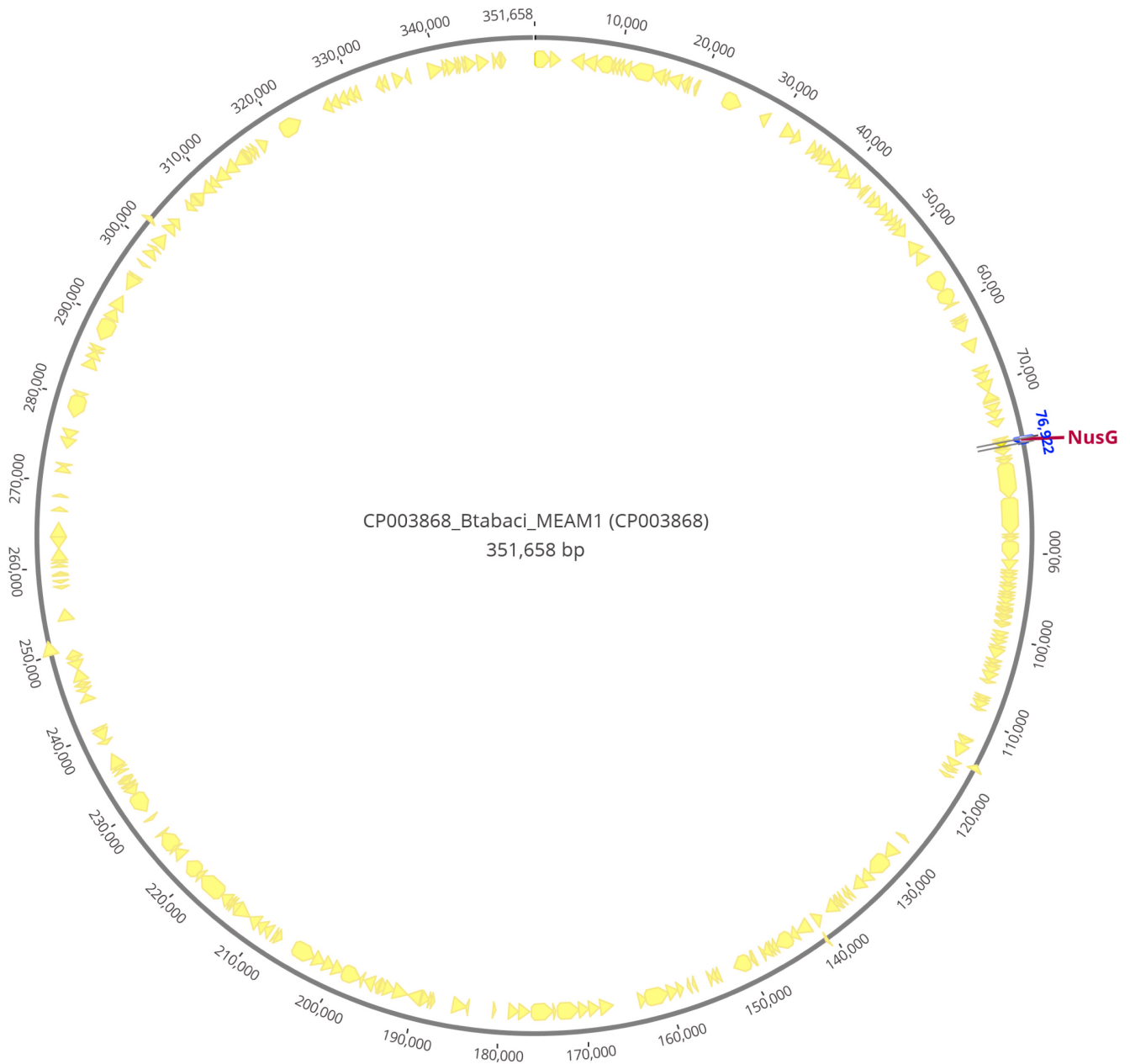
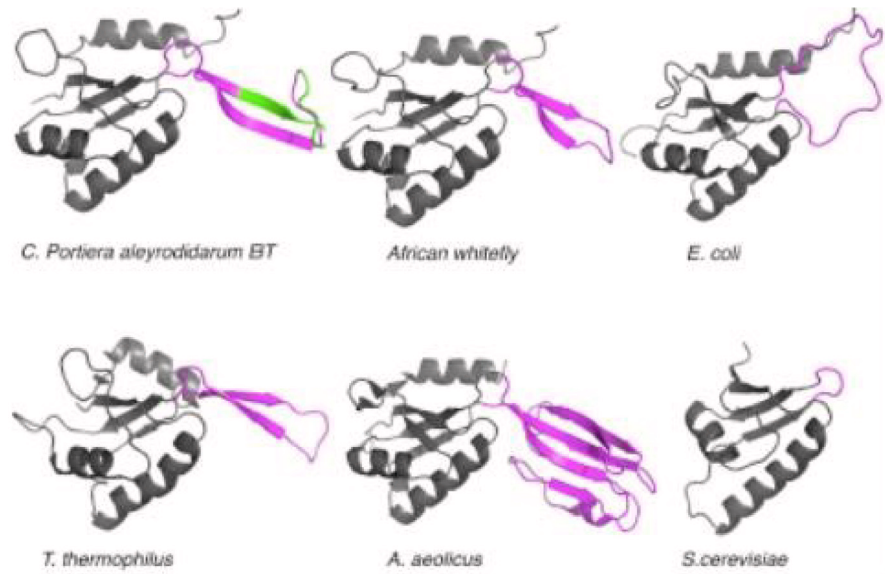
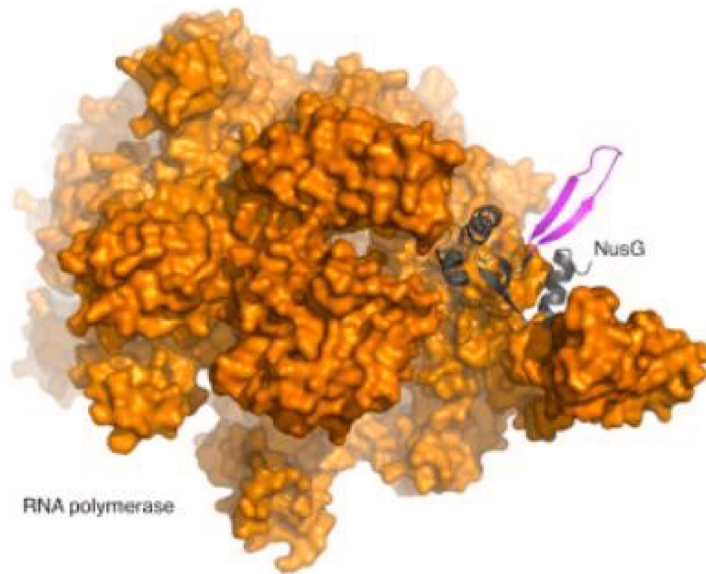


Figure 3. Primary endosymbiont *Portiera aleyrodidarum* whole genome from GenBank CP003868 showing the section of the NusG gene included in the analyses (position 76,922).



