Australian Government
Cotton Research and Development Corporation

FINAL REPORT 2017

Summary Details

Please use your TAB key to complete Parts 1 & 2.

CRDC Project Number: **UQ1305**

Project Title: Viruses, vectors and endosymbionts: Exploring interactions for control

Project Commencement Date: 23/09/2013 **Project Completion Date:** 31/08/2017

CRDC Research Program: 2 Industry

2.3.2 Supporting the industry's preparedness to deal with biosecurity threats
2.3.3 Researching the management of established, invasive and endemic insect pests, weeds and diseases
2.3.4 Assuring industry capacity to manage the stewardship of biotechnologies and crop protection products

Contact Details

Administrator: Mrs Carol Pomfret
Organisation: The University of Queensland
Postal Address: Room 311 Gehrmann Building, St Lucia, QLD 4072
Ph: 07 33654856 **Fax:** N/A **E-mail:** c.pomfret@uq.edu.au

Principal Researcher: Dr Sharon van Brunschot (23/09/2013 – 14/10/2016)
Dr Daisy Stainton (17/10/2016 – 31/08/2017)
Organisation: The University of Queensland
Postal Address: Goddard Building (8), The University of Queensland, St Lucia, Brisbane, QLD 4072
Ph: 07 33652769 **Fax:** N/A **E-mail:** s.vanbrunschot@uq.edu.au
d.stainton@uq.edu.au

Supervisor: Professor Gimme Walter
Organisation: The University of Queensland
Postal Address: Goddard Building (8), The University of Queensland, St Lucia, Brisbane, QLD 4072
Ph: 07 33652228 **Fax:** N/A **E-mail:** g.walter@uq.edu.au

Researcher 2: Dr James Hereward
Organisation: The University of Queensland
Postal Address: Goddard Building (8), The University of Queensland, St Lucia, Brisbane, QLD 4072
Ph: 07 33652755 **Fax:** N/A **E-mail:** j.hereward@uq.edu.au

Signature of Research Provider Representative: _____

Date Submitted: _____

Final Report

(The points below are to be used as a guideline when completing your final report.)

1. Background

1.1 Outline the background to the project.

Cotton leaf curl disease (CLCuD) presents a major biosecurity threat to the Australian cotton industry because it can decimate production. The disease is caused by a complex of one or more begomoviruses, and is spread by the whitefly (*Bemisia tabaci*). Devastating epidemics of CLCuD have been reported from a number of countries across Africa and southern Asia – an epidemic during the 1990's in Pakistan resulted in yield losses of more than US$5 billion in value.

Prior to this project, we had no knowledge of the ability of the whiteflies in Australian cotton to transmit viruses of the CLCuD complex. Also, a significant knowledge gap existed in our understanding of the complex interactions that occur between whiteflies and microbes. Microbial endosymbionts, the bacteria that live within the body or cells of another organism, are important (and often critical) to the biology of many insects. They can provide essential nutrients to their insect host, influence food digestion, sex ratios, reproduction and survival, and may even influence insecticide resistance and virus transmission.

This overarching aims of this project were to investigate the capacity of the whiteflies present in Australian cotton to transmit viruses of the CLCuD complex, the diversity of Australian whiteflies and their endosymbionts, and to develop diagnostic tools so that the risk of an incursion can be assessed and the necessary preparations made. To further support preparedness, virus diversity in Australia's northern neighbours was evaluated using new diagnostic tests and sequencing technologies.

Strong relationships were formed within this project, with much of the work done in collaboration with established CRDC researchers Dr Cherie Gambley (QDAF), Dr Murray Sharman (QDAF), Dr Jamie Hopkinson (QDAF) and Dr Lewis Wilson (CSIRO). In addition, new collaborative relationships were formed with external researchers at the University of Greenwich (United Kingdom), in particular the leader of the Bill and Melinda Gates funded African cassava whitefly project leader, Professor John Colvin.

2. Objectives

List the project objectives and the extent to which these have been achieved, with reference to the Milestones and Performance indicators.

2.1 Objective 1: *Assemble a collection of whiteflies representative of Australian cotton growing regions for molecular testing.*

2.1.1 **Milestone 1.1:** *Whitefly collection established.*

2.1.1 Performance indicator 1.1: *Whiteflies collected, and received from collaborators.*

Objective met:

This objective was met by assembling a comprehensive whitefly and whitefly-transmitted virus specimen collection, and annotated databases for each. In addition, live colonies of whitefly were collected in the field, and received from collaborators Jamie Hopkinson (DAF) and Paul De Barro (CSIRO), and these were maintained in controlled-environment insectary facilities at QDAF and CSIRO for experimental work.

The whitefly and whitefly-transmitted virus specimen collections contains samples collected during surveys undertaken in this project, and also samples that were gratefully received from collaborators from around Australia. This also includes a selection of historical collections of whiteflies (from 1993-2012), including the first populations of *Bemisia tabaci* Middle-East Asia Minor 1 (MEAM1, formerly known as B biotype) collected in Australia when it was first identified.

2.2 Objective 2: *Assess the diversity of endosymbionts infecting whiteflies from crops in Australian cotton growing regions.*

2.2.1 Milestone 2.1: *Next-generation sequencing of endosymbiont populations from representative Australian whitefly species.*

2.2.1 **Performance indicator 2.1:** *Sequencing of endosymbionts completed for at least two Australian species in the Bemisia tabaci complex.*

Objective met:

This objective was approached and met in two separate ways: (1) the endosymbiotic bacteria of more than 120 representative whiteflies collected from a range of crops, locations and time periods in Australia were interrogated using conventional molecular techniques, and (2) a deep analysis of the primary endosymbionts of 23 Australian whiteflies were assembled using next-generation sequencing technologies.

2.2.2 **Milestone 2.2:** *Next-generation sequencing of individual whiteflies to establish whitefly species, endosymbiont presence and virus presence.*

2.2.2 Performance indicator 2.2: *Data set developed and analysis of whitefly, endosymbiont and virus distribution updated 6 monthly.*

Objective met:

Illumina Miseq metagenomic sequencing has been completed and analysed for 22 separate plant and/or whitefly samples. The purpose of this work was to develop a new approach for the surveillance of whitefly and whitefly-transmitted viruses. Given that we currently have no whitefly-transmitted viruses

infecting cotton present in Australia, we decided to develop these molecular surveillance tools using a very closely related whitefly-transmitted virus that infects sweet potato in Australia. The overarching aim was to determine ways to improve the speed, reliability and cost-effectiveness of virus surveillance, to better protect the Australian cotton industry from a potential incursion of CLCuV. The methods developed also enabled the interrogation of the primary endosymbiont in whitefly samples, thereby meeting the specific goals of the objective.

2.3 Objective 3: *Research trip to the University of Greenwich to perform CLCuV transmission experiments utilizing several species of Bemisia tabaci and Australian cotton cultivars.*

2.3.1 Milestone 3.1 and 3.2: *Virus transmission experiments with northern Australian whiteflies, SLW-transmitted viruses, international whitefly populations and international SLW-transmitted viruses.*

2.3.1 Performance indicator 3.1 and 3.2: *Virus transmission experiments completed and protocols finalised for UK experiments.*

Objective met:

In order to answer the research question "Do Australian whiteflies transmit cotton leaf curl viruses", a research collaboration was established, because working with live exotic quarantine insect pests and viruses in Australia was not feasible due to quarantine restrictions. Experiments were conducted over a 6-month period in the quarantine facilities of the University of Greenwich (United Kingdom) in collaboration with Professor John Colvin. These experiments assessed the ability of Australian whiteflies, compared to a known vector species of *B. tabaci*, to transmit Cotton leaf curl Bangalore virus in Australian cotton.

2.4 Objective 4: *Targeted field collections of insecticide resistant and susceptible whiteflies, to generate cultures for experimental manipulation.*

2.4.1 Milestone 4.1: *Insecticide resistant and susceptible whiteflies collected, cultured and their endosymbionts characterised.*

2.4.1 Performance indicator 4.1: *Field collections complete and cultures of insecticide resistant and susceptible whiteflies harbouring different endosymbionts established.*

Objective met:

The aims of this research objective were to generate cultures of field populations of whitefly, to supplement existing ethanol collected specimens, for the molecular interrogation of their endosymbiont metacommunity populations. Specifically, we wanted to compare and contrast the endosymbiont communities of the known Australian cryptic species of *B. tabaci*, with specific examples of populations with known insecticide resistance status. The overarching questions for this section of the project were: (1) Which bacterial endosymbionts are correlated with each specific cryptic species of *B. tabaci* in Australia? and (2) Are particular bacterial endosymbionts associated with insecticide resistance in field collected populations. The results presented in this report for this milestone form part of a publication that is under preparation documenting the diversity of whitefly and their endosymbionts in Australia.

2.5 Objective 5: *Understand the positive and negative influences endosymbiont populations have on the biology of their whitefly hosts.*

2.5.1 Milestone 5.1: *Bioassays using whiteflies with differing endosymbiont backgrounds to determine relative fecundity, life span, survival, host-plant preference, virus transmission and insecticide resistance/susceptibility.*

2.5.1 Performance indicator 5.1: *Assessment of the potential for using bioassays to examine differences in SLW populations harbouring different endosymbionts.*

Objective met:

In order to adequately address this objective, preliminary work was done to collect, identify and maintain colonies of field-collected *B. tabaci* populations that had naturally differing endosymbiont profiles (but were genetically very similar). This task in itself proved near impossible, due to the fluctuating presence of secondary endosymbionts. Therefore, we decided (in agreement with Susan Maas) to alter the work done for this objective.

Therefore, biological experiments were done to assess the mating success among two genetically distinct, allopatric populations of native Australian whiteflies *B. tabaci* AUS1 (from Bundaberg) and AUSII (from Kununurra). These preliminary experiments generated important data and methodologies that were then expanded on by PhD student Ms Wanaporn Wongnikong (whose research is aligned with UQ1305). Ms Wongnikong has recently prepared a manuscript of this work for publication in a peer-reviewed journal.

2.6 Objective 6: *Develop new molecular methods for detecting viruses in the CLCuD complex, and species of Bemisia tabaci.*

2.6.1 Milestone 6.1: *Diagnostic development.*

2.6.1 **Performance indicator 6.1:** *Diagnostic tests evaluated for specificity, sensitivity and reproducibility.*

Objective met:

This milestone was successfully completed, and research was done for the development of improved molecular methods for the detection of cotton leaf curl viruses. An additional request in the final 6 months of the project was for the development a draft National Diagnostic Protocol for whitefly in Australia, and this is provided in this report in its entirety.

2.6.2 **Milestone 6.2:** *Provide diagnostic support to Dr Cherie Gambley (DAQ1405) and Murray Sharman (DAQ1201) for CLCuD surveillance in Australia (samples collected during survey collections).*

2.6.2 Performance indicator 6.2: *Perform molecular diagnostic testing when survey samples are made available.*

Objective met:

We provided continued diagnostic support to Cherie Gambley (DAQ1405), Jamie Hopkinson (DAQ1403), and Murray Sharman (DAQ1201) throughout the duration of the project. We performed molecular identification of whiteflies (and the viruses they transmit), as well as routine virus indexing of plant samples collected during surveys in Australia as part of their CRDC projects. The details of samples tested and their results (and an interpretation of these results) are provided in the body of this report.

2.7 Objective 7: *Research to support the activities of the African Cassava Whitefly Project (ACWP; Bill and Melinda Gates) at the Natural Resources Institute, University of Greenwich, as a join postdoc under the supervision of Prof John Colvin.*

2.7.1 Milestone 7.1: *Perform bioassay experiments to compare the ability of five different cryptic species of B. tabaci to utilise and detoxify host plant defense compounds.*

2.7.1 **Performance indicator 7.1:** *Bioassays completed, samples prepared and stored for chromatography and transcriptomics, report submitted to collaborators Prof John Colvin (University of Greenwich) and Dr Osnat Malka (The Hebrew University of Jerusalem).*

Objective met:

Work for this milestone was performed in collaboration with researchers of the ACWP Dr Osnat Malka and Prof Shai Morin (both from The Hebrew University of Jerusalem). I developed and performed experimental bioassays from August 2015 – April 2016, with data collated and presented in a report to Prof Morin before I left the University of Greenwich in May 2016. We received very positive feedback for the quality of the work done, from Prof Morin.

2.7.2 **Milestone 7.2:** *Coordinate and provide high-level support to several research themes within the ACWP.*

2.7.2 **Performance indicator 7.2:** *Management of insectaries and molecular laboratory, direct supervision of two research technicians, training of three PhD students in molecular biology techniques and report to Prof John Colvin.*

Objective met:

During my time as join-postdoc with the University of Greenwich, I performed various duties to support research in the ACWP, including supervising two research technicians, training seven PhD students in molecular biological techniques, and contributing to experimental research for various projects (both biological and molecular work).

2.7.3 **Milestone 7.3:** *Molecular characterisation of infectious clones from Danforth collection (USA) including Cotton leaf curl virus, and various cassava-infecting begomoviruses. Development of biolistic techniques for inoculation of tobacco.*

2.7.3 Performance indicator 7.3: *Infectious clones rescued, sequenced and report provided to Prof Sue Seal.*

Objective met:

In collaboration with Prof Sue Seal and Prof Maruthi Gowda (Uni of Greenwich), I completed a short project using molecular techniques to characterise 48 infectious clones held in a collection. I also developed and began work-up of modified biolistic inoculation procedures for inoculating experimental plants. I focussed on developing the *Cotton leaf curl virus* infectious clones, to develop a resource for future CRDC research initiatives, for example screening potential CLCuV-resistant plants. All results were compiled in a report submitted to Prof Seal before leaving the UK.

2.8 Objective 8: *Final analysis and publication.*

2.8.1 Milestone 8.1: *Analysis finalized*

2.8.1 **Performance indicator 8.1:** *Analysis finalized*

Objective met:

Analysis of the data generated in this project are mostly finalised, with much of it being prepared for inclusion in peer-reviewed journal articles.

2.8.2 **Milestone 8.2:** *Preparing results for publication in international journals.*

2.8.2 **Performance indicator 8.2:** *Two manuscripts submitted for publication in peer-reviewed international journals.*

Objective not yet met:

This milestone was delayed due to the changes in personnel on the project. We have one advanced draft of a paper detailing the development and application of new tools for the metagenomics detection of viruses and their vectors, that is due to be submitted to the journal in February 2018.

(1) van Brunschot S, Persley D, Dennien S, Hereward J, McMichael, L, Thomas JE, Walter G (2018). Viral metagenomics: a tool for virus discovery and analysis of diversity in plant and insect vectors. Viruses MDPI.

An additional manuscript is in its early stages (2), however the data and analysis for this paper are now complete. This manuscript is due for submission to the journal in May 2018.

(2) van Brunschot SL, Gambley C, Stainton D, Hereward JP, Walter GH (2018). A spatio-temporal analysis of an insect and its endosymbionts. Insect Science.

2.8.3 **Milestone 8.3:** *Communicate results to Australian cotton industry and researchers.*

2.8.3 **Performance indicator 8.3:** *Results presented at Australian meetings and industry publications.*

Objective met:

Throughout this project, results were continually presented at Australian research meetings and in industry publications. In total, eight separate presentations/posters were given for the work done in this project, and two full length industry articles were published in the Cotton Info Spotlight Magazine.

2.8.4 **Milestone 8.4:** *Final report development for CRDC.*

2.8.4 Performance indicator 8.4: *Report developed.*

Objective met:

The final report for this project was developed by Sharon van Brunschot and Gimme Walter, with some data and inputs provided by Daisy Stainton.

3. Methods

Detail the methodology and justify the methodology used. Include any discoveries in methods that may benefit other related research.

3.1 General methods

3.1.1 Whitefly collection and culturing

Laboratory colonies of three different cryptic species of *B. tabaci* (AUS1, AUS2 and MEAM1) were reared and maintained in Australia at the CSIRO and QDAF insectary facilities, Ecosciences Precinct. Colonies were maintained separately on *Euphorbia cyathophora* (Euphorbiaceae) and *Solanum melongena* cv. Black Beauty (eggplant, Solanaceae). Colonies were reared under standard conditions of 28 ± 2 °C, 60% humidity and a 14:10 h light:dark cycle. Colony purity was maintained by adhering to strict hygiene protocols, whereby only a single colony was opened on a given day, and yellow sticky traps were used to capture escaped whitefly in CER rooms.

Whitefly samples were collected in the field using a custom-made aspirator, the stored immediately in 90 % ethanol. Alternatively, they were transferred to a suitable clean host plant in a whitefly-proof cage for establishing a colony.

The purity of colonies was continually confirmed (every 2 months) by standard molecular methods (Dinsdale *et al*., 2010). Briefly, the 3' partial mitochondrial cytochrome oxidase 1 gene of ≥3 individuals of each cryptic species of *B. tabaci* was sequenced.

3.2 DNA extraction and sequencing

For standard PCR and sanger sequencing work, total nucleic acids were extracted from individual whitefly specimens from each of the populations using a modified chelex extraction method (White et al., 2009).

For next-generation sequencing of *B. tabaci* specimens, nucleic were extracted using a high-throughput spin column method. Briefly, single whitefly were homogenized using sterile zirconium beads in 190 µl of Lysis buffer (0.5% SDS, 250 mM NaCl, 25 mM EDTA, 200 mM Tris-HCL pH7.5) and 10 µl of Proteinase K (10mg/ml), then incubated at 55°C overnight using a dry heating block. DNAs were precipitated by adding 100 µl of Precipitation buffer (4M Ammonium acetate), with incubation at -20°C for 30 minutes. Then, 300 µl of Binding buffer (4M GuHCl, 75% Ethanol) was added to each tube, mixed by inversion, and a total of 520 µl of homogenate was transferred to spin column tubes or plates (Epoch BioLabs), and centrifuged at full speed for 4 minutes (at room temperature). The flow through was discarded and two membrane washes were done using 500 µl of Wash buffer (10 mM Tris-HCl pH7.5, 80% ethanol). DNAs were eluted twice using 50 µl of pre-warmed Elution buffer (10mM Tris-Cl, pH 7.5) to each tube or each well, then centrifuged at full speed for 2 minutes.

3.3 Mitochondrial cytochrome oxidase 1 gene PCR

PCR amplification of the 655 bp region of 3' barcoding region of the mitochondrial cytochrome oxidase I (COI) gene was done using standard methods, see Table 1 and Table 2 for details. (Simon et al., 1994). The partial nuclear pre-mRNA processing factor 8 (Prp8) gene was amplified using standard methods, see Table 3 (Hsieh et al., 2014a).

PCR products were verified by agarose gel electrophoresis and cleaned using 1U of Exonuclease I and Antarctic Phosphatase (New England Biolabs, Massachusetts, United States) under conditions at 37 °C for 20 min followed by 10 min at 80 °C. Then clean products were sequenced by Macrogen Inc. (Seoul, Republic of Korea).

For DNA extractions that had no amplification of this 3' region, the LCO1490 and HCO2198 primer pair designed to amplify the 5' region of the mtCOI (Folmer et al., 1994) was used.

Table 1: Details of PCR primers used to amplify mitochondrial and nuclear markers of Bemisia tabaci in this project

Region	Primer name	Primer Sequences (5'-3')	Amplicon size	Reference
prp8	prp8F	GCCTTGGGAGGTGTTGAAG	1060	(Hsieh et al., 2014a)
	prp8R	GGCTTGCATCCAGGGTACC		
	prp8seqMF	CTGGAGTTCTCATTGCGATC		
mtCOI (3')	CI-J-2195	TTGATTTTTTGGTCATCCAGAAGT	600	(Simon et al., 1994)
	TL2-N-3014	TCCAATGCACTAATCTGCCATATTA		
mtCOI (5')	LCO1490	GGTCAACAAATCATAAAGATATTGG	800	(Folmer et al., 1994)

	HC02198	TAAACTTCAGGCTGACCAAAAAATCA		

Table 2: Mitochondrial PCR conditions (Simon and Folmer primers)

PCR cycling protocol – Touch up
$95_{3min}(95_{30s},45_{30s},72_{1min})^{10}$, $(95_{30s},50_{30s},72_{1min})^{35}72_{10min}$

Reagent	Final amount (µL)
MyTaq Polymerase (5U/µL)	0.2
Primer F C1J-2195 or LCO1490 (10µM)	0.5
Primer R TL2-N-3014 or HC02198 (10µM)	0.5
MyTaq Buffer HS 5x	6.0
Nuclease-free water	20.8
Template	2.0
Total	30.0

Table 3: Nuclear pre-mRNA processing factor 8 (Prp8) gene PCR conditions

PrP8
PCR cycling protocol
$95_{2min}(95_{30s},55_{30s},72_{1min})^{35}72_{2min}$

Reagent	Final amount (µL)
MyTaq Polymerase (5U/ul)	0.2
Primer F/R (10µM each primer)	1.0
MyTaq Buffer HS 5x	6.0
Nuclease-free water	20.8
Template	2.0
Total	30.0

3.4 Primary and secondary endosymbiotic bacteria detection and identification

Primary and secondary endosymbiotic bacterial of *B. tabaci* samples were amplified using standard methods, see Table 4, Table 5, and Table 6 for details.

PCR products were verified by agarose gel electrophoresis and cleaned using 1U of Exonuclease I and Antarctic Phosphatase (New England Biolabs, Massachusetts, United States) under conditions at 37 °C for 20 min followed by 10 min at 80 °C. Then clean products were sequenced by Macrogen Inc. (Seoul, Republic of Korea).

Table 4: Endosymbiotic bacteria PCR primers used in this study.

Bacteria	Primer name	Primer Sequences (5'-3')	Amplicon size	Reference
Portiera sp.	28F	TGCAAGTCGAGCGGCATCAT	1000 - 1100	(Zchori-Fein & Brown, 2002)

	1098R	AAAGTTCCCGCCTTATGCGT		
Wolbachia sp.	81F	TGGTCCAATAAGTGATGAAGAAAC	590-632	(Braig et al., 1998)
	691R	AAAAATTAAACGCTACTCCA		
Rickettsia sp.	Rb-F	GCTCAGAACGAACGCTATC	900	(Gottlieb et al., 2006)
	Rb-R	GAAGGAAAGCATCTCTGC		
Hamiltonella sp.	92F	TGAGTAAAGTCTGGGAATCTGG	1250	(Zchori-Fein & Brown, 2002)
	1343R	CCCGGGAACGTATTCACCGTAG		
Cardinium sp.	CFB-F	GCGGTGTAAAATGAGCGTG	400	(Weeks & Breeuwer, 2003)
	CFB-R	ACCTMTTCTTAACTCAAGCCT		
Arsenophonus sp.	Ars23S-1	CGTTTGATGAATTCATAGTCAAA	600	(Thao & Baumann, 2004)
	Ars23S-2	GGTCCTCCAGTTAGTGTTACCCAAC		

Table 5: PCR conditions for Porteria sp., Arsenophonus sp., Wolbachia sp., Hamiltonella sp. and Rickettsia sp.

PCR cycling protocol

$95_{3min}(95_{30s},54_{30s},72_{1min})^{30}72_{2min}$

Reagent	Final amount (µL)
Mytaq polymerase (5U/µL)	0.15
Primer F/R (10µM each primer)	1.00
Mytaq Buffer HS 5x	4.00
Nuclease-free water	13.85
Template DNA	1.00
Total	20.00

Table 6: PCR conditions for Cardinium sp.

PCR cycling protocol

$95_{3min}(95_{30s},54_{30s},72_{1min})^{30}72_{2min}$

Reagent	Final amount (µL)
Mytaq Polymerase (5U/µL)	0.15
Primer F/R (10µM each primer)	3.00
Mytaq Buffer HS 5x	4.00
Nuclease-free water	11.85
Template DNA	1.00
Total	20.00

3.5 Plant DNA extraction and internal PCR control

Fresh and preserved (freeze-dried) plant samples were extracted using a modified standard method, detailed below (Doyle & Doyle, 1987). To ensure the extraction was successful, and that no PCR inhibitors such as phenols were interfering with the PCR, plant DNA barcode primers were used which target a 355bp region of *ITS2* gene. These primers ITS2-S2F (Chen et al., 2010a) and ITS4 (White et al., 1990b) have been used previously to barcode cotton species (Ashfaq et al., 2013). Pcr conditions are listed in Table 7.

CTAB DNA extraction method

1. Tissue (~40mg of fresh tissue or ~3mg of freeze dried) is ground in 1ml of 2% CTAB buffer (with 0.5% NaSO3 added fresh, e.g. 50mg in 10ml), in a 2ml tube (or powdered FD in 2ml tube before adding buffer), then incubated at 60ºC for 30 min with mixing by inversion every 5 min (or on the thermo-shaker).
2. Add 800 ul of chloroform-isoamyl alcohol (24:1), mix thoroughly by inversion / shaking. Do this in a fume hood.
3. Spin at 14 000 g for 5 min in a microcentrifuge at room temperature. Do this in a fume hood.
4. Using a pipette, transfer 400 ul of aqueous phase to 700 ul isopropanol in a new 1.5 ml tube (with final label) and mix by inversion.
5. Spin at 14 000 g for 15 min in a microcentrifuge at room temperature.
6. Using a pipette, remove and discard supernatant and wash pellet twice with ~700µl of 70% Ethanol.
7. Spin down pellets at 14 000 g for 2-5 min at room temperature for each wash.
8. Using a pipette, remove ethanol and vacuum dry pellet for approximately 10 min in speedivac at 40 ºC.
9. Resuspend pellets in 100 ul ddH2O (or more if large pellet). May need to heat / shake at 60C for 10min if rapid resuspension is required, or use next day after freezing.
10. Store at -20ºC.

Table 7: ITS2-S2F & ITS4 plant internal control PCR conditions

PCR cycling protocol

$95_{2min}(95_{30sec},50_{30sec},72_{1min})^{30}72_{2min}$

MyTaq Polymerase (5U/ul)	0.2ul
Primer F/R (10uM each primer)	1ul
MyTaq Buffer HS 5x	6ul
Water	21.8ul
Template	1ul
Total	30

3.6 Begomovirus detection and identification

Table 8: PCR primers used to detect begomoviruses in this study, including internal control assay for plant gene.

PCR name	Primer name	Primer Sequences (5'-3')	Amplicon size	Reference
Accore	ACCORE	GGRTTDGARGCATGHGTACAIGCC	550	(Wyatt & Brown, 1996a)
	AVCORE	GCCHATRTAYAGRAAGCCMAGRAT		
Beg	BEGF	CCAYTCTCTGCTTGAGBTGC	300	(van Brunschot et al., 2014)
	BEGR	ATCTTCCTNTGCAATCCAGG		
ITS	ITS2-S2F	ATGCGATACTTGGTGTGAAT	355	(Chen et al., 2010a)
	ITS4	TCCTCCGCTTATTGATATGC		(White et al., 1990b)

Table 9: Beg F/R PCR conditions

PCR cycling protocol

$95_{2min}(95_{30s},58_{30s},72_{1min})^{30}72_{2min}$

Reagent	Final amount (µL)
MyTaq Polymerase (5U/ul)	0.2
Primer F/R (10µM each primer)	1.0
MyTaq Buffer HS 5x	6.0
Water	20.8
Template	2.0
Total	30.0

Table 10: Accore/Avcore PCR conditions

PCR cycling protocol

$95_{2min}(95_{30s},50_{30s},72_{1min})^{30}72_{2min}$

Reagent	Final amount (µL)
MyTaq Polymerase (5U/µL)	0.2
Primer F/R (10µM each primer)	1.0
MyTaq Buffer HS 5x	6.0
Water	20.8
Template	2.0
Total	30.0

3.6 Sequence analysis

Forward and reverse sanger sequences were assembled using the denovo tool of Geneious v9.1.7 (Kearse et al., 2012a). Resulting contiguous sequences were trimmed and edited manually, then

compared with publicly available sequence data on GenBank using the nucleotide BLAST tool (Altschul et al., 1990), using the blastn algorithm against the nucleotide collection (nr/nt).

3.7 Phylogenetic inference

Whitefly mitochondrial COI sequence analyses

Representative whitefly 3' mitochondrial cytochrome oxidase 1 gene (mtCOI) sequences were downloaded from GenBank and used for phylogenetic reconstruction. All sequences were aligned using MAFFT (Katoh & Standley, 2013). The best-fit model of evolution (GTG+R) was determined from likelihood ratio tests performed using jModelTest (Posada, 2008). Phylogenetic relationships were reconstructed via Bayesian analyses using MRBAYES v3.1.2 (Ronquist & Huelsenbeck, 2003), implemented in Geneious v9.1.7 (Kearse et al., 2012a). Four Markov chains were run for a total of 1,100,000 generations, with chains sampled every 200 generations (the first 10 % of trees were discarded as burn-in). Posterior probabilities ≥ 0.95 were considered substantial node support. TRACER 1.5 (Rambaut & Drummond, 2003) was used to check that the effective sample size was not too low, indicating that the MCMC mixed well and that the samples were independent. Sequences from distant relatives of these taxa were included as outgroups.

Bacterial species-specific gene sequence analyses

Sequence data generated in this work for endosymbiotic bacteria of *B. tabaci* specimens was aligned with reference sequences downloaded from GenBank, using MAFFT (Katoh & Standley, 2013), implemented in Geneious v9.1.7 (Kearse et al., 2012a). Phylogenetic relationships were estimated using a distance-based approach, whereby a neighbour-joining phylogeny was generated using a HKY (Hasegawa, Kishino and Yano) model of nucleotide substitution, implemented in Geneious v9.1.7 (Kearse et al., 2012a). Reliability of nodes was assessed using 1,000 bootstrap iterations. Sequences from distant relatives of these taxa were included as outgroups.

Begomovirus DNA-A analyses

Begomovirus sequence data generated in this work was aligned with reference sequences downloaded from GenBank, using MAFFT (Katoh & Standley, 2013), implemented in Geneious v9.1.7 (Kearse et al., 2012a). Phylogenetic relationships were estimated using a distance-based approach, whereby a neighbour-joining phylogeny was generated using a HKY (Hasegawa, Kishino and Yano) model of nucleotide substitution, implemented in Geneious v9.1.7 (Kearse et al., 2012a). Reliability of nodes was assessed using 1,000 bootstrap iterations. Sequences from distant relatives of these taxa were included as outgroups.

4. Results

Detail and discuss the results for each objective including the statistical analysis of results.

4.1 Objective 1: *Assemble a collection of whiteflies representative of Australian cotton growing regions for molecular testing.*

The reference collection assembled for this project contained 607 distinct whitefly populations. These were collected across a range of agricultural and native landscapes, therefore with differing exposure to pesticides and whitefly-transmitted virus pressures. Over 90% of the samples were collected from locations within Australia – Queensland, New South Wales, the Northern Territory, Western Australia, South Australia and the Torres Strait Islands. The remaining samples were collected from Papua New Guinea, as well as countries where cotton leaf curl disease (CLCuD) is known or may potentially occur (Indonesia, China and Egypt).

Surveys were conducted during this project for collecting whiteflies and potential plant hosts of CLCuD in Kununurra (Western Australia) and Darwin (Northern Territory), in conjunction with Department of Agriculture staff (formerly NAQS). University of Queensland and CRDC PhD student Dean Brookes (UQ1403) joined these surveys to collect insects for his PhD research on *Nezara viridula* (green vegetable bug), see Figure 1. In both Kununurra and Darwin, we sampled in various community gardens, nurseries and larger farms and trial plots (including cotton). In total, we collected 32 plant samples (across 15 different hosts, including cotton and native relatives), and 27 whitefly populations (across 14 different hosts).

Figure 1: CRDC pest and disease survey for whiteflies (Bemisia tabaci), whitefly-transmitted viruses (begomoviruses) and green-vegetable bug (Nezara viridula) in northern Australia (Darwin and Kununurra, 2014). Left to right: Dave Nicholson (DAFWA), Luke Halling (NAQS entomologist), Sharon van Brunschot (UQ), Dean Brookes (UQ), John Westaway (NAQS botanist) and Gerard Worm, permaculturalist at Kununurra Community Garden. Photo by Jane Ray (NAQS).

An additional survey was conducted in conjunction with CRDC researchers Cherie Gambley (DAQ1405), Paul Campbell (DAQ1405), and Dean Brookes (UQ1403) with the assistance of Siva Subramaniam (QDAF). Together, we surveyed cropping regions, nurseries and community gardens in Townsville, Ayr, Gumlu and Bowen. Whiteflies were collected from a range of horticultural crops, ornamental hosts and weeds, including a single mature cotton crop in Bowen. In total, we collected 31 whitefly populations (across 9 different hosts).

A large number of whitefly populations were provided by Lewis Wilson and Tanya Smith (CSIRO), as part of a collaborative research project investigating which native and weed plant hosts contribute to pest whitefly build-up between cotton cropping seasons. A total of 193 whitefly populations, collected from a vast range of predominantly native and weed plant hosts, all from cotton growing regions in

NSW (2012-2016) were provided to us. In addition, Jamie Hopkinson (DAQ1403) and Paul Grundy (QDAF) provided 53 whitefly populations from cotton and native hosts, collected from QLD and NSW cotton growing regions from 2013-2017 seasons. Murray Sharman (DAQ1201) provided 14 populations of whiteflies collected during his own survey work in central QLD cotton growing regions, and from the Torres Strait Islands. Cherie Gambley (DAQ1405) provided critical samples that she collected during surveys of remote regional areas of Western Australia, and also tomato crops in South Australia. In total, Dr Gambley provided 24 populations across these locations, collected from crop and weed hosts.

Additional samples were gratefully received from various collaborators and colleagues around Australia, including Steven Coventry (Biological Services, South Australia), Diana Renfree (EE Muir and Sons, Katherine, NT), and both Luke Halling and Sally Cowan (NAQS).

For the historical samples, a total of 137 populations were provided by Paul De Barro (CSIRO), with most collected between 1993-1997 from crop and weed hosts across mainland Australia (18 distinct populations from cotton). This time-period is important for gaining an improved understanding of temporal changes in whitefly populations in Australia. This is because in October 1994, *Bemisia tabaci* Middle-East Asia Minor 1 (MEAM1; formerly known as B-biotype) was first identified in Australia, but investigations suggested it was introduced in late 1993 into Coffs Harbour (NSW), via legal importations of the ornamental plant *Euphorbia pulcherrima* (poinsettia) from California (Gunning et al., 1995).

A number of overseas samples were collected and contributed by collaborators. Whitefly samples from Papua New Guinea (2013) were provided by Cherie Gambley as part of a survey conducted in her CRDC project (DAQ1405). Whitefly and virus-infected plant samples from China were collected on a pest and disease survey conducted as part of this project (2013). In total, 28 virus-infected plant samples and 26 whitefly populations were collected from over 14 distinct sites throughout southern, central and northern China. Importantly, samples were obtained from the following CLCuV-infected hosts: *Gossypium hirsutum* (cotton), *Abelmoschus esculentus* (okra), *Hibiscus rosa-sinensis* (china-rose hibiscus) and nearby weed species, see Figure 2. In addition, symptomatic, putatively begomovirus-infected *Solanum lycopersicum* (tomato), *Capsicum sp.* (chilli), chinese medicinal herbs and weed species were collected. *B. tabaci* populations (some viruliferous) were collected from virus-infected host plants including: okra, hibiscus, eggplant, tomato, cucurbits and various weed species. Whitefly from Indonesia were collected by Sharon van Brunschot (privately), and other populations from Indonesia and Egypt were provided by Paul De Barro (CSIRO).

Figure 2: Gossypium hirsutum (cotton) infected with Cotton leaf curl Multan virus, in southern China. Note the symptoms of upward curling and cupping of leaves, vein darkening and vein thickening on the underside of leaves. Photo by Sharon van Brunschot.

In addition, a reference virus collection (and database) has been generated for the project containing more than 70 begomovirus-infected plant samples from Australia and overseas. This collection is stored as part of the Queensland Department of Agriculture, Fisheries and Forestry (QDAF) Herbarium under up-to-date AQIS importation permits and in an AQIS-certified Quarantine PC2 Facility, where required. This collection contains many CLCuV samples originally collected by Cherie Gambley, in addition to CLCuV and associated begomovirus samples that I collected in China as part of the current project, and virus isolates previously collected from around Australia by myself and our collaborators at QDAF, John Thomas and Denis Persley.

Together, this large collection of *B. tabaci* and begomovirus-infected plant samples assembled during this project formed an important basis for much of the research done in this project examining the diversity of whiteflies, endosymbionts and whitefly-transmitted viruses in Australia. The virus isolates were invaluable for the development and validation of improved diagnostic tests for CLCuD, done within this project by Daisy Stainton. Also, it will be a long-standing resource for future research done by PhD student Wanaporn Wongnikong, and any further CRDC-related research on whitefly and whitefly-transmitted viruses.

4.2 Objective 2: *Assess the diversity of endosymbionts infecting whiteflies from crops in Australian cotton growing regions.*

The diversity of Bemisia tabaci and their endosymbiotic bacterial metacommunities in Australia.

We examined the genetic diversity of whiteflies, and the primary and secondary endosymbionts they harbour, spanning a 20 year period in Australia.

The identity and diversity of whiteflies was investigated by conventional DNA sequencing of the partial mitochondrial cytochrome oxidase 1 gene (mtCOI). In addition, sequencing was done for representative samples (more than 1000) of the nuclear pre-mRNA processing factor 8 (PrP8), to support the mtCOI data.

The identity and diversity of the endosymbiotic communities of whiteflies in Australia was interrogated using a conventional PCR and sequencing approach. Six known endosymbiotic bacteria were checked for in representative whitefly populations, including the primary endosymbiont *Portiera*, and five secondary endosymbionts, *Arsenophonus, Cardinium, Hamiltonella, Rickettsia and Wolbachia*. Results for these sequencing projects are discussed individually below.

mtCOI barcoding of whiteflies in Australia

The partial mtCOI gene of 1585 whitefly individuals were sequenced, analysed, and compared to publicly available sequences on GenBank, to infer genetic relatedness and give a preliminary DNA sequence-based identification. All unique haplotypes (60 in total) are presented in a global phylogenetic tree presented in Figure 3. This phylogenetic tree contains sequences from 815 single whiteflies, from around the world, including all unique Australian haplotypes. Due to the complexity of this figure, it has been broken down into four sections that will be discussed individually.

Importantly, we decided in our work not to rely on mitochondrial DNA alone for uncovering evolutionary relationships within this cryptic species complex. We know that various factors, including fluctuating effective population sizes, gender-biased gene flow, and introgression affect mitochondrial and nuclear markers differently, and this implies that the use of mtDNA alone may result in biased estimates of evolutionary history (Ballard & Whitlock, 2004, Monsen & Blouin, 2003, Piertney et al., 2000, Chan & Levin, 2005, Pinho et al., 2007). Therefore, we sequenced the nuclear pre-mRNA processing factor 8 (prp8) gene of representative whitefly samples, to test for concordance of phylogenies.

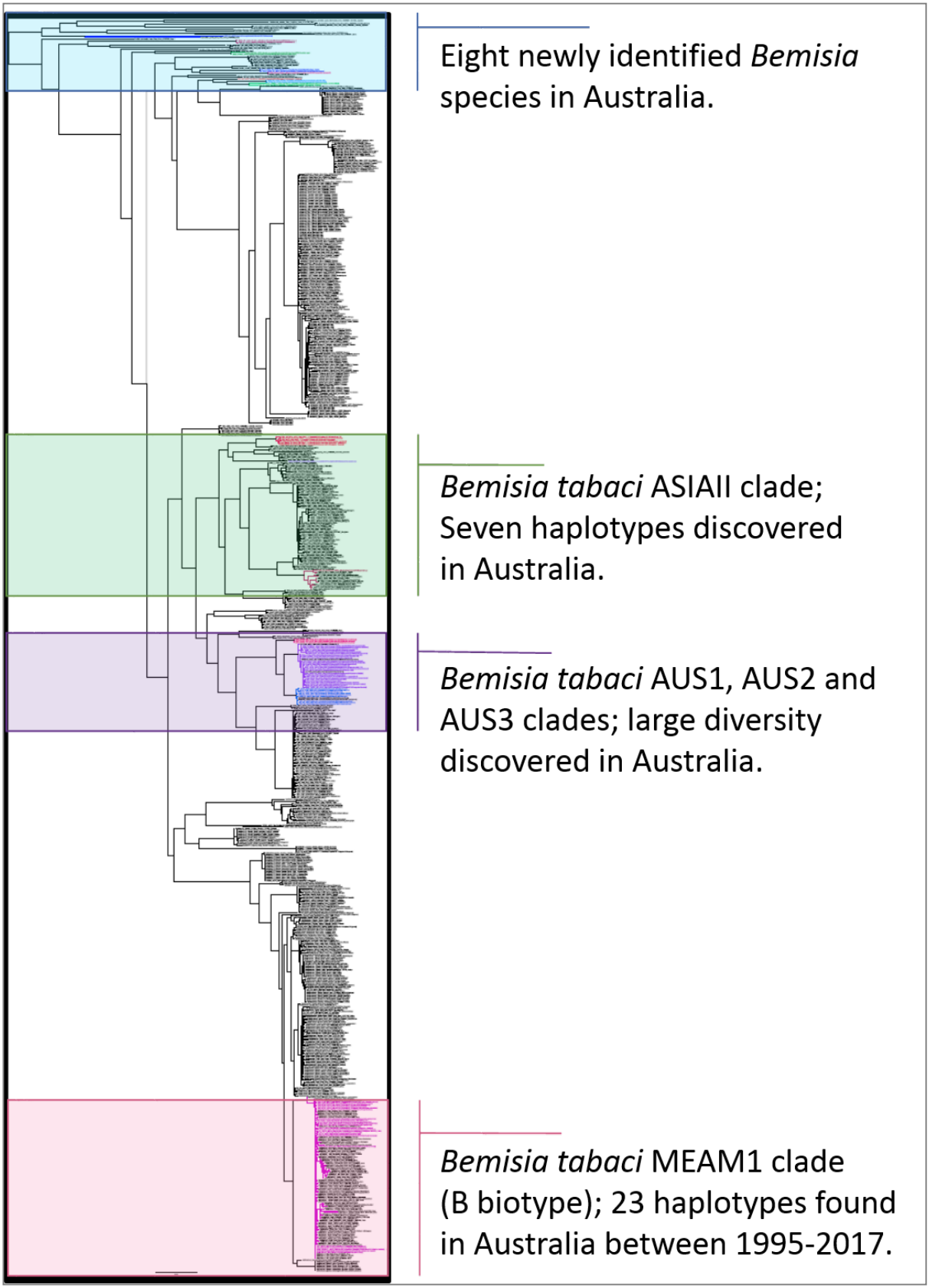


Figure 3: Bayesian phylogeny of global Bemisia tabaci species complex, constructed from mitochondrial cytochrome oxidase gene sequences (657 bp). The major clades that Australian whiteflies collected during this project belong to are highlighted in colour, and insets of each are shown in subsequent figures given the large size and complexity of the analysis.

mtCOI barcoding of basal Bemisia species in Australia

Eight *Bemisia* species, most of which remain unclassified, were identified (Figure 4). A population of *Bemisia subdecipiens* was identified from a *Melaleuca* sp. in Brisbane (Queensland, provided by Murray Sharman QDAF). Two haplotypes of an unclassified *Bemisia* sp. were identified from collections in Wyndham (far north of Western Australia, provided by Cherie Gambley QDAF), which shared no significant matches to any whitefly sequences on GenBank. Two haplotypes of an unclassified *Bemisia* sp. was identified by collections made by Tanya Smith (CSIRO) in cotton-growing regions in NSW, from sunflower and mung bean. A single population of *Bemisia atriplex* was collected by Jamie Hopkinson (DAF) from saltbush. A single mixed population of *B. tabaci* AUS1 and an unclassified *Bemsia* sp. was collected from *Sonchus oleraceus* in Bileola in 1996. A population of *Bemisia euphorbiae* was collected in Kununurra in 2014 during pest and disease surveys conducted in this region. An unclassified *Bemisia* sp. was collected on fireweed in Narrabri, near cotton. And a closely related unclassified *Bemisia* sp. was identified from *Euphorbia* sp. that grew in in Emerald in 1996.