A



B

Figure 4. Structure analysis of NusG from *P. aleyrodidarum* in *B. tabaci* and other endosymbionts. **A. Phyre2 based structure prediction of NusG from *Candidatus Portiera aleyrodidarum* in *B. tabaci* SSAI whitefly and comparisons to the structures of NusG from other bacterial species as indicated and of Spt4/5 from yeast. NusG is coloured in grey, the loop region in magenta and the 11-residue deletion is shown in green in the *C. Portiera aleyrodidarum* structure. **B.** A model of bacterial RNA polymerase (orange surface representation) bound to the N-terminal domain of the *T. thermophilus* NusG (grey cartoon representation).**

in *NusG* from the African whiteflies (WF2 and WF2b). A model of bacterial RNA polymerase (orange surface representation; PDB entry 2O5I) bound to the N-terminal domain of the *Thermus thermophilus NusG* shows that the loop region is not involved in the interaction between *NusG* and RNA polymerase (Figure 4B).

Discussion

In this study, we optimised a single whitefly RNA extraction method for field-collected samples. We subsequently successfully conducted RNA sequencing on individual Sub-Saharan Africa 1 (SSA1) *B. tabaci*, revealing unique genetic diversity in the bacterial endosymbionts as proof of concept. This is the first time a single whitefly transcriptome has been produced.

NusG deletion and implications within *P. aleyrodidarum* in SSA *B. tabaci*

We report the presence of the primary endosymbionts *P. aleyrodidarum* and several secondary endosymbionts within SSA1 transcriptome. Furthermore, *P. aleyrodidarum* in SSA1 *B. tabaci* was observed to have a deletion of 11 amino acids on the *NusG* gene that is associated with cellular transcriptional processes within another bacteria species. On the other hand, *P. aleyrodidarum* from NW2, MED and SSA1 (WF2a, WF1) *B. tabaci* species did not have this deletion (Figure 1). The deleted 11 amino acids were identified in a loop region of the N-terminal domain of *NusG* protein, resulting in a shortened loop in the SSA1 WF2b sample. This loop region has high variability in both structure and length across bacterial species, and is absent from archaea and eukaryotic species.

NusG is highly conserved and a major regulator of transcription elongation. It has been shown to directly interact with RNA polymerase to regulate transcriptional pausing and rho-dependent termination^{15,20,42,43}. Structural modelling of *NusG* bound to RNA polymerase indicated that the shortened loop region seen in the WF2b sample is unlikely to affect this interaction. Rho-dependant termination has been attributed to the C-terminal (KOW) domain region of *NusG*, therefore a shortening of the loop region in the N-terminal domain is also unlikely to affect transcription termination. Yet, there has been no function attributed to this loop region of *NusG*, and thus the effect of variability in this region across species is unknown. However, the deletion could represent the results of evolutionary species divergence. Further sequencing of the gene is required across the *B. tabaci* species complex to gain further understanding of the diversity.

Why the single whitefly transcriptome approach?

The sequencing of the whitefly transcriptome is crucial in understanding whitefly-microbiota-viral dynamics and thus circumventing the bottlenecks posed in sequencing the whitefly genome. The genome of whitefly is highly heterozygous⁴². Assembling of heterozygous genomes is complex due to the de Bruijn graph structures predominantly used⁴³. To deal with the heterozygosity, previous studies have employed inbred lines, obtained from rearing a high number of whitefly isolines^{35,44}. However, rearing whitefly isolines is time consuming and often

colonies may suffer contaminations, leading to collapse and failure to raise the high numbers required for transcriptome sequencing.

We optimised the ARCTURUS® PicoPure® kit (Arcturus, CA, USA) protocol for individual whitefly RNA extraction with the dual aim of determining if we could obtain sufficient quantities of RNA from a single whitefly for transcriptome analysis and secondly, determine whether the optimised method would reveal whitefly microbiota as proof of concept. Using our method, the quantities of RNA obtained from field-collected single whitefly samples were sufficient for library preparation and subsequent transcriptome sequencing. Across all transcriptomes over 30M reads were obtained. The amount of transcripts were comparable to those reported in other arthropoda studies from field collections³². However, we did not observe any difference in assembly qualities³²; probably due to the fact that our field-collected samples had degraded RNA based on RIN, and thus direct comparison with ³² was inappropriate.

Degraded insect specimen have been used successfully in previous studies⁴⁵. This is significant, considering that the majority of insect specimens are usually collected under field conditions and stored in ethanol with different concentrations ranging from 70 to 100%^{46–48} rendering the samples liable to degradation. However, to ensure good keeping of insect specimen to be used for mRNA and total RNA isolation in molecular studies, and other downstream applications such as histology and immunocytochemistry, it is advisable to collect the samples in an RNA stabilizing solution such as RNAlater. The solution stabilizes and protects cellular RNA in intact, unfrozen tissue, and cell samples without jeopardizing the quality, or quantity of RNA obtained after subsequent RNA isolation. The success of the method provided an opportunity to unmask vector-microbiota-viral dynamics in individual whiteflies in our study, and will be useful for similar studies on other small organisms.

Endosymbionts diversity across individual SSA1 *B. tabaci* transcriptomes

In this study, we identified bacterial endosymbionts (Table 2) that were comparable to those previously reported in *B. tabaci*⁴⁹ and more specifically SSA1 on cassava^{23,37}. Secondary endosymbionts have been implicated with different roles within *B. tabaci*. *Rickettsia* has been adversely reported across putative *B. tabaci* species, including the Eastern African region^{23,50,51}. This endosymbiont has been associated with influencing thermo tolerance in *B. tabaci* species⁴⁹. *Rickettsia* has also been associated with altering the reproductive system of *B. tabaci*, and within the females. This has been attributed to increasing fecundity, greater survival, host reproduction manipulation and the production of a higher proportion of daughters all of which increase the impact of virus⁴⁹. In addition, *Rickettsia* and *Hamiltonella* play a role in plant virus transmission in whiteflies²⁵ by protecting the safe transit of virions in the haemolymph of insects through chaperonins (*GroEL*) and protein complexes that aid in protein folding and repair mechanisms²⁰. However, *Hamiltonella* was reported to be absent in the indigenous whitefly populations

studied elsewhere^{15,50,52–54} and in Malawi, Nigeria, Tanzania and Uganda^{50,55} as also confirmed in our study. *Arsenophonus*, *Wolbachia*, *Arsenophonus* and *Cardinium spp* have been detected within MED and MEAM1 *Bemisia* species^{14,50}. In addition, 50 and 22 reported *Arsenophonus* within SSA1 *B. tabaci* in Eastern Africa that were collected on cassava. These endosymbionts have been associated with several deleterious functions within *B. tabaci* that include manipulating female-male host ratio through feminizing genetic males, coupled with male killing^{56,57}.

Within the context of SSA agricultural systems, the role of endosymbionts in influencing *B. tabaci* viral transmission is important. Losses attributed to *B. tabaci* transmitted viruses within different crops are estimated to be in billions of US dollars⁴⁶. Bacterial endosymbionts have been associated with influencing viral acquisition, transmission and retention, such as in tomato leaf curl virus^{58,59}. Thus, better understanding of the diversity of the endosymbionts provides additional evidence on which members of *B. tabaci* species complex more proficiently transmit viruses, and thus the need for concerted efforts towards the whitefly eradication.

Conclusions

Our study provides a proof of concept that single whitefly RNA extraction and RNA sequencing is possible and the method could be optimised and applicable to a range of small insect transcriptome studies. It is particularly useful in studies that wish to explore vector-microbiota-viral dynamics at individual insect level rather than pooling of insects. It is useful where genetic material is both limited, as well as of low quality, which is applicable to most agriculture field collections. In addition, the single whitefly RNA sequencing technique described in this study offers new opportunities to understand the biology, and relative economic importance of the several whitefly species occurring in ecosystems within which food is produced in Sub-Saharan Africa, and will enable the efficient development and deployment of sustainable pest and disease management strategies

to ensure food security in the developing countries. However, this method still requires further optimisation to recover viral reads, especially in cases with very low viral titre as observed in this study. Finally future studies could use freshly collected whiteflies on CBSD-affected plants to increase the detection of the causal viruses.

Data availability

The datasets used and/or analyzed during the current study are available from GenBank:

SRR5110306, SRR5110307, SRR5109958, KY548924, MG680297.

Competing interests

No competing interests were disclosed.

Grant information

This work was supported by Mikocheni Agricultural Research Institute (MARI), Tanzania through the “Disease Diagnostics for Sustainable Cassava Productivity in Africa” project, Grant no. OPP1052391 that is jointly funded by the Bill and Melinda Gates Foundation and The Department for International Development (DFID) of the United Kingdom (UK). The Pawsey Supercomputing Centre provided computational resources with funding from the Australian Government and the Government of Western Australia supported this work. J.M.W is supported by an Australian Award scholarship by the Department of Foreign Affairs and Trade (DFAT).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

The authors would like to thank Cassava Disease Diagnostics Project members.

Supplementary material

Supplementary Table 1. Distribution of the endosymbionts and number of contigs matching core genes present within each endosymbiont bacteria present in four SSA1 *B. tabaci* samples from this study.

[Click here to access the data.](#)

References

- De Barro PJ, Liu SS, Boykin LM, et al.: **Bemisia tabaci: a statement of species status.** *Annu Rev Entomol.* 2011; **56**: 1–19.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Polston JE, Capobianco H: **Transmitting plant viruses using whiteflies.** *J Vis Exp.* 2013; **8**(81): e4332.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jones DR: **Plant viruses transmitted by whiteflies.** *Eur J Plant Pathol.* 2003; **109**: 195–219.
[Publisher Full Text](#)
- Vassilyev DG, Vassilyeva MN, Perederina A, et al.: **Structural basis for transcription elongation by bacterial RNA polymerase.** *Nature.* 2007; **448**(7150): 157–162.
[PubMed Abstract](#) | [Publisher Full Text](#)