These *Bemisia* sp. have been extremely useful for our work, we have used them as basal outgroups in phylogenetic inference studies, and have also interrogated these samples to examine their endosymbiotic metacommunity diversity (in comparison to *B. tabaci*). A number of these specimens have been sent for formal taxonomic identification and to be lodged as vertebrate specimen vouchers by Peter Gillespie (curator of collections, NSW Biosecurity).

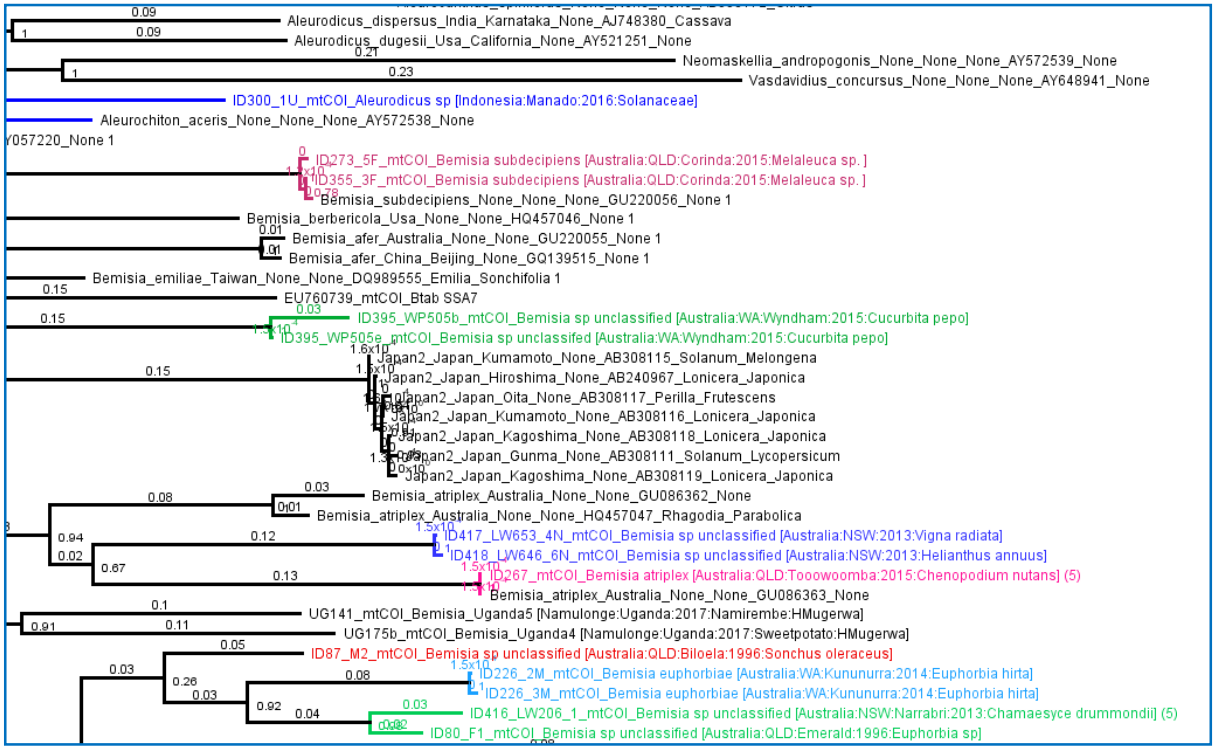


Figure 4: Inset of Bayesian phylogeny of global Bemisia tabaci species complex, constructed from mitochondrial cytochrome oxidase gene sequences (657 bp). This inset focuses on the eight newly identified Bemisia species in Australia, discovered during this project. These include: (1) Bemisia subdecipiens collected from Melaleuca sp. in Brisbane (QLD, 2015), (2) two unclassified Bemisia species collected from cucurbits in Wyndham (WA, 2015), (3) unclassified Bemisia species collected from Vigna radiata (mung bean) and Helianthus annuus (sunflower) in Narrabri (NSW, 2013), (4) Bemisia atriplex collected from Chenopodium nutans (saltbush) in Toowoomba (QLD, 2015), (5) an unclassified Bemisia species collected from Sonchus oleraceus in Bileola (QLD, 1996), (6) Bemisia euphorbiae collected from Euphorbia hirta in Kununurra (WA, 2014), (7) and unclassified Bemisia species collected from Chamaesyce drummondii (fireweed) in Narrabri (NSW, 2013), and (8) and unclassified Bemisia species collected from a Euphorbia weed in Emerald (QLD, 1996).

mtCOI barcoding of Bemisia tabaci ASIAII in Australia

An important discovery during this project, was the presence of *B. tabaci* ASIAII whiteflies in Australia, see Figure 5. These were first identified on a survey conducted as part of this project, in the Northern Territory in 2014. Specifically, seventeen individuals were collected across four locations in Darwin (on tomato, cucumber, pumpkin, eggplant and okra) and also at three locations in Kununurra (on cotton and sweet potato). Subsequent to this detection, a further seven individuals were detected in samples provided by Dr Cherie Gambley, collected on both *Physalis angulata* (solanaceous weed) and cucurbit plants in Wyndham and Kalumburu (Western Australia).

In total, we recovered six distinct haplotypes of *B. tabaci* ASIAII, and subsequently provisionally named these *B. tabaci* ASIAII_13. Sequence analysis showed they were significantly genetically different (more than 14% different over the sequenced COI gene fragment) to whiteflies previously reported in Australia.

Using the mtCOI barcode for comparison, the Australian *B. tabaci* ASIAII_13 were found to be most closely related to *B. tabaci* ASIAII_12 whitefly sample, which was collected on eggplant from Karawang, West Java, Indonesia (Firdaus et al., 2013). The Asia II species group contains the Asian H, K, ZHJ1, ZHJ2, G, Cv and P biotypes (for readers familiar with the biotype nomenclature). Importantly, the *B. tabaci* ASIAII species group contains the species *B. tabaci* ASIAII_1, which is known to be associated with epidemics of *Cotton leaf curl virus* (CLCuV) in Pakistan (Ahmed et al., 2011).

A single haplotype, subsequently provisionally named *B. tabaci* ASIAII_14 was identified from okra grown in Humpty Doo (NT). This haplotype is distinct from other *B. tabaci* in the ASIAII clade, and the full mitogenome has been sequenced using next-generation sequencing technologies during this project.

These findings may represent an introduction of this Asian species of whitefly, either recently or a very long time ago (potentially even millions of years ago). At the time of detection, we reported these identifications to both state (NT and WA) and federal biosecurity agencies, with the CCEP finding that no further action was required (with regard to eradication).

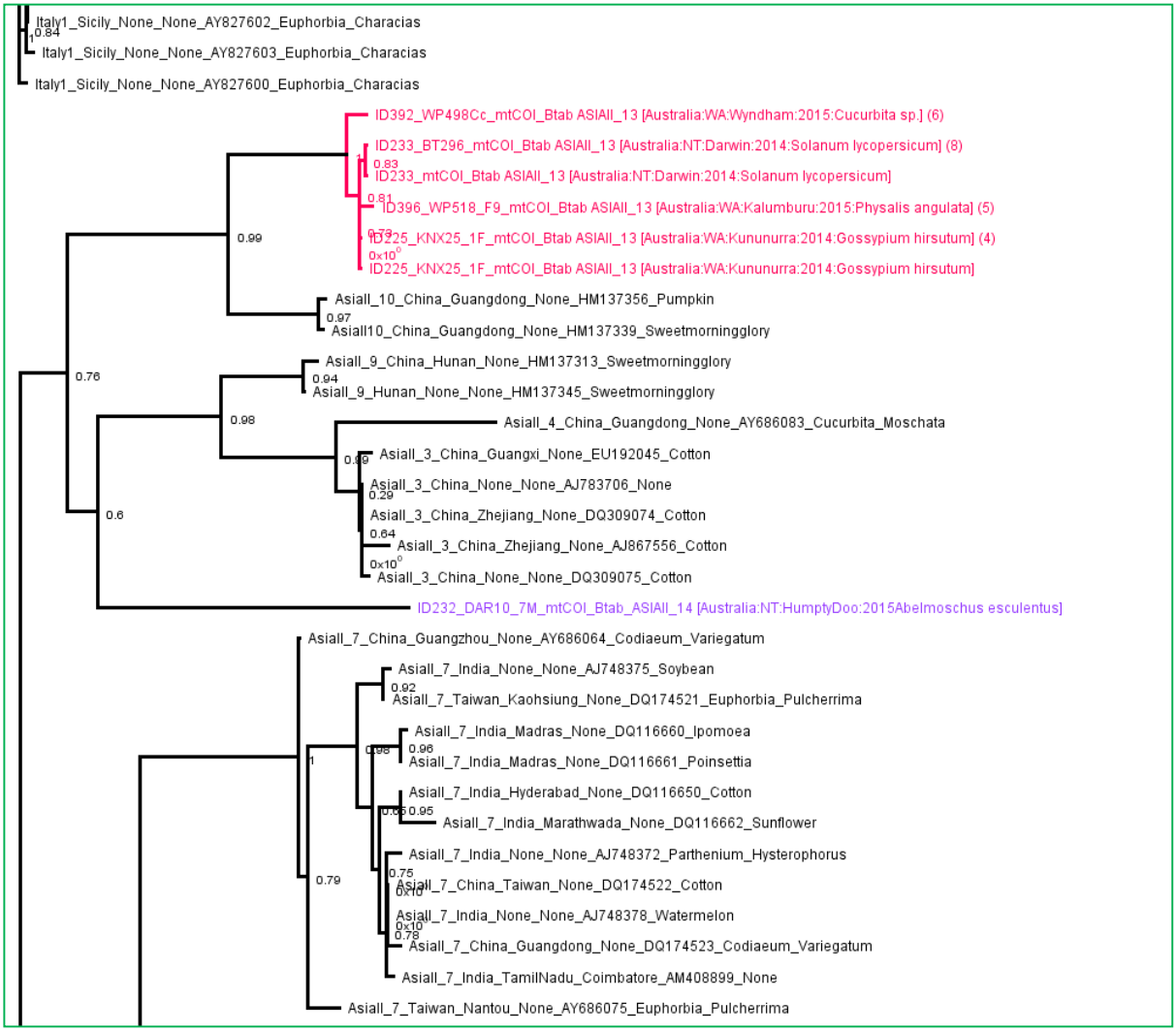


Figure 5: Inset of Bayesian phylogeny of global Bemisia tabaci species complex, constructed from mitochondrial cytochrome oxidase gene sequences (657 bp). This inset focuses on a subset of the Bemisia tabaci ASIAII clade. All haplotypes in the ASIAII clade originate from countries in south and south-east Asia, except for the newly identified haplotypes discovered in this project from Australia (coloured in pink and purple, numbers in parenthesis at the end of names indicate the total number of that haplotype that have been sequenced). Six distinct haplotypes, now named B. tabaci ASIAII_13 were collected in northern Australia on a variety of hosts including cotton. Also, a single whitefly collected from eggplant in Darwin (NT, 2015) was shown to be genetically distinct from the ASIAII_13 haplotypes, sharing only ~89% nt sequence identity.

mtCOI barcoding of Bemisia tabaci AUS1, AUS2 and AUS3 in Australia

An unexpected diversity of Australian native whiteflies was uncovered during this work. Twenty-six distinct haplotypes of *B. tabaci* AUS2 were identified, from a wide range of hosts including cotton, cucurbits, cassava, sweet potato and *Physalis* sp. weeds, collected between 1994 – 2014).

B. tabaci AUS1, previously known as the Australian native whitefly, was identified only in historical samples collected between 1995 – 1997) from Queensland and New South Wales. Three haplotypes of a genetically distinct population which forms a sister clade to AUS1 and AUS2, was collected from cassava grown on Mer Island, Torres Strait Islands, during a pest and disease survey conducted by Murray Sharman.

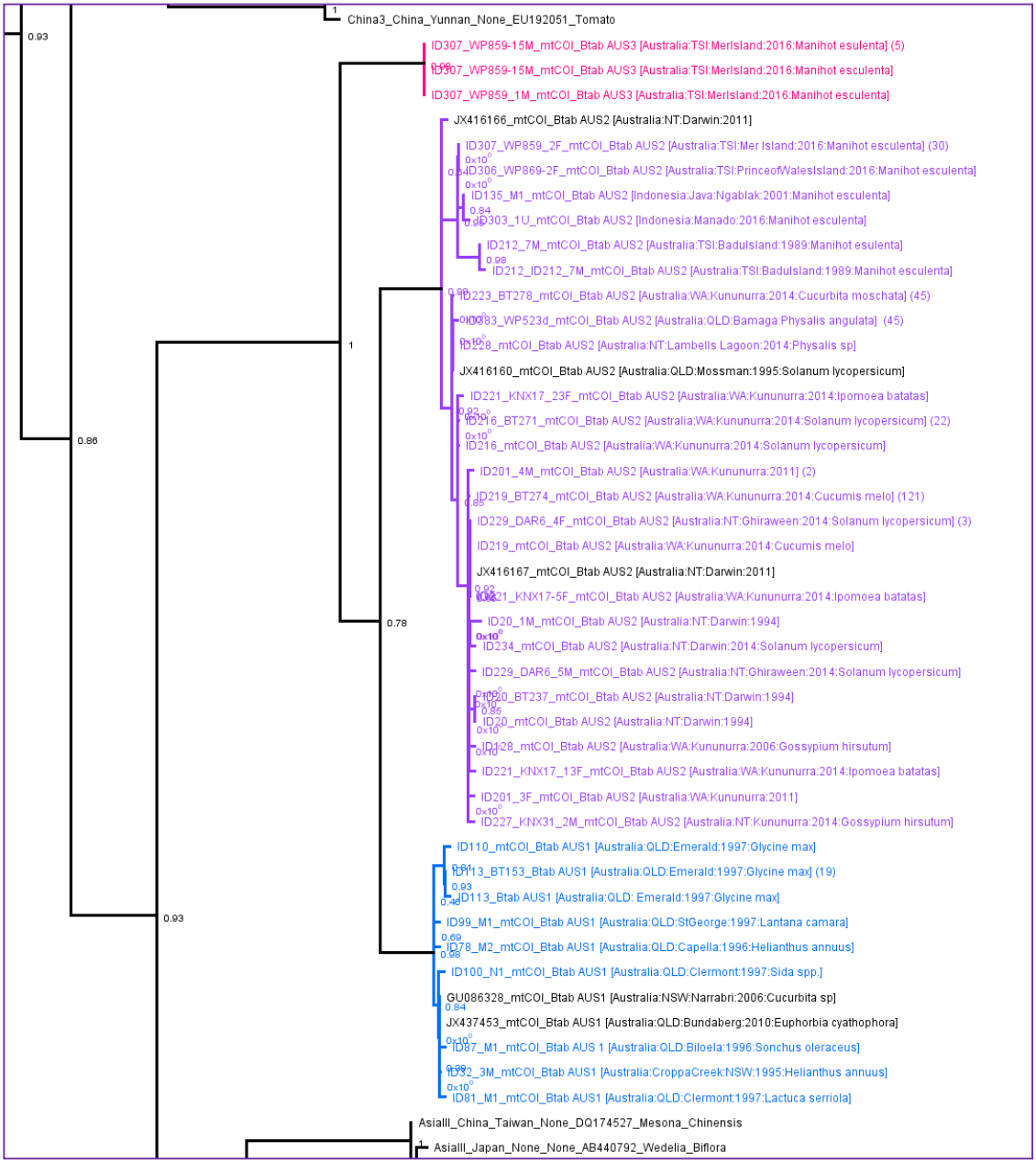


Figure 6: Inset of Bayesian phylogeny of global Bemisia tabaci species complex, constructed from mitochondrial cytochrome oxidase gene sequences (657 bp). This inset focuses on a subset of the Bemisia tabaci AUS1, AUS2 and newly identified AUS3 clades. All haplotypes in these AUS clades originate from Australia (coloured in pink, purple, and blue, numbers in parenthesis at the end of names indicate the total number of that haplotype that have been sequenced). Twenty-six distinct haplotypes of AUS2 were collected from a range of host plants in northern Australia (QLD, NT, WA), the Torres Strait Islands and also Indonesia. Nine distinct haplotypes of AUS1 were identified from historical samples, no current samples of AUS1 were found. Three haplotypes of a new clade, now named AUS3, were identified from the Torres Strait Islands only.

mtCOI barcoding of Bemisia tabaci MEAM1 in Australia

Some diversity of *B. tabaci* MEAM1 was identified across Australia, from between 1993 – 2017. A single haplotype was predominantly found, with 1098 individuals (98% of the MEAM1 specimens sequenced) having this exact haplotype sequence (coloured blue), see Figure 7. Genetically distinct haplotypes were identified, differing by 1-4 bp over the sequenced region were identified mostly from the mid 1990's. This result could indicate multiple introductions of *B. tabaci* MEAM1 into Australia, however more comprehensive analysis of NGS data is required to answer this question.

Figure 7: Inset of Bayesian phylogeny of global Bemisia tabaci species complex, constructed from mitochondrial cytochrome oxidase gene sequences (657 bp). This inset focuses on a subset of the Bemisia tabaci Middle-East Asia Minor 1 clade (formerly known as B biotype). Twenty-three distinct haplotypes of MEAM1 have been identified from collections made in Australia (1994-2014), from a wide range of host plants. In total, 98% of the samples sequenced were a single haplotype of MEAM1, depicted in blue.

Nuclear gene sequencing PrP8

The nuclear gene region (PrP8) of representative whitefly samples were sequenced and analysed, see Figure 8. The PrP8 gene region was shown to contain ~8 single nuclear polymorphisms (SNPs) differences between AUSI and AUSII, as well as ~20 SNPs between MEAM1 and the two Australian native species, allowing for further genetic differentiation between the Australian native species.

Phylogenetic analysis of these data sets is ongoing, however preliminary work has demonstrated that there is general concordance between these data and mtCOI data, and the phylogenies created with both data sets. This gives further support that the genetic groupings identified using the mtCOI barcoding method are consistent, and it will be useful to combine these genetic data with both biological and ecological data to infer cryptic species groupings within Australian *Bemisia tabaci* species.

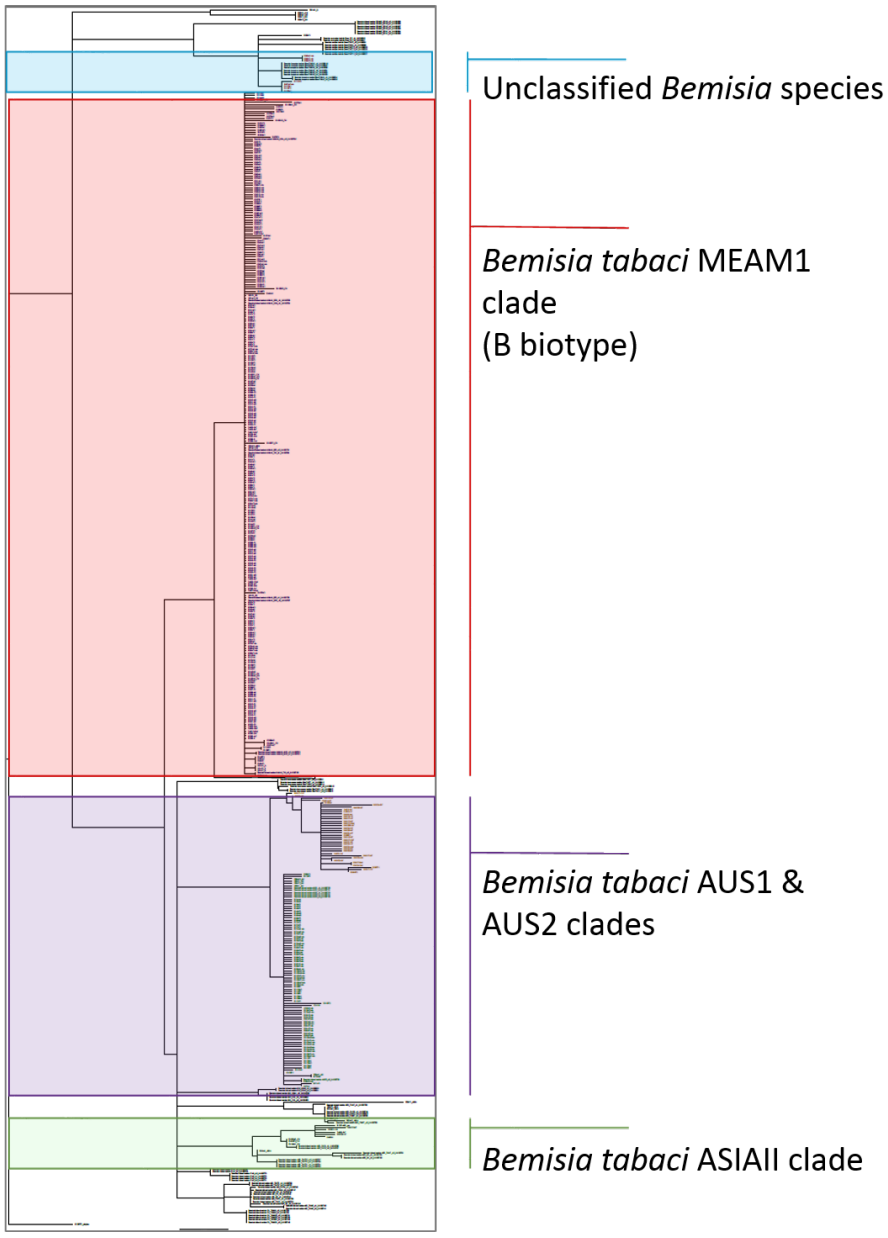


Figure 8: Bayesian phylogeny of global Bemisia tabaci species complex, constructed from nuclear pre-mRNA processing factor 8 (PrP8) sequences. The major clades that Australian whiteflies collected during this project belong to are highlighted in colour.

Endosymbiont sequencing

In order to effectively analyse the endosymbiont presence in the whitefly samples that were analysed, we optimized a multiplexed PCR approach to increase efficiency and decrease costs associated with analysing these specimens. This work is ongoing and is yielding useful data for a publication relating to the spatio-temporal comparisons of whiteflies, endosymbionts and viruses in Australia from 1993-present.

Six endosymbiont species were identified in *B. tabaci* species. The primary endosymbiont *Porteria* sp., and five secondary endosymbionts, *Arsenophonus* sp., *Wolbachia* sp., *Cardinium* sp., *Hamiltonella* sp. and *Rickettsia* sp. are present within *B. tabaci*. The endosymbiont primers used to screen whiteflies target different regions of the endosymbiont genome, with the *Porteria*, *Rickettsia*, *Hamiltonella* and *Cardinium* primers binding within a region of 16S rRNA, *Arsenophenus* within the 23S rRNA and *Wolbachia* primers within an outer surface protein (WSP).

Over 120 individuals were screened for the six endosymbiont species. No individual whitefly contained all six endosymbionts, with the majority containing 2-3 species. It was expected that all individuals would contain *Porteria* sp. as this is known to be the primary endosymbiont, interestingly, although *Porteria* sp. was identified in the majority of samples, not all samples showed evidence of *Porteria* sp. This could be due to a divergent *Porteria* sp. which is not identified with these primers or a different species acting as a primary endosymbiont. Trends of the secondary endosymbionts were seen within the *B. tabaci* species. For example, all *B. tabaci* AsiaII species contain Arsenphonous sp. with the majority also containing *Wolbachia* sp. (78%). The majority of AUSI individuals also contained *Arsenophonus* sp. (95%), however, only a few contained *Wolbachia* sp. (26%). AUSII had very few individuals containing *Arsenophonus* sp. (19%), instead a large number contained *Cardinium* sp. (59%), which (with one AsiaII exception) were the only *B. tabaci* species identified which contain *Cardinium* sp. The majority of MEAM1 species contained *Rickettsia* sp. (72%) and *Hamiltonella* sp. (78%), with only two individuals containing *Arsenophonus* sp., and none containing the other two species. Individuals which fall into an unclassified mtCOI clade, (this clade also contains individuals collected in Uganda), had *Porteria* sp., *Wolbachia* sp., and *Arsenophonus* sp. identified.

Portiera

The 16S rRNA sequences of *Portiera* sequenced from twenty distinct whitefly populations are presented in a phylogenetic analysis in Figure 9. This tree clearly shows that *Portiera* from *Bemisia tabaci* group closely together (blue), whereas *Portiera* from other *Bemisia* species e.g. *Bemisia subdecipiens* (red) and *B. euphorbiae* (purple) are genetically distinct and form their own separate clades.

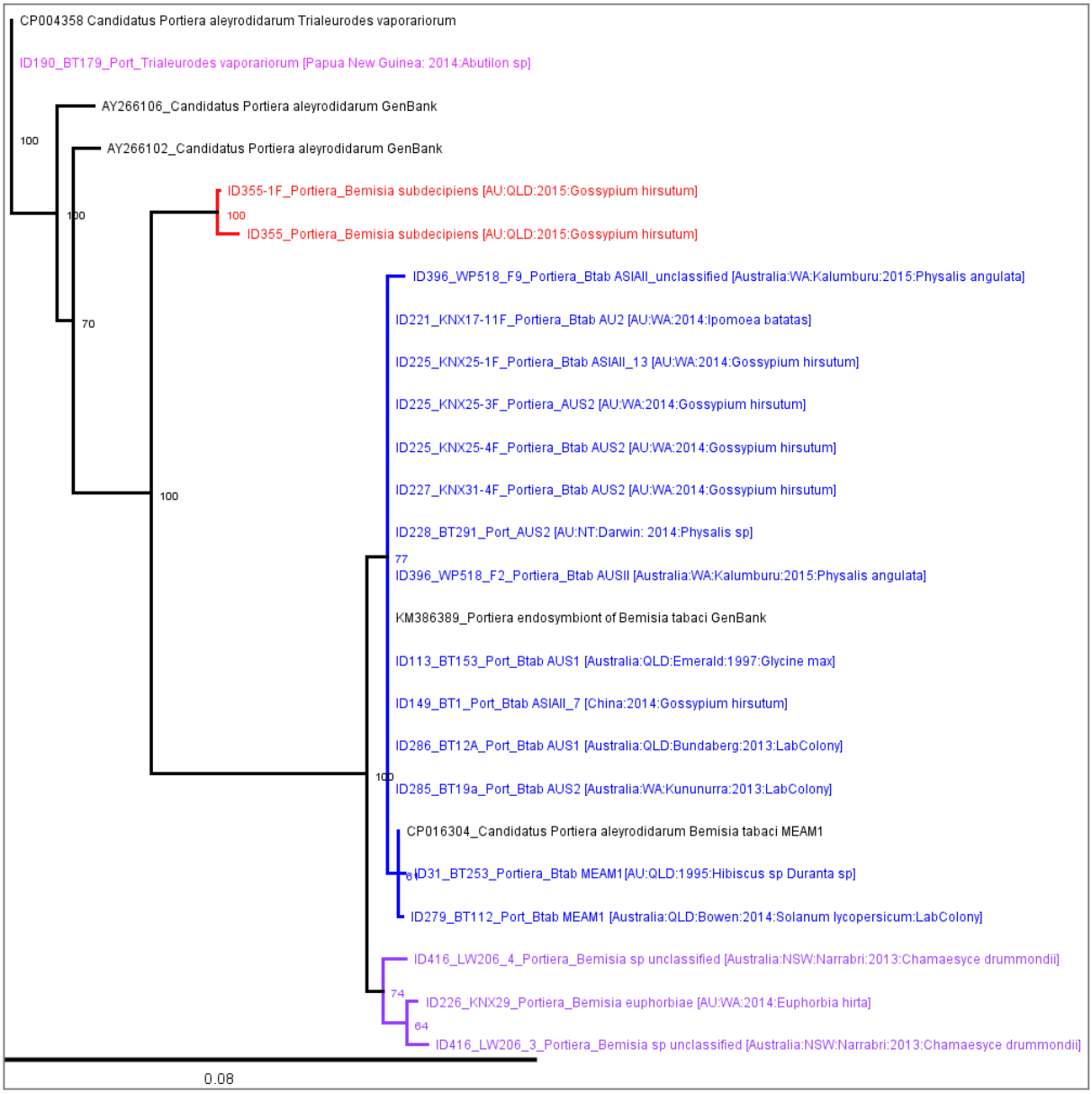


Figure 9: Neighbour-joining phylogeny of 16S rRNA gene sequences of Portiera species identified from whiteflies collected in Australia, compared to publicly available sequences on GenBank.

Arsenophonus

The 23S rRNA sequences of *Arsenophonus* sequenced from fifteen distinct whitefly populations (MEAM1, AUS1, ASIAII_13, AUS2 and *Bemisia atriplex*) are presented in a phylogenetic analysis in Figure 10. This tree clearly shows that *Arsenophonus* from *Bemisia tabaci* group together, and *Arsenophonus* sp. from other *Bemisia* species e.g. *Bemisia atriplex* and *B. euphorbiae* (red) are genetically distinct and form their clade.

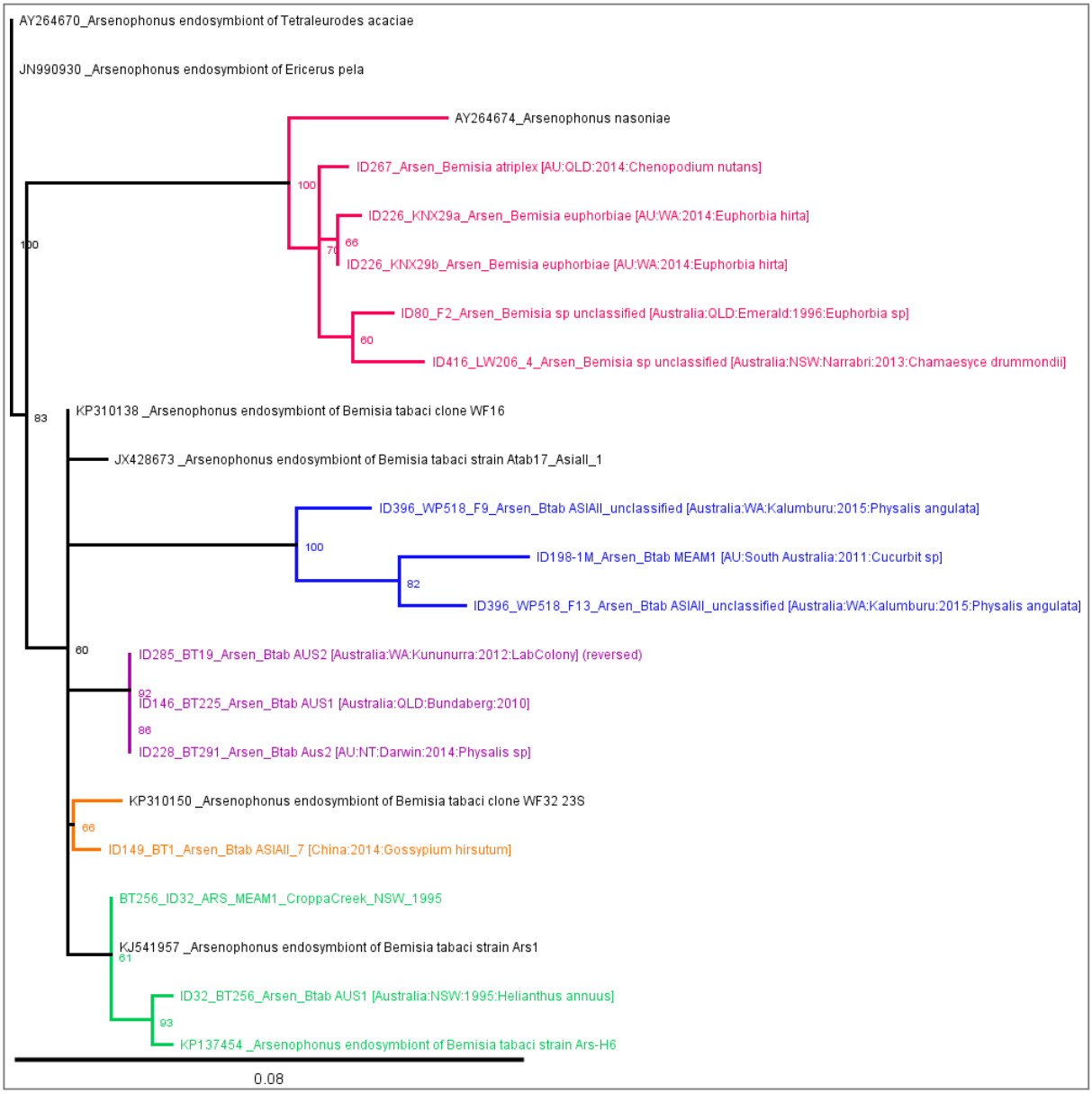


Figure 10: Neighbour-joining phylogeny of 23S rRNA gene sequences of Arsenophonus species identified from whiteflies collected in Australia, compared to publicly available sequences on GenBank.

Cardinium

The 16S rRNA sequences of *Cardinium* sequenced from fourteen distinct populations of *Bemisia tabaci* AUS2 are presented in a phylogenetic analysis in Figure 11. This tree shows that *Cardinium* from *Bemisia tabaci* group together in a large clade (purple), that interestingly also contains *Cardinium* sp. sequenced from the exotic and distantly related ash whitefly (*Siphoninus phillyreae*).

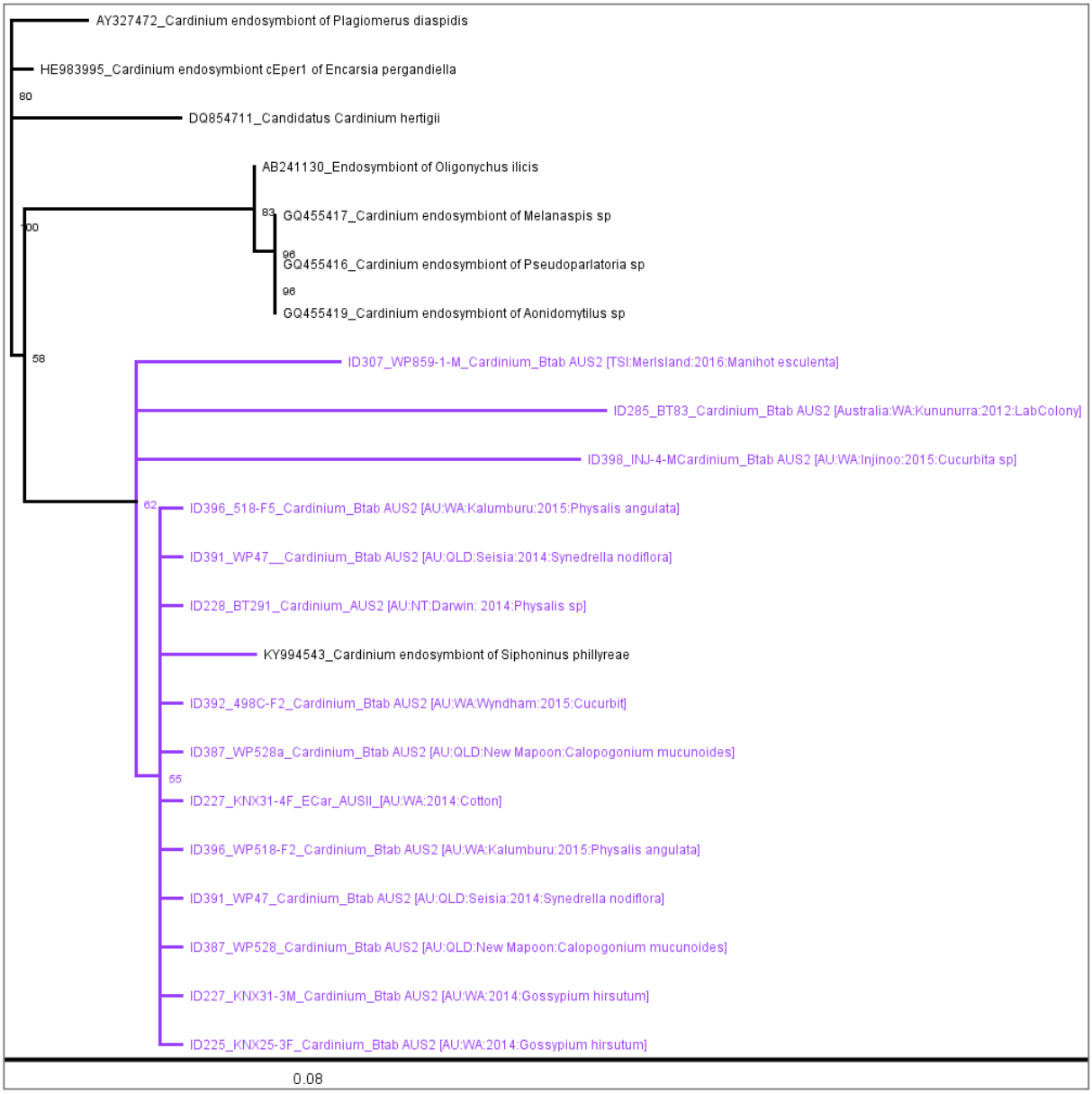


Figure 11: Neighbour-joining phylogeny of 16S rRNA gene sequences of Cardinium species identified from whiteflies collected in Australia, compared to publicly available sequences on GenBank.

Hamiltonella

The 16S rRNA sequences of *Hamiltonella* sequenced from sixteen distinct populations of *Bemisia tabaci* MEAM1 are presented in a phylogenetic analysis in Figure 12. This tree shows that *Hamiltonella* from *Bemisia tabaci* MEAM1 group together in a large clade (blue), that also contains *Cardinium* sp. sequenced from *B. tabaci* MEAM1 from other countries, *B. tabaci* MED (Q biotype) and interesting from Indian grain aphid (*Sitobion miscanthi*).

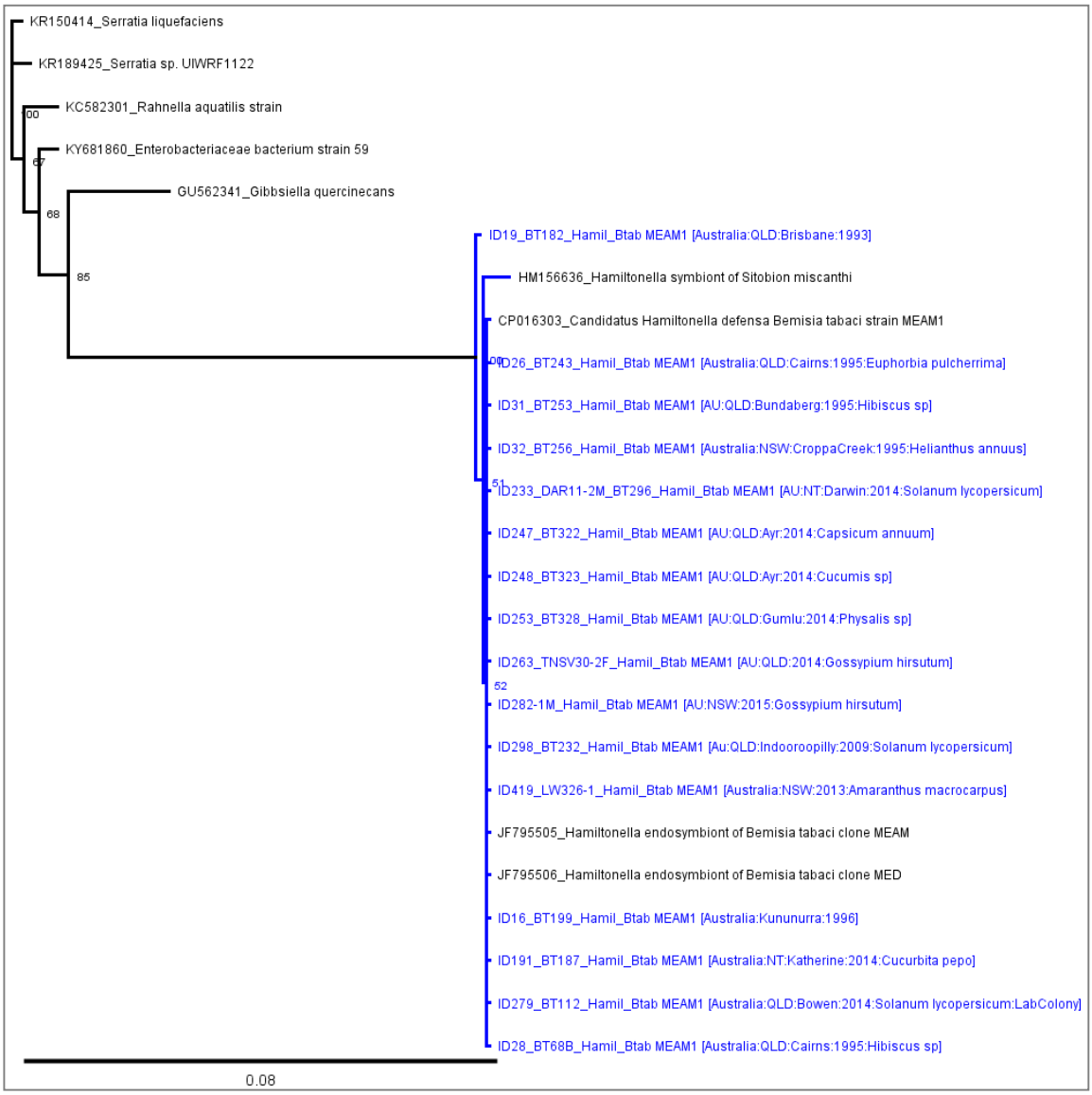


Figure 12: Neighbour-joining phylogeny of 16S rRNA gene sequences of Hamiltonella species identified from whiteflies collected in Australia, compared to publicly available sequences on GenBank.