5. Vassiliou V, Emmanouilidou M, Perrakis A, *et al.*: Insecticide resistance in *Bemisia tabaci* from Cyprus. *Insect Sci.* 2011; **18**(1): 30–39. [Publisher Full Text](#)
6. FAO: Save and Grow: Cassava. A Guide to Sustainable Production Intensification. 2013. [Reference Source](#)
7. Fauquet CM, Briddon RW, Brown JK, *et al.*: Geminivirus strain demarcation and nomenclature. *Arch Virol.* 2008; **153**(4): 738–821. [PubMed Abstract](#) | [Publisher Full Text](#)
8. Patil BL, Fauquet CM: Cassava mosaic geminiviruses: actual knowledge and perspectives. *Mol Plant Pathol.* 2009; **10**: 685–701. [PubMed Abstract](#) | [Publisher Full Text](#)
9. Legg JP, Owor B, Sseruwagi P, *et al.*: Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Adv Virus Res.* 2006; **67**(6): 355–418. [PubMed Abstract](#) | [Publisher Full Text](#)
10. Maruthi MN, Hillocks RJ, Mtunda K, *et al.*: Transmission of Cassava brown streak virus by *Bemisia tabaci* (Gennadius). *J Phytopathol.* 2005; **153**(5): 307–312. [Publisher Full Text](#)
11. Mware B, Narla R, Amata R, *et al.*: Efficiency of cassava brown streak virus transmission by two whitefly species in coastal Kenya. *J Gen Mol Virol.* 2009; **1**(4): 40–45. [Reference Source](#)
12. Ndunguru J, Sseruwagi P, Tairo F, *et al.*: Analyses of Twelve New Whole Genome Sequences of Cassava Brown Streak Viruses and Ugandan Cassava Brown Streak Viruses from East Africa: Diversity, Supercomputing and Evidence for Further Speciation. *PLoS One.* 2015; **10**(10): 1–18. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Alicai T, Ndunguru J, Sseruwagi P, *et al.*: Cassava brown streak virus has a rapidly evolving genome: implications for virus speciation, variability, diagnosis and host resistance. *Sci Rep.* 2016; **6**: 36164. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Gueguen G, Vavre F, Gnankine O, *et al.*: Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Mol Ecol.* 2010; **19**(19): 4365–4378. [PubMed Abstract](#) | [Publisher Full Text](#)
15. Marubayashi JM, Kliot A, Yuki VA, *et al.*: Diversity and Localization of Bacterial Endosymbionts from Whitefly Species Collected in Brazil. *PLoS One.* 2014; **9**(9): e108363. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. Thao ML, Baumann P: Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Appl Environ Microbiol.* 2004; **70**(6): 3401–3406. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Mooney RA, Schweimer K, Rösch P, *et al.*: Two structurally independent domains of *E. coli* NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. *J Mol Biol.* 2009; **391**(2): 341–358. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Yakhnin AV, Murakami KS, Babbitzke P: NusG Is a Sequence-specific RNA Polymerase Pause Factor That Binds to the Non-template DNA within the Paused Transcription Bubble. *J Biol Chem.* 2016; **291**(10): 5299–5308. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Zchori-Fein E, Brown JK: Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). *Ann Entomol Soc Am.* 2002; **95**(6): 711–718. [Publisher Full Text](#)
20. Gottlieb Y, Ghanim M, Chiel E, *et al.*: Identification and localization of a *Rickettsia* sp. in *Bemisia tabaci* (Homoptera: Aleyrodidae). *Appl Environ Microbiol.* 2006; **72**(5): 3646–3652. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Marubayashi JM, Yuki VA, Rocha KC, *et al.*: At least two indigenous species of the *Bemisia tabaci* complex are present in Brazil. *J Appl Entomol.* 2013; **137**(1–2): 113–121. [Publisher Full Text](#)
22. Ghosh S, Mitra PS, Loffredo CA, *et al.*: Transcriptional profiling and biological pathway analysis of human equivalence PCB exposure *in vitro*: Indicator of disease and disorder development in humans. *Environ Res.* 2015; **138**: 202–216. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Rosario K, Marr C, Varsani A, *et al.*: Begomovirus-Associated Satellite DNA Diversity Captured Through Vector-Enabled Metagenomic (VEM) Surveys Using Whiteflies (Aleyrodidae). *Viruses.* 2016; **8**(2): pii: E36. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Rosario K, Seah YM, Marr C, *et al.*: Vector-Enabled Metagenomic (VEM) Surveys Using Whiteflies (Aleyrodidae) Reveal Novel Begomovirus Species in the New and Old Worlds. *Viruses.* 2015; **7**(10): 5553–5570. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Kliot A, Cilia M, Czosnek H, *et al.*: Implication of the bacterial endosymbiont *Rickettsia* spp. in interactions of the whitefly *Bemisia tabaci* with tomato yellow leaf curl virus. *J Virol.* 2014; **88**(10): 5652–5660. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Poelchau MF, Coates BS, Childers CP, *et al.*: Agricultural applications of insect ecological genomics. *Curr Opin Insect Sci.* 2016; **13**: 61–69. [PubMed Abstract](#) | [Publisher Full Text](#)
27. Kapantaidaki DE, Ovčarenko I, Fytrou N, *et al.*: Low levels of mitochondrial DNA and symbiont diversity in the worldwide agricultural pest, the greenhouse whitefly *Trialeurodes vaporariorum* (Hemiptera: Aleyrodidae). *J Hered.* 2015; **106**(1): 80–92. [PubMed Abstract](#) | [Publisher Full Text](#)
28. Morin S, Ghanim M, Sobol I, *et al.*: The GroEL protein of the whitefly *Bemisia tabaci* interacts with the coat protein of transmissible and nontransmissible begomoviruses in the yeast two-hybrid system. *Virology.* 2000; **276**(2): 404–416. [PubMed Abstract](#) | [Publisher Full Text](#)
29. Xue J, Zhou X, Zhang CX, *et al.*: Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. *Genome Biol.* 2014; **15**(12): 521. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Kim JK, Won YJ, Nikoh N, *et al.*: Polyester synthesis genes associated with stress resistance are involved in an insect-bacterium symbiosis. *Proc Natl Acad Sci U S A.* 2013; **110**(26): E2381–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Wang XW, Zhao QY, Luan JB, *et al.*: Analysis of a native whitefly transcriptome and its sequence divergence with two invasive whitefly species. *BMC Genomics.* 2012; **13**(1): 529. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Kono N, Nakamura H, Ito Y, *et al.*: Evaluation of the impact of RNA preservation methods of spiders for *de novo* transcriptome assembly. *Mol Ecol Resour.* 2016; **16**(3): 662–672. [PubMed Abstract](#) | [Publisher Full Text](#)
33. Xue X, Li SJ, Ahmed MZ, *et al.*: Inactivation of *Wolbachia* reveals its biological roles in whitefly host. *PLoS One.* 2012; **7**(10): e48148. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Charlesworth D, Willis JH: The genetics of inbreeding depression. *Nat Rev Genet.* 2009; **10**(11): 783–96. [PubMed Abstract](#) | [Publisher Full Text](#)
35. Grabherr MG, Haas BJ, Yassour M, *et al.*: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011; **29**(7): 644–52. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Huelsenbeck JP, Andolfatto P, Huelsenbeck ET: Structurama: bayesian inference of population structure. *Evol Bioinform Online.* 2011; **7**: 55–9. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. Darriba D, Taboada GL, Doallo R, *et al.*: jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012; **9**(8): 772. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
38. Katoh K, Misawa K, Kuma K, *et al.*: MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002; **30**(14): 3059–3066. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Kelley LA, Mezulis S, Yates CM, *et al.*: The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015; **10**(6): 845–858. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Boykin LM, Shatters RG Jr, Rosell RC, *et al.*: Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol Phylogenet Evol.* 2007; **44**(3): 1306–1319. [PubMed Abstract](#) | [Publisher Full Text](#)
41. Hsieh CH, Ko CC, Chung CH, *et al.*: Multilocus approach to clarify species status and the divergence history of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. *Mol Phylogenet Evol.* 2014; **76**: 172–180. [PubMed Abstract](#) | [Publisher Full Text](#)
42. Xie W, Meng QS, Wu QJ, *et al.*: Pyrosequencing the *Bemisia tabaci* transcriptome reveals a highly diverse bacterial community and a robust system for insecticide resistance. *PLoS One.* 2012; **7**(4): e35181. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. Kajitani R, Toshimoto K, Noguchi H, *et al.*: Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 2014; **24**(8): 1384–1395. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Leshkowitz D, Gazit S, Reuveni E, *et al.*: Whitefly (*Bemisia tabaci*) genome project: analysis of sequenced clones from egg, instar, and adult (viruliferous and non-viruliferous) cDNA libraries. *BMC Genomics.* 2006; **7**: 79. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Gallego Romero I, Pai AA, Tung J, *et al.*: RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol.* 2014; **12**(1): 42. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
46. Legg JP, Sseruwagi P, Boniface S, *et al.*: Spatio-temporal patterns of genetic change amongst populations of cassava *Bemisia tabaci* whiteflies driving virus pandemics in East and Central Africa. *Virus Res.* 2014; **186**: 61–75. [PubMed Abstract](#) | [Publisher Full Text](#)
47. Wainaina JM, De Barro P, Kubatko L, *et al.*: Global phylogenetic relationships, population structure and gene flow estimation of *Trialeurodes vaporariorum* (Greenhouse whitefly). *Bull Entomol Res.* 2018; **108**(1): 5–13. [PubMed Abstract](#) | [Publisher Full Text](#)
48. Tajebe LS, Boni SB, Guastella D, *et al.*: Abundance, diversity and geographic distribution of cassava mosaic disease pandemic-associated *Bemisia tabaci* in Tanzania. *J Appl Entomol.* 2015; **139**(8): 627–637. [Publisher Full Text](#)

49. Rao Q, Luo C, Zhang H, *et al.*: **Distribution and dynamics of *Bemisia tabaci* invasive biotypes in central China.** *Bull Entomol Res.* 2011; **101**(1): 81–88.
[PubMed Abstract](#) | [Publisher Full Text](#)
50. Tajebe LS, Guastella D, Cavalieri V, *et al.*: **Diversity of symbiotic bacteria associated with *Bemisia tabaci* (Homoptera: Aleyrodidae) in cassava mosaic disease pandemic areas of Tanzania.** *Ann Appl Biol.* 2015; **166**(2): 297–310.
[Publisher Full Text](#)
51. Rao Q, Rollat-Farnier PA, Zhu DT, *et al.*: **Genome reduction and potential metabolic complementation of the dual endosymbionts in the whitefly *Bemisia tabaci*.** *BMC Genomics.* 2015; **16**(1): 226.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
52. Bing XL, Yang J, Zchori-fein E: **Characterization of a newly discovered symbiont of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae).** *Appl Environ Microbiol.* 2013; **79**(2): 569–75.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
53. Singh ST, Priya NG, Kumar J: **Diversity and phylogenetic analysis of endosymbiotic bacteria from field caught *Bemisia tabaci* from different locations of North India based on 16S rDNA library screening.** *Infect Genet Evol.* 2012; **12**(2): 411–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
54. Gnankine O, Mouton L, Henri H, *et al.*: **Distribution of *Bemisia tabaci* (Homoptera: Aleyrodidae) biotypes and their associated symbiotic bacteria on host plants in West Africa.** *Insect Conserv Divers.* 2013; **6**(3): 411–421.
[Publisher Full Text](#)
55. Ghosh S, Bouvaine S, Maruthi MN: **Prevalence and genetic diversity of endosymbiotic bacteria infecting cassava whiteflies in Africa.** *BMC Microbiol.* 2015; **15**(1): 93.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
56. Brumin M, Kontsedalov S, Ghanim M: ***Rickettsia* influences thermotolerance in the whitefly *Bemisia tabaci* B biotype.** *Insect Sci.* 2011; **18**(1): 57–66.
[Publisher Full Text](#)
57. Brumin M, Levy M, Ghanim M: **Transovarial transmission of *Rickettsia* spp. and organ-specific infection of the whitefly *Bemisia tabaci*.** *Appl Environ Microbiol.* 2012; **78**(16): 5565–5574.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
58. Rosario K, Capobianco H, Ng TF, *et al.*: **RNA viral metagenome of whiteflies leads to the discovery and characterization of a whitefly-transmitted carlavirus in North America.** *PLoS One.* 2014; **9**(1): e86748.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
59. Himler AG, Adachi-Hagimori T, Bergen JE: **Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias.** *Science.* 2011; **332**(6026): 254–6.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Referee Status:  

Version 3

Referee Report 15 March 2018

doi:[10.21956/gatesopenres.13874.r26304](https://doi.org/10.21956/gatesopenres.13874.r26304)



Henryk H. Czosnek

The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, The Hebrew University of Jerusalem, Jerusalem, Israel

I am satisfied with the latest version, especially with the amended title, which reflects much better the content than before.

I think this is a very good paper (it was also before – notwithstanding my comments), summarizing a very good piece of work and a true advance on the way to better understand the biology of whiteflies.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Referee Report 26 February 2018

doi:[10.21956/gatesopenres.13862.r26243](https://doi.org/10.21956/gatesopenres.13862.r26243)



Kai-Shu Ling

USDA-ARS (United States Department of Agriculture - Agricultural Research Service), Charleston, SC, USA

The presentation in this revised version has been improved. The technology in generating relatively high numbers of sequence reads through RNA sequencing from field-collected individual whiteflies is quite impressive, which may have a broader implication for other researchers who are working in the similar situation with limited amount of material from a small insect or other organism for transcriptome analysis.

As a proof of concept, authors focused on their sequence analysis primarily on sequences derived from endosymbionts. Besides the mtCOI gene, typical transcriptome analysis, including transcript expression and gene ontology relating to the whitefly genome, was not conducted in the current study. Therefore, although authors prefer to maintain the original title, I am still have a reservation on the title as it does not reflect well the main content of the current presentation.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 06 Mar 2018

Laura Boykin,

Thank you very much for your review. We have changed the title as you have suggested.

Competing Interests: No competing interests were disclosed.

Referee Report 21 February 2018

doi:[10.21956/gatesopenres.13862.r26244](https://doi.org/10.21956/gatesopenres.13862.r26244)



Henryk H. Czosnek

The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, The Hebrew University of Jerusalem, Jerusalem, Israel

I am still very impressed by the technical prowess consisting in developing a method that allows the analysis of the transcriptome of individual whiteflies. The manuscript has been improved by modifying some of the tables.

Again, and the authors have not followed me in this, I believe they should go beyond the analysis of endosymbiont transcripts; they have all the data to do so. This way, they could present the advantages and the limits of single insect transcriptome analyses in deciphering the many aspects of whitefly biology. Limiting the analysis to symbionts in this paper is a waste of analytical power offered by the technology. In addition, I stated in my review that the title is deceiving, since the focus is on the endosymbiont-related transcriptome. Again, the authors preferred to retain their title. I could suggest the following: "The first transcriptomes from field-collected individual whiteflies allow to identify specific endosymbiont gene expression". I believe this is critical for the "business card" of the paper.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 06 Mar 2018

Laura Boykin,

Thank you for your positive comments on our manuscript. We have changed the title of the paper to "**The first transcriptomes from field-collected individual whiteflies (*Bemisia tabaci*, Hemiptera: Aleyrodidae): a case study of the endosymbiont composition.**". All data is

available for other whitefly researchers to do the analyses suggested above. The main focus of our paper was to publish the laboratory method. We are pleased the second reviewer agrees.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 22 January 2018

doi:[10.21956/gatesopenres.13843.r26202](https://doi.org/10.21956/gatesopenres.13843.r26202)



Henry H. Czosnek

The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, The Hebrew University of Jerusalem, Jerusalem, Israel

In this paper, Sseruwagi *et al.* present a method of RNA preparation, which is suitable for the Illumina-based RNASeq analysis of the transcriptome of a single whitefly.

Using this method, the authors have sequenced the transcriptome of four individuals (one from Uganda and three from Tanzania) collected on cassava leaves with symptoms said to be produced by *Cassava brown streak virus* (CBSV), a (+)ssRNA virus belonging to the *Potyviridae* family, probably transmitted by *B. tabaci* (still questioned in the literature).

Indeed, this is an important technical feat. No doubt, it may help follow the movement of whiteflies and the diseases they transmit. In this context, it is interesting to note that the sequence and the functional analysis of the transcriptome of single cells has been published in several instances (recently reviewed by Liu and Trapnell, 2016).