Rickettsia

The 16S rRNA sequences of *Rickettsia* sequenced from fifteen distinct populations of *Bemisia tabaci* MEAM1 are presented in a phylogenetic analysis in Figure 13. This tree shows that *Rickettsia* from *Bemisia tabaci* MEAM1 group together in a large clade (blue), that also contains *Rickettsia* sp. sequenced from *B. tabaci* MEAM1 from another country.

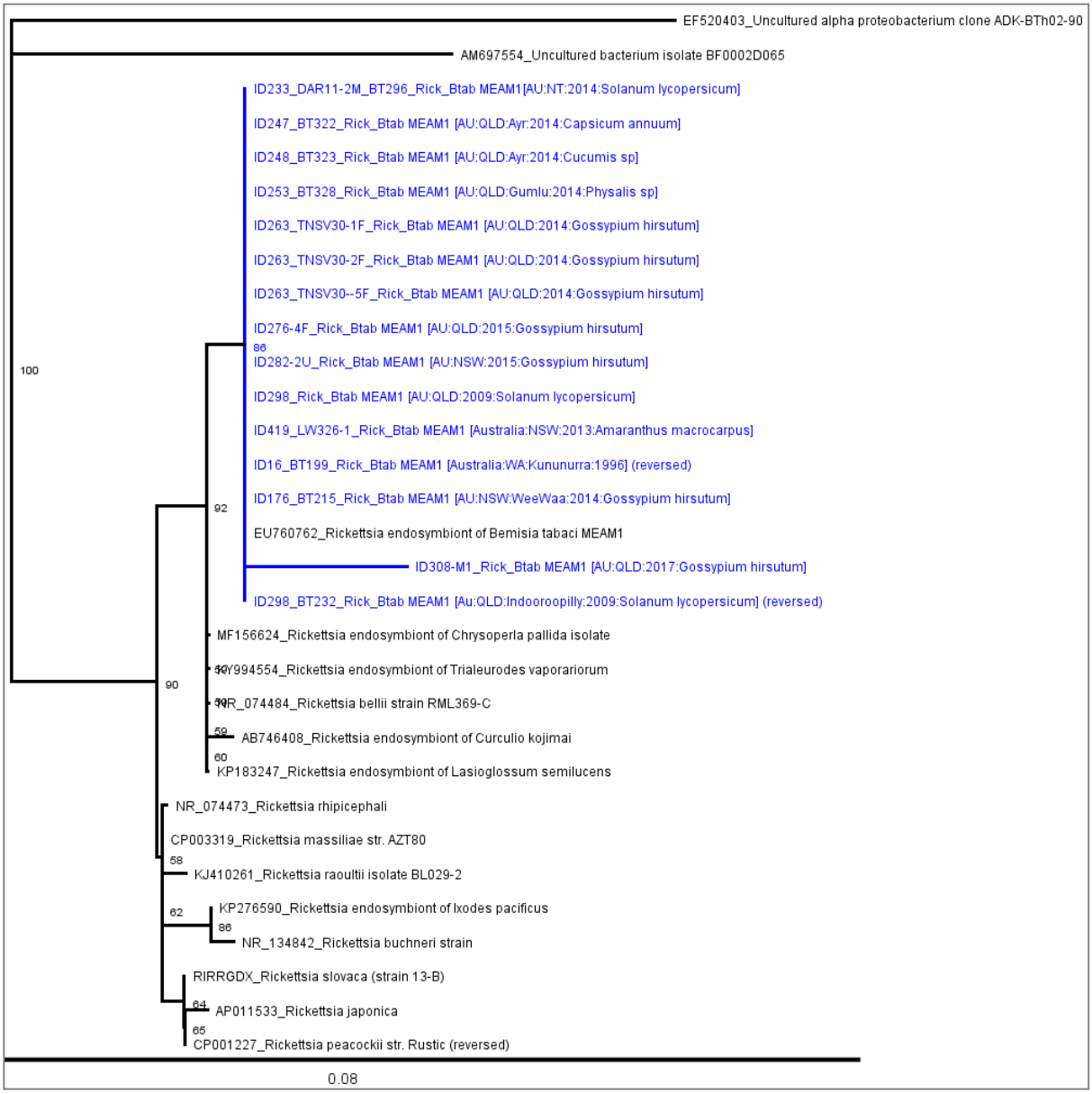


Figure 13: Neighbour-joining phylogeny of 16S rRNA gene sequences of Rickettsia species identified from whiteflies collected in Australia, compared to publicly available sequences on GenBank.

Wolbachia

The *Wolbachia* outer surface protein (wsp) gene sequences of *Wolbachia* sequenced from nineteen distinct populations of *Bemisia tabaci* (AUS1, AUS2, AUS3, ASIAII_13, *Bemsia atriplex*, *Bemisia euphorbiae* and unclassified *Bemisia* sp.) are presented in a phylogenetic analysis in Figure 14. This tree shows that *Wolbachia* from *Bemisia tabaci* AUS1, AUS2, AUS3, ASIAII_13 and interestingly

also *Bemisia atriplex*, group together (blue). *Wolbachia* from *Bemisia euphorbiae* and unclassified *Bemisia* sp. are genetically quite distinct and form their own separate clade (pink).

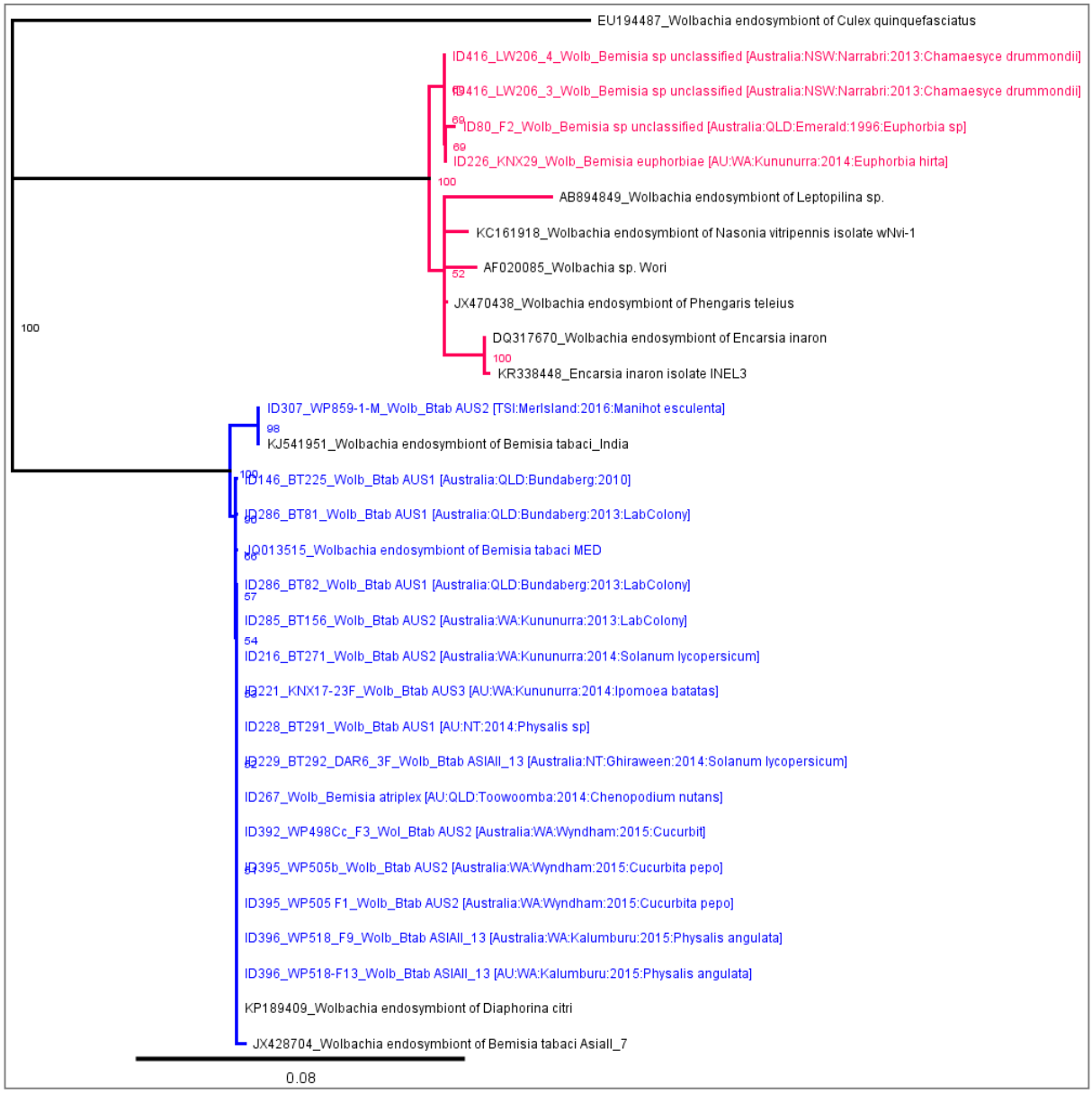


Figure 14: Neighbour-joining phylogeny of Wolbachia surface protein (wsp) gene sequences of Wolbachia species identified from whiteflies collected in Australia, compared to publicly available sequences on GenBank.

Next-generation sequencing of whiteflies, parasitoids and the whitefly primary obligate bacterial endosymbiont Candidatus Portiera aleyrodidarum

We have generated Illumina HiSeq next-generation sequencing (NGS) data for 23 single male whiteflies, representing the known genetic diversity of *Bemisia tabaci,* and inclusive of additional *Bemisia* sp. from Australia, see Table 11. The purpose of this research is to develop high quality data to document the whitefly and endosymbiont diversity in Australian whiteflies. We are currently preparing a journal article that will compare the divergence of these lineages to examine the evolutionary histories of host and obligate symbiont, and determine if these data assist in the reconstruction of molecular phylogenies of cryptic species within *B. tabaci*.

Two next-generation sequencing strategies were developed to characterise whitefly, virus and endosymbiont diversity: (1) Illumina HiSeq metagenomic sequencing, and (2) Illumina Miseq metagenomic sequencing. Each strategy was developed to answer different research questions, and this is expanded below.

(1) Illumina Hiseq metagenomic sequencing was completed for 23 whitefly individuals (and one whitefly parasitoid species) (Table 1). This was done to characterise the diversity of whitefly and their associated primary endosymbionts from Australia. Fourteen genetically distinct *B. tabaci* from Australia were sequenced, collected from various crop and weed hosts and locations, from as early as 1993 (first record of the invasive *B. tabaci* MEAM1 B biotype in Australia, provided by Dr Paul De Barro). We have assembled draft complete mitochondrial genomes (mitogenomes) for 10 of these whitefly species (~15 000 bp, see Figure 15 for an example), with the remaining mitogenomes in progress. We are also currently assembling the complete genomes for the obligate primary endosymbiont *Candidatus* Portiera aleyrodidarum (~360 kbp) for each sample (see Figure 16 for example). We are preparing a journal article that compares the divergence of these lineages to examine the evolutionary histories of host and obligate symbiont, and determine if these data assist in the reconstruction of molecular phylogenies of cryptic species within *B. tabaci*.

Our analyses to date have demonstrated that the general structure of divergence is shared between the *B. tabaci* mitogenome and Portiera genome data, indicative of co-divergence (Figure 17 and Figure 18).

For *B. tabaci* MEAM1 Aus (1995) compared to *B. tabaci* MEAM1 Aus (2014), complete mitochondrial genomes share 99.97% similarity but there is ~1.9% divergence between Portiera genomes. This level of divergence, in comparison with Portiera of other cryptic species of *B. tabaci*, could be indicative of more than one introduction of *B. tabaci* MEAM1 into Australia.

In addition, we have sequenced and provided support in assembling the complete mitochondrial genomes of four representative samples from the *B. tabaci* Mediterranean species group (MED; formerly known as the Q biotype), in a collaborative project with Prof's John Colvin, Sue Seal and PhD student Sona Vyskocilova (NRI, University of Greenwich). This research forms part of a high-impact research article comparing the biological characteristics (mating success, host-use) and genetic characteristics (complete mitochondrial genomes and nuclear markers) for genetically distinct populations of *B. tabaci* MED, thereby providing rigorous evidence for the existence of cryptic species within *B. tabaci*. This work was recently presented at the SciPlant 2017 conference (see Appendix 1).

These sample that we have sequenced include other *Bemisia* sp. from Australia collected from weeds in cotton growing regions, and the first specimen of *Bemisia tabaci* MEAM1 (B biotype) collected in Brisbane in 1993 (provided by Paul De Barro, CSIRO). Importantly, we have also sequenced four genetically distinct exotic populations of *Bemisia tabaci* Mediterranean species (MED) (formerly known as Q biotype), in a collaborative project with NRI, University of Greenwich. In Australia, we do not have *B. tabaci* MED but it is recognized as an "extreme" risk to the Australian cotton industry (as

outlined in the most recent Industry Biosecurity Plan), as both a pest and virus vector, particularly due to its ability to acquire such high levels of insecticide resistance. The purpose of this collaboration is to contribute to a better understanding of the genetic diversity of this exotic pest.

We are continuing analysis of these data to assemble and annotate the complete mitochondrial genomes (~15 000 bp), and also the genomes of the obligate primary endosymbiont *Candidatus* Portiera aleyrodidarum (~360 kbp). An annotated draft mitogenome is presented in Figure 15, and annotated draft Portiera genome in Figure 16.

In collaboration with Dr Jamie Hopkinson (DAQ1403), we have identified a previously unrecorded parasitic wasp species in Australia. This putative *Eretmocerus* species was first identified by Dr Hopkinson, parasitising a population of a species of whitefly that is genetically very closely related to *B. tabaci*, *Bemisia atriplex* in Toowoomba. In conjunction with Dr Wee Tek Tay (CSIRO) we have assembled a draft mitogenome for this wasp, and Dr Andrew Polaszek of the British Natural History Museum has completed morphological and taxonomic studies, together these data will published in a peer-reviewed journal article. This research will contribute to an improved understanding of whitefly parasitoids in Australia.

Table 11: Details of whitefly samples that have been sequenced using next-generation sequencing approaches in this project. Samples are diverse, and include representatives of Australian whiteflies AUS1, AUS2, ASIAII_13, and MEAM1. Additional sequencing has been done on newly identified Bemisia species collected in Australia, related to Bemisia tabaci. Overseas samples of representative isolates of B. tabaci MED (formerly known as Q biotype) have also been sequenced.

ID	Species	Host	City	State	Country	Date of Collection	Collector
19	*Bemisia tabaci* MEAM1	Unknown	Brisbane	Queensland	Australia	1/07/1993	Unknown
32	*Bemisia tabaci* MEAM1	*Helianthus annuus* (sunflower)	Croppa Creek	New South Wales	Australia	12/12/1995	Unknown
113	*Bemisia tabaci* AUS1	*Glycine max* (soybean)	Emerald	Queensland	Australia	24/02/1997	Paul J De Barro
192	*Bemisia tabaci* MEAM1	*Cucurbita moschata* (pumpkin)	Katherine	Northern Territory	Australia	16/07/2014	Di Renfree
221	*Bemisia tabaci* AUS2	*Ipomoea batatas* (sweet potato)	Kununurra	Western Australia	Australia	10/09/2014	S. L. van Brunschot
226	*Bemisia euphorbiae*	*Euphorbia hirta*	Kununurra	Western Australia	Australia	12/09/2014	S. L. van Brunschot
227	*Bemisia tabaci* ASIAII_13	*Gossypium hirsutum* (cotton)	Kununurra	Western Australia	Australia	12/09/2014	S. L. van Brunschot
229a	*Bemisia tabaci* ASIAII_13	*Solanum lycopersicum* (tomato)	Ghiraween	Northern Territory	Australia	13/09/2014	S. L. van Brunschot
229b	*Bemisia tabaci* ASIAII_13	*Solanum lycopersicum* (tomato)	Ghiraween	Northern Territory	Australia	13/09/2014	S. L. van Brunschot
232a	*Bemisia tabaci* ASIAII_13	*Abelmoschus esculentus* (okra)	Humpty Doo	Northern Territory	Australia	13/09/2014	S. L. van Brunschot
232b	*Bemisia tabaci* MEAM1	*Abelmoschus esculentus* (okra)	Humpty Doo	Northern Territory	Australia	13/09/2014	S. L. van Brunschot
233	*Bemisia tabaci* ASIAII_13	*Solanum lycopersicum* (tomato)	Darwin	Northern Territory	Australia	13/09/2014	S. L. van Brunschot
307	*Bemisia tabaci* AUS2	*Manihot esculenta* (cassava)	Mer Island	Torres Strait Islands	Australia	15/06/2016	M. Sharman
308	*Bemisia tabaci* AUS2	*Manihot esculenta* (cassava)	Saibai Island	Torres Strait Islands	Australia	16/06/2016	M. Sharman
309	*Eretmocerus aleyrodesii*	*Bemisia atriplex*	Toowoomba	Queensland	Australia	1/11/2016	Jamie Hopkinson
345	*Bemisia tabaci* MED Spain	Squash (spp unknown)	Malaga	Andalusia	Spain	2013	Jesus Navas-Castillo
346	*Bemisia tabaci* MED Israel	*Gossypium hirsutum* (cotton)	N/A	N/A	Israel	2013	Shai Morin
347	*Bemisia tabaci* MED Uganda F7	*Abelmoschus esculentus* (okra)	Gayaza	Wakiso	Uganda	2014	Habibu Mugerwa
348	*Bemisia tabaci* MED Sudan S	*Gossypium hirsutum* (cotton)	N/A	N/A	Sudan	1978	Ciba-Geigy
355	*Bemisia subdecipiens*	*Melaleuca* sp. (tea tree)	Corinda	Queensland	Australia	08/02/2015	Murray Sharman
400	*Bemisia tabaci* MEAM1	*Gossypium hirsutum* (cotton)	Croppa Creek	New South Wales	Australia	1/02/2015	M. Stone
415	*Bemisia atriplex*	*Chenopdodium giganteum* (tree spinach)	Toowoomba	Queensland	Australia	1/10/2016	Jamie Hopkinson
416	*Bemisia* sp. unclassified	*Chamaesyce drummondii* (fireweed)	Narrabri	New South Wales	Australia	28/02/2013	Tanya Smith

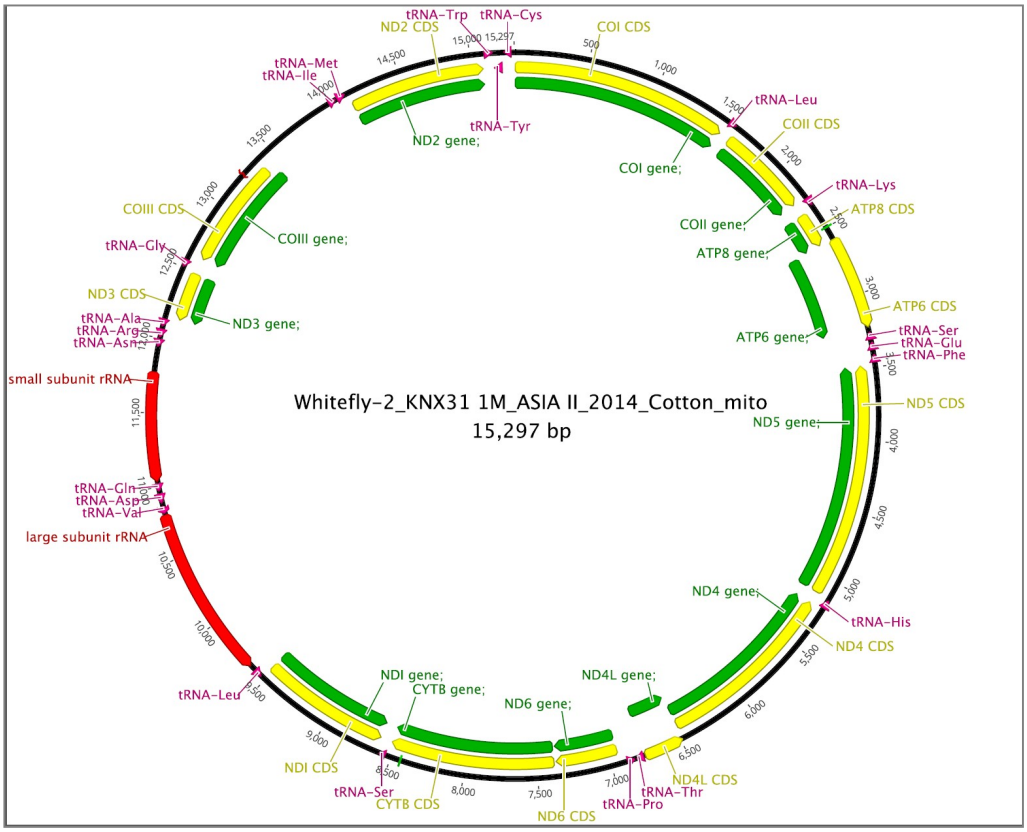


Figure 15: Annotated full mitochondrial genome of a representative sample of Bemisia tabaci ASIAII_13, collected from a cotton crop grown in Kununurra 2014, assembled from NGS sequence data.

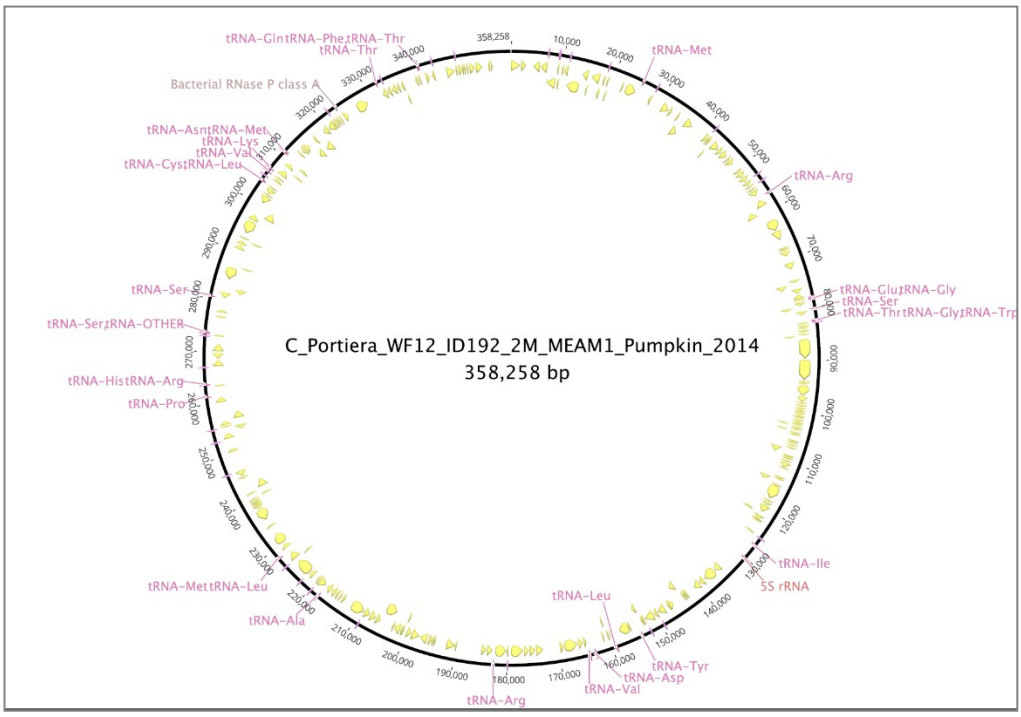


Figure 16: Annotated complete genome of Candidatus Portiera aleyrodidarum of a Bemisia tabaci MEAM, collected from pumpkins frown in Katherine (Northern Territory) in 2014.

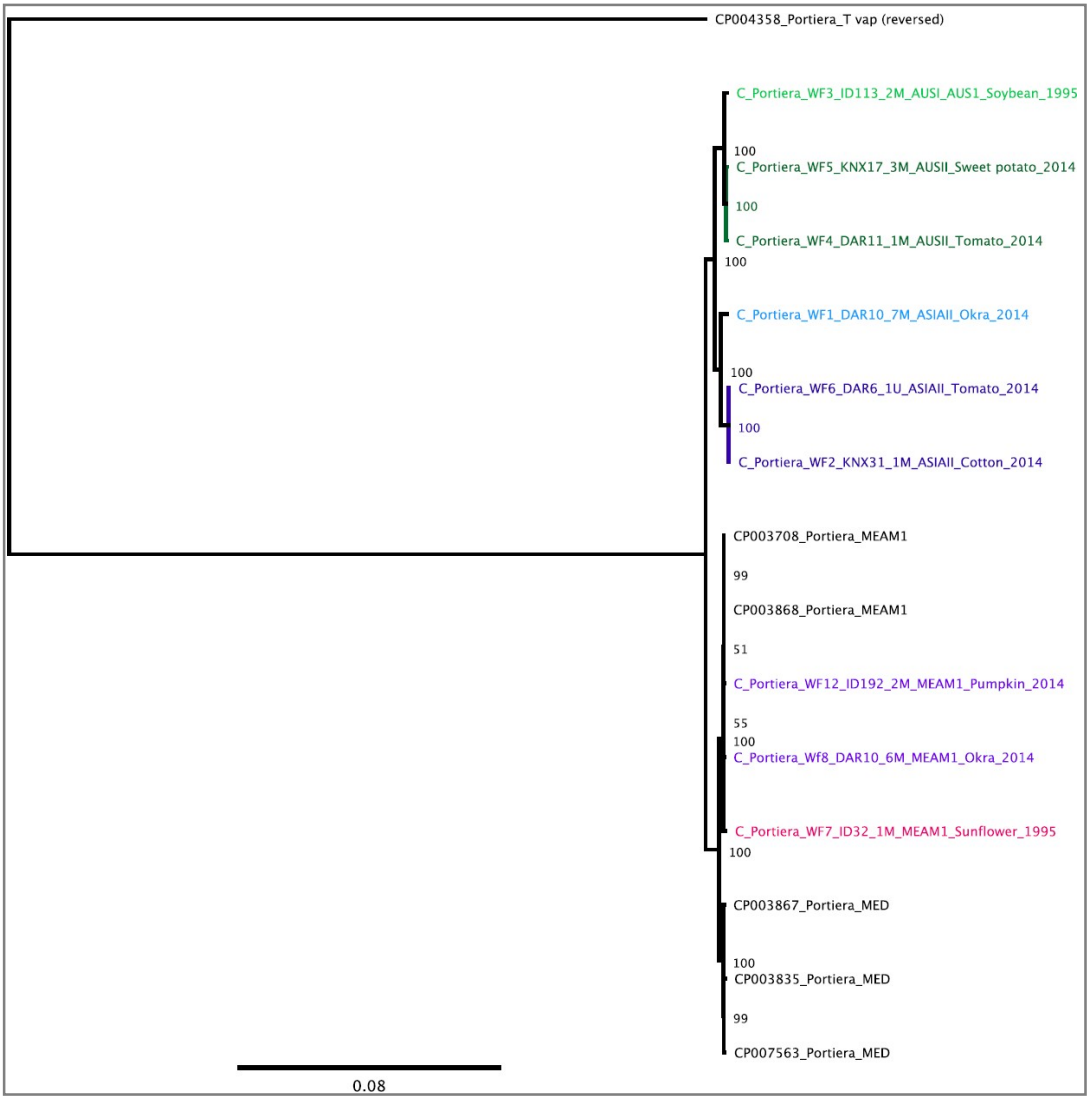


Figure 17: Preliminary phylogenetic analysis comparing the relatedness of complete genomes of representative Portiera (primary obligate symbiont) sequenced from Australian whiteflies, compared to publicly available data on GenBank.

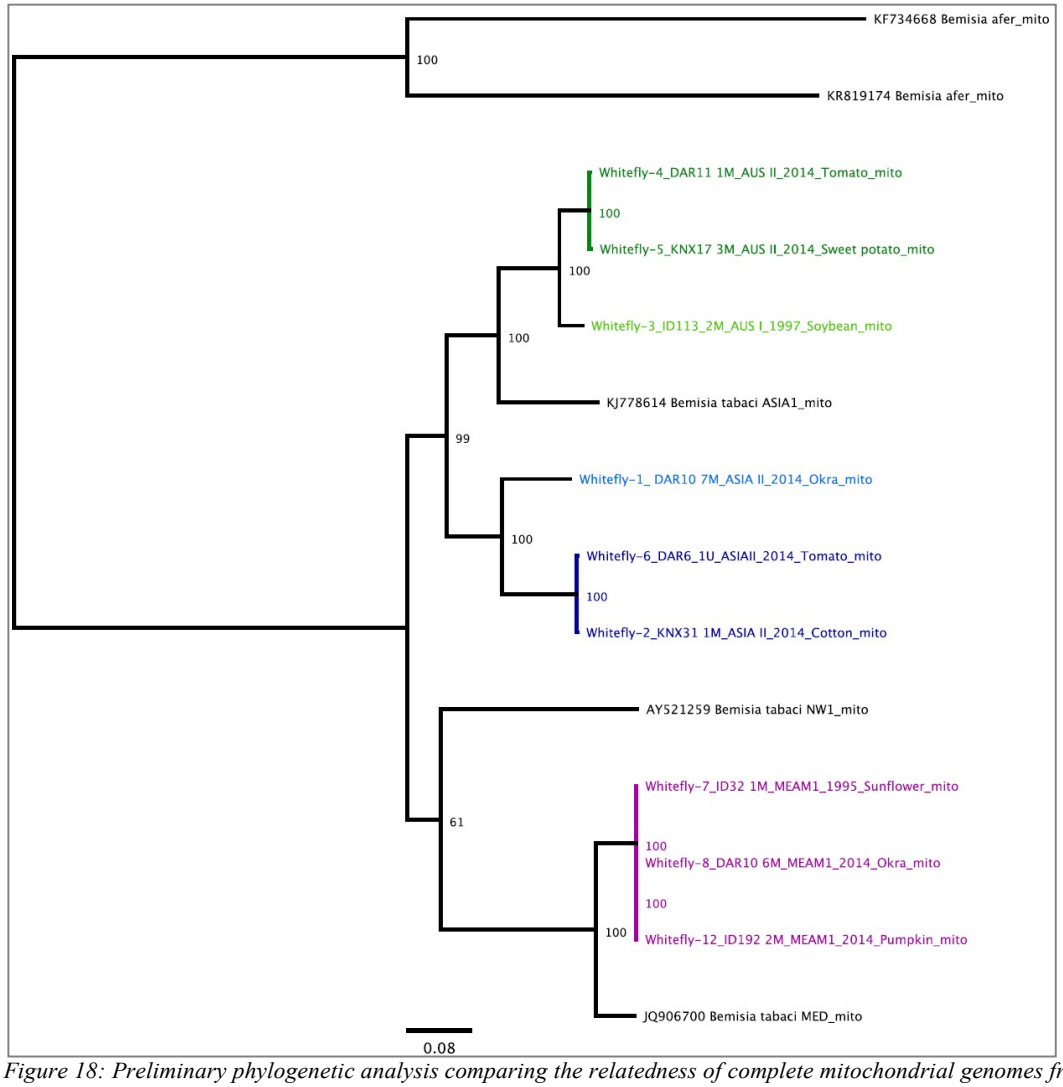


Figure 18: Preliminary phylogenetic analysis comparing the relatedness of complete mitochondrial genomes from representative Australian whitefly, compared to publicly available data on GenBank.

Viral metagenomics: next generation sequencing reveals begomovirus diversity in both plant and insect vectors

Illumina Miseq metagenomic sequencing has been completed and analysed for 22 separate plant and/or whitefly samples. The purpose of this work was to develop a new approach for the surveillance of whitefly and whitefly-transmitted viruses. Given that we currently have no whitefly-transmitted viruses infecting cotton present in Australia, we decided to develop these molecular surveillance tools using a very closely related whitefly-transmitted virus that infects sweet potato in Australia. The overarching aim was to determine ways to improve the speed, reliability and cost-effectiveness of virus surveillance, to better protect the Australian cotton industry from a potential incursion of CLCuV.

Each sample was enriched for circular DNA prior to NGS, and the resulting data was used to reconstruct full virus genome sequences, and in the case of whitefly samples – partial whitefly mitochondrial genomes, partial endosymbiont genomes and full virus genomes. This work has been prepared as a manuscript that is currently under internal revision, for submission to the journal Viruses, entitled “Viral

metagenomics: a tool for virus discovery and analysis of diversity in plant and insect vectors", see Appendix 2.

Using this approach, we detected and identified, for the first time in Australia, whitefly-transmitted viruses infecting field-grown sweet potato. Specifically, we identified mixed infections of *Sweet potato leaf curl virus* (SPLCV), *Sweet potato leaf curl Spain virus* and *Sweet potato leaf curl China virus* from asymptomatic sweet potato samples from two locations in Australia. The high depth of coverage (103-14501 times) for these genomes, in addition to the long sequence read length strategy used, gives confidence to the genome assemblies generated in our work.

We also concurrently identified and sequence-characterized (i) SPLCV, and (ii) the identity of the whitefly vector *Bemisia tabaci* ASIAII, in field-collected whitefly collected from the infected sweet potato crop.

In addition, we identified partial genome sequences of *Sweet potato badnavirus B* (*Caulimoviridae, dsDNA virus)* in a single sample, which may indicate the presence of active infections of this virus, or alternatively an integrated form.

Taken together, this is the first report of (i) whitefly-transmitted viruses infecting sweet potato in Australia, (ii) putative badnavirus-infection of sweet potato, (iii) the use of vector–enabled metagenomics (VEM) to survey begomovirus diversity directly from whiteflies collected across crop plants from a single geographic region.

Sweet potato is a vegetatively propagated, so it is likely that this is the predominant mode of transmission of these viruses in Australia. Furthermore, it is likely that the route of introduction of these viruses was via the importation of asymptomatic infected planting material.

This study demonstrates the power of NGS for describing highly variable, mixed infections of virus species *in planta*. It also demonstrates the use of VEM for describing the circulating viral community in a given region, whilst concurrently identifying the vector identity, giving insight into plant virus and vector diversity in a given region. This information facilitates plant virus surveillance and management of viral diseases, and these new diagnostic tools and expertise significantly increase the preparedness of Australian cotton in the event of an introduction of cotton-infecting whitefly-transmitted viruses.

4.3 Objective 3: *Research trip to the University of Greenwich to perform CLCuV transmission experiments utilizing several species of Bemisia tabaci and Australian cotton cultivars.*

Fundamental research was done to investigate of the capacity of Australian whiteflies to transmit viruses of the CLCuD complex. It has been shown in recent studies that the virus-vector relationships between *B. tabaci* and the virus species they are able to acquire and/or transmit can be species-specific (Gottlieb et al., 2010, Jiang et al., 2004). In order to evaluate the risk that CLCuD presents to Australian cotton, it is important to understand the ability of the whitefly populations present in Australia to transmit cotton leaf curl viruses.

In order to answer the research question "Do Australian whiteflies transmit cotton leaf curl viruses", a research collaboration needed to be established, because working with live exotic quarantine insect pests and viruses in Australia was not feasible due to quarantine restrictions. Experiments were

conducted over a 6-month period in the quarantine facilities of the University of Greenwich (United Kingdom) in collaboration with Professor John Colvin. Dr Warwick Stiller (CSIRO) kindly arranged for seed of cotton cultivars Sicot 71 & Sicot 75 under an arranged MTA to be sent from Australia to the University of Greenwich for experimental use.

A preliminary set of experiments was run three months into the research trip, an important first step in developing the hands-on techniques and methodologies for this work. Once completed, a final set of experiments was run two months later using these refined methodologies (see Figure 19). With appropriate replications, we have tested the transmission of *Cotton leaf curl Bangalore virus* and cognate satellites (i) *Ludwigia leaf distortion betasatellite* and *Guar leaf curl alphasatellite*, by the following whitefly cryptic species:

1. *B. tabaci* AUSI (Australia)
2. *B. tabaci* AUSII (Australia)
3. *B. tabaci* MEAM1 (Australia; B biotype)
4. *B. tabaci* ASIAI (India)

Briefly, virus-infected source plants were generated by feeding *B. tabaci* ASIAII_7 (India) on infected plants and transferring to healthy Sicot 71, and allowing the striking symptoms to develop (Figure 20 & Figure 21). Source plants were treated after 72 hours with a physical insecticide Breaker (BioTech). All transmission experiments were assessed using symptomology, molecular indexing methods and DNA sequencing.

To summarise results from our work, we found that all four cryptic species of *B. tabaci* tested successfully transmitted *Cotton leaf curl Bangalore virus* to Sicot 71 with similar efficiency, with transmission confirmed by symptoms and molecular tests (PCR). Initial PCR-based screening using a degenerate primers (Deng et al., 1994) produced non-specific results, instead amplifying cotton gDNA (data not shown). Molecular indexing methods were then changed to use a primer set that exhibited no non-specific amplification (Idris et al., 2014). Molecular indexing results showed that 1/6 of control transmissions for the *B. tabaci* AUSII (where no virus was used) gave a positive PCR signal (DNA-A only), however no symptoms were observed. Therefore, the molecular indexing results need to be further confirmed to ensure the validity of this dataset and interpretations of this work.

Molecular indexing results for betasatellites and alphasatellites showed that instances of the transmission of these molecules along with the cognate DNA-A were rare by all *B. tabaci* cryptic species. This result was reflected in the comparatively low severity of symptoms seen in test plants, compared with source plants.

Figure 19: Whitefly cultures generated for virus transmission experiments, at The University of Greenwich (UK).

Figure 20: Transmission of Cotton leaf curl Bangalore virus to Sicot 71 by whitefly-mediated transmission (by B. tabaci ASIAII_7). When compared to healthy (left), the symptoms of severe stunting, leaf cupping and chlorosis are obvious.

Figure 21: Transmission of Cotton leaf curl Bangalore virus to Sicot 75 by whitefly-mediated transmission (by B. tabaci ASIAII_7). When compared to healthy (left), the symptoms of severe stunting, reduced flowering, leaf cupping and leaf distortion are obvious.

4.4 Objective 4: *Targeted field collections of insecticide resistant and susceptible whiteflies, to generate cultures for experimental manipulation.*

The aims of this research objective were to generate cultures of field populations of whitefly, to supplement existing ethanol collected specimens, for the molecular interrogation of their endosymbiont metacommunity populations. Specifically, we wanted to compare and contrast the endosymbiont communities of the known Australian cryptic species of *B. tabaci*, with specific examples of populations with known insecticide resistance status. The overarching questions for this section of the project were: (1) Which bacterial endosymbionts are correlated with each specific cryptic species of *B. tabaci* in Australia? and (2) Are particular bacterial endosymbionts associated with insecticide resistance in field collected populations.

To answer these questions, we collaborated with Dr Jamie Hopkinson (SAQ1403) who kindly provided us with four whitefly colonies with quantified insecticide resistance levels. In addition, in addition we examined three other colonies that had been challenged with various classes of insecticides (data not shown).

We then developed molecular techniques to efficiently interrogate the endosymbiont metaccomunities of these colonies, and also representative whiteflies from our ethanol preserved whitefly collection (see Table 12).

To briefly summarise the results, we observed consistent patterns in the presence of endosymbiotic bacteria with respect to *B. tabaci* cryptic species. These patterns are in-keeping with the knowledge that

secondary endosymbionts are rarely fixed in populations, whereas the primary endosymbiont *Portiera* is (essential for metabolism, always retained). In *B. tabaci* AUS1, secondary endosymbionts *Wolbachia* and *Arsenophonus* were consistently observed. For *B. tabaci* AUS2, secondary endosymbionts *Wolbachia* and *Cardinium* were consistently observed, and occasionally *Arsenophonus*. For *B. tabaci* MEAM1, secondary endosymbionts *Rickettsia* and *Hamiltonella* were consistently observed. For *B. tabaci* ASIAII, secondary endosymbionts *Wolbachia* and *Arsenophonus* were consistently observed. For the other *Bemisia* sp. that were examined, *Wolbachia* and *Arsenophonus* were consistently observed, with a one population LW206 (*Bemisia* sp unclassified [Australia:NSW:Narrabri:2013:Chamaesyce drummondii]) interestingly containing *Rickettsia*. Some samples contained unusual positive amplifications for some endosymbionts, and these are marked in yellow and will be further confirmed using DNA sequencing, before publication.

No significant trends were observed in relation to insecticide resistance status and endosymbiont presence. A recent study reported the observation of the presence of *Rickettsia* infections was associated with increases susceptibility to a wide range of insecticides (Kontsedalov et al., 2008), including pyriproxifens. Our results clearly demonstrate the presence of Rickettsia in all of the resistant populations we examined, contradicting the results reported in Kontsedalov (2008).

The results presented here form part of a publication that is under preparation documenting the diversity of whitefly and their endosymbionts in Australia.

Table 12: Endsoymbiont metacommunity profiling of whitefly samples, both preserved and live colonies.