This paper is a bit disappointing, to me. The title claims that the authors have produced the first transcriptome of a single whitefly. This is true on its face value; and the results have been posted in GenBank (although, raw data). It should be relatively easy to identify transcripts since the sequence of *B. tabaci* MEAM1 and MED are known. Nonetheless, this reviewer is expecting some valuable information on gene expression in the whole animal. Is the lack of data in the paper due to a low number of reads of transcripts of cellular genes? Could the author identify, say, transcripts of housekeeping genes or genes involved in sugar metabolism or else? This would underline the power and the limits of the one-insect-one-transcriptome analysis.

Instead, the authors have chosen to focus on the whitefly primary (P) and secondary (S) endosymbionts, especially on the NusG gene of the primary endosymbiont *Portiera aleyrodidarum* (4 figures). This gene might be interesting but it is more a structural than a functional study, which to my point of view lessens the importance of the study.

The supplementary Table 1 is interesting but does not tell us the endosymbiont composition of the four individuals scrutinized. Is it *Portiera* (P), and *Arsephonus*, *Wolbachia* and *Rickettsia*? Also, the title of Table 1 is not clear; what is the meaning of “number of genes in endosymbionts bacteria”? Is it the number of genes with homologies to others?

It is interesting that Hamiltonella sequences have not been found, knowing that this is the symbiont that produces GroEL, which binds to the CP of begomoviruses, and facilitates the transit of the virus in the hemolymph. It is also interesting that CBSV sequences were not found, although the whiteflies have been collected on symptomatic plants. Is it that, after all, *B. tabaci* is not the vector of this virus?

Altogether, I expected much more from the title. I suggest to lower expectations of the reader by amending the title to something like "Analysis of the endosymbiont transcriptome from individual whiteflies".

I recommend publication after relating to the points mentioned above.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Referee Expertise: Molecular biology of virus-plant-insect interactions

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Feb 2018

Laura Boykin,

Your comments have greatly improved our manuscript and an updated version is now available for review. Thank you.

Reviewer 2

Comments

In this paper, Sseruwagi *et al.* present a method of RNA preparation, which is suitable for the Illumina-based RNASeq analysis of the transcriptome of a single whitefly.

Using this method, the authors have sequenced the transcriptome of four individuals (one from

Uganda and three from Tanzania) collected on cassava leaves with symptoms said to be produced by *Cassava brown streak virus* (CBSV), a (+)ssRNA virus belonging to the *Potyviridae* family, probably transmitted by *B. tabaci* (still questioned in the literature).

Indeed, this is an important technical feat. No doubt, it may help follow the movement of whiteflies and the diseases they transmit. In this context, it is interesting to note that the sequence and the functional analysis of the transcriptome of single cells has been published in several instances (recently reviewed by Liu and Trapnell, 2016).

This paper is a bit disappointing, to me. The title claims that the authors have produced the first transcriptome of a single whitefly. This is true on its face value; and the results have been posted in GenBank (although, raw data). It should be relatively easy to identify transcripts since the sequence of *B. tabaci* MEAM1 and MED are known. Nonetheless, this reviewer is expecting some valuable information on gene expression in the whole animal. Is the lack of data in the paper due to a low number of reads of transcripts of cellular genes? Could the author identify, say, transcripts of housekeeping genes or genes involved in sugar metabolism or else? This would underline the power and the limits of the one-insect-one-transcriptome analysis.

Response:

The initial experiment was meant to determine if we could obtain sufficient RNA and conduct RNAseq analysis on individual field collected B. tabaci. Our primary aim was to unravel the microbiota within individual transcriptome. Conducting gene expression analysis is still a challenge with the current method, mainly due to variation in starting RNA concentration of the whiteflies. Secondly, we did not achieve adequate ribosomal depletion, which may have hindered successful gene expression analysis. However, based on this method (ongoing) we have indeed identified nuclear genes and single copy orthologs.

Comments

Instead, the authors have chosen to focus on the whitefly primary (P) and secondary (S) endosymbionts, especially on the NusG gene of the primary endosymbiont *Portiera aleyrodidarum* (4 figures). This gene might be interesting but it is more a structural than a functional study, which to my point of view lessens the importance of the study.

Response:

We focused on the NusG mainly due to the unique deletion observed on what should be highly conserved proteins that are reported to be crucial in bacterial replication. It highlights the unique features of the endosymbionts from SSA species of B. tabaci compared to other putative species of B. tabaci and further highlights the difference within this species within the species complex. These findings were possible because we studied individual whitefly transcriptomes, and may probably not have been discovered by transcriptomes generated from pooled isolines.

Comments

The supplementary Table 1 is interesting but does not tell us the endosymbiont composition of the four individuals scrutinized. Is it *Portiera* (P), and *Arsephonus*, *Wolbachia* and *Rickettsia*? Also, the title of Table 1 is not clear; what is the meaning of “number of genes in endosymbionts bacteria”? Is it the number of genes with homologies to others?

Response:

We have revised and clarified the legends and content of both Tables 1 and 2 and supplementary Table1.

Comments

It is interesting that Hamiltonella sequences have not been found, knowing that this is the symbiont that produces GroEL, which binds to the CP of begomoviruses, and facilitates the transit of the virus in the hemolymph.

Response:

Candidatus Hamiltonella defensa has been reported to be absent in whiteflies in Africa. Our study found very negligible numbers of contigs in only one of the whiteflies (WF2a) studied. However, the literature (lines 117 to 120 in this paper) indicates that *Rickettsia* spp. is also involved in virus transmission, and is among the predominant endosymbionts detected in our study. We have added the results of *Hamiltonella* and *Rickettsia* to clarify the reviewer's concerns.

Comment

It is also interesting that CBSV sequences were not found, although the whiteflies have been collected on symptomatic plants. Is it that, after all, *B. tabaci* is not the vector of this virus?

Response:

RNA viruses such as CBSVs are picked up and kept for short periods in the whitefly stylet, unlike the DNA viruses that build-up and keep long in the midgut, and are more likely to be detected if present in the whitefly under study. It is also possible that the whitefly were not viruliferous considering that less than 10% of field collected whiteflies are viruliferous despite them feeding on infected.

Secondly, a recent publication (Ateka E, Alicai T, Ndunguru J, Tairo F, Sseruwagi P, Kiarie, S., et al. (2017) Unusual occurrence of a DAG motif in the Ipomovirus Cassava brown streak virus and implications for its vector transmission. PLoS ONE 12(11): e0187883 reported the presence of a DAG motif within CBSVs indicating they could be aphid-transmitted viruses rather than by whiteflies.

Comment

Altogether, I expected much more from the title. I suggest to lower expectations of the reader by amending the title to something like "Analysis of the endosymbiont transcriptome from individual whiteflies".