Whitefly sample name	***B. tabaci species (COI)***	***Porteria* sp.**	***Rickettsia* sp.**	***Wolbachia* sp.**	***Hamiltonella* sp.**	***Arsenophonus* sp.**	***Cardinium* sp.**	***Insecticide resistance profile S = susceptible R = resistant***
BT81	AUS1	+		+				S
BT256	AUS1	+			+	+		
ID82_M1	AUSI	+		+		+		
ID82_M2	AUSI	+				+		
ID82_F1	AUSI	+				+		
ID82_F2	AUSI	+				+		
ID78_M1	AUSI	+				+		
ID78_M2	AUSI	+				+		
ID78_F1	AUSI	+				+		
ID78_F2	AUSI	+				+		
ID86_F1	AUSI	+				+		
ID86_F2	AUSI	+				+		
ID86_F3	AUSI	+		+		+		
ID85_F1	AUSI					+		
ID85_F2	AUSI			+		+		
ID83_U	AUSI	+				+		
ID83_F1	AUSI	+				+		
ID83_F2	AUSI	+				+		
ID75_F1	AUSI	+		+				
ID106_F1	AUSI	+		+		+		
ID106_F2	AUSI	+				+		
BT19	AUS2	+		+		+		S
WP518-F5	AUS2	+					+	

Whitefly sample name	***B. tabaci species (COI)***	***Porteria* sp.**	***Rickettsia* sp.**	***Wolbachia* sp.**	***Hamiltonella* sp.**	***Arsenophonus* sp.**	***Cardinium* sp.**	***Insecticide resistance profile S = susceptible R = resistant***
BT271	AUS2	+		+		+		
ID20-1 M	AUS2			+				
BT291	AUS2	+		+		+	+	
WP528	AUS2						+	
WP528	AUS2	+				+	+	
WP532	AUS2						+	
WP47a	AUS2	+	+			+	+	
WP47b	AUS2	+				+	+	
WP505	AUS2	+		+		+		
498C-M1	AUS2	+						
498C-F2	AUS2	+					+	
498C-F3	AUS2	+		+				
WP505-F1	AUS2	+		+				
WP500-F9	AUS2	+		+				
WP518-F2	AUS2	+					+	
DAR11-1 M	AUS2	+					+	
KNX25-3 F	AUS2	+					+	
KNX25-4 F	AUS2	+						
KNX31-2 M	AUS2	+		+				
KNX31-3 M	AUS2	+					+	
KNX31-4 F	AUS2	+					+	
KNX25-2 F	AUS2	+						
WP859-1-M	AUS2			+			+	
KNX17-11-F	AUS2	+		+				
WP859-15-M	AUS2						+	

Whitefly sample name	***B. tabaci species (COI)***	***Porteria* sp.**	***Rickettsia* sp.**	***Wolbachia* sp.**	***Hamiltonella* sp.**	***Arsenophonus* sp.**	***Cardinium* sp.**	***Insecticide resistance profile S = susceptible R = resistant***
INJ-4-M	AUS2						+	
KNX17-23-F	AUS2	+		+			+	
INJ-2-F	AUS2	+						
ID298	MEAM1	+	+		+			
BT253	MEAM1	+			+			
ID192-2 M	MEAM1		+					
DAR11-2 M	MEAM1	+	+		+			
ID282-2 U	MEAM1		+					
ID276-4 F	MEAM1	+	+					
ID282-1 M	MEAM1	+	+		+			R
TNSV30-1-F	MEAM1	+	+		+			
ID288-5-F	MEAM1	+	+		+			
ID198-1-M	MEAM1	+				+		
TNSV30-2-F	MEAM1	+	+		+			
TNSV30-5-F	MEAM1	+	+		+			
WP869-1-F	MEAM1	+	+			+		
WP859-5-F	MEAM1	+	+		+			
ID308_M1	MEAM1	+	+		+			
ID308_F1	MEAM1	+	+		+			
ID308_F2	MEAM1	+	+		+			
ID320_M1	MEAM1	+	+					
ID320_F1	MEAM1	+	+		+			
ID316_M1	MEAM1	+	+		+			
ID316_M2	MEAM1	+			+			
ID316_F1	MEAM1	+	+		+			

Whitefly sample name	***B. tabaci species (COI)***	***Porteria* sp.**	***Rickettsia* sp.**	***Wolbachia* sp.**	***Hamiltonella* sp.**	***Arsenophonus* sp.**	***Cardinium* sp.**	***Insecticide resistance profile S = susceptible R = resistant***
ID316_F2	MEAM1	+	+		+			
ID44_M1	MEAM1	+			+			
ID44_M2	MEAM1	+	+		+			
ID44_F1	MEAM1	+	+		+			
ID44_F2	MEAM1	+			+			
ID93_M1	MEAM1	+	+		+			
ID93_F1	MEAM1	+			+			
ID93_F2	MEAM1	+			+			
ID75_F2	MEAM1	+			+			
ID320_M2	MEAM1		+					
ID320_F2	MEAM1	+	+		+			
ID93_M2	MEAM1	+			+			
BT257	MEAM1	+			+			
BT213	MEAM1		+					
BT243	MEAM1	+			+			
BT323	MEAM1	+	+		+			
BT328	MEAM1	+	+		+			
BT322	MEAM1	+	+		+			
ID298	MEAM1	+	+		+			
ID277	MEAM1	+	+					R
ID280	MEAM1	+			+			R
BT112	MEAM1	+	+		+			S
ID281	MEAM1	+			+			R
DAR6-1	ASIAII	+		+		+		
DAR6-2 F	ASIAII	+		+		+		

Whitefly sample name	***B. tabaci species (COI)***	***Porteria* sp.**	***Rickettsia* sp.**	***Wolbachia* sp.**	***Hamiltonella* sp.**	***Arsenophonus* sp.**	***Cardinium* sp.**	***Insecticide resistance profile S = susceptible R = resistant***
DAR6-3 F	ASIAII	+		+		+		
KNX17-5 F	ASIAII	+				+		
KNX25-1 F	ASIAII	+				+		
BT300	ASIAII	+		+		+		
BT301	ASIAII	+		+		+		
BT296	ASIAII	+		+		+		
BT298	ASIAII	+		+		+	+	
WP518-F13	ASIAII	+		+		+		
WP518-F9	ASIAII	+		+		+		
WP518-F10	ASIAII	+		+		+		
ID267	*Bemisia atriplex*	+		+		+		
KNX29a	*Bemisia euphorbiae*	+		+		+		
KNX29b	*Bemisia euphorbiae*	+		+		+		
LW206-3	*Bemisia* sp.	+	+	+		+		
LW206-4	*Bemisia* sp.	+	+	+		+		
ID87-M2	*Bemisia* sp. (UG-like)	+		+		+		
ID80-M1	*Bemisia* sp. (UG-like)	+		+		+		
ID80-F1	*Bemisia* sp. (UG-like)	+		+		+		
ID80-F2	*Bemisia* sp. (UG-like)	+		+		+		
ID80-U	*Bemisia* sp. (UG-like)	+		+		+		
SVBID273	*Bemisia subdecipiens*	+				+		
SVBID273	*Bemisia subdecipiens*	+				+		

4.5 Objective 5: Understand the positive and negative influences endosymbiont populations have on the biology of their whitefly hosts.

In order to adequately address this objective, preliminary work was done to collect and identify *B. tabaci* populations that had naturally differing endosymbiont profiles (but were genetically very similar). This task in itself proved near impossible, due to the fluctuating presence of secondary endosymbionts. Therefore, we decided (in agreement with Susan Maas) to alter the work done for this objective.

Firstly, biological experiments were done to assess the mating success among two genetically distinct, allopatric populations of native Australian whiteflies *B. tabaci* AUS1 (from Bundaberg) and AUSII (from Kununurra). The overarching aim of this research was to test if, under laboratory conditions, if individuals from each population were able to recognise each other as suitable mating partners and produce female progeny (*B. tabaci* are haplodiploid). This work therefore was aimed at resolving the species status of native whiteflies in Australia.

Preliminary experiments were done to develop methodologies and generate preliminary data to guide future experiments. Reciprocal crossing experiments were performed using *B. tabaci* AUS1 (from Bundaberg) and AUSII (from Kununurra) native whiteflies, to assess their life history traits, fecundity and mating compatibility, have been completed. These were done to trial experimental conditions, protocols and equipment. The comprehensive experiments with large numbers of replicates (>10 per treatment, 4-8 treatments run over a 70 day time period) will be performed from January-March 2015.

Notable results from these experiments included: (1) Female progeny were produced in both of the reciprocal crosses, (2) The sex ratio (female:male) were significantly different between the control (0.6:0.4) and reciprocal crosses (0.3:0.7), and (3) The number of successful intromissions remained similar between controls and crosses. The skewed sex-ratio in the reciprocal crosses indicated two possibilities: (1) The pairs took longer to mate, resulting in less female progeny, and/or (2) The success of fertilization was decreased in reciprocal crosses, resulting in less females. The data derived from these preliminary experiments were then used to design further experiments to

This work was then comprehensively expanded on by PhD student Ms Wanaporn Wongnikong, and she has prepared a manuscript of this work for publication (see Appendix 3). Ms Wongnikong is supervised by Prof Gimme Walter, Sharon van Brunschot and James Hereward and her research is aligned with UQ1305. Ms Wongnikong began her whitefly research in August 2016, and has been developing methods to record and document the courtship and mating behaviour, fecundity and life span of *B. tabaci* AUSI and AUSII, which harbour different endosymbionts. This research therefore extends on the preliminary results obtained for this milestone. Ms Wongnikong has successfully passed her mid-term review milestone with The University of Queensland, for her research toward "An investigation of the biology, ecology and species limits within the cryptic species complex of whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Australia". Her work over the past 16 months has been to accurately document, record and assess the courtship and mating behaviour, fecundity and life span of the Australian whiteflies AUSI and AUSII, which harbour different endosymbionts.

Ms Wongnikong is now building further on this research by examining virus transmission (using the introduced whitefly-transmitted virus *Tomato yellow leaf curl virus*) by these species and *B. tabaci* MEAM1 in the coming 6 months, and the results of this work will be reported to CRDC and ultimately published in a peer-reviewed journal article.

4.6 Objective 6: Develop new molecular methods for detecting viruses in the CLCuD complex, and species of *Bemisia tabaci*.

This research objective details the development of improved molecular methods for the detection of cotton leaf curl viruses. An additional request in the final 6 months of the project was to develop a draft National Diagnostic Protocol for whitefly in Australia, and this is provided in this report in its entirety (see Appendix 4).

Introduction

The begomovirus genus (Geminiviridae) contains over 300 species of viruses, all of which are vectored by the whitefly, *Bemisia tabaci* a cryptic species complex. Begomovirus genomes consist of either one (monopartite) or two (bipartite) circular components. These viruses infect a wide range of dicotyledonous plants, with many considered important agricultural pathogens.

A number of begomovirus species have been found associated with cotton leaf curl disease (CLCuD) of cotton. Of these, five monopartite core species have been identified from outbreaks of disease *Cotton leaf curl Gezira virus* (CLCuGeV), *Cotton leaf curl Kokhran virus* (CLCuKoV), *Cotton leaf curl Alabad virus* (CLCuAV), *Cotton leaf curl Bangalore virus* (CLCuBaV) and *Cotton leaf curl Multan virus* (CLCuMuV). These core species have also been found infecting a number of non-cotton hosts, including other Malavaceae such as Hibiscus and Okra, and Solanaceae Tomato and Eggplant, although not all species have been found in each host (Table 13). The core species are also found in different geographical regions, with CLCuKoV and CLCuAV found in Pakistan and India, and Bangalore only in India. Whereas, CLCuMuV and CLCuGeV have a much wider geographical distribution with isolates of CLCuGeV found in Africa, the Middle East and Pakistan, and CLCuMuV found in India, Pakistan and South East Asia Table 14).

No CLCuD causing begomoviruses have been identified in Australia, but one has been identified as close as the Philippines (Table 14). In 2014 a CLCuD causing virus CLCuMuV was identified in Hibiscus (*Hibiscus rosa-sinensis*) in the Philippines (Dolores, 2014).

CLCuD symptoms in cotton include vein thickening, upward leaf curling and enations (leaf like growths) on the underside of the leaves. Epidemics can result in large economic losses, with an epidemic in Pakistan resulting in an estimated five billion (USD) loss over five years. CLCuKoV and CLCuMuV are associated with a number of epidemics in India and Pakistan.

A bipartite begomovirus *Cotton leaf crumple virus* (CLCrV), which is found only in the Americas, also infects cotton, however symptoms of infection are considered mild (Briddon & Markham, 2000). Further Geminiviruses, *chickpea chlorotic dwarf virus*, *Tomato leaf curl New Delhi virus*, *Tomato leaf curl Bangalore virus*, *Okra enation leaf curl virus* and *Papaya leaf curl virus* have also been found associated with the core species in cotton.

Methods and Results

Identifying regions of similarity in core species for primer design
In order to design primers which will detect all core CLCuD causing species, all available sequences were downloaded from GenBank and aligned in Geneious, starting at the nonanucleotide motif. A single region which shared similarities across all core species was identified and primers, CLCu5 were designed within this region to amplify a 355 bp product. Alignments suggest these primers should detect all five core species. Although these primers were designed to detect the core species of CLCuD causing viruses, there is the potential that these primers will also detect other begomoviruses.

Screening with previously designed primers

Accore and BEG primers were being used as screening primers for Milestone 4.2, a further begomovirus primer set SPG1-2 was also identified. The primers SPG, BEG and CLCu5 primers span a section of three genes, the replication enhancer gene, transcriptional activator protein gene and the replication-association protein gene, with the SPG primer also spanning a fourth gene, C4 which is thought to counter host responses. The Accore primer pair binds in a different region of the genome, spanning the end of the movement protein gene and a section of the coat protein gene. Primer sequences and PCR protocol information is provided in Table 15.

All PCRs were carried out using MyTaq™ Hot Start DNA Polymerase (Bioline) in a 30ul reaction (Table 16). After the corresponding PCR protocol (Table 16), 10ul was loaded on a gel, with the remaining 20ul of positive samples, cleaned and sent for sequencing.

As multiple species of begomoviruses cause CLCuD, and these viruses have also been found in a number of different hosts, a variety of plant species extracts were used to test the four primer sets. The plant extracts were from overseas material which was thought to be infected with a Begomovirus.

Internal plant control
Plant extracts had been extracted previously using the CTab method. To ensure the extraction was successful, and that no PCR inhibitors such as phenols were interfering with the PCR, plant DNA barcode primers were used which target a 355bp region of *ITS2*. These primers ITS2-S2F (Chen et al., 2010b) and ITS4 (White et al., 1990a) have been used previously to barcode cotton species (Ashfaq, 2013). Of the 70 plant extracts tested, 65 samples were positive for *ITS2*.

Specificity
The 65 plant extracts were from a variety of plant samples including okra, hibiscus, tomato and papaya as well as 24 cotton samples, from both cultivated and wild cotton (full list Table 17). All four primer sets were optimised and run on primer specific PCR protocols. PCR products were run on agarose gels, with clean single bands identified for all four primer sets. Positive samples were cleaned and sent to Macrogen Inc (Korea) for Sanger sequencing. NCBI BLASTn was used to identify the PCR fragment for each positive sample. Top Blastn hits for each sample are provided in Table 17.

Begomoviruses were detected in samples from 13 different hosts. As you can see in Table 17, not all samples amplify with all primers, with Accore and SPG identifying viruses in the least number of samples, 24 samples each, versus CLCu5 and BEG which identified 43 and 50 infected samples respectively. Isolates with top blast hits to the five core CLCuD causing viruses are shown in bold.

It should be noted, that although ClCu5 primer pair was designed for the five core species of CLCuD causing viruses, this primer set also detected a number of other begomoviruses including; Tomato yellow leaf curl virus, Tomato leaf curl viruses, Malvastrum yellow vein Changa Manga virus, Alternanthera yellow vein virus, Chili leaf curl virus and Papaya leaf curl virus.

In total 54 samples were found to be positive for Begomoviruses, with core CLCuD causing viruses detected in 18 cotton samples, and 12 non-cotton samples. Other begomoviruses which have been found in mixed infections with the core species such as PaLCV and tomato leaf curl New Delhi viruses were also identified. Three core species were detected in cotton CLCuMuV, CLCuGeV, ClCuKoV with core species also detected in a Chili sample (n=1), Hibiscus (n=4), Melon (n=1), Okra (n=2), a Parthenium relative (n=1), Sodom apple (n=2) and a Tomato (n=1). No CLCuBaV or CLCuAV were identified in any samples. While all primer sets detected at least one core species, all samples which were found to contain core species, were identified by either one or both of the primers ClCu5 and/or BEG.

In some instances, the top BLASTn hits of the different primer pairs returned different virus species. There is the possibility that these plants are infected with more than one begomovirus, with mixed begomoviruses infections identified previously. Sequence alignment and comparisons across shared regions further suggest that the majority are mixed infections.

Sequences of CLCu5 primer region of the core species, the CLCu5 isolates sequenced as part of this study, and the top BLASTn hits to those isolates were trimmed and aligned in Muscle. A neighbour joining HKY phylogenetic tree, bootstrap 1000, of these sequences was constructed in Geneious.

All isolates from CLCu5 which were identified as non-core begomovirus species; TYLCV, ChiLCV, AlYVV, ToLCTWV, PaLCuCNV, ToLCNDV cluster with reference sequences of those species,

separate to the core species (-

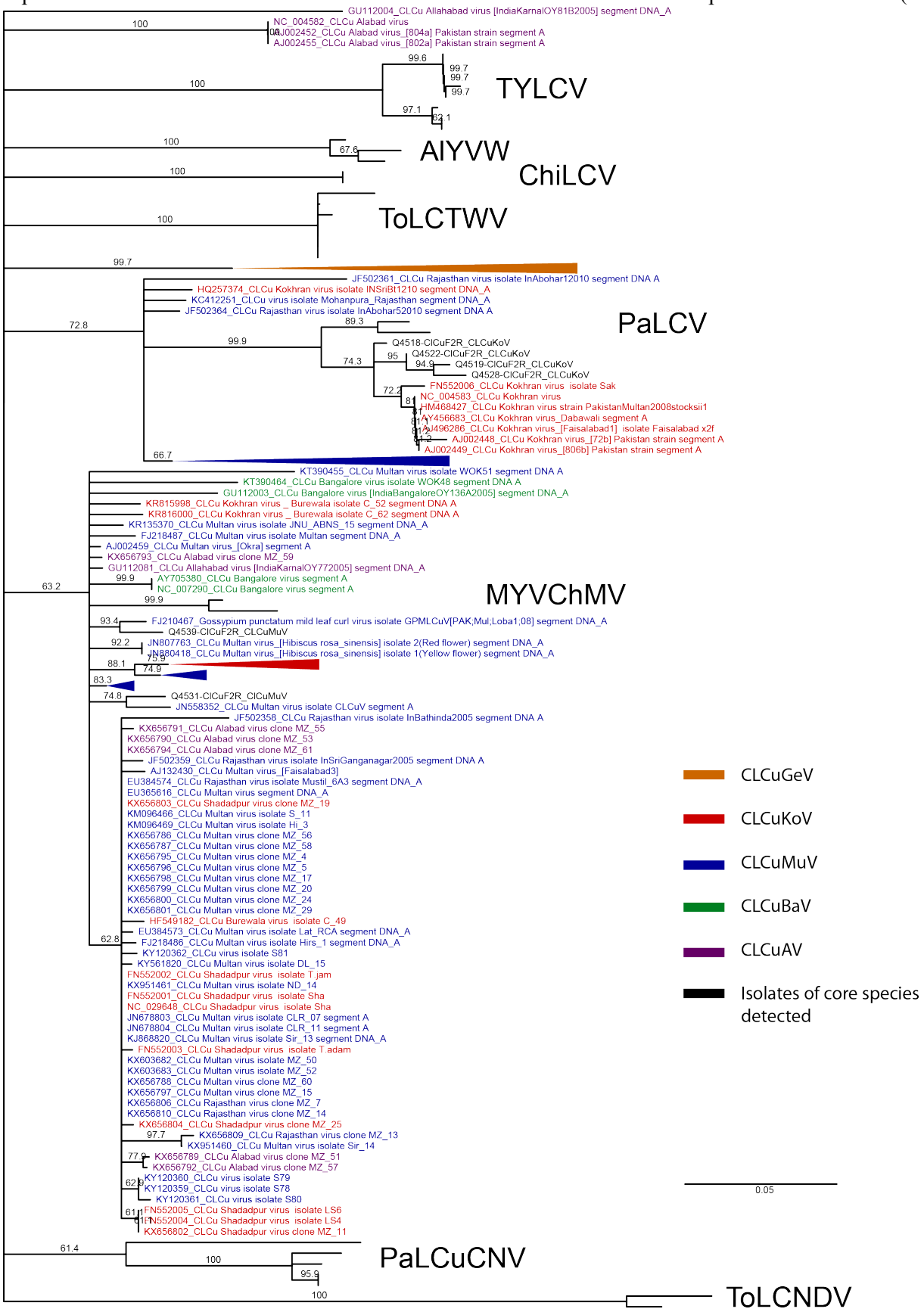


Figure 22). Interestingly the PaLCV sequences fall basal to one of the CLCuKoV groups. However,

the PaLCV sequences are still distinct to the CLCuKoV sequences, allowing for the differentiation of the sequences of the two species.

All Gezira sequences cluster into a single clade, along with the CLCuGeV isolates identified in this study (orange collapsed clade). The CLCuAV and CLCuBaV do not form strong clades. The CLCu regions of the other 4 species fall into more than 4 separate clades. This is not entirely surprising as even at the full genome level, CLCuMoV and CLCuKoV fall into a number of separate clades. The majority of sequences still fall within species specific clades, (Blue or red collapsed clades), however within some clades there are a small number of sequences which are a different species, for example two CLCuMuV sequences are found within a CLCuKoV clade (collapsed red clade). This is likely due to recombination between species of this section of the genome as Begomoviruses have high recombination rates. However, all the CLCuMuV and CLCuKoV isolates, identified in this study fall within clades with the corresponding species present as the majority (majority of the sequenced isolates

are found within the collapsed clades in -

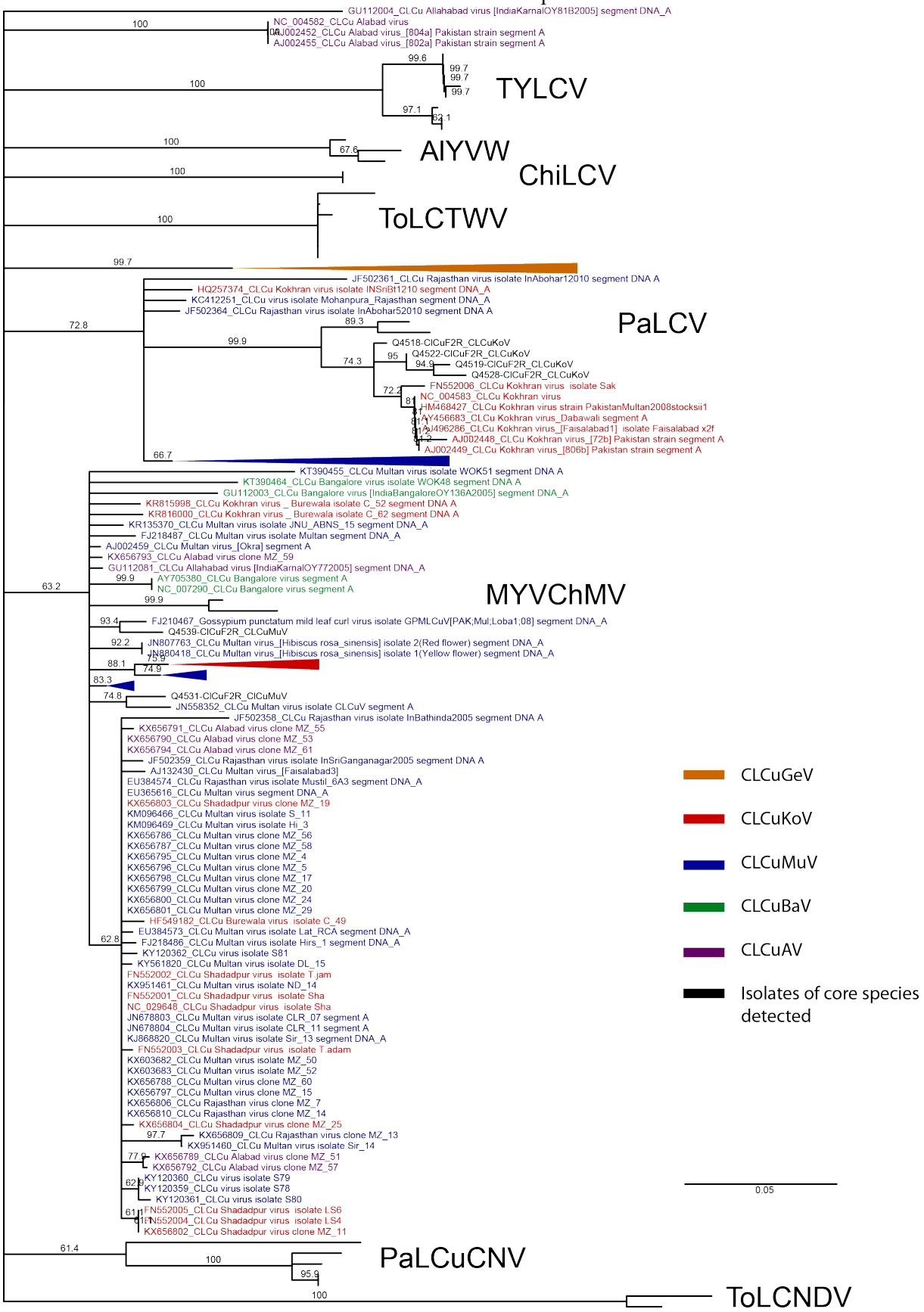


Figure 22).

Sensitivity

A subset of samples from the specificity test were chosen for dilution tests to investigate the relative sensitivity of the four primer sets. The viral load in plants can differ for a number of reasons, such as the time since infection, host type, host age, host stress, temperate, as well as the tissue sampled. As viral loads can differ, identifying primers which are able to detect low viral loads is an important consideration.

Serial dilutions from 1/10 to 1/100,000 were prepared from a subset of 13 samples (Table 18). Once again different plant extractions were used, including cotton, tomato, okra and hibiscus. Samples containing different Begomoviruses, were chosen, including CLCuMuV, CLCuKoV, PaLCuCNV, AlYVV, ToLCTWV and TYLCV. All six serial dilutions from each sample were screened with the four primer sets ClCu5, BEG, SPG and Accore, with the same protocols as above. Samples were run on a gel and scored based on whether a band was visually detected. Each primer set was scored from 0-6, with 0 not detected and 6 detected at the highest dilution of 1/100,000. Figure 23 shows an example of two samples Q4886 and Q4525.

Table 18shows each sample with the number corresponding to the highest dilution detected. Numbers in bold are the most sensitive across the four primer sets. CLCu5 detects virus at the highest or equal highest dilution across all 13 samples. BEG detects viruses at the highest equal dilution in 8/13 samples, followed by SPG (6/13), then Accore (2/13).

Betasatellite primers

Along with the four primers mentioned above, all of which bind to the main genome DNA-A, a further primer set was tested which binds to the satellite molecule, betasatellite. Begomoviruses do not require betasatellites for infection, however the presence of a betasatellite is associated with an increase in symptom severity in the host. Betasatellites, require a DNA-A genome for replication and movement, so cannot cause infection alone. Betasatellites are also promiscuous, and are found associated with different DNA-A genomes, therefore they cannot alone be used to identify the species causing infection in the plant. Only one example of a non-begomovirus associated betasatellite has been found (infecting a closely related genus, Mastrevirus (Kumar et al., 2014)). Therefore, the premise is that if a betasatellite is present within a plant sample, then it is highly likely that a begomovirus is also present. In this way, betasatellite detection can be used to suggest further investigation into identifying the begomovirus is likely present.

A betasatellite primer BetaUni was tested. Even after attempts to optimise these primers, multiple sized bands were produced (~400-1300bp) including double bands. It was thought that these primers were likely mispriming to plant genome, which was the case with a previous set of primers. Therefore, although this primer did not produce clean bands, PCR products of varying sizes from eighteen samples were sent for sequencing. Interestingly none of the sequences were plant genomes, all were betasatellite genomes (Table 19). The multiple bands may instead be due to differences in genome lengths eg defective molecules, or the primers are binding to more than one region in each betasatellite component.

All betasatellites were detected from samples which were positive for DNA-A components. In total eighteen samples had betasatellites identified. BLASTn analyses identified Cotton leaf curl Multan betasatellite (n=12), Tomato yellow leaf curl China betasatellite (n=2), Cotton leaf curl Kokhran betasatellite (n=1), Tomato leaf curl Ranchi betasatellite (n=1), Croton yellow vein mosaic virus betasatellite (n=1) and Ageratum yellow vein China betasatellite (n=1).

Cotton leaf curl Multan betasatellite was the most prevalent, within cotton samples this betasatellite was associated with CLCuKoV (6 samples) and CLCuMuV (1 sample). The Cotton leaf curl Kokhran betasatellite was identified from a cotton sample infected with CLCUKoV. The other five Cotton leaf curl Multan betasatellites were identified in Hibiscus (n=4) and a Malvastrum (n=1). As expected, there

was no correlation with the betasatellite identified and the begomovirus species it was associated with. Therefore, the identification of a particular betasatellite suggests a begomovirus detection but cannot be used to identify the begomovirus species.

Rolling Circle Amplification

As Begomoviruses have a circular genome, rolling circle amplification (RCA) can be utilised to increase detection when viral concentration is low. RCA uses a strand displacement polymerase which preferentially amplifies circular DNA. After amplification through RCA the resulting DNA can be testing with the PCR primers as above. All 65 samples were amplified through RCA using an Illustra Templiphi kit (GE Healthcare) with the resulting product tested with the four DNA-A primers as above.

Of the 65 samples, a further nine samples were found to contain begomoviruses which had tested negative with extract DNA (Table 20). A further two samples, which were positive with the straight extract for some of the four primers, had begomoviruses detected with an additional primer (Accore).

This extra step is very sensitive to contamination and also requires a special kit/polymerase to carry out. The RCA reaction also requires an overnight (16hr) step extending the time required for a result. Therefore, it is unlikely to be used as a routine diagnostic tool, however it is worth mentioning for samples of high importance, that RCA is an option.

A further use of RCA is identifying a virus which has been unable to be detected by any PCR methods. This non-sequence specific method uses restriction enzymes to cut the amplified concatenated virus into singular linear fragments which can be run on a gel, with those producing virion sized fragments excised and sequenced through cloning. Cloning requires specialised equipment and skills as well as the appropriate regulatory certifications and lab containment facilities. CLCuD causing viruses are well studied with a raft of sequence information available therefore it is unlikely that RCA followed by restriction enzyme digests would be necessary. RCA followed by PCR however, may be useful in certain circumstances.

Full genomes

If further genetic information is required for any of the begomoviruses detected with screening primers, then back-to-back or abutting primers can be designed. As begomovirus genomes are circular, the use of back-to-back primers will allow the elucidation of the full virus genome. By designing the back-to-back primers within the sequenced region of the screening primers, no further genetic information is needed to elucidate the full genome.

Conclusion

Begomoviruses were detected from a number of different plant hosts, including cotton. Sequencing of positive samples should be undertaken as these primers will also amplify other begomoviruses. No single primer pair detected all begomoviruses within the samples tested. Individually BEG and CLCu5 detected begomoviruses from all infected samples, including three of the core species which are associated with cotton leaf curl disease. All samples from which a core species was detected, was also identified with CLCu5 and/or BEG. CLCu5 primer was also the most sensitive of the primer pairs tested. Detecting begomoviruses at the highest or equal highest concentration compared to the other primers. In one sample CLCu5 was still able to detect a begomovirus at a 1/100,000 sample dilution. Phylogenetically the CLCu5 isolates of the five core species fell within clades of the corresponding species.

Although no CLCuBaV or CLCuAV were identified from the samples available, sequence information suggest that CLCu5 primers should amplify (with no primer binding mismatches) both species and Beg is also likely to amplify these species (BegF has 1-4 mismatches and BegR has 1-2 mismatches).

Sequence information for CLCrV, the bipartite begomovirus which infects cotton but causes mild symptoms, is unlikely to be amplified by CLCu5 (7 mismatches for CLCuR). However, CLCrV should be detected by Beg (2 mismatches for both BegF and BegR). The Accore primer set should also detect CLCrV (Avcore no mismatches, Accore 1 mismatch).

These results suggest that for a comprehensive screen for begomoviruses, both CLCu5 and BEG should be employed. These primers have been shown to detect a number of begomoviruses including three of the core species which cause CLCuD; CLCuMuV, CLCuGeV and CLCuKoV. However, since neither primer pair alone detected begomoviruses in all samples, both primers should be used.

The betasatellite primer sets can be further used to determine if a sample is infected with a begomovirus. However, the begomovirus which is associated with the betasatellite cannot be determined by the betasatellite sequence alone due to the promiscuious nature of betasatellites. Therefore, if an infection is still suspected after negative begomovirus screening, a betasatelllite screen can be used to identify if further tests need to be carried out. Rolling circle amplification can be used to amplify the begomovirus DNA in order to increase the amount of DNA available for PCR using Beg and CLCu5FR.

-

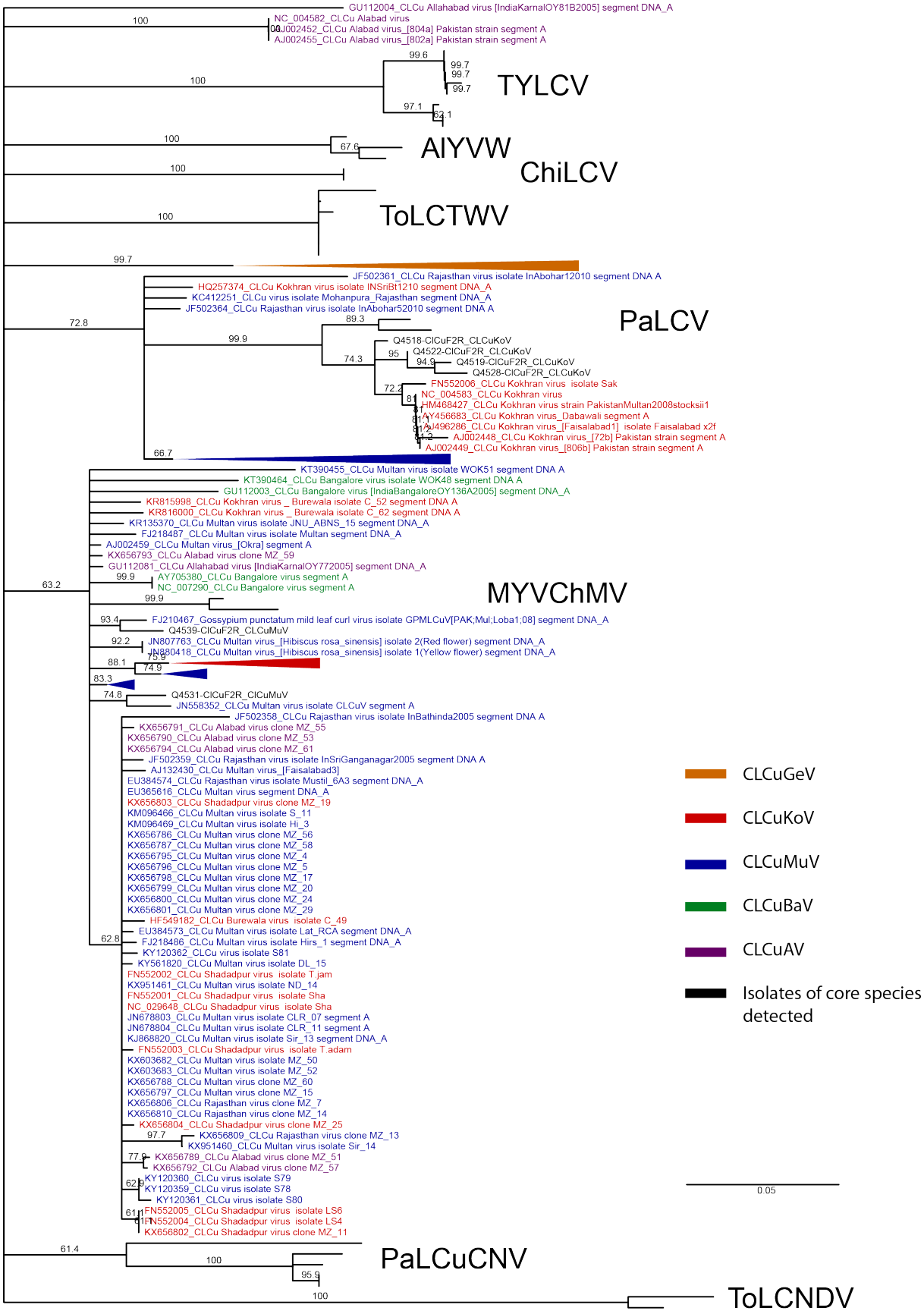


Figure 22: Neighbour joining phylogenetic tree of a section of the CLCu5 primer binding region. Sequences of the core CLCuD causing begomovirus, other begomovirus species detected, and isolates sequenced during this study are included. Large clades of the core species have been collapsed.

Figure 23: Dilution series of two samples Q4886 and Q4525 with four primer sets, Accore, BegFR, CLCu5FR, SPG1-2. Samples were 1-6; full strength, 1/10, 1/100, 1/1000, 1/10,000, 1/100,000 dilution.

Table 13: Non-cotton plant hosts which have had full genomes of core CLCuD causing viruses identified and deposited in GenBank (simplified from geminivirus.org)

Host	Family	Begomovirus species
Hibiscus rosa-sinensis	Malvaceae	*Cotton leaf curl Multan virus*
Hibiscus rosa-sinensis	Malvaceae	*Cotton leaf curl Kokhran virus*
Hibiscus cannabinus	Malvaceae	*Cotton leaf curl Multan virus*
Hibiscus cannabinus	Malvaceae	*Cotton leaf curl Bangalore virus*
Malvaviscus penduliflorus	Malvaceae	*Cotton leaf curl Multan virus*
Alcea rosea (Hollyhock)	Malvaceae	*Cotton leaf curl Gezira virus*
Malachra capitata	Malvaceae	*Cotton leaf curl Kokhran virus*
Abelmoschus esculentus (Okra)	Malvaceae	*Cotton leaf curl Gezira virus*
Abelmoschus esculentus (Okra)	Malvaceae	*Cotton leaf curl Bangalore virus*
Abelmoschus esculentus (Okra)	Malvaceae	*Cotton leaf curl Alabad virus*
Abelmoschus esculentus (Okra)	Malvaceae	*Cotton leaf curl Multan virus*
Sida	Malvaceae	*Cotton leaf curl Gezira virus*
Chili pepper	Solanacea	*Cotton leaf curl Burewala virus*
Solanum melongena (Eggplant)	Solanacea	*Cotton leaf curl Burewala virus*
Solanum lycopersicum (Tomato)	Solanacea	*Cotton leaf curl Multan virus*
Solanum lycopersicum (Tomato)	Solanacea	*Cotton leaf curl Gezira virus*
Solanum lycopersicum (Tomato)	Solanacea	*Cotton leaf curl Kokhran virus*
Digera arvensis	Amaranthaceae	*Cotton leaf curl Multan virus*

Emilia sonchifolia	Asteracea	*Cotton leaf curl Multan virus*
Papaya	Caricaceae	*Cotton leaf curl Multan virus*
Papaya	Caricaceae	*Cotton leaf curl Burewala virus*
Luffa cylindrica	Cucurbitaceae	*Cotton leaf curl Kokhran virus*
Ricinus communis (caster-oil plant)	Euphorbiacea	*Cotton leaf curl Kokhran virus*

Table 14: Countries which have had full genomes of core CLCuD causing viruses identified and deposited in GenBank (simplified from geminivirus.org)

Description	Country
Cotton leaf curl Alabad virus	India
Cotton leaf curl Alabad virus	Pakistan
Cotton leaf curl Bangalore virus	India
Cotton leaf curl Gezira virus	Burkina Faso
Cotton leaf curl Gezira virus	Cameroon
Cotton leaf curl Gezira virus	Côte d'Ivoire
Cotton leaf curl Gezira virus	Egypt
Cotton leaf curl Gezira virus	Israel
Cotton leaf curl Gezira virus	Jordan
Cotton leaf curl Gezira virus	Niger
Cotton leaf curl Gezira virus	Nigeria
Cotton leaf curl Gezira virus	Oman
Cotton leaf curl Gezira virus	Pakistan
Cotton leaf curl Gezira virus	Saudi Arabia
Cotton leaf curl Gezira virus	Sudan
Cotton leaf curl Gezira virus	United Arab Emirates
Cotton leaf curl Kokhran virus	India
Cotton leaf curl Kokhran virus	Pakistan
Cotton leaf curl Multan virus	China
Cotton leaf curl Multan virus	India
Cotton leaf curl Multan virus	Pakistan
Cotton leaf curl Multan virus	Philippines
Cotton leaf curl Multan virus	Thailand

Table 15: Primers used in this study. PCR protocols have been optimised with MyTaq™ Hotstart DNA Polymerase (Bioline)

Primer	Primer pairs	PCR protocol	Sequence	Binding region Nonanucleotide start aligned dataset	Amplification size (bp)	
CLCu5	CLCuF	$95_{2min}(95_{30sec},52_{30sec},72_{1min})^{30}72_{2min}$	GTTCCCCTGTGCGWGAATC	1683-2039	346	This study
	CLCuR		CTTTAARGARTTCATGGGKGC			
Beg	BEGF	$95_{2min}(95_{30sec},58_{30sec},72_{1min})^{30}72_{2min}$	CCAYTCTCTGCTTGAGBTGC	1655-1950	300	Van Brunschot
	BEGR		ATCTTCCTNTGCAATCCAGG			
SPG	SPG1	$95_{2min}(95_{30sec},50_{30sec},72_{1min})^{30}72_{2min}$	CCCCKGTGCGWRAATCCAT	1686-2619	912	Li et al. (2004)
	SPG2		ATCCVAAYWTYCAGGGAGCTAA			
Accore	ACCORE	$95_{2min}(95_{30sec},50_{30sec},72_{1min})^{30}72_{2min}$	GGRTTDGARGCATGHGTACAIGCC	530-1228	550	Wyatt and Brown (1996b)
	AVCORE		GCCHATRTAYAGRAAGCCMAGRAT			
BetaUni	BetaUniFor	$95_{2min}(95_{30sec},58_{30sec},72_{1min})^{30}72_{2min}$	AGCCTTAGCTACGCCGGAGC	Back-to-back full genome	~1.35kb	van Brunschot et al. (2014)
	BetaUniRev		CTGCTGCGTAGCGTAGTGGT			
	Plant primer					
ITS2	ITS2-S2F	$95_{2min}(95_{30sec},50_{30sec},72_{1min})^{30}72_{2min}$	ATGCGATACTTGGTGTGAAT	ITS region of plant genome	355	Chen et al. (2010b)
	ITS4		TCCTCCGCTTATTGATATGC			White et al. (1990a)

Table 16: PCR reaction mix for each sample

MyTaq Polymerase (5U/ul)	0.2ul
Primer F/R (10uM each primer)	1ul
MyTaq Buffer HS 5x	6ul
Water	21.8ul
Template	1ul
Total	30

Table 17: Host information for all plant samples positive for begomoviruses. Top blast hit of all PCR products, with corresponding accession number and percentage pairwise identity are shown. Core five species are shown in Bold.

Host	Location	Sample ID	CLCu5 primer region BLASTn Hit	Accession number	% Pairwise Identity	BEG primer region BLASTn hit	Accession number	% Pairwise Identity	SPG primer region BLASTn hit	Accession number	% Pairwise Identity	Accore primer region BLASTn hit	Accession number	% Pairwise Identity
Cotton	Unknown	Q3773				ToLCTWV	KP195719	97%						
Cotton	Pakistan	Q4506	CLCuKoV	JN678805	99%	CLCuKoV	FR750324	98%	CLCuKoV	HG937518	74%	CLCuKoV	LN852667	99%
Cotton	Pakistan	Q4507	CLCuKoV	JN678805	100%	CLCuKoV	FR750324	97%	CLCuKoV	FR837934	96%	CLCuKoV	LN852667	99%
Cotton	Pakistan	Q4508	CLCuKoV	JN678805	99%	CLCuKoV	FR750324	96%	CLCuKoV	FR837934	92%			
Cotton	Pakistan	Q4510				CLCuMuV	AJ002458	98%						
Cotton	Pakistan	Q4511	CLCuKoV	JN678805	99%	CLCuKoV	FR750324	98%						
Cotton	Pakistan	Q4514				CLCuMuV	AJ002458	97%						
Cotton	Pakistan	Q4521	CLCuKoV	FR750322	90%	CLCuKoV	FR750324	97%				CLCuKoV	LN852667	99%
Cotton	Pakistan	Q4523				CLCuKoV	HM461866	92%						
Cotton	Pakistan	Q4524	CLCuKoV	JN678805	100%	CLCuKoV	FR750324	98%	CLCuKoV	FR837934	96%	CLCuKoV	JF502366	99%
Cotton	Pakistan	Q4525	CLCuKoV	JN678805	100%	CLCuKoV	HG937518	98%				CLCuKoV	JF502366	98%
Cotton	Pakistan	Q4526	CLCuKoV	JN678805	99%	CLCuKoV	FR750324	98%						
Cotton	Pakistan	Q4534	CLCuKoV	JN678805	100%	CLCuKoV	FR750324	97%	CLCuKoV	FR837934	98%	CLCuKoV	LN852667	99%
Cotton	Pakistan	Q4535				CLCuKoV	FR750320	97%						
Cotton	Pakistan	Q4540	ChiLCV	KR779820	100%	ChiLCV	KR779820	96%	ChiLCV	KR779820	92%	ToLCKV	HM134235	96%
Cotton	China	Q4882	CLCuMuV	KP762786	100%	CLCuMuV	GQ503175	97%	CLCuMuV	KF766949	99%			
Cotton	China	Q4893	ToLCTWV	KP195719	98%									
Cotton	China	Q4894				ToLCTWV	KP195719	98%						
Cotton	China	Q4895				ToLCTWV	KP195719	98%						
Cotton	China	Q4910	TYLCV	JX128100	99%									
Cotton Australian variety	Pakistan	Q4531	CLCuMuV	JN558352	97%	CLCuMuV	KF766951	93%	CLCuMuV	JN558352	94%	CLCuMuV	KX068709	98%
Cotton *Gossypium barbadense*	Pakistan	Q4530	AlYVV	FJ015062	97%	CLCuKoV	AH013912	95%						
Cotton Wild	Pakistan	Q4537				CLCuKoV	AH013912	86%						
Cotton Wild	Pakistan	Q4539	CLCuMuV	FJ210467	95%	CLCuKoV	AH013912	94%						
Chilli	Egypt	Q2535	CLCuGeV	GU945265	98%	CLCuKoV	AH013912	93%				CLCuGeV	FJ030874	88%

Host	Location	Sample ID	CLCu5 primer region BLASTn Hit	Accession number	% Pairwise Identity	BEG primer region BLASTn hit	Accession number	% Pairwise Identity	SPG primer region BLASTn hit	Accession number	% Pairwise Identity	Accore primer region BLASTn hit	Accession number	% Pairwise Identity
Chilli	China	Q4905	PaLCuCNV	KU892671	98%	PaLCuCNV	KU892671	98%				PaLCuCNV	AJ558123	99%
Chinese medicine herb	China	Q4899				EmYVV	JQ247188	96%	EmYVV	KJ016240	93%			
Citrus	China	Q4890				ToLCTWV	KP195719	98%						
Hibiscus	China	Q4887	CLCuMuV	KP762786	99%	ToLCTWV	KP195719	99%	CLCuMuV	KP762786	99%			
Hibiscus	China	Q4896	CLCuKoV	JF502353	74%				CLCuMuV	KP762786	99%			
Hibiscus	China	Q4900	CLCuMuV	KY432574	100%	CLCuMuV	AJ02459	74%	CLCuMuV	KP762786	100%	CLCuMuV	KY432576	96%
Hibiscus	China	Q4901	CLCuMuV	KP762786	99%	ToLCTWV	GU723730	76%	CLCuMuV	KP762786	98%	CLCuMuV	KY432576	99%
Malvastrum	Pakistan	Q4517	MYVChMV	FR715681	97%	MYVChMV	FR715681	96%	MYVChMV	FR715681	96%			
Melon	Egypt	Q2545	CLCuGeV	GU945265	100%	ChiLCV	KR779820	98%				CLCuGeV	AY036010	99%
Okra	China	Q4883	CLCuMuV	KP762786	99%	CLCuMuV	GQ503175	98%	CLCuMuV	KF766949	99%	CLCuMuV	KY432576	97%
Okra	China	Q4884	ToLCTWV	KP195719	99%									
Okra	China	Q4898	CLCuMuV	KP762786	100%	CLCuMuV	GQ503175	99%	CLCuMuV	KP762786	99%			
Papaya	Pakistan	Q4533	ToLCNDV	HM989845	95%	CYMV	HF565179	92%	ToLCNDV	HM989845	97%	CYMV	LK028570	93%
Parthenium relative	Pakistan	Q4528	CLCuKoV	AH013912	97%	CLCuKoV	AH013912	96%						
Poinsettia	Pakistan	Q4532	PaLCV	AJ436992	96%	AEV	LT716984	95%	PaLCV	KY978407	94%	ToLCPKV	FM164938	98%
Sodom apple	Pakistan	Q4518	CLCuKoV	AH013912	97%	CLCuKoV	AH013912	95%						
Sodom apple	Pakistan	Q4519	CLCuKoV	AH013912	99%	CLCuKoV	AH013912	95%						
Tomato	Pakistan	Q4522	CLCuKoV	AH013912	98%	CLCuKoV	AH013912	96%						
Tomato	China	Q4886	ToLCTWV	KP195719	88%	ToLCTWV	KP195719	98%	ToLCTWV	KP195719	97%	ToLCTWV	KP195719	99%
Tomato	China	Q4902	PaLCuCNV	AM691554	86%	PaLCuCNV	KU892671	98%	AYVCNV	KU954387	97%	PaLCuCNV	AJ558123	84%
Tomato	China	Q4903	PaLCuCNV	KU892671	91%	PaLCuCNV	KU892671	98%	AYVCNV	KU954387	98%	ToLCCNV	KU980921	81%
Tomato	China	Q4906	TYLCV	JX128100	99%	TYLCV	FN256256	98%	TYLCV	JX128100	91%	TYLCV	LC099965	100%
Tomato	China	Q4907	TYLCV	JX128100	99%	TYLCV	KX034551	98%	TYLCV	MF590742	95%	TYLCV	MF590742	100%
Tomato	China	Q4908	TYLCV	KC138543	99%	TYLCV	KX347127	98%				TYLCV	MF590732	99%
Tomato	China	Q4909	TYLCV	KC138543	99%	TYLCV	KX347127	98%	TYLCV	MF590743	96%	TYLCV	MF590732	100%
Unknown weed	Pakistan	Q4529	AlYVV	KT717678	92%	AlYVV	LC316183	96%				AlYVV	LC316183	97%
Unknown weed	China	Q4888	ToLCTWV	KP195719	99%	ToLCTWV	KP195719	98%						
Unknown weed	China	Q4889	ToLCTWV	JX128098	84%	ToLCTWV	KP195719	98%						

Host	Location	Sample ID	CLCu5 primer region BLASTn Hit	Accession number	% Pairwise Identity	BEG primer region BLASTn hit	Accession number	% Pairwise Identity	SPG primer region BLASTn hit	Accession number	% Pairwise Identity	Accore primer region BLASTn hit	Accession number	% Pairwise Identity
Unknown weed	China	Q4897				ToLCTWV	KP195719	98%						

All abbreviations for all tables: AEV - Ageratum enation virus, ALYVV - Alternanthera yellow vein virus, AYVCNB - Ageratum yellow vein China betasatellite, AYVCNV - Ageratum yellow vein China virus, ChiLCV - Chilli leaf curl virus, CLCuGeV - Cotton leaf curl Gezira virus, CLCuKoB Cotton leaf curl Kokhran betasatellite, CLCuKoV - Cotton leaf curl Kokhran virus, CLCuMuB - Cotton leaf curl Multan betasatellite, CLCuMuV - Cotton leaf curl Multan virus CroYVMB - Croton yellow vein mosaic virus-associated betasatellite, CYMV - Catharanthus yellow mosaic virus, EmYVV - Emilia yellow vein virus, MYVChMV - Malvastrum yellow vein Changa Manga virus, PaLCuCNV - Papaya leaf curl China virus, PaLCV - Papaya leaf curl virus, PeLCuV - Pedilanthus leaf curl virus, TYLCCNB - Tomato yellow leaf curl China betasatellite, ToLCCNV - Tomato leaf curl China virus, ToLCKV - Tomato leaf curl Karnataka virus, ToLCNDV - Tomato leaf curl New Delhi virus, ToLCPKV - Tomato leaf curl Pakistan virus, ToLCRnB - Tomato leaf curl Ranchi betasatellite, ToLCTWV, - Tomato leaf curl Taiwan virus, TYLCV - Tomato yellow leaf curl virus.

Table 18: Dilution series of samples with four primer sets, Accore, BegFR, CLCuFR, SPG1-2. Lowest dilution detected was recorded. Dilutions were 1-6; full strength, 1/10, 1/100, 1/1000, 1/10,000, 1/100,000 dilution. With 0 indicating no PCR product was detected. Dilution in bold are the lowest dilution detected for that sample across the four primer sets.

Host	Sample ID	Dilution detected CLCu5F/R1	BLASTn hit	Dilution detected BegF/R	BLASTn hit	Dilution detected SPGF/F	BLASTn hit	Dilution detected Accore	BLASTn hit
Cotton	Q4506	5	CLCuKoV	4	CLCuKoV	5	CLCuKoV	3	CLCuKoV
Cotton	Q4525	4	CLCuKoV	4	CLCuKoV	4	Positive PCR but Sequence failed	3	CLCuKoV
Cotton	Q4540	4	ChiLCV	4	ChiLCV	4	ChiLCV	4	ToLCKV
Cotton *Gossypium barbadense*	Q4530	1	AlYVV	1	CLCuKoV	0		0	
Cotton Australian variety	Q4531	4	CLCuMuV	2	CLCuMuV	4	CLCuMuV	3	CLCuMuV
Chilli	Q4905	1	PaLCuCNV	1	PaLCuCNV	0		0	PaLCuCNV
Hibiscus	Q4900	3	CLCuMuV	1	CLCuMuV	3	CLCuMuV	0	CLCuMuV
Okra	Q4883	5	CLCuMuV	1	CLCuMuV	4	CLCuMuV	1	CLCuMuV
Parthenium relative	Q4528	1	CLCuKoV	1	CLCuKoV	0		0	
Sodom apple	Q4518	1	CLCuKoV	1	CLCuKoV	0		0	
Tomato	Q4886	5	ToLCTWV	5	ToLCTWV	4	ToLCTWV	3	ToLCTWV
Tomato	Q4902	6	PaLCuCNV	5	PaLCuCNV	5	AYVCNV	5	PaLCuCNV
Tomato	Q4907	4	TYLCV	4	TYLCV	4	TYLCV	4	TYLCV

Table 19: Betasatellites identified using BetaUni primers with the begomoviruses detected previously.

Host	Location	Sample ID	Betasatellite primer Region Blastn hit	Accession number	% pairwise identity	CLCu5 primer region BLASTn hit	BEG primer region BLASTn hit	SPG primer region BLASTn hit	Accore primer region BLASTn hit
Cotton	Pakistan	Q4506	CLCuMuB	LN845929	92%	CLCuKoV	CLCuKoV	CLCuKoV	CLCuKoV
Cotton	Pakistan	Q4507	CLCuMuB	KP015741	93%	CLCuKoV	CLCuKoV	CLCuKoV	CLCuKoV
Cotton	Pakistan	Q4510	CLCuMuB	HE978344	97%		CLCuMuV		
Cotton	Pakistan	Q4511	CLCuMuB	JF502398	99%	CLCuKoV	CLCuKoV		
Cotton	Pakistan	Q4521	CLCuMuB	JF502398	99%	CLCuKoV	CLCuKoV		CLCuKoV
Cotton	Pakistan	Q4525	CLCuMuB	LN831985	91%	CLCuKoV	CLCuKoV		CLCuKoV
Cotton	Pakistan	Q4534	CLCuKoB	HF549186	91%	CLCuKoV	CLCuKoV	CLCuKoV	CLCuKoV
Cotton	Pakistan	Q4535	ClCuMuB	JF502398	99%		CLCuKoV		
Hibiscus	China	Q4887	CLCuMuB	KP762787	96%	CLCuMuV	ToLCTWV	CLCuMuV	
Hibiscus	China	Q4896	CLCuMuB	JQ943409	99%	CLCuKoV		CLCuMuV	
Hibiscus	China	Q4900	CLCuMuB	AY704663	94%	CLCuMuV	CLCuMuV	CLCuMuV	CLCuMuV
Hibiscus	China	Q4901	CLCuMuB	GQ906588	93%	CLCuMuV	ToLCTWV	CLCuMuV	CLCuMuV
Malvastrum	Pakistan	Q4517	CLCuMuB	GH932562	93%	MYVChMV	MYVChMV	MYVChMV	
Papaya	Pakistan	Q4533	CroMTVB	AM410551	95%	ToLCNDV	CYMV	ToLCNDV	CYMV
Poinsettia	Pakistan	Q4532	ToLCRnB	GQ994096	82%	PaLCV	AEV	PaLCV	ToLCPKV
Tomato	China	Q4902	AYVCNB	HF569260	95%	PaLCuCNV	PaLCuCNV	AYVCNV	PaLCuCNV
Tomato	China	Q4906	TYLCCNB	AJ457820	70%	TYLCV	TYLCV	TYLCV	TYLCV
Tomato	China	Q4907	TYLCCNB	AJ536625	79%	TYLCV	TYLCV	TYLCV	TYLCV

*Table 20: Viruses detected from plant samples only after circular DNA was amplified using rolling circle amplification (RCA). Samples with * had isolates previously identified, with further primers also detecting begomoviruses after RCA. Core species which cause CLCuD are in bold..*

Host	Location	Sample ID	CLCu Primer Region BLASTn hit	Accession number	% Pairwise Identity	BEG Primer Region BLASTn hit	Accession number	% Pairwise Identity	SPG Primer Region BLASTn hit	Accession number	% Pairwise Identity	Accore Primer Region BLASTn hit	Accession number	% Pairwis Identity
Cotton	Unknown	Q3371	CLCuKoV	JN678805	0.99	CLCuKoV	AH013912	95%	CLCuKoV	AM774303	90%	CLCuKoV	JF502366	99%
Cotton	Unknown	Q3772	CLCuKoV	AH013912	0.96				CLCuKoV	HM468427	95%	CLCuKoV	KR816000	98%
Cotton	Pakistan	Q4515	CLCuMuV	AJ002458	0.99	CLCuMuV	AJ002458	97%	CLCuMuV	AJ002447	90%			
Cotton wild	Pakistan	Q4536	PaLCV	KX302713	0.95	PeLCuV	LT795118	96%	PeLCuV	HF568781	93%	CLCuMuV	FJ210467	97%
False amaranth	Pakistan	Q4520	CLCuKoV	AH013912	0.96	CLCuKoV	AH013912	95%	CLCuKoV	HM468427	96%	CLCuMuV	KY120362	91%
False amaranth	Pakistan	Q4527	CLCuKoV	AH013912	0.96	CLCuKoV	AH013912	97%	CLCuKoV	HM468427	87%	CLCuKoV	HF549182	86%
Hibiscus	China	Q4891	CLCuMuV	KP762786	0.99				CLCuMuV	KP762786	98%			
Hibiscus	China	Q4904	CLCuMuV	KP762786	1.00	PaLCuCNV	KU892671	99%	CLCuMuV	KP762786	91%	PaLCuCNV	KU892671	77%
Okra	China	Q4885	ToLCTWV	KP195720	0.87	ToLCTWV	KP195719	98%						
Chinese medicine herb	China	Q4899				EmYVV*	JQ247188*	96%*	EmYVV*	KJ016240*	93%*	EmYVV	KC878472	82%
Cotton	China	Q4893	ToLCTWV*	KP195719*	0.98*							ToLCTWV	KP195719	89%

Ashfaq, M., Asif, M., Anjum, Z. I., Zafar, Y., 2013. Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (Gossypium: Malvaceae). Molecular Ecology Resources 13, 573-582.

Briddon, R.W., Markham, P.G., 2000. Cotton leaf curl virus disease. Virus Research 71, 151-159.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y., Leon, C., 2010. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. PLOS ONE 5, e8613.

Dolores, L.M., Gonzales, M. C., Magdalita, P. M., 2014. Isolation and Identification of a Begomovirus Associated With the Leaf Curl Disease of Gumamela (Hibiscus rosa-sinensis) in the Philippines. Asian Journal of Agriculture and Food Sciences 02.

Kumar, J., Kumar, J., Singh, S.P., Tuli, R., 2014. Association of satellites with a mastrevirus in natural infection: Complexity of Wheat dwarf India virus disease. Journal of Virology 88, 7093-7104.

Li, R., Salih, S., Hurtt, S., 2004. Detection of Geminiviruses in Sweetpotato by Polymerase Chain Reaction. Plant Disease 88, 1347-1351.

van Brunschot, S.L., Bergervoet, J.H.W., Pagendam, D.E., de Weerdt, M., Geering, A.D.W., Drenth, A., van der Vlugt, R.A.A., 2014. A bead-based suspension array for the multiplexed detection of begomoviruses and their whitefly vectors. Journal of Virological Methods 198, 86-94.

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, PCR Protocols. Academic Press, San Diego, pp. 315-322.

Wyatt, S.D., Brown, J.K., 1996. Detection of subgroup III geminivirus isolates in leaf extracts by degenerate primers and polymerase chain reaction. Phytopathology. 86.

Provision of diagnostic support to Dr Cherie Gambley (DAQ1405) and Murray Sharman (DAQ1201) for CLCuD surveillance in Australia (samples collected during survey collections)

We provided continual diagnostic support to Cherie Gambley (DAQ1405), Jamie Hopkinson (DAQ1403), and Murray Sharman (DAQ1201) throughout the duration of the project. We performed molecular identification of whiteflies (and the viruses they transmit), as well as routine virus indexing of plant samples collected during surveys in Australia as part of their CRDC projects.

Samples tested for Dr Gambley included 73 representative samples from fourteen whitefly populations (separate locations and host plants) collected in northern Australia. Most of the whiteflies collected were identified as *B. tabaci* AUSII. No *B. tabaci* Middle East Asia Minor 1 (MEAM1; B biotype) or *B. tabaci* AUS1 (Australian native whitefly 1) were identified from the samples. The whitefly collected from cucurbit plants at survey site ORIA_WP498 (Wyndham, Western Australia) belong to the *B. tabaci* ASIAII species grouping. This finding of *B. tabaci* ASIAII in Western Australia was a new record for this state. I prepared a detailed report for this work and in conjunction with Dr Gambley (DAQ1405) our results were reported to state and federal biosecurity agencies in June 2015.

I have also provided diagnostic support to Jamie Hopkinson (DAQ1403), and Murray Sharman (DAQ1201) for samples of whiteflies they have collected from various hosts including cotton, saltbush (*Einadia nutans*), sunflower, native rosella in Queensland (Emerald, Clermont, Capella, Toowoomba and Darling Downs).

Diagnostic testing of 79 plant samples from Timor-Leste and the Torres Strait Islands (TSI) collected by Murray Sharman (DAQ1201) were completed. No evidence of begomoviruses have been detected using two commonly utilised methods. These methods are designed to identify a range of begomovirus species. Further testing of these samples will be done. In addition, 5 representative *B. tabaci* individuals from five populations, which

Murray collected from different islands within the TSI were identified using DNA sequencing methods. DNA sequencing results showed that the whitefly collected were a mix of *B. tabaci* MEAM1 and AUSII.

Overseas samples provided by Dr Murray Sharman as part of the project DAQ1201 were screened for the presence of begomoviruses using two degenerate primers. Optimisation of broad spectrum begomovirus primers was carried out. Some primer sets resulted in non-specific binding or double bands. Sequencing of a number of these products identified that these primers were also able to bind to plant DNA, including cotton DNA. These primers were therefore rejected due to the possibility of false positives. Two primers; BegF/R and Accore/avcore gave single strong bands with known positives, with no mispriming identified. These primers bind to two different regions of the DNA-A component and were utilised for the screening of samples for begomoviruses. Like the majority of screening primers, BegF/R and Accore/avcore target a small section of the genome. As begomoviruses have a circular genome, if further sequence information is required then Back-to-back primers can be designed. These back-to-back primers are designed within the sequence elucidated by the screening primers, meaning no further genetic information is required to amplify a larger section, or even the full genome of the virus.

102 overseas samples were provided and were screened with two broad spectrum begomovirus primers. Of these samples 68 were cotton, upland, sea island and tree cotton, 13 samples were ornamental Hibiscus, with a number of weed species also tested. None of the overseas samples were positive for ClCuD causing viruses.
These screening primers identified only one sample which was positive, sequencing of this sample detected the begomovirus *Ageratum yellow vein virus*, which does not infect cotton. Back-to-back primers were designed for two additional samples which were first sequenced by Murray Sharman in order to further verify these samples. These were confirmed as *Ageratum yellow vein virus* and *Kudzu Mosaic Virus*, both begomoviruses, neither of which infect cotton.
A plant with potential viral disease symptoms, including abnormal growth and cupping leaves was detected at a post entry quarantine plant facility. Extracts of this plant were provided by Dr Sharman and were tested with the begomovirus screening primers. This plant (not cotton) is not a known host for begomoviruses. No begomoviruses were detected by either Dr Sharman or Dr Stainton, however due to the symptoms it was still recommended that these plants were destroyed.

Two broad spectrum begomovirus primers were optimised which gave strong single bands against known positives. Primers which had the potential to produce false positives were removed from screening. Internal plant primers were used to verify the extractions as Cotton is known to contain phenols and other chemicals which can interfere with PCR. This allows for robust screening of overseas samples for begomoviruses such as those which cause CLCuD.

No CLCuD causing begomoviruses have been detected in the overseas samples provided from project DAQ1201

4.7 Objective 7: *Research to support the activities of the African Cassava Whitefly Project (ACWP; Bill and Melinda Gates) at the Natural Resources Institute, University of Greenwich, as a join postdoc under the supervision of Prof John Colvin.*

Research was done in collaboration with Prof John Colvin, to support two Aims of the ACWP, including: Aim 2 (Super abundance and virus outbreaks), Aim 3 (Resistant and Detoxification), plus additional supervision and support of staff and students of the project.

Experimental research for Aim 2 was related to the characterization of whitefly-transmitted virus infectious clones. This work was done in collaboration with Prof Sue Seal (University of Greenwich), and completed to a high standard. A brief summary of this work is provided here.

Experimental research for Aim 3 was to support research done by Dr Osnat Malka and Prof Shai Morin (The Hebrew University of Jerusalem). A core component of their research is understanding the how some cryptic

species of *B. tabaci* are able to detoxify plant defense compounds, to identify gene targets for potential interference using transgenic approaches. I provided support for their research, performing bioassays, assisting with transcriptomics-based research and general molecular work. My contributions to their work have been recognised as co-authorship on two publications, the first which is currently under review with Molecular Ecology (see Appendix 5).

Lastly, high level supervision and training was provided to two full-time staff (Ms Rebecca Grimsley and Ms Kirsty Malpas) and three PhD students (Mr Joachim Nwezeobi, Mr Habibu Murgerwa, and MS Sona Vyskocilova). Three molecular training sessions were provided (each lasting around 1 week) to the staff and students, and continued training was provided in insectary and glasshouse methods and techniques.

Bioassays to compare ability of Bemisia tabaci species to detoxify host defense compounds

Aims:
Bioassays were performed to assess feeding and reproductive host-use by six putative cryptic species of *Bemisia tabaci*, on four different host plants.

Methods:

Plants
Table 21: Host plants used for bioassays

Plant common name	**Latin Name**	**Variety**
Eggplant	*Solanum melongena*	cv. Black Beauty, Kings Seed.
Sweet pepper	*Capsicum annuum*	cv. California Wonder, Chiltern Seeds.
Kale	*Brassica oleracea*	cv Dwarf Green Curled, Kings Seed, Borecole
Cassava	*Manihot esculenta*	col22

Colonies
Table 22: Colonies used for bioassays

Whitefly species	**Colony host/s**	**Colony ID#**	**COI ID#, checked 20.03.2017**
MEAM1-Peru	Solanum melongena cv. Black Beauty	1	148
SSA1 SG3	Solanum melongena cv. Black Beauty	58	150
Bemisia tabaci MED Spain Q1	Solanum melongena cv. Black Beauty	173	149
Bemisia tabaci MED Uganda Okra ASL	Abelmoschus esculentus, Okra and Solanum melongena cv. Black Beauty	263	151
NW2	Solanum melongena cv. Black Beauty and Gossypium hirsutum, Cotton	91	144
ASIAII_1	Solanum melongena cv. Black Beauty	56	145

Observation data
Collate host-dependent observations from Osnat, John and Sona from previous experiences. Combine these with the binary reproductive success assessments (yes/no) from these experiments.

Bioassays

- Set up 2 LL per treatment (in case of plant death) and transfer ~100 adults to each pot (depending on vigour of colony). See Table 3.
- Remove (and count at same time) adults at 2-3 weeks stage, before nymphs eclose.

– Check for eclosion after 3-5 weeks (allow extra time for developmental variability on different host plants, compared to eggplant). Count adults. Keep colonies in LL pots indefinitely.

Table 23: Bioassays for assessing plant host feeding and reproduction of different whitefly species.

Whitefly species	Bioassays set up	Number of whitefly per LL pot	Removed and counted original test adults that could survive and feed on host*	Counted adults that eclosed during bioassay
MEAM1-Peru	20.03.2017	100	10.04.2017 (20 days)	28.04.2017 (39 days)
SSA1 SG3	21.03.2017	50	10.04.2017 (20 days)	27.04.2017 (37 days)
Bemisia tabaci MED Spain Q1	24.03.2017	100	11.04.2017 (18 days)	29.04.2017 (39 days)
Bemisia tabaci MED Uganda Okra ASL	25.03.2017	70	11.04.2017 (17 days)	29.04.2017 (34 days)
NW2	29.03.2017	40	15.04.2017 (17 days)	30.04.2017 (30 days)
ASIAII_1	30.03.2017	50	16.04.2017 (17 days)	01.05.2017 (30 days)

* done when late-stage nymphs visible on eggplant bioassay for each whitefly species

Results

Bioassays

General notes

- The MEAM1-Peru colony was not strong, there were few adults in stock cage which are slow to reproduce, and so overall numbers were low in this assay.
- NW2 colony has a habit of boom and bust – seems to only boom when on young eggplant and at this point the stock colony had crashed a little and adults seemed slow and lethargic.
- When adults were observed on kale, they were always settled on the old lower leaves (seemed unusual).
- Observations and counts to assess feeding and survival in no-choice assay are presented in Table 4.
    - A single rep for MEAM1/Pepper did not have any adults, plant was healthy.
    - A single rep for NW2/Kale did not have any adults, plant was healthy.
    - A single rep for MED Uganda/Pepper did not have any adults, plant was healthy.
- Observations and counts to assess reproduction in no-choice assay are presented in Table 5, with some notes below:

MEAM1-Peru

- Did not reproduce on pepper, but could feed and survive on this host.
- Results on other hosts were as expected.

SSA1-SG3

- Fed and reproduced on all hosts
- Performed best on eggplant and cassava

MED Spain Q1

- Fed and reproduced on Eggplant, Pepper, Kale and not cassava
- Was OK on kale, which differs a little from Sona's results where she says "poor host"

MED Uganda

- Fed and survived on Eggplant, pepper and kale (not cassava)
- Reproduced only on eggplant and pepper

NW2

- Fed and reproduced on Eggplant, Pepper, Kale and not cassava

ASIAII_1

- Fed and survived on Eggplant, pepper and kale (not cassava)
- Some late-stage nymphs were found on one cassava plant (1 rep only), but these stalled in development.

Table 24: Counts of original test adults that fed and survived 2-3 weeks after transfer to new hosts (no-choice assay).

Whitefly species	Rep	Eggplant	Pepper	Kale	Cassava
MEAM1-Peru	1	12	3	10	0
	2	13	0	2	0
SSA1 SG3	1	40	30	26	21
	2	42	21	17	20
Bemisia tabaci MED Spain Q1	1	23	8	4	0
	2	25	11	17	0

Bemisia tabaci MED Uganda Okra ASL	1	37	4	7	0
	2	34	19	9	0
NW2	1	11	1	0	0
	2	9	3	2	0
ASIAII_1	1	30	9	11	0
	2	33	18	18	0

Results in disagreement to collated observations are highlighted in orange.

Table 25: Counts of adults that had eclosed on each different host plant, taken 30-40 days post oviposition whereby the original test adults were removed at 14-21 days.

Whitefly species	Rep	Eggplant	Pepper	Kale	Cassava
MEAM1-Peru	1	10	0	5	0
	2	2	0	4	0
SSA1 SG3	1	200+	13	3	100+
	2	200+	14	15	100+
Bemisia tabaci MED Spain Q1	1	100+	4	30+	0
	2	100+	7	30+	0
Bemisia tabaci MED Uganda Okra ASL	1	100+	6	0	0
	2	100+	0	0	0
NW2	1	10	0	2	0
	2	3	2	2	0
ASIAII_1	1	100+	100+	2	0
	2	100+	100+	3	0

Results in disagreement to collated observations are highlighted in orange.

Observations

Observational data was collated (Table 6), with most observation data being well aligned. Any observations that didn't match are highlighted in orange and include:

Table 26: Observations of reproductive success collated from whitefly researchers

Whitefly species	**Observer**	**Eggplant**	**Pepper**	**Kale**	**Cassava**
MEAM1-Peru	Osnat	+	+	+	-
	John[#]	good	ok	good	no
	Sona[#]	unknown	unknown	unknown	unknown
	Sharon[*]	+	-	+	-
SSA1 SG3	Osnat	+	unknown	unknown	+
	John[#]	good	unknown	unknown	good
	Sona[#]	unknown	unknown	unknown	unknown
	Sharon[*]	+	+	+	+
***Bemisia tabaci* MED Spain Q1**	Osnat	+	+	+	-
	John[#]	good	good	poor	no
	Sona[#]	good	ok	poor	no

	Sharon*	+	+	+	-
***Bemisia tabaci* MED Uganda Okra ASL**	Osnat	+	unknown	+	no
	John#	good	no	unknown	no
	Sona#	good	no	poor	no
	Sharon*	+	+	-	-
NW2	Osnat	+	unknown	unknown	-
	John#	poor	no	no	no
	Sona#	unknown	unknown	unknown	unknown
	Sharon*	+	+	+	-
ASIAII_1	Osnat	+	unknown	unknown	unknown
	John#	+	unknown	-	-
	Sona#	unknown	unknown	unknown	unknown
	Sharon*	+	+	+	-

*From these experiments

#John and Sona's ratings are (0) No, (1-2) Poor, (3-4) OK, (5) Good.

Results in disagreement are highlighted in orange.

Discussion

- Observations of feeding, survival and reproduction on different hosts are presented.
- It was interesting to me to see that the percentage survival did not necessary reflect reproductive success (e.g. ASIA II_1/Pepper contrast with SSA1 SG3/Pepper).

Characterisation of begomovirus infectious clones

Introduction

We have partially characterized the DNA-A components of the infectious clones (IC's) received from Danforth Centre, using conventional Sanger sequencing methods. This was done as an initial check to determine if the IC's matched the lists provided, prior to undertaking more complete characterization of the IC's.

Methods

Samples

Infectious clones were provided by Danforth Centre.

Rolling circle amplification

Infectious clones were separately amplified by rolling-circle amplification (RCA) with random priming using Phi29 DNA polymerase (TempliPhi™, GE Healthcare, Uppsala, Sweden). For each reaction, clones were scraped directly from tubes stored at -80C and used as template for RCA. For plasmid minipreps, 1 ul of DNA was used as template for RCA.

Degenerate DNA-A PCRs

PCR amplification of the core region of the begomovirus coat protein gene (V2/V1 junction) was performed using the degenerate primers AVCORE 5'-GCCHATRTAYAGRAAGCCMAGRAT-3′ and ACCORE 5′-GGRTTDGARGCATGHGTACANGCC-3′ (Allie et al., 2014, Idris et al., 2014). Amplicons of the expected size (~579 bp) were sequenced using conventional sequencing.

Results

Partial DNA-A sequences were obtained for all 24 ICs sequenced to date, results are summarized in Table 1. All sequence data provided as FASTA files.

Table 1: Summary of infectious clone sequence information to date

Infectious clone #	Clone name	Sequence name	Sequence length bp	BLAST match	Sequence result
1	ACMV -[CM] A	IC1	540	99% nt identity African cassava mosaic virus isolate West Kenyan 844 segment DNA1, complete sequence GenBank: J02057.1	ACMV
2 ^	ACMV -[IC] B	IC2			
3	ACMV -[KE:844:82] A	IC3	536	99% nt identity African cassava mosaic virus isolate West Kenyan 844 segment DNA1, complete sequence GenBank: J02057.1	ACMV
4 ^	ACMV -[KE:844:82] B	IC4			
5	EACMV-[K24] A	IC5	542	99% nt identity East African cassava mosaic virus segment DNA A, complete sequence, isolate EACMV-[K24] GenBank: AJ717557.1	EACMV
6	EACMV-[K24] B	IC6		100% nt identity East African cassava mosaic virus segment DNA B, complete sequence, isolate EACMV-[K24] GenBank: AJ704936.1	EACMV
7	EACMV-KE2[K27] A	IC7	541	99% nt identity East African cassava mosaic virus-KE2 segment DNA A, complete sequence, isolate EACMV-KE2[K27] GenBank: AJ717537.1	EACMV
8 ^	EACMV-KE2[K27] B	IC8			
9	EACMV-KE2[K29] A	IC9	540	99% nt identity East African cassava mosaic virus segment DNA A, complete sequence, isolate EACMV-[K29] GenBank: AJ717551.1	EACMV
10 ^	EACMV-KE2[K29] B	IC10			
11	EACMV-KE2[K268] A	IC11	539	99% nt identity East African cassava mosaic virus segment DNA A, complete sequence, isolate EACMV-[K268] GenBank: AJ717558.1	EACMV
12	EACMV-KE2[K268] B	IC12			

13*	EACMV-[K48] A	IC13	542	99% nt identity East African cassava mosaic virus-KE2 segment DNA A, complete sequence, isolate EACMV-KE2[K48] GenBank: AJ717542.1	EACMV
14#	EACMV-[K48] B				
15	EACMV-[K201] A	IC15	542	99% nt identity East African cassava mosaic virus-KE2 segment DNA A, complete sequence, isolate EACMV-KE2[K201] GenBank: AJ717541.1	EACMV
16	EACMV-[K201] B	IC16	657	100% nt identity East African cassava mosaic virus-KE2 segment DNA B, complete sequence, isolate EACMV-KE2[K201] GenBank: AJ704953.1	EACMV
17	EACMKV-[K229] A	IC17	539	99% nt identity East African cassava mosaic Kenya virus segment DNA A, complete sequence, isolate EACMKV-[K229] GenBank: AJ717578.1	EACMKV
18	EACMKV-[K229] B	IC18	659	100% nt identity East African cassava mosaic Kenya virus segment DNA B, complete sequence, isolate EACMKV-[K229] GenBank: AJ704968.1	EACMKV
19	EACMKV-[K298] A	IC19	540	99% nt identity East African cassava mosaic Kenya virus segment DNA A, complete sequence, isolate EACMKV-[K298] GenBank: AJ717572.1	EACMKV
20 ^	EACMKV-[K298] B	IC20			
21	EACMKV-[K308] A	IC21	541	99% nt identity East African cassava mosaic Kenya virus segment DNA A, complete sequence, isolate EACMKV-[K308] GenBank: AJ717574.1	EACMKV
22 ^	EACMKV-[K308] B	IC22			
23*	EACMV-Ug[K66] A	IC23	540	99% nt identity East African cassava mosaic virus DNA A complete sequence, isolate KE:mtw:CMD-MI95:12 GenBank: HG530116.1	EACMV
24#	EACMV- Ug[K66] B				
25*	EACMV-Ug[K115] A	IC25	540	99% nt identity East African cassava mosaic virus-Uganda variant segment DNA A, complete sequence, isolate EACMV-UG[K136] GenBank: AJ717520.1	EACMV-Ug
26#	EACMV- Ug[K115] B				

27	EACMV-Ug[K282] A	IC27	559	99% nt identity East African cassava mosaic virus DNA A complete sequence, isolate KE:mtw:CMD-MI88:12 GenBank: HG530113.1	EACMV-Ug
28	EACMV- Ug[K282] B	IC28	669	100% nt identity East African cassava mosaic virus-Uganda variant segment DNA B, complete sequence, isolate EACMV-UG[K282] GenBank: AJ704961.1	EACMV-Ug
29	EACMZV-[K3] A	IC29	537	99% nt identity East African cassava mosaic Zanzibar virus segment DNA A, complete sequence, isolate EACMZV-[K3] GenBank: AJ717560.1	EACMZV
30 ^	EACMZV-[K3] B	IC30			
31	EACMZV-[K10] A	IC31	539	99% nt identity East African cassava mosaic Zanzibar virus segment DNA A, complete sequence, isolate EACMZV-[K10] GenBank: AJ717567.1	EACMZV
32 ^	EACMZV-[K10] B	IC32			
33	EACMZV-[K19] A	IC33	539	99% nt identity East African cassava mosaic Zanzibar virus segment DNA A, complete sequence, isolate EACMZV-[K19] GenBank: AJ717562.1	EACMZV
34	EACMZV-[K19] B	IC34	636	100% nt identity East African cassava mosaic Zanzibar virus segment DNA B, complete sequence, isolate EACMZV-[K19] GenBank: AJ704942.1	EACMZV
35	EACMZV-[K275] A	IC35	541	99% nt identity East African cassava mosaic Zanzibar virus segment DNA A, complete sequence, isolate EACMZV-[K275] GenBank: AJ717564.1	EACMZV
36	EACMZV-[K275] B	IC36	560	100% nt identity East African cassava mosaic Zanzibar virus segment DNA B, complete sequence, isolate EACMZV-[K275] GenBank: AJ704948.1	EACMZV
37	EACMV-Ug[Cas005]-A	IC37	539	99% nt identity East African cassava mosaic virus (Uganda variant) segment A, isolate EACMV-UG [Ca055] GenBank: FN668377.1	EACMV-Ug
38	EACMV-Ug[Cas005]-B	IC38	619	100% nt identity East African cassava mosaic virus (Uganda variant) segment B, isolate EACMV-UG [Ca055]	EACMV-Ug
39	EACMV-UG[UG:Svr2:97] A	IC39	542	99% nt identity East African cassava mosaic virus DNA A complete sequence, isolate KE:mtw:CMD-MI95:12 GenBank: HG530116.1	EACMV

40	EACMV-UG[UG:Svr3:97] B	IC40	631	100% nt identity East African cassava mosaic virus-Uganda3 Severe DNA-B strain EACMV/Ug3//Svr BV1 (BV1) and BC1 (BC1) genes, complete cds GenBank: AF126807.1	EACMV
41	EACMV-CM[CI:98] A	IC41	542	99% nt identity East African cassava mosaic Cameroon virus component A, complete sequence GenBank: AF112354.1	EACMV-CM
42	EACMV-CM[CI:98] B	IC42	210	99% nt identity East African cassava mosaic Cameroon virus-[Ivory Coast] DNA-B, complete sequence GenBank: AF259897.1	EACMV-CM
43	ICMV -IN[IN:Mah:88] A	IC43	460	100% nt identity Indian cassava mosaic virus-[Maharashtra] DNA-A, complete genome, isolate ICMV-Mah GenBank: AJ314739.1	ICMV -IN
44	ICMV-IN[IN:Mah:88] B	IC44			
45	SLCMV -LK[LK:Col:98] A	IC45	261	99% nt identity Sri Lankan cassava mosaic virus-[Colombo] DNA-A, complete genome, isolate SLCMV-Col GenBank: AJ314737.1	SLCMV -LK
46	SLCMV -LK[LK:Col:98] B	IC46			
47	SACMV-[ZA] A	IC47	547	99% nt identity South African cassava mosaic virus segment A, complete sequence GenBank: AF155806.1	SACMV
48	SACMV-[ZA] B	IC48			

***** Plasmid miniprep DNA

[#] Empty tubes (no sample provided)

^ Amplified cloning vector using primers PCRc1/PBL1v2040 (Rojas et al., 1993)

[+] Did not amplify using primers PCRc1/PBL1v2040 (Rojas et al., 1993)

Discussion

Partial DNA-A sequencing showed that all IC's analyzed matched the exhibit list provided.

Full characterization is required to verify that infectious clones are complete (at least full-length, if not partial repeats) and intact. These will then be used for bombardment experiments utilizing an RCA-mediated approach whereby RCA-derived multimeric products will be introduced into host plants via biolistic inoculation using the Helios gene gun system.

DNA-B components will be characterized using degenerate PCR primers PBL1v2040 5'-GCCTCTGCAGCARTGRTCKATCTTCATACA-3′ and PCRc1 5'-CTAGCTGCAGCATATTTACRARWATGCCA-3′ (Rojas et al., 1993). Full-length sequencing will be done using primers EB03/4 and VNF021/22 (Fondong et al., 2000).

Full length DNA-As will be sequenced using primers VNF003/4 and VNF007/8 (Fondong et al., 2000).

4.8 Objective 8: *Final analysis and publication.*

Throughout this project, results were continually presented at Australian research meetings and in industry publications. In total, eight separate presentations/posters were given for the work done in this project, and two full length industry articles were published in the Cotton Info Spotlight Magazine.

The final analysis and manuscript development has been delayed due to the changes in personnel on the project. We have one advanced draft of a paper detailing the development and application of new tools for the metagenomics detection of viruses and their vectors, that is due to be submitted to the journal in February 2018.

(1) van Brunschot S, Persley D, Dennien S, Hereward J, McMichael, L, Thomas JE, Walter G (2018). Viral metagenomics: a tool for virus discovery and analysis of diversity in plant and insect vectors. Viruses MDPI.

An additional manuscript is in its early stages (2), however the data and analysis for this paper are now complete. This manuscript is due for submission to the journal in May 2018.

(2) van Brunschot SL, Gambley C, Stainton D, Hereward JP, Walter GH (2018). A spatio-temporal analysis of an insect and its endosymbionts. Insect Science.