Response:

We appreciate the reviewers comment regarding the title but we prefer the current title as our study is the first transcriptome generated from field collected whiteflies- the analyses pipelines can be investigated and expanded upon with future studies.

Competing Interests: n/a

Referee Report 16 January 2018

doi:10.21956/gatesopenres.13843.r26170



Kai-Shu Ling

USDA-ARS (United States Department of Agriculture - Agricultural Research Service), Charleston, SC, USA

Sseruwagi and colleagues in this manuscript described a method to effectively generate a high throughput RNA-seq dataset using purified total RNA extracted from each individual field-collected adult whitefly, *Bemisia tabaci*, which generated 39-42 million raw reads per library using Illumina sequencing.

Because the genome sequence of cassava whitefly *B. tabaci* SSA-1 is yet available, through *de novo* assembly of cleaned reads, high number of contigs (65,000-162,000) from each library were generated. Functional prediction to profile the generated transcripts of *B. tabaci* SSA1 were not performed. However, sequences to the mitochondrion cytochrome I oxidase (mtCOI) gene were identified from each of the four RNA-seq libraries. Phylogenetic analysis of mtCOI confirmed its close relationship to the cassava whitefly *B. tabaci* SSA1 clade. In addition, these RNA-seq datasets also contained sequences relating to five endosymbiont bacteria. Although authors claimed to have transcriptomes for these endosymbionts, extensive analysis to functionally profile the identified RNA sequences of these endosymbionts was not conducted in the current study. Individual analysis through amino acid alignment of the identified Nus G gene sequences in the primary *Portiera aleydidarum* from four RNA-seq datasets revealed an eleven amino acid residue deletion in two of the four individual whitefly libraries. Although this finding is interesting, a validation test would be necessary to confirm the missing sequences in those individuals through Sanger sequencing of amplicons generated using Nus G specific primers on the original RNA preparations. It is also surprising that not a single sequence relating to cassava-infecting viruses although these whiteflies were supposedly collected from cassava plants infected with cassava brown streak virus which has a poly-A tail in its RNA genome. It would be an interest to test the original RNA preparations to determine which viruses may be in these individual whiteflies.

Specific comments and suggestions:

Title: As mentioned in the general comments, this is more like a method paper in doing RNA sequencing on little RNA extracted from individual whitefly, not an extensive transcriptome analysis. I would suggest to change the title to something like this: Effective RNA sequencing using little RNA extracted from field-collected individual whiteflies (*Bemisia tabaci*) useful for transcriptome analysis.

IN ABSTRACT:

- Page 1: change 65,000-162,000 transcripts to contigs.
- Page 1: the compound sentence starting with “BLASTn searches ...” This compound sentence is too long and the meaning is not clear, need to rewrite.

IN INTRODUCTION:

- Page 3: need to modify the sentence ending with “...is hampered by both DNA and RNA (transmitted) virus”, either deleting “transmitted” or change the sentence to “is hampered by whitefly-transmitted DNA and RNA viruses”.
- Page 3: The sentence starting as “Relevant to this study are two RNA Potyviruses: ...” Potyviruses should be replaced with “ipomoviruses, in the family Potyviridae”: ...
- Page 3: in the same paragraph as above, you may want to elaborate a little bit more on virus species and genetic diversity, such as which viruses have been proven to be transmitted and which have not be proven to be transmitted by the SSA1 whiteflies and their efficiency.
- Page 3: Should be consistent in using P-endosymbiont, rather than P-symbiont and in other locations in the same document using P-endosymbiont.
- Page 3: You stated “Seven facultative S-endosymbionts...,” but only six endosymbionts are actually listed, need to add the missing one.
- Page 3: In the sentence starting with “It is against this background that we...” Your goal was to study whitefly-virus interaction, it is rather surprising that not a single virus sequence read was detected in these RNA-seq datasets in field-collected whiteflies. Rather than speculating these individual whitefly did not carry the target viruses, why not doing some tests by RT-PCR to confirm the lack of target viruses in these RNA preparations?

IN METHODS:

- Page 4: Since whiteflies were collected from CBSD-symptomatic cassava plants in Uganda and Tanzania, it might still be possible to conduct RT-PCR tests to determine the presence of viruses in the purified whitefly RNA preparations.
- Page 4: The rationale in using the Brazilian whitefly sample? Also it appears to me that there were no analysis of sequences from this Brazilian dataset in the result section.

IN RESULTS:

- Page 5: There were 65,550 to 162,487 contigs generated from individual RNA-seq datasets. How all these sequences could be assigned to? Any ideas on what proportion of the sequences belonging to whitefly *B. tabaci* genome, what proportion of the sequences to endosymbionts?
- Page 5: The citation to Table 2 seems to point to the listing of five endosymbionts, however the content in the Table 2 showing the origin of whiteflies collected.
- Page 5: What are these incidences (74.8%, 71.2%, 54.1% and 58.5%) mean? The meaning is not clear, a proportion of total endosymbiont sequences that assigned to *P. aleyrodidarum*?
- Page 5: The (data not shown) is not a good idea, it would be good to present it as a supplementary file.

IN DISCUSSION:

- Page 8: As mentioned in the general comment, you developed a method to effectively conduct RNA-sequencing on little RNA from individual whitefly and sequence analysis on specific genes in endosymbionts and COI gene of whitefly, I wouldn't over use the term transcriptome analysis.
- Page 8: The deletion of 11 amino acids on the *Nus G* gene, this could be resulted from sequence assembly from short reads, therefore a validation of the deletion sequence on these whiteflies by RT-PCR and Sanger sequencing would be necessary, as pointed out by the authors that field collected samples may have resulted in RNA degradation in RNA purification, cDNA library preparation and/or sequencing.
- Page 9: The sentence starts: "In addition, 51 and 21 reported...", what do these numbers (51 and 21) mean here? If there are citations to references, you need to use author name followed by the reference number.
- Page 9: Tomato leaf curl virus should be changed to tomato yellow leaf curl virus.

IN CONCLUSION:

- Page 9: Suggest to change transcriptome sequencing to RNA sequencing.
- Page 9: In this study, you only worked with whiteflies, in the sentence you need to change "is" to "could be optimized and applicable to ..."
- Page 9: When you mentioned it is useful to study vector-microbiota-viral dynamics, but why not a single viral sequence read was detected in these datasets? Although there are some advantage in using RNA sequencing of individual whitefly, you may also want to point out there are still some room for improvement.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Referee Expertise: Virology, vector (whitefly) biology, epidemiology and genomics

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Feb 2018

Laura Boykin,

Thank you for your comments, they have greatly improved our manuscript. A new version is ready for review and our responses to each comment are listed below.

Reviewer 1

Specific comments and suggestions:

Title: As mentioned in the general comments, this is more like a method paper in doing RNA sequencing on little RNA extracted from individual whitefly, not an extensive transcriptome analysis. I would suggest changing the title to something like this: Effective RNA sequencing using little RNA extracted from field-collected individual whiteflies (*Bemisia tabaci*) useful for transcriptome analysis.