Two further publications are planned from the work done in this project, which are related to (1) Evidence for the co-divergence of the primary bacterial endosymbiont and host cryptic species of the whitefly *Bemisia tabaci*, and (2) Analysis of host-use by *Bemisia tabaci* between cotton cropping seasons: bridging hosts that contribute to pest outbreaks.

5. Outcomes

Describe how the project's outputs will contribute to the planned outcomes identified in the project application. Describe the planned outcomes achieved to date.

- We have documented the diversity of *Bemisia tabaci* in Australia revealed from our sampling and historical samples collected across the country. During this project, we have documented for the first time the presence of *B. tabaci* ASIAII populations in Australia and have also

identified a new population which we have tentatively named *B. tabaci* AUS3 in the Torres Strait Islands. This fundamental knowledge is essential for evaluating risks associated with the potential introduction of cotton leaf curl virus.

- We have documented the diversity of primary and secondary bacterial endosymbionts present in Australian *B. tabaci* populations, using both a shallow screening approach to examine large numbers of whitefly, and also a deep analysis using next-generation sequencing to sequence complete genomes of these bacteria. These data will help to inform future research initiatives targeting the control of whiteflies and whitefly-transmitted viruses. This future research will benefit industry as it could offer a cost-effective and environmentally sound alternative pest management approach.

- We have tested the ability of Australian whiteflies to transmit cotton leaf curl viruses. Our research results should assist in a re-assessment risk of establishment of CLCuD, should the virus enter Australia. Our work will help to inform future biosecurity prevention and response programs, to ultimately better prepare cotton growers for the potential incursion of CLCuD.

- This project has developed and evaluated improved diagnostic approaches for the detection of whiteflies and whitefly-transmitted viruses. This includes both conventional approaches, novel metagenomic approaches, and also a draft National Diagnostic Protocol. Importantly, this project has provided further training of two skilled plant pathologists (Sharon van Brunschot and Daisy Stainton), and new diagnostic tests for detecting the viruses and vectors responsible for CLCuD epidemics. The outcomes of this work will be submitted for publication in February 2018, and the National Diagnostic Protocol can be further drafted and completed.

- This project provided continued expert support to other CRDC researchers (Cherie Gambley, Murray Sharman and Jamie Hopkinson). We performed molecular identification of whiteflies (and the viruses they transmit), as well as routine virus indexing of plant samples collected during surveys in Australia as part of their CRDC projects. This work supported their projects and enabled appropriate biosecurity actions to be taken for samples tested.

- Research was done to support the activities of the African Cassava Whitefly Project (ACWP; Bill and Melinda Gates) at the Natural Resources Institute, University of Greenwich, as a joint postdoc under the supervision of Prof John Colvin. This whitefly-focussed research was mostly focussed on biological assays examining the ability of different whiteflies to detoxify plant defense compounds. In addition, the role was to provide training and mentorship to junior staff and PhD students on the project. This joint post-doc role enabled the strengthening of collaborative relationships between CRDC, UQ and the Natural Resources Institute.

5.1 Discuss the technical advances achieved (eg commercially significant developments, patents applied for or granted licenses, etc.);

Technical advances achieved in this project include the development of novel diagnostic technologies for the detection and identification of whiteflies and whitefly-transmitted viruses. Once it is up taken in routine surveillance strategies in Australia, these technologies will provide a more cost-efficient, high-throughput, reliable and robust approach for identifying these damaging vectors and the viruses they transmit. Early detection is always the cornerstone of good biosecurity, so therefore improving our changes of detecting an incursion before it becomes established, could have enormous cost benefits to the Australian cotton industry.

5.2 Discuss any other information developed from research (eg discoveries in methodology, equipment design, etc.); and

- We developed novel next-generation sequencing methods for examining whiteflies, the viruses they transmit and the bacterial endosymbionts they harbour. This includes a novel metagenomic approach for the surveillance of whitefly-transmitted viruses (and their vectors) in-field.

- We have developed novel next-generation sequencing approaches for single whiteflies, a technique applicable for the sequencing of complete mitochondrial genomes of small insects.
- We have developed new methods for high-throughput screening and identification of whiteflies in Australia.

- We have established and developed techniques and methodologies for maintaining pure cultures of multiple cryptic species of B. tabaci in Australia (with no contamination).

- We have further developed conventional diagnostic assays for the detection of cotton leaf curl viruses.

- We have developed a draft National Diagnostic Protocol for whitefly in Australia.

5.3 Discuss any required changes to the Intellectual Property register.

Not applicable.

6. Conclusion

6.1 Provide an assessment of the likely impact of the results and conclusions of the research project for the cotton industry. What are the take home messages?

- *Improved knowledge on biosecurity risks*: This study demonstrated the ability of Australian whiteflies to transmit cotton leaf curl viruses with a similar efficiency to their exotic whitefly vector species. These data will be communicated to industry and the wider scientific community in order to provide a basis for a re-examination of the risk analysis for the introduction of cotton leaf curl disease into Australia.

- *Improved knowledge on the diversity of whitefly in Australia*: To date, little has been known of the diversity of *B. tabaci* affecting cotton and horticultural crops in Australia. We have documented the diversity of *B. tabaci* populations across mainland Australia, from the early 1990s until 2017. Knowledge of the identity and diversity of these populations, across hosts and geographic regions, will provide invaluable information to assist in decision making processes for managing these pests. This is particular true in the context of cross-sector management, in geographic regions where multiple crops are grown but are affected by the same *B. tabaci* species.

- *The development of novel technologies for the improved detection of whiteflies and the viruses they transmit:* Various methods and approaches exist for the detection of *B. tabaci* and whitefly-

transmitted viruses, however a thoroughly road-test and unified approach is presented in the work generated in this project. Cost efficient, high-throughput, reliable methods have been developed for routine surveillance and identification. In addition, next-generation technologies have been developed that utilise next-generation sequencing technologies for simultaneously detecting whiteflies and the viruses they carry, even in asymptomatic crops.

- *Documented baseline data on the diversity of bacterial endosymbionts present in Australian whiteflies:* We have examined the diversity of primary and secondary bacterial endosymbionts present in Australian *B. tabaci* populations, using both a shallow screening approach to examine large numbers of whitefly, and also a deep analysis using next-generation sequencing to sequence complete genomes of these bacteria. These data will help to inform future research initiatives targeting the control of whiteflies and whitefly-transmitted viruses. This future research will benefit industry as it could offer a cost-effective and environmentally sound alternative pest management approach.

- *Training of early career scientists:* This project supported two early career scientists (Dr Sharon van Brunschot and Dr Daisy Stainton) to further develop their research, project management and supervision skills. Importantly, this project has boosted the human capacity available to the Australian cotton industry to continue preparing for the potential introduction of the devastating cotton leaf curl disease.

6.2 Extension Opportunities

Detail a plan for the activities or other steps that may be taken:

6.2.2 Detail plans to further develop or to exploit the project technology.

Diagnostic technologies

- I suggest that holding a meeting (Prof Walter, Susan Maas, Sharon van Brunschot) would be a useful start to discuss the outcomes of the project in relation to diagnostic development, to develop a plan together for extension of these protocols to state government personnel and laboratories for use.

- Discussions are needed for further refinement of the draft National Diagnostic Protocol and how/when this can be done.

6.2.3 Details plans for the future presentation and dissemination of the project outcomes, and for future research.

Publications

- As outlined in section 4.8, a publication plan is in place to ensure the communication of the results of this project to the wider scientific community. This plan has been developed in conjunction with Prof Walter who will provide support in the preparation of manuscripts.

Conference

- An international whitefly conference is being held in Perth in 2018, and could be a good opportunity for both Dr van Brunschot and also PhD student Wanaporn Wongnikong to present their whitefly research to an international audience.

Future research

- A successive project UQ1801 was successfully commenced in 01/01/2017, which is a mostly desktop study investigating novel transgenic approaches to control silverleaf whitefly. This pilot project will investigate a novel and sustainable solution, based on RNAi technology, to control SLW and therefore protect the industry from CLCuD. Together with researchers at the

University of Greenwich (UK), we will assess transgenic approaches and other developing technologies for producing whitefly-resistant plants, perform research to demonstrate feasibility for use and identify pathways for adoption and use in Australian cotton.

- I would like to build on UQ1801 by strengthening partnerships with Prof P.K. Singh from the National Botanical Research Institute (Lucknow, India). Prof Singh recently presented his impressive research at an ACWP meeting, where he has developed two types of transgenic cotton, resistant to whitefly. Prof Singh is very keen to collaborate with research partners in Australia, I think this could be a unique opportunity for the Australian cotton industry to investigate novel approaches for protecting their industry into the future from the potentially devastating CLCuD.

- In the beginning of 2018, I will be co-supervising Dr Pan Li-Long, a postdoctoral researcher from Professor Liu Shu-Sheng's lab (Zhejiang University, China). Dr Pan will join the Natural Resources Institute and will be devoting a 12-month period to examining the "Transmission of cotton leaf curl disease related begomoviruses by different species of the *Bemisia tabaci* whitefly complex". He will benefit from the knowledge I gained during this project, and will extend on my research (and Australian whiteflies will be used in his project). This work will be published jointly and benefit the Australian cotton industry, as he will use the most damaging cotton leaf curl viruses as his study model (*Cotton leaf curl Burewala virus*).

9. Publications

List the publications arising from the research project and/or a publication plan.

9.1 Refereed journal publications

Published

Mabvakure B, Martin DP, Kraberger S, Cloete L, van Brunschot S, Geering AD, Thomas JE, Bananej K, Lett J-M, Lefeuvre P (2016) Ongoing geographical spread of *Tomato yellow leaf curl virus*. Virology 498:257-264

Under review

Malka O, Feldmesser E, Santos-Garcia D, Sharon E, Krause Sakate R, Delatte H, **van Brunschot S**, Patel M, Mugerwa H, Seal S, Colvin J, Morin S (2017). Species-complex evolution and host-plant associations in *Bemisia tabaci*: a plant-defense, detoxification perspective revealed by RNAseq analyses. Molecular Ecology.

In preparation

van Brunschot S, Persley D, Dennien S, Hereward J, McMichael, L, Thomas JE, Walter G (2018). Viral metagenomics: a tool for virus discovery and analysis of diversity in plant and insect vectors. Viruses MDPI.

van Brunschot SL, Gambley C, Stainton D, Hereward JP, Walter GH (2018). A spatio-temporal analysis of an insect and its endosymbionts. Insect Science.

Wongnikong W, van Brunschot S, De Barro P.J., Walter G (2018). Testing mate recognition through reciprocal crosses of two native populations of the whitefly Bemisia tabaci (Gennadius) in Australia. Journal of Insect Behavior.

9.2 Thesis

Not applicable.

9.3 Conference seminars and presentations

Vyskocilova S, Visendi P, van Brunschot S, Seal S, Colvin J. Molecular, biological and ecological assessment of cryptic species diversity within the Mediterranean species of the *B. tabaci* complex. Proceedings of the 1st SciPlant Conference, 26th-28th September 2017, Brisbane, Queensland, Australia. See Appendix 1.

Stainton D, van Brunschot S, Hereward J, Walter G. (2017) Invasive and native whiteflies in Australia: cryptic species complex, endosymbionts and begomoviruses. Proceedings of the International Fuscom Meeting, 8th-9th August, Toowoomba, Queensland, Australia.
See Appendix 6.

Gambley C, Campbell P, Grundy P, Sharman M and van Brunschot S. (2016) Cotton leaf curl disease: how prepared are we? Proceedings of the International Fuscom Meeting, 12th-13th April, Toowoomba, Queensland, Australia.

See Appendix 8.

van Brunschot S, Walter G, Hereward J, Gambley C, De Barro P, Seal S, and Colvin J (2016). Whiteflies and begomoviruses: an Australian perspective. Plant Health Australia Workshop on Whitefly and Whitefly-Transmitted Viruses, 1st September, Brisbane, Queensland, Australia.
See Appendix 9.

van Brunschot SL, Walter G, Hereward J, Gambley C, Campbell P, De Barro P, Howie L (2014) Investigating the interactions of viruses, whiteflies and endosymbionts. Proceedings of the International Fuscom Meeting, 3rd-5th November, Toowoomba, Queensland, Australia.

See Appendix 11.

van Brunschot SL, Hereward J, Gambley C, De Barro P, Walter G (2014) Investigating the interactions of viruses, vectors and endosymbionts. Proceedings of the 11th Australasian Plant Virology Workshop, 13th-15th August, Brisbane, Queensland, Australia.

See Appendix 12.

van Brunschot SL, Hereward J, Gambley C, De Barro P, Walter G (2013) Viruses, vectors and endosymbionts: exploring interactions for the control of cotton leaf curl disease. Cotton Science Conference, 8th-11th September, Narrabri, New South Wales, Australia.

See Appendix 14.

van Brunschot SL, Hereward J, Gambley C, De Barro P, Walter G (2013) Exploring the interactions of whiteflies, endosymbionts and Cotton leaf curl virus. Northern Farming Systems IPM Researchers Forum, 30th July, Toowoomba, Queensland, Australia.

See Appendix 15.

9.4 Cotton industry reports

Two research updates were produced for this project, and published by the Cotton Info Team in their Spotlight Magazine (See Appendix 10 and 13).

9.5 Online Resources

Two research updates were published online, and were made available by the Cotton Info Team at their Spotlight Magazine online resource.

Links are provided below:

1. http://www.cottoninfo.com.au/publications/spotlight-magazine-autumn-2015
2. http://www.cottoninfo.com.au/publications/spotlight-magazine-winter-2014

10. Final Report Executive Summary

Provide a one page Summary of your research that is not commercial in confidence, and that can be published on the World Wide Web. Explain the main outcomes of the research and provide contact details for more information. It is important that the Executive Summary highlights concisely the key outputs from the project and, when they are adopted, what this will mean to the cotton industry.

Cotton leaf curl disease (CLCuD) presents a major biosecurity threat to the Australian cotton industry because it can decimate production. The disease is caused by a complex of one or more begomoviruses, and is spread by the whitefly (*Bemisia tabaci*). Devastating epidemics of CLCuD have been reported from a number of countries across Africa and southern Asia – an epidemic during the 1990's in Pakistan resulted in yield losses of more than US$5 billion in value.

This overarching aims of this project were to investigate the capacity of the whiteflies present in Australian cotton to transmit viruses of the CLCuD complex, the diversity of Australian whiteflies and their endosymbionts, and to develop diagnostic tools so that the risk of an incursion can be assessed and the necessary preparations made. To further support preparedness, virus diversity in Australia's northern neighbours was evaluated using new diagnostic tests and sequencing technologies.

In collaboration with the University of Greenwich (Prof John Colvin, leader of the Bill and Melinda Gates funded African Cassava Whitefly Project), we were able to test the ability of Australian whiteflies to transmit cotton leaf curl virus (CLCuV) to Australian cotton cultivars. This research was done in their secure quarantine facilities in the United Kingdom, with no risk to the Australian cotton industry. We found that of the three Australian putative cryptic species of *B. tabaci* tested, all were able to transmit CLCuV with a similar efficiency to the cognate exotic whitefly vector species.

Using a combination of biological, molecular and ecological approaches, we have created the most comprehensive picture of the diversity of whiteflies, their microbial endosymbionts and the viruses they transmit in Australia. In total, we documented five putative cryptic species of *B. tabac*i, of which two were new to science, collected over a 20-year time period. We have examined the diversity ofprimary and secondary bacterial endosymbionts present in Australian *B. tabaci* populations, using both a shallow screening approach to examine large numbers of whitefly, and also a deep analysis using next-generation sequencing to sequence complete genomes of these bacteria. These data will help to inform future research initiatives targeting the control of whiteflies and whitefly-transmitted viruses. This future research will benefit industry as it could offer a cost-effective and environmentally sound alternative pest management approach.

We developed novel next-generation sequencing methods for examining whiteflies, the viruses they transmit and the bacterial endosymbionts they harbour. This includes a novel metagenomic approach for the surveillance of whitefly-transmitted viruses (and their vectors) in-field. In addition, we have developed new methods for high-throughput screening and identification of whiteflies in Australia and novel next-generation sequencing approaches for single whiteflies, a technique applicable for the sequencing of complete mitochondrial genomes of small insects.

In summary, this project has generated an enormously useful resource of biological and genetic information, relating to virus transmission by whiteflies, whitefly and endosymbiont diversity, and the development of novel diagnostic technologies for the efficient and reliable detection and identification of whiteflies and the viruses they transmit. The data generated in this project will support decision making and risk analysis around exotic whiteflies and whitefly-transmitted viruses in Australia, and importantly will inform decision-making processes for managing endemic pest populations. The human capacity and expertise built in this project, combined with the improved diagnostic technologies, will significantly increase the preparedness of the Australian cotton industry for the potential introduction of cotton leaf curl disease.

11. Appendices

Appendix 1: Conference poster

Vyskocilova S, Visendi P, van Brunschot S, Seal S, Colvin J. Molecular, biological and ecological assessment of cryptic species diversity within the Mediterranean species of the *B. tabaci* complex. Proceedings of the 1st SciPlant Conference, 26th-28th September 2017, Brisbane, Queensland, Australia.

Appendix 2: Peer-reviewed journal article (in preparation)

van Brunschot S, Persley D, Dennien S, Hereward J, McMichael, L, Thomas JE, Walter G (2017). Viral metagenomics: a tool for virus discovery and analysis of diversity in plant and insect vectors. Viruses MDPI.

Viral metagenomics: a tool for virus discovery and analysis of diversity in plant and insect vectors.

Sharon van Brunschot [1, 2, 3]
Sandra Dennien [3]
James Hereward [1, 2]
Denis Persley [3]
Lee McMichael[3]
John E Thomas[3]
Gimme Walter [1, 2]

[1] School of Biological Sciences, The University of Queensland
[2] Cotton Research and Development Corporation
[3] Horticulture and Forestry Science, Agri-Science Queensland, Department of Agriculture and Fisheries

Abstract

Using next-generation sequencing approaches, we identified mixed infections of *Sweet potato leaf curl virus* (SPLCV), *Sweet potato leaf curl Spain virus* and *Sweet potato leaf curl China virus* in sweet potato samples from two locations in Australia. In addition, we identified partial genome sequences of *Sweet potato badnavirus B* in a single sample, which may indicate the presence of active infections of this virus, or alternatively an integrated form. We also report on the use of vector–enabled metagenomics to survey begomoviruses directly from whiteflies collected across crop plants from a single geographic region. Simultaneously, we detected SPLCV and the identity of the whitefly vector *Bemisia tabaci* ASIAII, in field-collected whitefly. This study describes the first report of sweepoviruses in Australia, and the first use of VEM to study virus diversity via insect populations in Australia.

Introduction

Sweetpotato (*Ipomoea batatas*, family *Convolvulaceae*) is a nutritious food crop. The Australian sweet potato industry achieves the highest yields of commercial production systems anywhere in the world, with a farm gate value estimated at approximately $90 million (Henderson, 2015).

Queensland is the biggest producer with over 70% of production, centered mainly on Bundaberg. The second major producing area is around Cudgen in northern New South Wales, with remaining producers situated in Mareeba, Atherton and Rockhampton (QLD), Murwillumbah (NSW), Perth, Carnarvon and Kununurra (WA) (Australian Sweetpotato Growers Inc, 2016).

One of the major limitations to production worldwide, is the cumulative effect of virus infection leading to yield decline in this vegetatively propagated crop. More than 30 viruses are known to infect sweet potato, from the genera *Badnavirus*, *Begomovirus*, *Carlavirus*, *Cavemovirus*, *Crinivirus*, *Cucumovirus*, *Enamovirus*, *Ipomovirus*, *Nepovirus*, *Potyvirus*, and *Tospovirus* (Valverde et al., 2007, Cuellar et al., 2011, Kreuze et al., 2009, Kashif et al., 2012). In addition to vegetative propagation, important insect vectors associated with the spread of these viruses include whitefly (*Bemisia tabaci*), thrips and aphids. The severity of symptom expression of many of these viruses is highly variable depending on the host sweet potato cultivar, virus species and environmental conditions.

Currently, there is a paucity of information relating to the diversity of viruses infecting sweet potato in Australia. The most common and widespread virus infecting sweet potato is *Sweet potato feathery mottle virus* (genus *Potyvirus*) (Persley & Coleman, 2010). In addition, *Sweet potato virus 2* (genus *Potyvirus*) has also been described from Western Australia (Tairo et al., 2006).

The aims of this research were to use a broad molecular strategy to detect and identify the presence of sweet potato begomoviruses (sweepoviruses) in two *Ipomoea* sp. plants collected from Queensland and Western Australia. Concurrently, we also trialed the use of a new research strategy, termed vector-enabled metagenomics (VEM), to survey for circular DNA viruses present in *Bemisia tabaci* collected across various crop hosts in Kununurra, Western Australia.

Methods

Samples

Plant material and DNA extraction

Leaf material was collected and immediately lyophilized for preservation, then stored at −20°C (Table 1). Total nucleic acids were extracted from 0.01 g of lyophilized leaf tissue using a TissueLyser bead mill (QIAGEN, Valencia, CA, USA) and a BioSprint 15 DNA Plant Kit (QIAGEN), both used according to the manufacturer's instructions.

Whitefly samples and DNA extraction

Adult *B. tabaci* were collected from various crops in Kununurra, Western Australia, using a sweep net and aspirator. Samples were immediately preserved in 90-95% ethanol (Table 2).

Nucleic acids from single *B. tabaci* adults were extracted separately using a high-throughput silica spin column plate method developed by James Hereward (Walter Lab, UQ). Briefly, single whiteflies were placed in 1.5 ml eppendorf tubes containing 0.5 mm zirconium oxide beads, and residual ethanol was allowed to evaporate for 30 mins. Samples were homogenized in a Bullet Blender (Next Advance, Averill Park, NY) for 4 minutes at setting 8, in lysis buffer (0.5% SDS, 250 mM NaCl, 25 mM EDTA, 200 mM Tris-Hcl pH7.5, 0.5 mg/ml proteinase K). Homogenized samples were incubated overnight at 55°C in a shaking incubator. Precipitation buffer (4M Ammonium acetate) was added and samples mixed by inversion. Binding buffer was then added (4M GuHCl, 75% Ethanol), and samples were loaded onto silica spin plates and spun (~13500 rpm) using Sigma Centrifuge 4-15C, for 4 mins. The plate was washed twice with wash buffer (10 mM Tris-HCl pH7.5, 80% ethanol), with the final spin extended to 15 mins to dry the plate. Nucleic acids were eluted from the plate by the addition of pre-warmed elution buffer (10mM Tris-Cl, pH 8.5), in two separate additions. All extracts were stored at −20°C.

TABLE 1: Plant samples used in this study

QDAF Virology accession number	Host plant	Place of Collection	City	State	Country	Date of collection	Collector	Notes and diagnostic information
5177	*Ipomoea batatas* scions grafted onto rootstock of *Ipomoea setosa*	Gatton Research Station	Gatton	Queensland	Australia	29/05/2015	S. Dennien	This cultivar of *I. batatas* is often asymptomatic (from Sandra) Previous samples of this cultivar were subjected to virus diagnostic testing, including: (1) Nitrocellulose membrane ELISA (Sandra Dennien) gave negative results. (2) Degenerate begomovirus PCRs done by Lee McMichael 2008/2009, positive results for begomovirus (SPGprimers). (3) Positive for SPLCV using qPCR (Gatton, Sandra).
5624	*Ipomoea batatas*	Peter Pegg's Farm	Kununurra	Western Australia	Australia	10/09/14	S. L. van Brunschot	Healthy field crop. *Bemisia tabaci* present in crop at low level, likely an Australian native species due to low incidence. RCA + BamHI restriction digestion 5/06/2015 Book 3 – strong RFLP pattern, consistent with begomvorius infection.

Table 2: Whitefly samples used in this study

ID	Description	Species*	Host Plant	Place of Collection	City	State	Country	Date of Collection	Collector	Additional Notes
216	*Bemisia tabaci*	AUSII	*Solanum lycopersicum,* Tomato	Frank Wise Institute	Kununurra	Western Australia	Australia	9/09/2014	S. L. van Brunschot	KNX5, young tomatoes green, v few whitefly photos 173,4,5,6,7.
217	*Bemisia tabaci*	AUSII	*Solanum lycopersicum,* Tomato	Frank Wise Institute	Kununurra	Western Australia	Australia	9/09/2014	S. L. van Brunschot	KNX6 older, full fruit set, a failed *Alternaria* sp. Trial, so very healthy.
218	*Bemisia tabaci*	AUSII	*Salvia hispanica,* Chia	Frank Wise Institute	Kununurra	Western Australia	Australia	9/09/2014	S. L. van Brunschot	KNX8 very few whiteflies, photos 178, 9, 80.
219	*Bemisia tabaci*	AUSII	*Cucumis melo reticulatus*, Rockmelon	Arcadia Farms, Mulligans Rd	Kununurra	Western Australia	Australia	10/09/2014	S. L. van Brunschot	KNX14 very few whiteflies.
220	*Bemisia tabaci*	AUSII	*Phaseolus vulgaris,* Green beans	Peter Peggs Farm	Kununurra	Western Australia	Australia	10/09/2014	S. L. van Brunschot	KNX15, there were no whiteflies on tomato or capsicum nearby.
221	*Bemisia tabaci*	AUSII	*Ipomoea batatas,* Sweet potato	Peter Peggs Farm	Kununurra	Western Australia	Australia	10/09/2014	S. L. van Brunschot	KNX17, larger population for KNX (relatively), Photos 238/9/40/41/42/43.
222	*Bemisia tabaci*	AUSII	*Cicer arietinum,* Chickpea	Oasis Farm	Kununurra	Western Australia	Australia	11/09/2014	S. L. van Brunschot	KNX20B, moderate density (relatively), 235/6/7.
223	*Bemisia tabaci*	AUSII	*Cucurbita moschata,* Butternut pumpkin	Oasis Farm	Kununurra	Western Australia	Australia	12/09/2014	S. L. van Brunschot	KNX21, v low density, healthy pumpkin.
224	*Bemisia tabaci*	AUSII	*Salvia hispanica,* Chia	Frank Wise Institute	Kununurra	Western Australia	Australia	11/09/2014	S. L. van Brunschot	KNX24, fairly high density (relatively).
225	*Bemisia tabaci*	AUSII	*Gossypium hirsutum,* Cotton (GM)	Frank Wise Institute	Kununurra	Western Australia	Australia	12/09/2014	S. L. van Brunschot	KNX25, fairly high density, though less than chia and sweet potato, photo 281.

227	*Bemisia tabaci*	AUSII	*Gossypium hirsutum,* Cotton (GM)	Frank Wise Institute	Kununurra	Western Australia	Australia	12/09/2014	S. L. van Brunschot	KNX31, v low density of whiteflies.

* Species confirmation done by Sanger sequence confirmation using COI barcoding

NGS library preparation

Total nucleic acid extracts of plant samples 5177 and 5264 were separately amplified by rolling-circle amplification (RCA) with random priming using Phi29 DNA polymerase (TempliPhi™, GE Healthcare, Uppsala, Sweden), to enrich for small circular templates such as begomovirus genomes.

For VEM, total nucleic acids extracted from single whitefly were pooled, concentrated, then subjected to RCA. Briefly, 5 ul of each extracts from single whiteflies from each location (11 whitefly in total, Table 2) were combined and nucleic acids concentrated using Agencourt AMPure XP Purification beads (Beckman Coulter), using a 2x protocol. DNA bound to beads was eluted in 10 ul of elution buffer (10mM Tris-Cl, pH 8.5). To account for amplification bias, 1 ul of concentrated DNA was used as the template for two separate RCA reactions, done using double volumes for optimal amplification. Resulting RCA reactions were combined. Then, 1 ul of the RCA-amplified concatamers were digested with *Bam*HI, and a RFLP pattern was observed indicating successful RCA amplification.

Multiplexed NGS libraries were and prepared in-house using the NEBNext Ultra DNA Library Prep Kit (New England Biolabs). Libraries sequenced in this work were part of a 22-plex multiplexed library. RCA products mentioned previously were kept separate, and subjected to shearing by sonication using an ultrasonic waterbath for 10 minutes, to generate sheared fragments of ~ 500 bp. Sheared RCA fragments were subjected to size selection (400-600 bp) using a 2% cartridge of a Pippin Prep™ electrophoresis instrument (ThermoFisher Scientific). Size-selected fragments were then subjected to multiplexed library preparation (Dual Index Primer Set 1 and 2, New England Biolabs) according to manufacturer's instructions, with minor modifications. Library purifications were performed using Agencourt AMPure XP beads (Beckman Coulter). Purified libraries were quantified using a Microchip Electrophoresis System (Simadzu) and Qubit fluorometer (ThermoFisher Scientific). Equimolar pools of each library were mixed, then mixed libraries were size selected again using the PippinPrep Instrument (400-600 bp) and purified using 1x Agencourt AMPure XP beads. Eluted size-selected pooled libraries were again quantified using a Qubit fluorometer, with concentration adjusted to 10 nM. The multiplexed library was sequenced by a commercial facility using an Illumina Miseq system (250 bp paired end version 2 sequencing chemistry, Illumina).

PCR assay and conventional Sanger sequencing to confirm whitefly species 96

The mitochondrial COI gene of *B. tabaci* samples (Table 2) were amplified using the PCR primers C1-J-2195/ TL2-N-3014 (Simon et al., 1994), using optimized PCR conditions. Amplicons were subjected to conventional Sanger sequencing by Macrogen (Korea).

Metagenomic data analysis and genome assembly

Approximately 1 million metagenomic reads were obtained for each multiplexed library sequenced. Raw read quality was assessed using FastQC (Andrews, 2010). Adapter sequences and low- quality reads were trimmed using Trimmomatic (Bolger et al., 2014).

De novo assemblies were performed using Geneious version R7 (Kearse et al., 2012). Consensus sequences (0% threshold) from resulting contigs were compared against the GenBank non-redundant database using MegaBLAST (e-vale <0.001). Matches were perused manually and matches to plant

viruses (<70% nt identity) were subsequently used to identify representative sequences to use as scaffolds for map-to-reference assembly strategies. Mapping and alignment to sequence scaffolds were done using Bowtie2 and Geneious assemblers (end-to-end and local recruitment), implemented in Geneious. Consensus sequences were generated using 0% threshold (most prevalent nt at each position) and compared, with manual editing where necessary. For virus sequences, this generated a single representative sequence from the pool of quasi species present in the virus population. The reliability of assemblies was checked by back-mapping reads to created consensus sequences. Details of genome assemblies, including comparisons to reference sequences (nt and amino acid comparisons) are presented in Table 3 and Figures 1 and 2.

Virus genome sequence analysis

Pairwise comparisons

All pairwise comparisons were performed using the MUSCLE algorithm (Edgar, 2004) implemented in Genious v7, with manual checking and editing where necessary.

Phylogenetic inference

A maximum likelihood (ML) phylogenetic tree was constructed to infer relationships between DNA-A genomes of novel begomovirus species sequences reconstructed in this work, with described species from GenBank (n=x). For this purpose, sequences of relevant vetted species of sweepoviruses, and related Old World and New World begomoviruses. A ML phylogenetic tree was then reconstructed using RAxML v8.2.x using the GTR GAMMA substitution model (Stamatakis, 2006) and 1000 bootstrap replicates to assess branch support.

Representative sequences of the global genetic diversity of *Bemisia tabaci* documented in the literature to date were used for phylogenetic reconstruction. Sequences were aligned using MUSCLE. The best-fit model of evolution (GTG+R) was determined from likelihood ratio tests performed using jModelTest (Posada, 2008). Phylogenetic relationships were reconstructed via Bayesian analyses using MRBAYES v3.1.2 (Ronquist & Huelsenbeck, 2003) implemented in Geneious v7 (Kearse et al., 2012). Four Markov chains were run for a total of 1,100,000 generations, with chains sampled every 200 generations (the first 10 % of trees were discarded as burn-in). Posterior probabilities ≥ 0.95 were considered substantial node support. TRACER 1.5 (Rambaut & Drummond, 2003) was used to check that the effective sample size was not too low, indicating that the MCMC mixed well and that the samples were independent. Sequences from close relatives of these taxa, *Bemisia atriplex*, were included as outgroups.

Results

Sequence analysis of full-length sweepovirus genomes from plant samples

Six full-length monopartite DNA-A genomes of sweepoviruses were obtained through NGS of three separate libraries of circular DNA enriched templates (Figure 1). BLAST analysis of assembled genome sequences revealed high genome-wide nucleotide similarity with sweepovirus representative

sequences. Using the current begomovirus species demarcation criteria of 91% genome-wide pairwise identity (Brown et al., 2015, Adams et al., 2014), we have attributed putative species names to the virus genomes recovered (Figure 1, Table 3).

5177 (Queensland)

The complete nucleotide sequences of two distinct begomovirus species were identified. Full-length DNA-A genomes of these monopartite begomoviruses were assembled (Table 3). Genome 5177_Begomovirus_1 shared 98% genome-wide pairwise identity with the DNA-A of *Sweet potato leaf curl Spain virus* (EF456743). Genome 5177_Begomovirus_2 shared 97% genome-wide pairwise identity with the DNA-A of *Sweet potato leaf curl virus* (JX961672). Genome organization, and pairwise identity of gene regions (as compared to representative reference sequences) are shown in Figure 1. Depiction of sequence read coverage is shown in Figure 2.

The genome organization of these begomovirus DNA-A's was typical of Old World begomoviruses, and shared high nucleotide identity with representative sequences of each species as indicated in Figure 1.

Using the de novo assembly approach, no contigs were produced matching DNA-B, alphasatellite, betasatellite nor the newly discovered deltasatellites (Lozano et al., 2016).

Using the mapping assembly approach, no reads mapped to scaffolds representing DNA-B, alphasatellites, betasatellites or deltasatellites.

The majority of sequence reads mapped to the chloroplast genome of *Ipomoea batatas* (>20 000 nt, ~99% nt identity), and also the mitochondrial genome of *Nicotiana tabacum* (~3450 nt, 98% match), data not shown.

5624 (Western Australia)

The complete nucleotide sequences of three distinct begomovirus species were identified. Full-length DNA-A genomes of these monopartite begomoviruses were assembled (Table 3). Contig 5264_Begomovirus_3 shared 95% genome-wide pairwise identity with the DNA-A of *Sweet potato leaf curl Spain virus* (EF456743). Contig 5264_Begomovirus_4 shared 99% genome-wide pairwise identity with the DNA-A of *Sweet potato leaf curl virus* (JX961672). Contig 5264_Begomovirus_5 shared 95% genome-wide pairwise identity with the DNA-A of *Sweet potato leaf curl China virus* (JX91673). Genome organization, and pairwise identity of gene regions (as compared to representative reference sequences) are shown in Figure 1. Depiction of sequence read coverage is shown in Figure 2.

The genome organization of these begomovirus DNA-A's was typical of Old World begomoviruses, and shared high nucleotide identity with representative sequences of each species as indicated in Figure 1.

Using the de novo assembly approach, no contigs were produced matching DNA-B, alphasatellite, betasatellite or deltasatellites.

Using the mapping assembly approach, no reads mapped to scaffolds representing DNA-B, alphasatellites, betasatellites or deltasatellites.

The majority of sequence reads mapped to the chloroplast genome of *Ipomoea batatas* (>18 500 nt, ~99% nt identity), data not shown.

Table 3: Virus and whitefly genome sequences identified in this study.

Library	Contig name	Contig Length	Number of mapped sequence reads	Mean insert size (min-max)	Coverage at each nt position (min-max)	Best match (% identity)/ accession number/origin)
5177	NGS Begomovirus 1	2778 (full genome)	48107	414 (207-621)	189-11448	*Sweet potato leaf curl Spain virus* SPLCESV [ES:CI:BG5:02] (98%/ EF456743/Spain)
5177	NGS Begomovirus 2	2829 (full genome)	63493	383 (192-574)	160-14501	*Sweet potato leaf curl virus* SPLCV-US [KR:Hae1:09] (97%/JX961672/Korea)
5177	NGS BadnaB 1	750 (partial gene)	4863	439 (220-658)	31-697	Sweet potato badnavirus B isolate Yanshu1 China 74(2); partial match to polyprotein gene cds (85%/KM009092/Yanshu, China)
5177	NGS BadnaB 2	1722 (partial gene)	3332	438 (219-657)	53-596	Sweetpotato badnavirus B isolate TZ:BK59:12; partial match to hypothetical protein gene cds and polyprotein gene junction (77%/ KF836892/Tanzania)
5264	NGS Begomovirus 3	2778 (full genome)	13712	421 (211-631)	101-3403	*Sweet potato leaf curl Spain virus* SPLCESV [ES:CI:BG5:02] (95%/ EF456743/Spain)
5264	NGS Begomovirus 4	2829 (full genome)	54736	367 (184-550)	137-12177	*Sweet potato leaf curl virus* SPLCV-US [BR:Est1:07] (99%/FJ969834/Brazil)
5264	NGS Begomovirus 5	2789 (full genome)	26724	422 (211-633)	103-6499	*Sweet potato leaf curl China virus* SPLCCNV [KR:MuaGE21:11] (95%/JX91673/Korea)
WFKNX	VEM Begomovirus 1	2829 (full genome)	663	432 (216-648)	6-176	*Sweet potato leaf curl virus* SPLCV-US [BR:Est1:07] (99%/FJ969834/Brazil)
WFKNX	Btabaci mitogenome 1	15243	1798	406 (203-609)	6-71 (with 6 gaps = 0 coverage)	*Bemisia tabaci* Asia I mitochondrion, complete genome Partial mapping to full genome (some gaps) (82%/KJ778614)
WFKNX	Btabaci partial COI	657 (complete barcode region)	1798	406 (203-609)	20-71	*Bemisia tabaci* Australia II Isolate W16 cytochrome c oxidase subunit 1 gene, partial cds; mitochondrial (100%/JX416167)

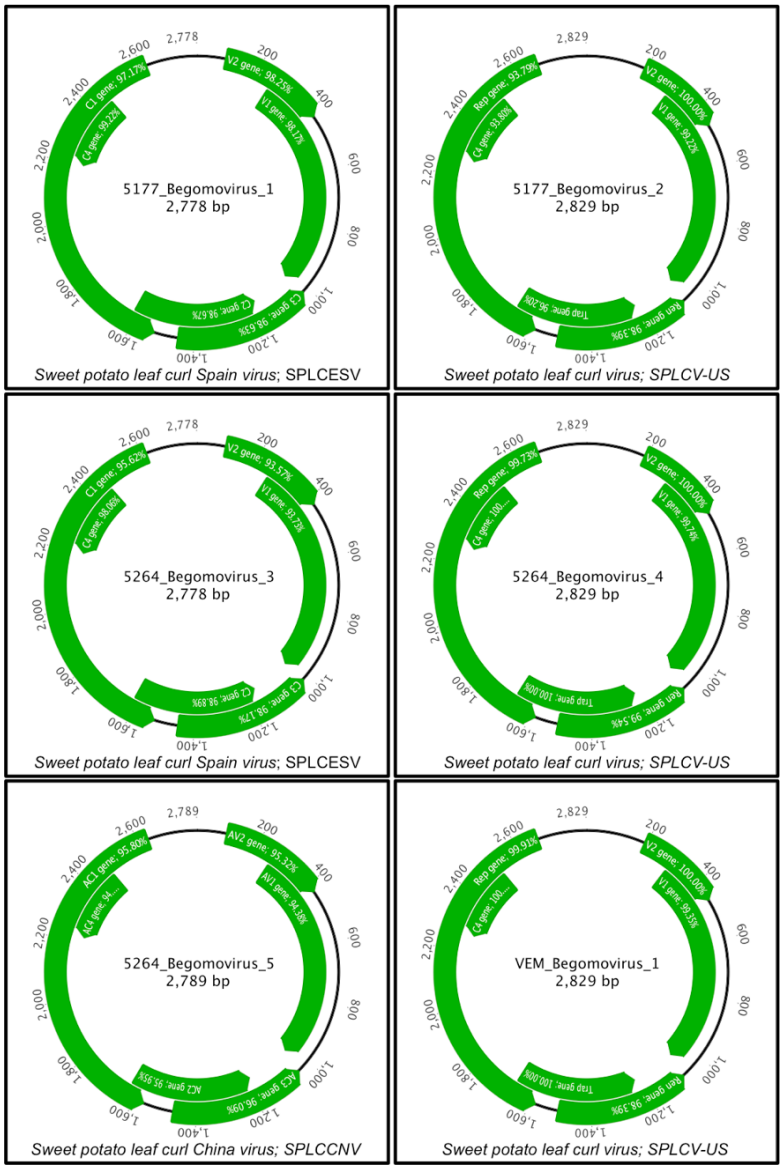


Figure 1: Schematic depicting the genome organization and features observed in sweepovirus genomes assembled in this study. Monopartite begomovirus genomes encode a pre-coat protein (V2 gene), coat protein (V1 gene), replication enhancer protein (REn gene), transcriptional activator protein (TrAP gene), replication-associated protein (Rep gene), and virulence factor (C4 gene). Percent nucleotide similarities over gene regions, as compared to appropriate reference sequences, are displayed. *Sweet potato leaf curl Spain virus* reference sequence EF456743, *Sweet potato leaf curl virus* – US strain reference sequence FJ969834, *Sweet potato leaf curl China virus* reference sequence JX91673. The current species demarcation criteria (91% cut off, ICTV) were used to infer putative species of the assembled genomes, as denoted at base of each genome image.