Response:

We appreciate the reviewers comment regarding the title but we prefer the current title as our study it is the first transcriptome generated from field-collected whiteflies- the analyses pipelines can be investigated and expanded upon with future studies.

IN ABSTRACT:

Page 1: change 65,000-162,000 transcripts to contigs.

Response:

Done

Page 1: the compound sentence starting with "BLASTn searches ..." This compound sentence is too long and the meaning is not clear, need to rewrite.

Response:

Done

For research articles we ask that you structure your abstract under subtitles of background, methods, results and conclusions. For more information please see section 3 of our research

article guidelines.

Response:

Done

IN INTRODUCTION:

Page 3: need to modify the sentence ending with "...is hampered by both DNA and RNA (transmitted) virus", either deleting "transmitted" or change the sentence to "is hampered by whitefly-transmitted DNA and RNA viruses".

Response:

Corrected to read" is hampered by whitefly-transmitted DNA and RNA viruses"

Page 3: The sentence starting as "Relevant to this study are two RNA Potyviruses: ..." Potyviruses should be replaced with "ipomoviruses, in the family Potyviridae": ...

Response:

Corrected as suggested

Page 3: in the same paragraph as above, you may want to elaborate a little bit more on virus species and genetic diversity, such as which viruses have been proven to be transmitted and which have not been proven to be transmitted by the SSA1 whiteflies and their efficiency.

Response:

Done as suggested. Current knowledge shows that all the cassava mosaic begomoviruses (CMBs) in Africa are transmitted by SSA1 and studies are on-going to study the efficiency of other SSA species.

Page 3: Should be consistent in using P-endosymbiont, rather than P-symbiont and in other locations in the same document using P-endosymbiont.

Response:

P-endosymbiont has been added throughout the text.

Page 3: You stated "Seven facultative S-endosymbionts," but only six endosymbionts are actually listed, need to add the missing one.

Response:

Orientia-like organism added as the seventh endosymbionts

Page 3: In the sentence starting with "It is against this background that we..." Your goal was to study whitefly-virus interaction, it is rather surprising that not a single virus sequence read was detected in these RNA-seq datasets in field-collected whiteflies. Rather than speculating these individual whitefly did not carry the target viruses, why not doing some tests by RT-PCR to confirm the lack of target viruses in these RNA preparations?

Response:

Due to very little RNA (~17 uL) all was used in the cDNA library preparation; thus no further experiments could be done after library preparation. Additionally, the fact that we were analysing a single whitefly, the very low viral titre may not have been detectable. It is also known that RNA viruses such as CBSVs are picked up and kept for short periods in the whitefly stylet, unlike the DNA viruses that build-up and keep long in the midgut, and are more likely to be detected if present in the whitefly under study. Studies have shown that less than 10% of field whiteflies are viruliferous in any sample, and therefore it is possible that we missed the infected individuals during sampling. The suggestion would be to use freshly-collected samples in the future.

IN METHODS:

Page 4: Since whiteflies were collected from CBSD-symptomatic cassava plants in Uganda and Tanzania, it might still be possible to conduct RT-PCR tests to determine the presence of viruses in the purified whitefly RNA preparations.

Response:

It may not be possible to detect CBSVs even with RT-PCR because of the reasons provided above.

Page 4: The rationale in using the Brazilian whitefly sample? Also it appears to me that there were no analysis of sequences from this Brazilian dataset in the result section.

Response:

We have added further information regarding the Brazilian whitefly to the end of the first paragraph in the methods section.

IN RESULTS:

Page 5: There were 65,550 to 162,487 contigs generated from individual RNA-seq datasets. How all these sequences could be assigned to? Any ideas on what proportion of the sequences belonging to whitefly *B. tabaci* genome, what proportion of the sequences to endosymbionts?

Response:

We have added an additional line to Table 1 to show the number of endosymbiont contigs- thank you for this suggestion.

Page 5: The citation to Table 2 seems to point to the listing of five endosymbionts, however the content in the Table 2 showing the origin of whiteflies collected.

Response:

Corrected Table 2 with the prevalence of the endosymbionts included.

Page 5: What are these incidences (74.8%, 71.2%, 54.1% and 58.5%) mean? The meaning is not clear, a proportion of total endosymbiont sequences that assigned to *P. aleyrodidarum*?

Response:

For clarity, the absolute counts of the contigs associated with each endosymbiont have been provided rather than percentages and are linked directly to Table 2.

Page 5: The (data not shown) is not a good idea, it would be good to present it as a supplementary file.

Response:

The mtCOI sequences were extracted from the transcriptomic data and then whiteflybase.org was used to verify the species ID.

IN DISCUSSION:

Page 8: As mentioned in the general comment, you developed a method to effectively conduct RNA-sequencing on little RNA from individual whitefly and sequence analysis on specific genes in endosymbionts and COI gene of whitefly, I wouldn't over use the term transcriptome analysis.

Response:

RNA sequencing has been used instead of transcriptome sequencing where possible.

Page 8: The deletion of 11 amino acids on the *Nus G* gene, this could be resulted from sequence

assembly from short reads, therefore a validation of the deletion sequence on these whiteflies by RT-PCR and Sanger sequencing would be necessary, as pointed out by the authors that field collected samples may have resulted in RNA degradation in RNA purification, cDNA library preparation and/or sequencing.

Response:

Previous studies and Genbank records (CP004358, LN649236, LN649255, LN734649) – from independent studies- have proven this is a legitimate deletion. Therefore, confirmation via further experiments is not needed in our opinion.

Page 9: The sentence starts: “In addition, 51 and 21 reported...”, what do these numbers (51 and 21) mean here? If there are citations to references, you need to use author name followed by the reference number.

Response:

Rephrased to read “In addition Arsenophonus has been reported within SSA1 B. tabaci in Eastern Africa collected from cassava [20, 50].”

Page 9: Tomato leaf curl virus should be changed to tomato yellow leaf curl virus.

Response:

Corrected as suggested.

IN CONCLUSION:

Page 9: Suggest changing transcriptome sequencing to RNA sequencing.

Response:

Done

Page 9: In this study, you only worked with whiteflies, in the sentence you need to change “is” to “could be optimized and applicable to ...”

Response:

Rephrased as requested.

Page 9: When you mentioned it is useful to study vector-microbiota-viral dynamics, but why not a single viral sequence read was detected in these datasets? Although there are some advantage in using RNA sequencing of individual whitefly, you may also want to point out there are still some room for improvement.

Response:

A sentence has been added on future improvements “However this method still requires further optimisation to recover viral reads especially in cases with very low viral titre as observed in this study”.

Competing Interests: n/a