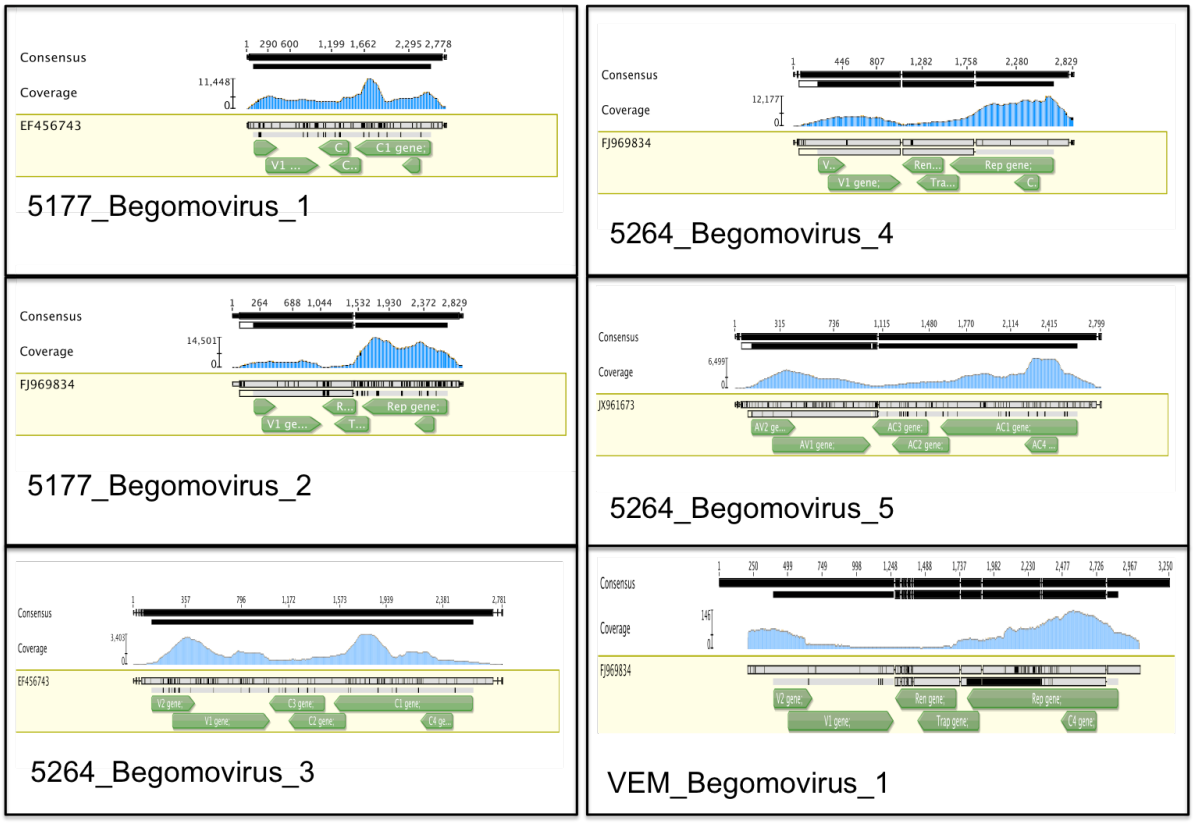


Figure 2. Schematic of sequence mapping coverage of begomovirus genomes assembled in this study.

Badnavirus partial genome reconstruction

For sample 5177, de novo assembly strategies revealed two contig assemblies that were partial matches to two separate isolates of *Sweet potato badnavirus B* (full length of genome ~ 8000 bp, ds DNA). This was further investigated by performing mapping against these representative sequences. Resulting assemblies are described in Table 2, with alignments shown in Figures 3 and 4.

A contig NGS_BadnaB1 was assembled that was 750 bp in length, and shared 85% nt identity to *Sweet potato badnavirus B* isolate Yanshu1 from China (KM009092). The alignment in Figure 3 demonstrates the high level of amino acid conservation over the polyprotein gene coding sequence.

A second contig NGS_BadnaB2 was assembled that was 1722 nt in length, with some gaps, and shared 77% match to *Sweet potato badnavirus B* isolate TZ:BK59:12 from Tanzania (KF836892). The alignment in figure 4 demonstrates the moderate level of amino acid conservation over the protein coding regions depicted.

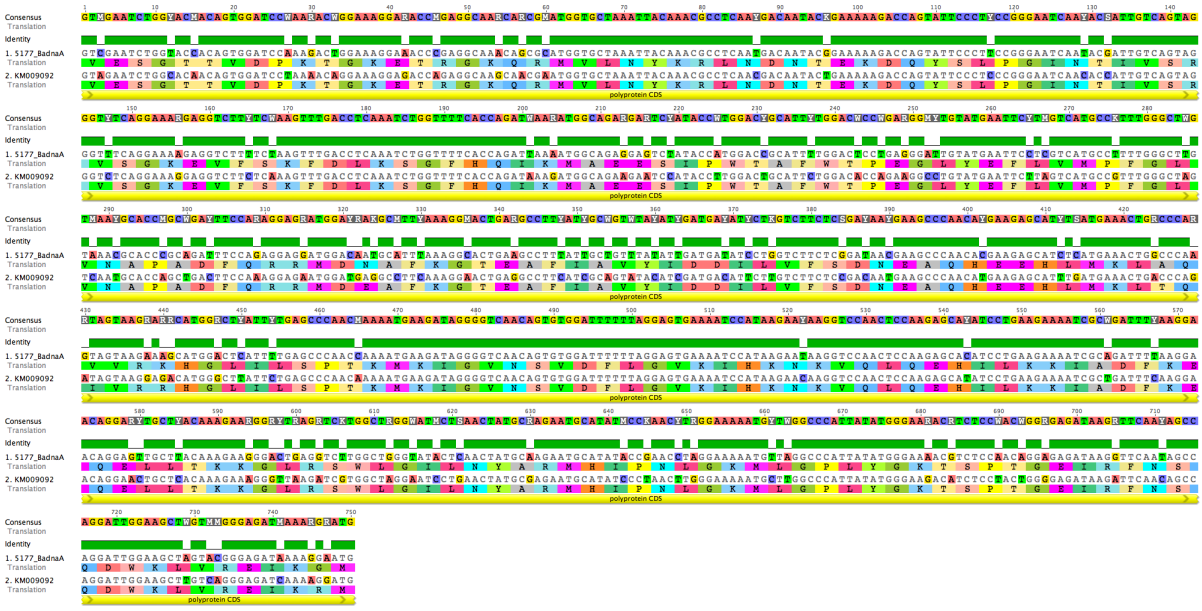


Figure 3: Alignments of NGS BadnaB1 assembly depicting conservation in amino acid sequences over protein coding sequences.

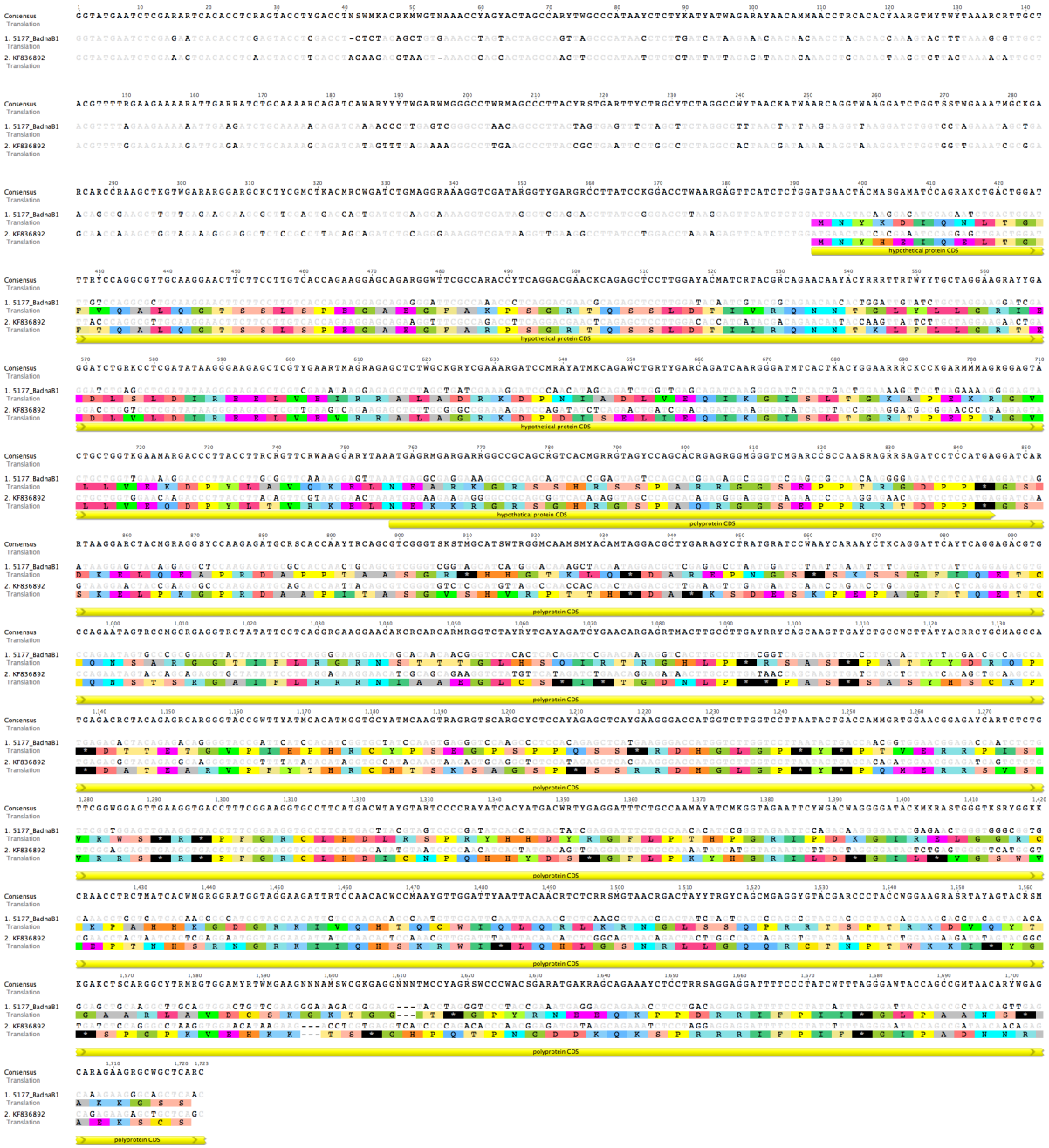


Figure 4: Alignments NGS BadnaB2 assembly depicting conservation in amino acid sequences over protein coding sequences.

Vector enabled metagenomics

A complete, full-length DNA-A genome of a monopartite begomovirus was assembled with a relatively low coverage (Table 3). Genome VEM_Begomovirus_1 shared 99% genome-wide pairwise identity with the DNA-A of *Sweet potato leaf curl virus* (FJ969834).

A partial mitochondrial genome (circular, 15 243 bp) was assembled and mapped to the published mitogenome of *Bemisia tabaci* ASIAI (KJ778614). Some gaps were evident in both mapped and denovo assemblies. A high level of coverage over the COI gene of the mitogenome was obtained, and from this the COI barcode was extracted, which showed 100% nt identity to *Bemisia tabaci* AUSII.

For this library, the majority of sequence reads mapped to plant chloroplast and mitochondrial genes, and the primary symbiont genome Candidatus Portiera aleyrodidarum.

Phylogenetic inference

Sweepoviruses

Pairwise nucleotide comparisons were performed using MUSCLE on the full-length genomes discovered in this study, and full-length genomes available on GenBank. Current accepted sweepovirus species were used where appropriate (Brown et al., 2015, Albuquerque et al., 2012).

The phylogenetic relationships among these viruses were inferred (Figure 5), with all sweepoviruses discovered in this work grouping with their proposed species/strains and were separated from both Old and New Wold begomoviruses as was expected.

Badnaviruses

Pairwise nucleotide comparisons were performed using MUSCLE on the partial badnavirus protein coding sequence (no gaps) reconstructed in this study, and corresponding partial sequences of closely related representative badnaviruses available on GenBank.

The phylogenetic relationships among these viruses were inferred (Figure 6), with the badnavirus sequence reconstructed in this study grouping with other known *Sweet potato badnavirus B* isolates.

Whitefly mtCOI

Conventional sequencing was done on single whitefly DNA extracts that were used for pooled VEM NGS sequencing (geographic location – Kununurra, Western Australia). Resulting sequences were aligned with

All of the 11 whitefly samples were shown by conventional sequence to share 99-100 % nt identity with *Bemisia tabaci* AUSII species grouping (Figure 7).

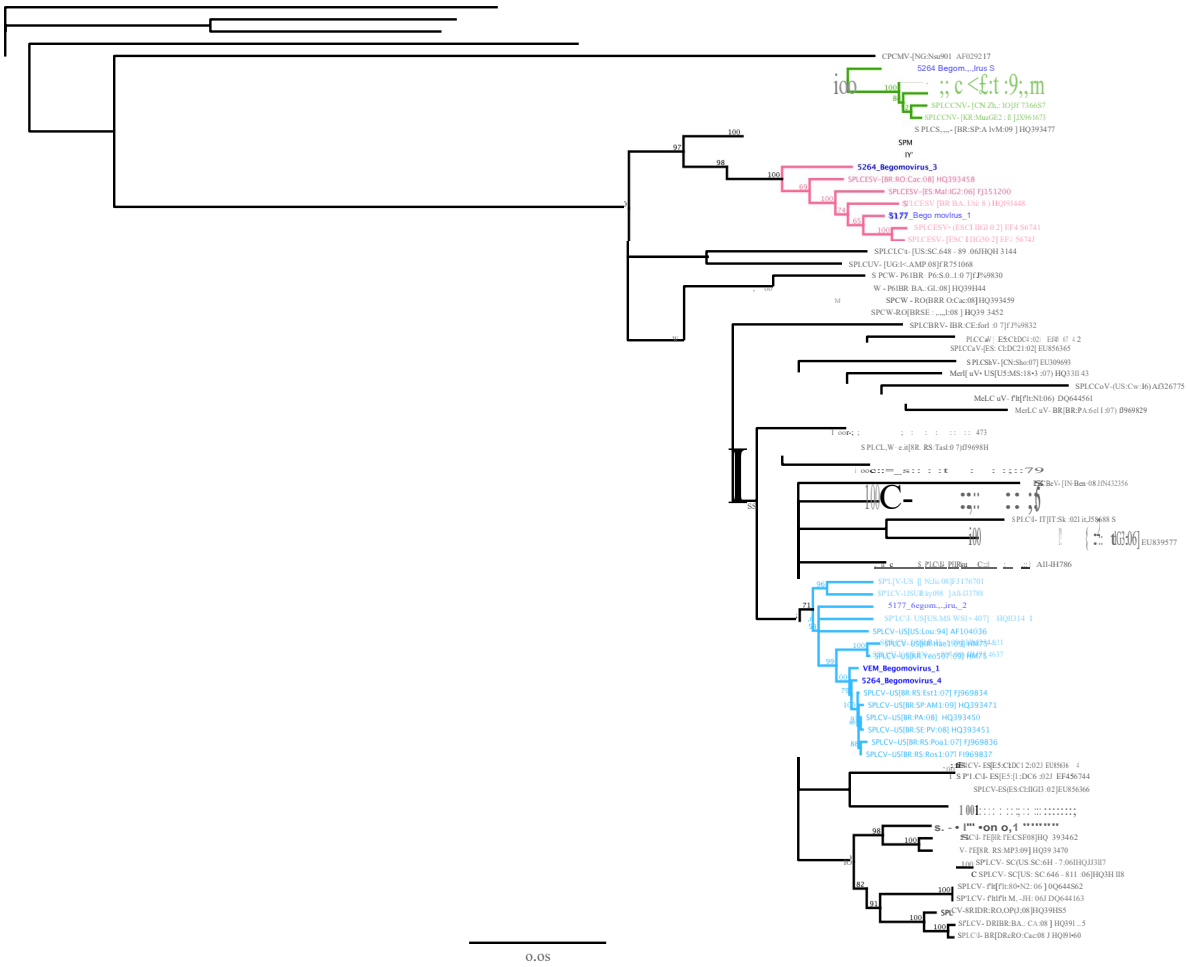


Figure 5: Maximum-likelihood (ML) phylogenetic tree based on the multiple alignment of the complete sequences of sweepoviruses discovered in this work (in bold blue) and representative sequences available in GenBank.

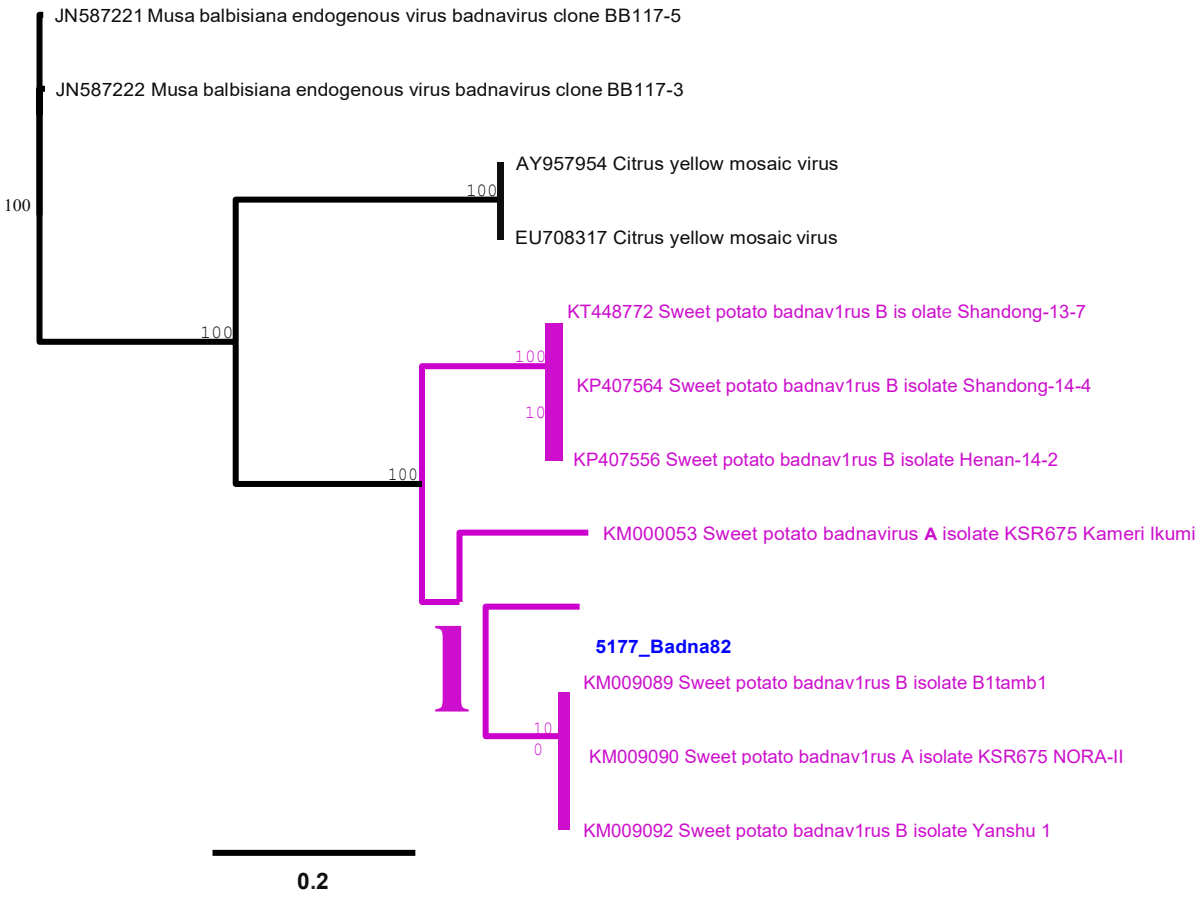


Figure 6: Maximum-likelihood (ML) phylogenetic tree based on the multiple alignment of the partial polyprotein gene coding sequence of a putative badnavirus discovered in this work (in bold blue) and representative sequences available in GenBank.


Bemisia_atriplex2_GU086363
ID267_Bemisia atrip lex _QLD_2015
SubSahAfl_U9anda_AY05 7181
SubSahAfl_Uganda_Soroti_AY563657
SubSahAfl_SouthAfrica_AF344264
SubSahAfl_Uganda_Busu ku ma_AY90 3487
SubSahAfl_Congo_DQI30054
SubSahAfl_Uganda_H44_AY05 7169
SubSahAf2_Mali_AY827605
SubSahAf2_Uganda_AY05 7146
NewWorld2 Argentina AF340213
Indian0cean _Seychel l es_AJ550 1 82
IndianOcean Reunion EU760735
Mediterranean_Cameroon_AF344258
Mediterranean_Uganda_Hap24_AY903540
Mediterranean _Ghana_SubSahAfSL_AY8 2 7 590
Mediterranean_Czech _B1_Hap4_GU0863 3O
Mediterranean_Spain _AF342 769
Mediterranean _Sudan _AY8 2 7613
Mediterranean_Franee_AM691O5 5
Mediterranean _China _4_Hap2 _GU0863 33
Mediterranean _Algeria _AM176574
Mediterranean _Crete _Hap18 _DQ365856
Mediterranean _Morocco _AM176572
Mediterranean _Taiwan _AOP_GU086339
Mediterranean _China _Fuzhou _GU168793
Mediterranean _China _Hubei _EF694112
Mediterranean _Egypt _Hap14_FJ02 5 79 7
Mediterranean _Croat ia Il a GU086334
Mediterranean_Turkey _Hap3_A¥827617
Medite rra nean_Burkina _Faso_FJ766429
MidEastAml _Pakistan _Ghotki _Sindh _GU977267
Mid EastAm I _China _Jia ng su_EU263624
Mid EastAml _Israel_DQ174536
MrdEastAm1_TrinidadTobago _FN821791
MrdEastAml _USA_Arrzona _AY05 7123
ID180_MEAM1 _Emerald _2014
ID189 _Btab _COI_MEAM1_Arctu rus_2013
ID193 _Btab _COI_MEAM1_Katherme _2014
MidEastAml _Australia _2014 _NSW H aplotype _LW
M1dEastAml _Yemen _GU086360
Asia11_7_China_Guangzhou_AY686064
Asiall 7 India Mararathwada OQ116662
Asiall_10 _China_Guangdong _HM137339
Asia ll_9 _China_Hunan_HM 13 7313
Asiall_9_Hunan_HM137345
Asiall_4_China_Guangdong_AY686083
Asiall_3 _China_Gua ngx i_EU19204 5
Asi all_3 _China_Zhejiang _OQ309074
Australia Indonesia GU086327
Australia Indonesia HQ457045
AUSII _Mo ssman_1995 _JX416160
ID 2 16 _Btab _COI _AUS1I_Kununurra _2014
ID2 2 1 _Btab _COI _Kununurra _2014
ID 222 _Btab _COI _Kununurra _201 4
ID 223 _Btab _COI _Kununurra _2014
ID 20 _Btab _AUSII _Darwin _19 94
ID224 Btab COi Kununurra 2014
ID217_Btab _COI _Kununurra _201 4
ID 218 _Btab _CO I_Kunu nurr a_20 14
ID 219 _Btab _COI _Kununurra _20 14
ID 220 _Btab _COI _Kununurra _20 14
ID 22 S_Btab _CO I_Kununu rr a_20 14
ID 227 Btab COi Kununurra 2014
0.6i9 AUSI _Elund ; berg= 2010J X43 7 453
AUSI Emerald 1997
Asialll Taiwan DQ174528
Asiall-I Ch ina Taiwan DQ174527
A ia1_china _Gu ngdong_DQ9869 59
Asial _India _Karnata k a_HM590153
Asial _India _HM590154
Asia1_Bang lad esh_Bioug ra _Al 74 83 88
Asia1_India _Uttar_Pradesh _HM590 16 3
As ial _Thailand _EF5285 52
Asia1_1ndonesia_A8248262
Asia l _China _Guangxi _EU 192044
Asial _Ind one sia_HE65 3 704
Asial _Singa pore _AY686095
0.06

Figure 7: Bayesian phylogeny of global representative species of the *Bemisia tabaci* species complex, reconstructed from partial 3' mitochondrial cytochrome oxidase gene sequences. *B. tabaci* AUSII species group is denoted in pink, with samples from this study shown in blue.

Discussion

This study describes the first identification of begomoviruses infecting sweet potato in Australia. We describe the use of next generation sequencing (NGS) approaches to identify and characterize the presence of three sweepoviruses in two separate samples, and the presence of badnavirus genomic sequences in one of these samples. Lastly, we describe the use of vector-enabled metagenomics (VEM) to survey whiteflies collected from a single geographic region, across several plants hosts, for the presence of begomoviruses.

Using our custom enriched and multiplexed NGS sequencing approach, we identified a mixed infection of two sweepoviruses *Sweet potato leaf curl virus* and *Sweet potato leaf curl Spain virus* in a sweet potato breeding line from Queensland. We also identified a mixed infection of three sweepoviruses in an asymptomatic sweet potato crop grown in Western Australia; *Sweet potato leaf curl virus*, *Sweet potato leaf curl Spain virus* and *Sweet potato leaf curl China virus*. To our knowledge, these viruses have not been reported in Australia to date. The high depth of coverage (103-14501 times coverage) for these genomes, in addition to the long sequence read length strategy used, gives confidence to the assemblies derived in this study. The frequency and importance of recombination in geminivirus evolution, particularly its contribution to plant disease epidemics is well known (refs). Our discovery of the presence of mixed infections of sweepoviruses, particularly in the breeding lines which are propagated extensively, has the potential to lead to the emergence of new species and strains of sweepoviruses in Australia.

In addition, we reconstructed two sequence assemblies that had good coverage and significant similarity to reported isolates of *Sweet potato badnavirus B*. To our knowledge, *Sweet potato badnavirus B* has not been reported in Australia. The detection of a badnavirus in this sample using DNA sequencing is not clear evidence of badnavirus infection. The genomes of other badnaviruses are known to be integrated into their host plant species, as such the sequences we have detected in this study could be integrated virus sequences and not active infections.

The VEM approach we employed exploits ability of mobile adult whitefly to feed on various plant hosts over time and space, effectively subsampling numerous plants for viruses. This study utilised VEM to describe viral communities from whiteflies collected from several crops in Kununurra, Western Australia. We were successful in identifying and assembling a complete genome of Sweet potato leaf curl virus from our sample, with relatively low coverage. The majority of sequences derived from this approach matched plant and whitefly symbiont sequences, not begomoviruses. Importantly, no other begomoviruses (including DNA-B, betasatellite, alphasatellite and deltasatellite) molecules were detected in this pooled sample. It is important to note, that detecting these viruses in their vectors provides clear indication of virus acquisition during feeding, but gives no information on the transmissibility of the viruses by these particular vector species. Separate biological virus transmission experiments would need to be performed to determine this. Most begomoviruses are transmitted in a circular, non-propagative manner by *B. tabaci*. However, not all begomovirus species can be transmitted by various whitefly species within the species-complex of *B. tabaci*; these interactions are specific, complex and likely have arisen due to the long co-evolution of certain viruses with certain *B. tabaci* species.

Interestingly, the method we employed simultaneously amplified the whitefly mitochondrial genome, with near complete coverage, allowing us to concurrently infer the identity of the whitefly carrying the SPLCV that was detected. We therefore showed that *Bemisia tabaci* AUSII species, which we previously reported to be present in Kununurra (van Brunschot et al., 2013), was able to acquire SPLCV from a host plant during

feeding. To further confirm the results of the whitefly COI assemblies derived from this NGS experiment, we used conventional sequencing to check the samples used to create the pooled DNA and showed that they were all *B. tabaci* AUSII, confirming our NGS results.

By capturing viruses directly from insect vectors, researchers can gain insight into the diversity of viruses in a given geographical region (across host plants that the vector visits). However, this method does provide information on the transmissibility of these viruses in the given vector. Viruses will be amplified that are acquired during feeding – they may not be transmitted successfully to another plant by these vectors.

The NGS approach we used employed a rolling circle amplification (RCA) process, to broadly enrich samples for circular DNA templates. For plant samples, this approach has the advantage that it is able to enrich for circular DNA viruses, however circular plant chloroplast and mitochondrial genomes will also be amplified (though the RCA enzyme is more efficient at amplifying smaller genomes). In whiteflies, using RCA enriches for virus genomes, but will also amplify whitefly mitochondrial genomes in addition to bacterial symbiont genomes. We chose the Illumina Miseq 250 PE sequencing approach to achieve insert sizes of ~ 400 bp (excluding adapter sequences), to minimize assembly errors when reconstructing begomovirus genomes, this is particularly important for begomoviruses which often harbor recombinant genomes within their quasispecies pools. Despite this, reconstructing and inferring virus genomes from viral quasispecies in NGS data remained a significant computational challenge. In order to simplify analysis, we chose to reconstruct the single most predominant genome of each virus species identified, as opposed to reconstructing all coexistent variants (which were numerous). We cannot exclude the likelihood of recombinants virus genomes in these samples, particularly due to the presence of mixed infections in both samples increasing the likelihood of the presence of recombinant genomes that we were not able to reconstruct using our assembly methods. We employed several strategies to check for the presence of DNA-B, alphasatellite, betasatellite and deltasatellite molecules in all RCA-enriched plant and whitefly samples, with none found. This indicates the sweepoviruses detected in this study were true monopartite begomvoiruses, in accordance with the majority of Old World sweepoviruses reported.

This study demonstrates the power of NGS for describing highly variable, mixed infections of virus species *in planta*. It also demonstrates the use of VEM for describing the circulating viral community in a given region, whilst concurrently identifying the vector identity, giving insight into plant virus and vector diversity in a given region. This information facilitates plant virus surveillance and management of viral diseases.

While the yield losses caused by these viruses remains to be studied, the data from this study are of practical importance in terms of regional and international exchange of sweetpotato germplasm.

References

Adams M, Lefkowitz E, King Q, Carstens E, 2014. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2014). *Archives of Virology* **159**, 2831.

Albuquerque LC, Inoue-Nagata AK, Pinheiro B, Resende RO, Moriones E, Navas-Castillo J, 2012. Genetic diversity and recombination analysis of sweepoviruses from Brazil. *Virology Journal* **9**, 1.

Andrews S, 2010. FastQC: A quality control tool for high throughput sequence data. *Reference Source*.

Australian Sweetpotato Growers Inc 2016. Key Facts: Australian Sweetpotato Industry http://www.aspg.com.au On-line [URL] Year/Electronic Resource Number|.

Bolger AM, Lohse M, Usadel B, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.

Brown JK, Zerbini FM, Navas-Castillo J*, et al.*, 2015. Revision of Begomovirus taxonomy based on pairwise sequence comparisons. *Archives of Virology* **160**, 1593-619.

Cuellar WJ, De Souza J, Barrantes I, Fuentes S, Kreuze JF, 2011. Distinct cavemoviruses interact synergistically with sweet potato chlorotic stunt virus (genus Crinivirus) in cultivated sweet potato. *Journal of General Virology* **92**, 1233-43.

Edgar RC, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-7.

Henderson C 2015. Sweet potato: growers enthusiastic about supporting focussed research. *HortLink Spring 2015,* http://hortlink.horticulture.com.au/Spring2015/4/ On-line [URL] Year/Electronic Resource Number|.

Kashif M, Pietilä S, Artola K*, et al.*, 2012. Detection of viruses in sweetpotato from Honduras and Guatemala augmented by deep-sequencing of small-RNAs. *Plant Disease* **96**, 1430-7.

Kearse M, Moir R, Wilson A*, et al.*, 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-9.

Kreuze JF, Perez A, Untiveros M*, et al.*, 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* **388**, 1-7.

Lozano G, Trenado HP, Fiallo-Olivé E*, et al.*, 2016. Characterization of non-coding DNA satellites associated with sweepoviruses (genus Begomovirus, Geminiviridae) - definition of a distinct class of begomovirus-associated satellites. *Frontiers in Microbiology* **7**.

Persley DM, Coleman E, 2010. Sweetpotato. In: Persley DM, Cooke T, House S, eds. *Diseases of vegetable crops in Australia.* Victoria, Australia: CSIRO Publishing.

Posada D, 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-6.

Rambaut A, Drummond A, 2003. Tracer: MCMC trace analysis tool. *University of Oxford*.

Ronquist F, Huelsenbeck JP, 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-4.

Rosario K, Marr C, Varsani A*, et al.*, 2016. Begomovirus-Associated Satellite DNA Diversity Captured Through Vector-Enabled Metagenomic (VEM) Surveys Using Whiteflies (Aleyrodidae). *Viruses* **8**, 36.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P, 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* **87**, 651-701.

Stamatakis A, 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-90.

Tairo F, Jones RaC, Valkonen JPT, 2006. Potyvirus Complexes in Sweetpotato: Occurrence in Australia, Serological and Molecular Resolution, and Analysis of the Sweet potato virus 2 (SPV2) Component. *Plant Disease* **90**, 1120-8.

Valverde RA, Clark CA, Valkonen JP, 2007. Viruses and virus disease complexes of sweetpotato. *Plant Viruses* **1**, 116-26.

Van Brunschot SL, Gambley CF, De Barro PJ*, et al.*, 2013. Panel of real-time PCRs for the multiplexed detection of two tomato-infecting begomoviruses and their cognate whitefly vector species. *Plant Pathology* **62**, 1132-46.

Appendix 3: Peer-reviewed journal article (in preparation)

Wongnikong W, van Brunschot S, De Barro P.J., Walter G (2017). Testing mate recognition through reciprocal crosses of two native populations of the whitefly Bemisia tabaci (Gennadius) in Australia. Journal of Insect Behavior.

Testing mate recognition through reciprocal crosses of two native populations of the whitefly *Bemisia tabaci* (Gennadius) in Australia

W. Wongnikong[1, 2], S. van Brunschot[3], J. Hereward[1], P.J. De Barro[4], G.H. Walter[1]

[1] School of Biological Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia
[2] Plant Protection Research & Development Office, Department of Agriculture, Bangkok 10900, Thailand
[3] Agriculture, Health & Environment Department, Natural Resources Institute, University of Greenwich, Medway Campus, Central Avenue, Chatham Maritime ME4 4TB, United Kingdom
[4] CSIRO Health & Biosecurity, GPO Box 2583, Brisbane QLD 4001, Australia

*Corresponding author: Wanaporn Wongnikong, School of Biological Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia
E-mail: wanaporn.wongnikong@uqconnect .edu.au

Possible Journal: Journal of Insect Behavior

Introduction

The way in which male and female whiteflies interact with one another in nature is still unclear, mainly because it is so difficult to track and observe such tiny insects in the field. Experiments are necessarily conducted under laboratory conditions. Reciprocal cross-mating experiments are used to test the species status of different populations of these insects. Surprisingly, perhaps, the interpretation of results from such cross-mating test can be ambiguous because several issues need to be considered in the design of the test which sometimes it is not clear what has actually been assessed, for example, if there is no mating in control crosses, the problems for this issue could be an observation arena not appropriate or introducing wrong population in the control crosses.

The central issue is whether the males and females that are exposed to one another actually recognise one another as potential mating partners. This requires an experimental design that is appropriate for this purpose, as demonstrated by some tests on tiny parasitic Hymenoptera and thrips (Fernando and Walter 1997; Rafter and Walter 2013). These latter tests were designed with the appreciate control ones that ensure that it is the recognition process that is being scrutinized.

Bemisia tabaci (Gennadius) (Hemiptera: Aleyrodidae) is a global pest on a range of crop species (Mound & Halsey 1978; Muniz, 2000; Oliveira et al. 2001) and a range of evidence that this taxon represents a cryptic species complex including the supporting data derived from genetic diversity studies (De Barro, 2012; De Barro et al., 2011; Dinsdale et al., 2010; Hu et al., 2015; Lee et al., 2013; Liu et al., 2012; Firdaus et al., 2013; Stansly & Naranjo, 2010). Dinsdale et al. 2010 stated that this taxon contained 11 higher genetic groups and at least 24 morphologically indistinguishable species. Qin et al (2016) hypothesized that there are at least 35 cryptic species present within the known diversity of *B. tabaci*. Taken together, these studies all demonstrate that extensive genetic diversity exists within the known global diversity of *B. tabaci.*

There is little clarity as to which recognisable populations, designated ecologically by host plant or virus transmission abilities (See Walter 2003, table 7, p.179), actually represent independent species. And this situation prevails despite the clarity introduced by Perring et al. (1993) a quarter of a century ago. They demonstrated convincingly that New World *B. tabaci* (commonly known as the A biotype; hereafter New World) did not mate in small cage with Middle East-Asia Minor 1 *B. tabaci* (MEAM 1, commonly known as the B biotype; hereafter MEAM1). Allozymic frequency analyses showed that at least two distinct species.

Several of the reciprocal crossing experiments have been conducted over the past two decades, most of including De Barro and Hart 2000; Li et al. 2012; Maruthi, Colvin and Seal 2001; Maruthi et al. 2004; Omondi et al. 2005; Qin et al. 2016; Sun et al. 2011; Xu et al. 2011 and Zang and Liu 2007. These have been conducted at the population level (more than one male or one female) in a cage (e.g., Omondi et al. 2005; Zang and Liu 2007) that means the interpretation of

the result is likely to be ambiguous. Also, the duration of the exposure of the males and females to one another has usually not been justified. The period of exposure should be appropriate to the assessment of whether the males and females recognise one another as potential mating partners.

We therefore address the question of whether the AUSI and AUSII *B. tabaci* populations are distinct species. We do so by the way in which males and females from two native Australian populations of *Bemisia tabaci* (*sensu lato*) interact with one another to determine if the males and females from these populations recognise one another as potential mating partners, and mate with one another. In these tests, the results from control crosses are crucial to interpreting the results of the cross-mating tests. Microsatellite markers were also employed to confirm that any hybrids produced in the cross mating. Tests were indeed hybrids. Because we recorded all behaviours we can also describe (in quantified form) the mating behaviour of whiteflies from each population in ethograms. The outcomes from this research will keep the resolve the species status of these two populations and will also contribute to an understanding of the mate recognition systems in the *B. tabaci* species complex more generally.

Materials and Methods

Whitefly cultures and adult collection

Bemisia tabaci AUSI and AUSII were collected on *Euphorbia cyathophora* Murray (painted spurge) from Bundaberg (coastal Queensland) and Kununurra (northern inland of West Australia), Australia respectively. Each population was maintained in a separate room at 25.5±0.5 °C, a 14h: 10h photoperiod, and 60±4 % relative humidity. The whiteflies used in the experiment were collected from culture at the fourth instar (red eyes, raised body) by removing a leaf with the insects on it and then cutting sections such that each held a single pupa. Each was placed in its own stoppered vial. After adult emergence, each insect was sexed under the microscope without narcosis. Before conducting cross-mating experiment, five female adult whiteflies in each colony was randomly sampled to confirm their identity by COI barcoding (see below) every four weeks.

Cross-mating tests

Six treatments (Table 1) of 24hr old virgin adults were prepared because whiteflies less than 12h old do not mate (Li et al. 1989). Two of these treatments comprised only unmated females, as extra controls, to make sure that the unmated females in each treatment produced only male progeny. Each replicate in each cross-mating treatment contained one male and one female, which were introduced into a clip cage (adapted from Muniz and Nombela (2001)), so that the insects were on the underside of a painted spurge leaf still attached to the plant. The mating behaviour of whiteflies in each replicate was recorded for 9 hours during the day under the same conditions as colony maintenance, using a Panasonic HD Camcorder HC-V380 (Osaka, Japan) with continuous function. On each day, five replicates were run for the same population to avoid contamination across the cultures. To maintain purity of cultures, reciprocal crosses were carried out in a separate room from the culture rooms. The specific methods used for controls and reciprocal crosses are expanded below. After all tests had been completed, the videos were replayed and behaviours were timed and otherwise qualified.

When recording had been completed, the male in each clip cage was removed using an aspirator. The female was allowed five days to oviposit, and was then removed using an aspirator. Eggs were allowed to hatch and the emerged nymphs were allowed to complete their life cycle within the clip cage. Once the F1 generation insects completed their life cycle, the clip cages were removed and the leaf was detached and held at 4°C for 10 minutes so that the adults could be counted and sexed under a stereomicroscope to determine mating success (fertilized eggs produce females, males develop from unfertilized eggs) and also to calculate the offspring sex ratio. The identities of the parents and F1 progeny from all experiments that produced progeny were confirmed by COI barcoding, and then hybrids were tested with microsatellite markers to confirm their status as hybrids (see below)

Table 1 Mating behaviour and reciprocal crossing treatments involving AUSI and AUSII populations of *Bemisia tabaci* (*sensu lato*)

Crosses	Replications	Male	Female
Control cross (AUSI)	29	AUSI	AUSI
Control cross (AUSII)	35	AUSII	AUSII
Reciprocal cross	20	AUSI	AUSII
Reciprocal cross	20	AUSII	AUSI
Unmated female	20	-	AUSI
Unmated female	20	-	AUSII

DNA extraction and gene sequencing

DNA was extracted from *B. tabaci* specimens using a modified chelex extraction, adapted from White et al. (2009). Single whiteflies were homogenized using zirconium beads in 1.5 ml tubes containing 6 µl of 10 mg/ml Proteinase K and 50 µl of Chelex solution (10% Chelex in 10 mM Tris H-Cl and 1 mM EDTA pH 8.0), then incubated at 37°C for 1 hour, followed by incubation at 96°C to inactivate the Proteinase K.

PCR amplification of the 655 bp region of 3' barcoding region of the mitochondrial cytochrome oxidase I (COI) gene was done using the primers C1-J-2195 (5′-TTGATTTTTTGGTCATCCAG AAGT-3′) and L2-N-3014 (5′-TCCAATGCACTAATCTGC CATATTA-3′) (Simon et al., 1994). Each 30 µl reaction contained 2 µl DNA template, 1U MyTaq Polymerase (Bioline, Australia), 0.2 µM of each PCR primer, and 1 x MyTaq buffer (Bioline, Australia).

PCR reaction conditions consisted of an an initial denaturation at 95 °C for 3m; followed 10 cycles of 30s at 95 °C, annealing at 45 °C for 30s, and 1m extension at 72 °C; then 30 cycles of 30s at 95 °C, annealing at 50 °C for 30s, and 1m extension at 72 °C; final extension was at 72 °C for 10m. PCR products were verified by agarose gel electrophoresis and cleaned using 1U of Exonuclease I and Antarctic Phosphatase (New England Biolabs, Massachusetts, United States) under conditions at 37 °C for 20 min followed by 10 min at 80 °C. Then clean products were sequenced using forward and reverse primers by Macrogen Inc. (Seoul, Republic of Korea).

The 21 sequences (two pairs of parents and 17 progenies produced from both pairs) obtained from the analysis were assembled and then aligned together with a further 37 sequences using MUSCLE which were performed on Genious version 9.1.5 (Kearse et al. 2010, 2012). Representative mtCOI haplotypes of *B. tabaci* (focus only collections from Australia) were downloaded from National Center for Biotechnology Information (NCBI), and additional new mtCOI sequence information was provided by collaborators (van Brunschot, S. L. unpublished, Murgerwa, H. unpublished). The Neighbor-Joining phylogenetic tree were constructed on Genious version 9.1.5 (Kearse et al. 2010, 2012) using a bootstrap analysis of 10000 replications.

The Pre-mRNA processing factor 8 region (1060 bp) protocols followed those developed by Hsieh et al. (2014) with minor modification; the forward, reverse primers and forward direction sequencing primers were prp8F (5′-GCCTTGGGAG GTGTTGAAG-3′), prp8R (5′-GGCTTGCATCCAGGGTACC-3′), and prp8seqMF (5′-CTGGAGTTCTCATTGCGATC-3′) respectively. The PCR reactions comprised 2 µl DNA template, 1U MyTaq Polymerase (Bioline, Australia), 0.2 µM forward primer, 0.2 uM reverse primer, and 1x MyTaq Buffer (Bioline, Australia). PCR conditions consisted of initial denaturation at 95 °C for 3m followed by 30 cycles of 30 s at 95 °C, annealing at 55 °C for 30s, and extension at 72 °C for 1m final extension was at 72 °C for 3m. PCR products were visualised and cleaned (as above). Then clean products were sequenced using forward primers and forward direction sequencing primers (Hsieh et al. 2014) by Macrogen Inc. (Seoul, Republic of Korea). The 21 sequences from this study with further 80 sequences which were downloaded from GenBank were assembled and aligned (as above). The phytogenic tree was constructed with the same parameter as above.

Microsatellite development and genotyping

Some microsatellite loci were selected from the literature (De Barro et al. 2003; Hadjistylli et al. 2014), but most of the published ones that were tested did not provide good genotypes (only two out of six loci did). Microsatellite loci were therefore developed from next generation sequencing of *B. tabaci* AUSI and AUSII (125 bp paired end Illumina sequencing). Data for AUSI was de-novo assembled using SOAP (ref) and microsatellites were identified using QDD program (Version 3) (Meglecz et al. 2010, 2014) which uses primer3 to design primers (Untergasser et al. 2012). Forty-eight primer pairs were screened across 15 individuals of AUSI (14 females and one male) and AUSII (30 individuals in total), females were screened to test for null alleles and males were used to test for the specificity of microsatellite amplification (because males are haploid). Eleven primers were selected that amplified well in both populations and these were used to screen the putative hybrids produced in the reciprocal-crossing experiment and their parents. An M13 tail was attached to each primer (GTAAAACGACGGCCAG) at the 5' end of the forward primer (Schuelke 2000) and PIG tail (GTTTCTT) at the 5' end of the reverse primers (Brownstein et al. 1996). The 12 µl of PCR reaction was composed of 2 µl DNA template, 0.5 U MyTaq Polymerase (Bioline, Australia), 0.1 µM of forward primer, 0.2 µM of reverse primer, 1 x MyTaq Buffer (Bioline, Australia), 0.2 µM M13labelled primer with different fluorescent dye: 6-FAM, VIC, PET or NED. PCRs were performed under the following conditions: initial denaturation at 95 °C for 2m; 35 cycles of 15s at 95 °C, annealing at 57 °C for 25s, extension at at 72 °C for 30s; followed by 10 cycles of 15s at 95 °C, annealing at 54 °C for 25s, extension at 72 °C for 30s; final extension was 10 min at 72 °C. Before sending the DNA to Macrogen for genotyping, the quality of the PCR products was checked on a microchip electrophoresis machine, MultiNA™ (Shimadzu Corporation, Japan), and then were cleaned using 1U of Exonuclease I and Antarctic Phosphatase.

Microsatellite analysis

The peaks were analysed using Geneious version 9.1.5 (Kearse et al., 2010, 2012). The basic population genetic statistics, including Hardy-Weinberg Equilibrium, were calculated using Genepop version 4.6 (Rousset 2008). Null allele frequency was estimated with the EM algorithm (Dempster et al. 1977) implemented in FreeNA (Chapuis and Estoup 2007). The locus-specific statistics (number of different alleles (N_a), the observed (H_O), and expected (H_E) heterozygosity) were calculated using GenAlEx 6.5 (Peakall and Smouse 2006, 2012).

The population assignment of the parents and progeny produced in reciprocal-crossing experiment was analysed using structure program version 2.3.4 (Pritchard, Stephens and Donnelly 2000; Falush, Stephens and Pritchard 2003, 2007;

Hubisz et al. 2009). Only one individual progeny from each cross was used in this analysis because Structure is designed to assign natural populations rather than full-sib cohorts. A principal coordinates analysis (PCOA), based on genetic distance, was calculated using GenAlEx 6.5 (Peakall and Smouse 2006, 2012).

Data Analyses

The duration of each phase in the mating sequence, the numbers of male and female progeny produced in each treatment, and the proportion of female progeny were analysed by one-way ANOVA, and the specific differences between the treatments were analysed with the least significant difference test (LSD) at $P= 0.05$. Those data that did not meet the requirements of normality and homogeneity of variance were square root transformed before analysis. For numbers of female progeny and proportion of female progeny, only those mated whiteflies that produced female progeny were included in the analysis. All data were analysed using R version 3.0.2 (R core team 2013).

Results

Observations on mating behaviour

From 261 hours of recording, with 29 pairs of AUSI whiteflies, there were 18 copulation events. However, one pair did mate twice, so only the mating behaviour from the first successful mating was analysed i.e., 17 copulation events. For *B. tabaci* AUSII, 35 pairs were filmed. From 315 hours of recording there were only ten copulation events. Twenty replications of each cross-mating: AUSI-M x AUSII-F and AUSII-M x AUSI-F were filmed. There were only three mating events, all of which involved AUSI-M x AUSII-F, while no mating events occurred with AUSII-M x AUSI-F (Table 2).

Table 2 Number of mated in each cross involving *Bemisia tabaci* AUSI and AUSII

Crosses	No. pairs	No. mated (%)
Control cross (AUSI)	29	17 (58.6)
Control cross (AUSII)	35	10 (28.6)
Reciprocal cross (AUSI-M x AUSII-F)	20	3 (15)
Reciprocal cross (AUSII-M x AUSI-F)	20	-

When the whiteflies were initially introduced into the clip cages; both males and females rested and fed on the leaf. They moved around sometimes and the female moved less than the male. Therefore, once the male started to walk randomly around the leaf and approach the female, it implied he was searching for the female. This was confirmed by his eventually making contact with the female in each case. The sequence of mating behaviour in the *B. tabaci* AUSI and AUSII populations can be defined in four phases (Figure 1). The first phase is "searching and approaching". Searching refers to when the male starts walking randomly around the leaf and ultimately approaches the female to within 2-3 mm and close enough to contact the female. Males are considered, therefore, to be searching for the female. If the male did not approach female in this way, mating was never initiated.

The second stage is called "contact" phase when the male contacts the female with his antennae or tarsi. It is thought that the male sent a specific signal to the female. The initial contact was made on different parts of the female's body, which are categorized as: posterior (abdomen area), middle (thorax region), and anterior (head) region. *Bemisia tabaci* AUSI, AUSII and AUSI-M x AUSII-F showed a similar pattern with mostly making the initial contact at anterior region at 70.6 % (n = 12), 50.0 % (n = 5), and 70.0 % (n=2) respectively, followed by first contacting at posterior end and the middle of female's body (Table 3)

After first contact, the male demonstrated several movements before moving to the parallel position phase. Three movements were showed in *B. tabaci* in every populations; firstly, moving from anterior region, then directly making parallel orientation to female with at 47.1% (n = 8) in AUSI, at 30.0% (n = 3) in AUSII and at 33.3% (n=1) in AUSI-M x AUSII-F populations (Table 2). Secondly, moving from the posterior end of the female, then male went directly up to the side of the female at 11.8% (n=2) in AUSI, at 40% (n=4) in AUSII, and at 33.3% (n=1) in AUSI-M x AUSII-F populations. The last one was that male travelled from the anterior region to posterior end and then directly up the side of the female at 17.6% (n=3) in AUSI, at 20% (n=2) in AUSII and at 33.3% (n=1) in AUSI-M x AUSII-F population. Some movement was found only in the AUSI population, including moving from middle part of female's body, and then making parallel orientation to female (5.9%, n=1), moving from the posterior end of the female, then male went directly up to the side of the female (5.9%, n=1), and circling the female (11.8%, n=2). The movement which was observed only in *B. tabaci* AUSII was that the male first contacted the middle of the female, then moved to contact the posterior end, then directly up the side of the female, and finally sat parallel to female (10.0%, n=1)

After contacting the female, the male positions himself parallel to the female which is called "parallel orientation" (Figure 2A), in AUSI species 70.6% (n=12) of males had a parallel position to the female with the same side as initial contact whereas 29.4% (n=5) the male positioned different side as initial contact. AUSII and AUSI-M x AUSII-F populations, the male only had parallel position to the female on the same side as the first contact.

The last phase of the mating behaviour of *B. tabaci* AUSI and AUSII is "male positioning and copulation". Before copulation, the male raised his wings nearest to the female, covering her body. Male used four wings to cover female's

body (the nearest wings covered the top of female's abdomen, and the farther ones cover the hind abdomen) (Figure 2-B). During the wing hanging, the male had a sporadic continuous flicking movement. After that, the male positioned himself below the female's abdomen. The copulation began when the tip of the male abdomen contacted the tip of female abdomen and ended when the abdomens were no longer in contact with each other.

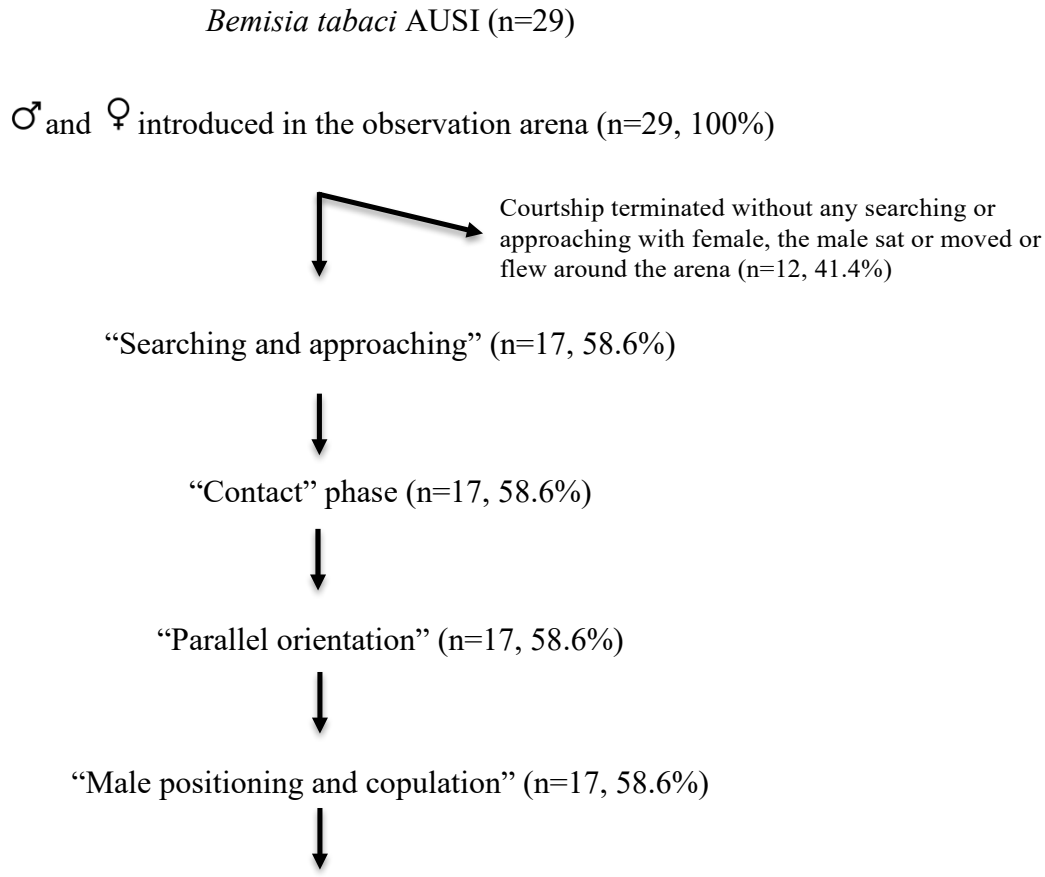
Successful mating

Mating success in *B. tabaci* AUSI was 58.6% (17/29). In 94.1% of those events (n = 16), the couple went directly from the searching and approaching phase to the male positioning himself and copulation phase in a single attempt, whereas 5.9 % (n = 1) had to attempt courtship five times before they could achieve parallel orientation (and so completed successful mating on the sixth attempt). For *B. tabaci* AUSII, mating success was 28.6 % (10/35) and 100 % of these completed successful mating in a single attempt. For AUSI-M x AUSII-F, mating success was 15% (3/20), 33.3% of mating success (n=1) had to attempt courtship twice before they could achieve parallel orientation (and so completed successful mating on the third attempt). There was no successful mating event occurred in AUSII-M x AUSI-F**.**

Unsuccessful mating

The unsuccessful courtship sequence of *B. tabaci* AUSI and AUSII was at 41.4% (12/29) and 71.4% (25/35), respectively. For reciprocal cross, AUSI-M x AUSII-F, unsuccessful mating was at 85% (17/20), and there was at 100% in AUSII-M x AUSI-F. There were two types of unsuccessful mating: not involving mating behaviour and involving mating behaviour. The former referred to the involvement of feeding, resting or grooming or moving around the observation arena but did not show any searching or approaching. These behaviours were shown by all crosses: control AUSI, AUSII and cross mating between AUSI-M x AUSII-F and vice versa. The cross control AUSI also showed unsuccessful mating from attempts at parallel orientation or positioning of the female; however, those sequences happened after successful mating.

For cross mating between AUSI-M x AUSII-F, four unsuccessful mating events which related to courtship sequences was observed, two of them attempted to position female; however, they did not success in intromission; another two unsuccessful mating ended up at parallel orientation phase, and then male or female went away. AUSII-M x AUSI-F, five out of 20 replications were related in courtship sequences; four events were failed to position female. One pair terminated the mating sequence at parallel orientation or positioning stage.



♂ departs ♀ after copulation

Figure 1: Description of the mating sequences observed for *Bemisia tabaci* AUSI populations (n=29).

Bemisia tabaci AUSI (n=35)

♂ and ♀ introduced in the observation arena (n=35, 100%)

Courtship terminated without any searching or approaching with female, the male sat or moved or flew around the arena (n=25, 71.43%)

"Searching and approaching" (n=10, 28.6%)

"Contact" phase (n=10, 28.6%)

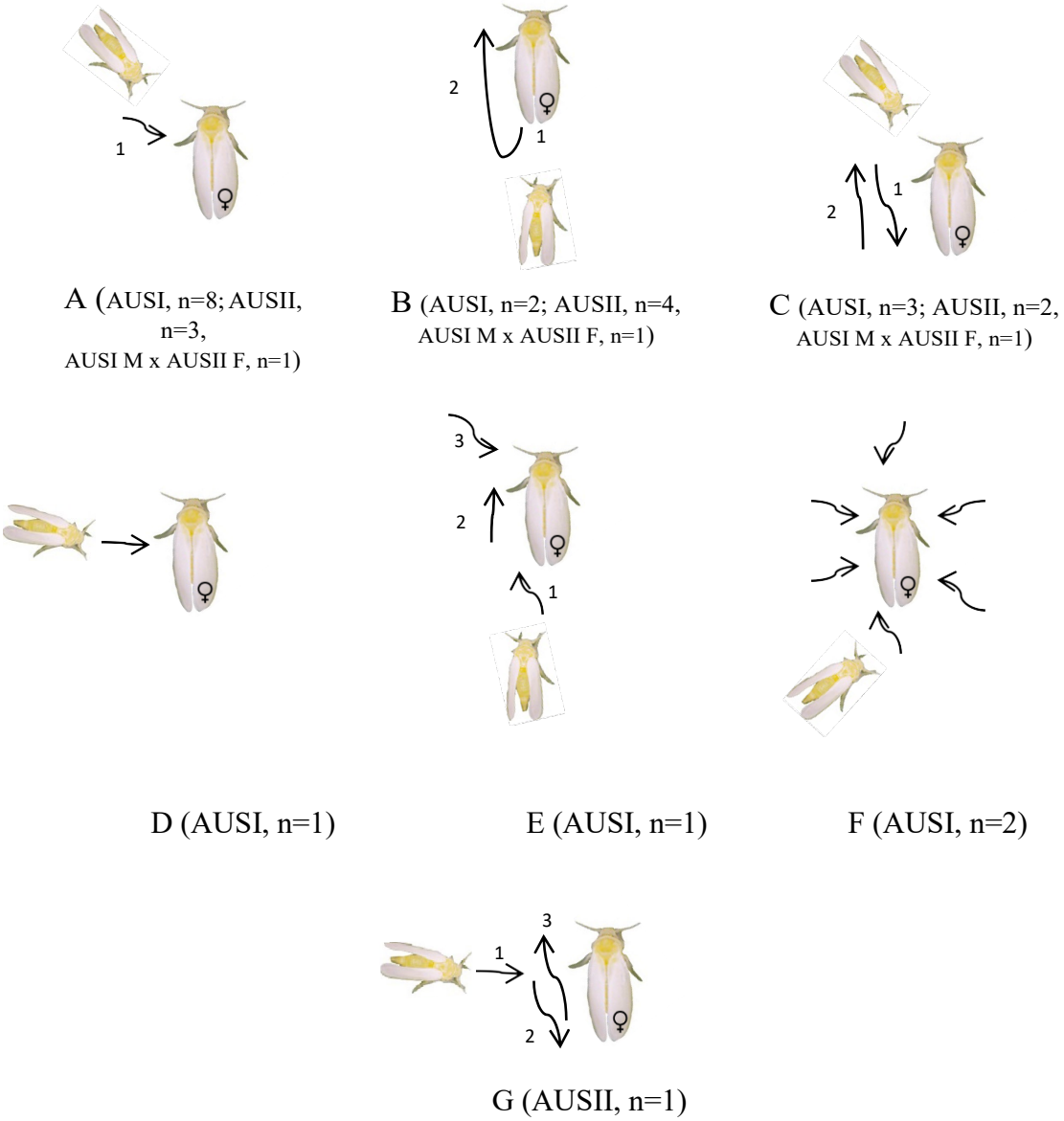
"Parallel orientation" (n=10, 28.6%)

"Male positioning and copulation" (n=10, 28.6%)

♂ departs ♀ after copulation

Figure 2: Description of the mating sequences observed for *Bemisia tabaci* AUSII populations (n=35).

Figure 3: Description of the mating sequences observed for *Bemisia tabaci* AUSI and AUSII populations. (A) Parallel orientation: the male sat next to the female facing the same direction (B) Male positioning and copulation, the nearest wings of the male covered the top of female's abdomen, and the farther ones cover the hind abdomen.



A (AUSI, n=8; AUSII, n=3, AUSI M x AUSII F, n=1)

B (AUSI, n=2; AUSII, n=4, AUSI M x AUSII F, n=1)

C (AUSI, n=3; AUSII, n=2, AUSI M x AUSII F, n=1)

D (AUSI, n=1)

E (AUSI, n=1)

F (AUSI, n=2)

G (AUSII, n=1)

Figure 4: The variation in behaviours of male whitefly when positioning around female whitefly, prior to moving parallel to the female. (Arrow means the direction of male whitefly moving around female)

Timing in each mating phrase in *Bemisia tabaci* AUSI, AUSII and AUSI-M x AUSII-F

In each population, they showed different timing in each stage. There were similarities between control crosses (AUSI and AUSII) including searching, contact, parallel, and intromission time. However, there was different between control crosses and reciprocal cross especially parallel time. There was significant difference in duration from introduction into observation arena till searching in three crosses: control cross (AUSI), control cross (AUSII) and reciprocal cross (AUSI-M x AUSII-F) (Table 4, F=6.67, df=2, *P*= 0.00443) which ranged from 300 to 28860 seconds (9557.6±1606.5, n=17), from 3724 to 32400 seconds (21312±2965, n=10) and from 2851 to 32670 (20110.3±8920.3, n=3) seconds, respectively. In contrast, there was no significant difference between three crosses in searching time till initial contact (F=1.971, df=2, *P*=0.159). Control cross (AUSI), control cross (AUSII) and reciprocal cross (AUSI-M x AUSII-F), the duration ranged from 2 to 405 seconds (125.0±27.1, n=17), from 10 to 171 seconds (73±17.2, n=10) and from 1 to 52 seconds (28.3±14.8, n=3) respectively (Table 4). Similarly, there was no significant difference in duration from initial contact to parallel orientation (F=0.173, df=2, *P*=0.842) which ranged from 7 to 46 seconds (23.2±2.9, n=17) in AUSI, from 10 to 45 seconds (21.5±3.2, n=10) in AUSII and from 13 to 24 seconds (19.7±3.4, n=3) in AUSI-M x AUSII-F populations. The duration in parallel orientation in reciprocal cross (AUSI-M x AUSII-F) was significant higher than control cross (AUSI), control cross (AUSII) (Table 4, F=7.884, df=2, *P*=0.00201). The duration ranged from 624 to 4730 seconds (2088.7±1323.3, n=3) in reciprocal cross (AUSI-M x AUSII-F), from 286 to 943 seconds (501.2±39, n=17) in AUSI, and from 322 to 1078 seconds (551.2±73.6, n=10) in AUSII. However, there was no significant difference in intromission duration (Table 4, F=0.485, df=2, *P*=0.621). In duration of mating behaviour (from searching and approaching stage to copulation phase), there was significant difference between in in reciprocal cross (AUSI-M x AUSII-F) and control cross (AUSI and AUSII) (F=7.049, df=2, *P*=0.00344), however, there was no significant difference between two control crosses (Table 4)

Reciprocal-crossing experiments

Control cross AUSI and AUSII produced both male and female progeny. There was statistically significant in male progeny (F=11.55, df=5, *P*=2.51e-09), however, there was no significant in female progeny (F=1.963, df=2, *P*=0.161). The mean proportion of female progeny in first generation (F=2.53, df=2, *P*=0.0991) ranged from 0.5 to 0.6 (Table 5).

Female progeny were produced from reciprocal cross AUSI-M x AUSII-F, however, there was no significant in female progeny when comparison to control cross AUSI and AUSII (Table 5, F=1.963, df=2, *P*=0.161). The mean proportion of female progeny in first generation was significantly difference from control cross AUSII (F=2.53, df=2, *P*=0.0991) (Table 5).

Species validity using the mitochondrial cytochrome oxidase I (mtCOI) region Pre-mRNA processing factor 8

All individuals of parent and progeny of cross mating (AUSI-M x AUSII-F) were checked for species validity by using mtCOI barcoding. The Phylogenetic tree based on mtCOI (655bp) showed that all parent were different species; father and mother were *B. tabaci* AUSI and AUSII respectively meaning that no contamination in the cross mating. Moreover, all progeny was *B. tabaci* AUSII the same as female parent (Figure 5). The Phylogenetic tree which based on Pre-mRNA processing factor 8 (Figure 6) also supported the COI region that there were two clade between parent and progeny from cross mating.

Microsatellite analysis of hybrids from reciprocal cross mating

Two couples from reciprocal cross AUSI-M x AUSII-F that produced female progeny and all progeny (21 individuals in total) were tested for the parentage by using 11 microsattatile markers (Table 6). The result showed that the female progeny were true hybrid, except one male progeny was not hybrid (Figure 7 and 8).

Table 3 The point of initial contact on females by males, and male movement before assuming the parallel position in *Bemisia tabaci* AUSI, AUSII and cross-mating individuals.

	Initial contact area (%)			Male movements before parallel orientation (%)							Parallel orientation (%)	
Pop. *	Anterior	Middle	Posterior	Anterior region then parallel	Middle and then parallel	Posterior, go up and parallel [a/]	Anterior, go along the side to anterior, go up and parallel [b/]	Posterior, go along the side, contact anterior and parallel [c/]	Male circled the female	Middle, then posterior end, go up and parallel	Same side as first contact	Different side as first contact
AUSI (n=17)	70.6(n=12)	5.9 (n=1)	23.5 (n=4)	47.1 (n=8)	5.9 (n=1)	11.8 (n=2)	17.7 (n=3)	5.9 (n=1)	11.8(n=2)	0	70.6(n=12)	29.4(n=5)
AUSII (n=10)	50.0(n=5)	20.0 (n=2)	30.0 (n=3)	30.0 (n=3)	0	40.0 (n=4)	20.0 (n=2)	0	0	10.0 (n=0)	100 (n=10)	0
AUSI-M x AUSII-F (n=3)	66.7 (n=2)	0	33.3(n=1)	33.3 (n=1)	0	33.3 (n=1)	33.3 (n=1)	0	0	0	100 (n=3)	0

***** *B. tabaci* AUSI, 17 mating events from 29 replications; *B. tabaci* AUSI, 10 mating events from 35 replications; AUSI-M x AUSII-F, three mating events from 20 replications; AUSII-M x AUSI-F, there was no mating event in this cross-mating

[a/] after contact from posterior end, then go directly up the side of the female and sit parallel to the female

[b/] contact female from anterior, then go along the side to anterior, directly up the side of female and sit parallel to the female

[c/] first contact from the posterior, then go along the side of female and contact anterior and then parallel

Table 4 Mean time of courtship and copulation of *Bemisia tabaci* AUSI, AUSII and cross-mating populations.

Crosses	Number of successful mating (%)	Time from introduction until searching [b/] (s)	Time from searching until first contact [c/] (s)	Time from first contact to parallel orientation [c/] (s)	Parallel orientation time [c/] (s)	Time of intromission [c/] (s)	Time from searching to finish copulation
Control cross (AUSI, n=29)	17 (58.6)	9557.6±1606.5[b] (300-28860)	125±27.1 [a] (2-405)	23.2±2.9 [a] (7-46)	501.2±39[b] (286-943)	234.4±8.7 [a] (198-320)	883.8±51.3[b] (493-1714)
Control cross (AUSII, n=35)	10 (28.6)	21312±2965[a] (3720-32400)	73±17.2 [a] (10-171)	21.5±3.2 [a] (10-45)	551.2±73.6[b] (322-1078)	221.7±10 [a] (171-277)	867.4±78.4[b] (514-1571)
AUSI-M x AUSII-F (n=20)	3 (15)	20100.3±8920.3[ab] (2851-32670)	28.3±14.8 [a] (1-52)	19.7±3.4 [a] (13-24)	2088.7±1323.3[a] (624-4730)	223.3±15 [a] (197-249)	2360±1297.1[a] (938-4950)

In AUSI-F x AUSII-M, there was no mating event in this cross-mating

[b/] searching refers to when the male started to walk randomly toward the female

[c/] Mean in this column followed by the same letters do not differ significantly (P>0.05)

Table 5 The average number of male and female progeny of *Bemisia tabaci*

Crosses	Replications	Number of successful mating (%)	Mean no. (±SE) of progeny		Mean proportion (±SE) of female progeny in F1 [a,b/]
			♂	♀ [a,b/]	
Control cross (AUSI)	29	17 (58.6)	18.2±2.8[b]	17.3±2.3 [a]	0.6±0.1[ab]
Control cross (AUSII)	35	10 (28.6)	10.8±1.4[c]	10.9±3.1 [a]	0.5±0.1[b]
AUSI-M x AUSII-F	20	3 (15)	11.3±2.0[c]	8±4.0 [a]	0.9±0.1[a]
AUSII-M x AUSI-F	20	0	27.7±2.6[a]	0	0
Unmated AUSI-F	20	0	31.4±3.7[a]	0	0
Unmated AUSII-F	20	0	19.8±2.1[b]	0	0

In AUSI-M x AUSII-F, there were only two pairs that produced female progeny, in AUSII-M x AUSI-F and there was no mating event in this cross-mating

[a/] only the mated whiteflies and produced female progeny was analysed

[b/] Mean in this column followed by the same letters do not differ significantly ($P>0.05$)

Figure 5: Neighbor-Joining consensus tree based on the 3' end of the mitochondrial cytochrome oxidase I (COI) gene (655 bp) of representative parent and progeny from reciprocal cross (AUSI-M x AUSII-F). Horizontal branch length is drawn to scale; the bar indicates 0.03 nt replacements per site, the support for nodes using a bootstrap analysis of 10 000 replications are shown. The outgroup is *Bemisia atriplex*.

Figure 6: Neighbor-Joining consensus tree based on Pre-mRNA processing factor 8 (1060 bp) represented that parent and progeny from reciprocal cross (AUSI-M x AUSII-F). Horizontal branch length is drawn to scale; the bar indicates 0.01 nt replacements per site, the support for nodes using a bootstrap analysis of 10 000 replications.

Table 6 Primer sequences and characteristics of 11 microsatellite loci using for testing the parentage of offspring produced in a reciprocal-crossing experiment with *Bemisia tabaci* AUSI and AUSII

Locus name	Primers (5'-3' direction)	Repeat motif	Size range (bp)	Fluorescent dye colour	N_a		H_O		H_E		PHW		Null	
					AUSI	AUSII	AUSI	AUSII	AUSI	AUSII	AUSI	AUSII	AUSI	AUSII
WFAUS2	F:TGGAGATAGGAGTAATGATAGAGAGG R: GAACTCCCAGTTGGAAGCAA	$(AAG)_{20}$	280-298	6-FAM	3	2	0.429	0.143	0.357	0.133	1.0000	1.0000	0.0000	0.0001
WFAUS3	F: AAGAATCACTCGTTTCAACCAA R: TTTACTTACTCACTCGCTTGCAT	$(AAGT)_{19}$	84-180	6-FAM	5	3	0.769	0.714	0.737	0.533	0.7331	0.2814	0.0000	0.0000
WFAUS10	F: CACTGCACAGGTTCGGAGAT R: AGTTGCCTTTGAACTCGACG	$(AAG)_{18}$	295-346	6-FAM	6	4	0.571	0.357	0.689	0.403	0.0660	0.0879	0.0235	0.0001
WFAUS11	F: TGCAACGTCATTCAGGTACG R: CGTGGTAAGGAAACGCTCAC	$(ACT)_{18}$	316-409	VIC	8	8	0.786	0.667	0.778	0.771	0.5399	0.2582	0.0000	0.0351
WFAUS12	F: TTGTCTGTCCTGGGACCCTA R: CGCTGGGATACCATCATCTG	$(AAG)_{17}$	230-239	VIC	2	2	0.214	0.167	0.191	0.278	1.0000	0.2547	0.0000	0.1123
WFAUS17	F: GCTAGGAAGCCGAACAGATG R: AATTCCGGAGCTACTCTGCC	$(AAC)_{15}$	419-452	NED/PET	7	3	0.714	0.071	0.737	0.135	0.3600	0.0370	0.0175	0.0096
WFAUS19	F: TTGTGCTCAGAAGAACACAGAA R: GGCAGAATGGAATTTCAAGG	$(ACT)_{13}$	157-190	PET	2	1	0.429	0.000	0.408	0.000	1.0000	-	0.0000	0.0010
WFAUS39	F: TCTTTCTTCAACGCTGCGA R: TAGGTGGCCATACACCGATT	$(AAG)_{10}$	303-309	NED	3	3	0.500	0.143	0.401	0.449	1.0000	0.0025	0.0000	0.2349
WFAUS40	F: AGCGGGAAATTAACATTGGC R: TGAAGTGAGACAGGGTGAAACC	$(AAAC)_{10}$	336-372	NED/PET	6	4	0.714	0.286	0.704	0.566	0.8446	0.0103	0.0000	0.1846
WF1B11 (43) (Hadjistylli et al. 2010)	F: GCAATGAACAGTTTTCTGCATGCGCG R: GCACACAGCTCTCCAAAAGAAAGGTC	$(CCTGA)_{12}$ imp	137-177	NED/PET	4	3	0.786	0.429	0.686	0.439	1.0000	1.0000	0.0000	0.0090
BEM15 (47) (De Barro et al. 2003)	F: AGCAGCATCAACAGGCTC R: CTAGATTCTGCTTGAGAGG	$(CAA)_6$ $(CAG)_4$ $(CAA)_4$	197-212	NED/PET	4	3	0.357	0.571	0.403	0.426	0.6290	0.6290	0.0406	0.0000

imp, imperfect

N_a = number of different Alleles (N_a), H_O = observed heterozygosity, H_E = expected heterozygosity, PHW = Hardy–Weinberg probability test

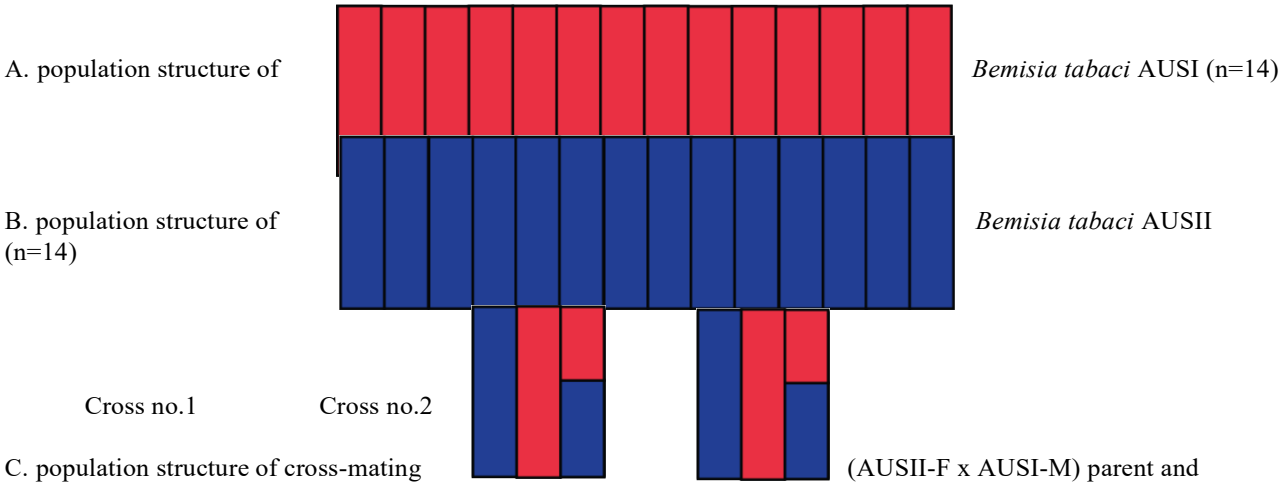
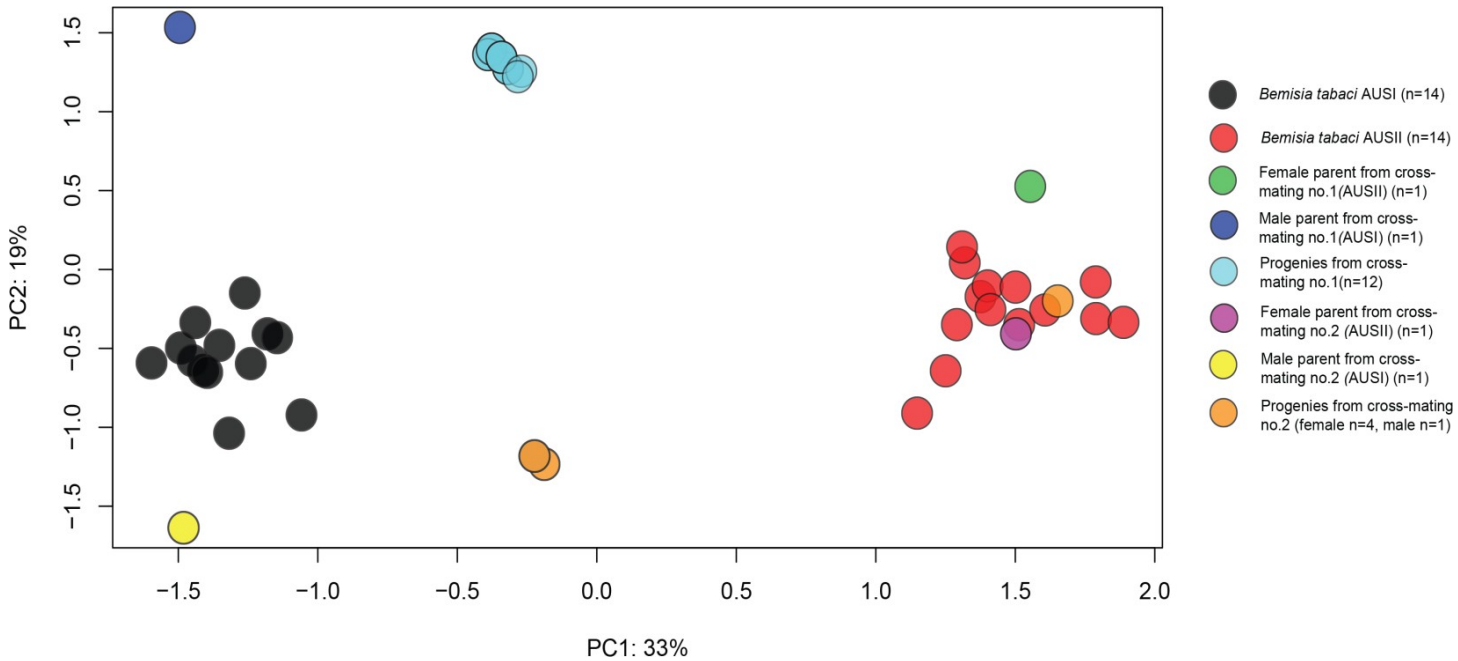


Figure 7: Bayesian clustering analysis performed in structure programme using representative from AUSI, AUSII populations, parent and progeny from reciprocal cross showed that the progeny from reciprocal cross were hybrid.

Figure 8: A Principal Coordinates analysis was produced using 11 microsatellite loci from 49 individuals (AUSI 14, AUSII 14, parent and progeny from reciprocal cross 21 individuals) across eight populations showed that the progeny



from reciprocal cross were hybrid, except male progeny from.

Discussion

Main finding

- There are four phases in mating behaviour of cross control AUSI and AUSII
- AUSI and AUSII share mating behaviour
- AUSI showed mating successful more than AUSII
- Cross mating AUSI-M x AUSII-F did mate (3 pairs) and produced female progeny
- Using microsatellite to validate hybrid from cross mating and confirmed that they were true hybrid

Future direction

- An acoustic signal method should be employed to determine whether AUSI and AUSII utilize specific acoustic signals before mating
- Observation on male's behaviour, how male behaves when live without male? Does male try to find an appropriate mating partner?

References

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping Biotechniques 20:1004-1006, 1008-1010

Chapuis M-P, Estoup A (2007) Microsatellite Null Alleles and Estimation of Population Differentiation Molecular Biology and Evolution 24:621-631 doi:10.1093/molbev/msl191

De Barro P, Scott K, Graham G, Lange C, Schutze M (2003) Isolation and characterization of microsatellite loci in *Bemisia tabaci* Molecular Ecology Resources 3:40-43

De Barro PJ (2012) The *Bemisia tabaci* Species Complex: Questions to Guide Future Research Journal of Integrative Agriculture 11:187-196 doi:10.1016/S2095-3119(12)60003-3

De Barro PJ, Hart PJ (2000) Mating interactions between two biotypes of the whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Australia Bulletin of Entomological Research 90:103-112 doi:10.1017/S0007485300000201

De Barro PJ, Liu S-S, Boykin LM, Dinsdale AB (2011) *Bemisia tabaci*: A statement of species status Annual Review of Entomology 56:1-19 doi:10.1146/annurev-ento-112408-085504

Dempster AP, Laird NM, Rubin DB (1977) Maximum Likelihood from incomplete data via the EM algorithm J R Stat Soc Ser B-Methodol 39:1-38

Firdaus S, Vosman B, Hidayati N, Jaya Supena ED, G.F. Visser R, van Heusden AW (2013) The *Bemisia tabaci* species complex: Additions from different parts of the world Insect Science 20:723-733 doi:10.1111/1744-7917.12001

Hadjistylli M, Schwartz SA, Brown JK, Roderick GK (2014) Isolation and Characterization of Nine Microsatellite Loci from *Bemisia tabaci* (Hemiptera: Aleyrodidae) Biotype B Journal of Insect Science (Tucson) 14:1-3

Hsieh CH, Ko CC, Chung CH, Wang HY (2014) Multi locus approach to clarify species status and the divergence history of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex Molecular Phylogenetics and Evolution 76:172-180 doi:10.1016/j.ympev.2014.03.021

Hu J, Chen Y-D, Jiang Z-L, Nardi F, Yang T-Y, Jin J, Zhang Z-K (2015) Global haplotype analysis of the whitefly *Bemisia tabaci* cryptic species Asia I in Asia Mitochondrial DNA 26:232-241 doi:10.3109/19401736.2013.830289

Kearse M et al. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data Bioinformatics 28:1647-1649 doi:10.1093/bioinformatics/bts199

Kearse M et al. (2010) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data Bioinformatics 28:1647-1649

Lee W, Park J, Lee G-S, Lee S, Akimoto S-I (2013) Taxonomic Status of the *Bemisia tabaci* Complex (Hemiptera: Aleyrodidae) and Reassessment of the Number of Its Constituent Species PLoS ONE 8:e63817 doi:10.1371/journal.pone.0063817

Li X-x, Li S-j, Xue X, Ahmed MZ, Ren S-x, Cuthbertson AGS, Qiu B-l (2012) The Effects of Space Dimension and Temperature on the Cross Mating of Three Cryptic Species of the *Bemisia tabaci* Complex in China Journal of Integrative Agriculture 11:197-205 doi:10.1016/S2095-3119(12)60004-5

Liu S-s, Colvin J, De Barro PJ (2012) Species Concepts as Applied to the Whitefly *Bemisia tabaci* Systematics: How Many Species Are There? Journal of Integrative Agriculture 11:176-186 doi:10.1016/S2095-3119(12)60002-1

Maruthi MN, Colvin J, Seal S (2001) Mating compatibility, life-history traits, and RAPD-PCR variation in *Bemisia tabaci* associated with the cassava mosaic disease pandemic in East Africa Entomologia Experimentalis et Applicata 99:13-23 doi:10.1046/j.1570-7458.2001.00797.x

Maruthi MN, Colvin J, Thwaites RM, Banks GK, Gibson G, Seal SE (2004) Reproductive incompatibility and cytochrome oxidase I gene sequence variability amongst host-adapted and geographically separate *Bemisia tabaci* populations (Hemiptera: Aleyrodidae) Systematic Entomology 29:560-568 doi:10.1111/j.0307-6970.2004.00272.x

Meglecz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects Bioinformatics 26:403-404 doi:10.1093/bioinformatics/btp670

Meglecz E et al. (2014) QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate Mol Ecol Resour 14:1302-1313 doi:10.1111/1755-0998.12271

Muñiz M (2000) Host suitability of two biotypes of *Bemisia tabaci* on some common weeds Entomologia Experimentalis et Applicata 95:63-70 doi:10.1046/j.1570-7458.2000.00642.x

Muñiz M, Nombela G (2001) *Bemisia tabaci*: A new clip-cage for biological studies. European Whitefly Studies Network (EWSN). http://digital.csic.es/bitstream/10261/11859/1/Bemisia%20Clip-cage.pdf. Accessed 12 January 2016 2016

Omondi BA, Sseruwagi P, Obeng-Ofori D, Danquah EY, Kyerematen RA (2005) Mating interactions between okra and cassava biotypes of *Bemisia tabaci* (Homoptera: Aleyrodidae) on eggplant International Journal of Tropical Insect Science 25:159-167 doi:10.1079/IJT200578

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update Bioinformatics 28:2537-2539 doi:10.1093/bioinformatics/bts460

Peakall ROD, Smouse PE (2006) genalex 6: genetic analysis in Excel. Population genetic software for teaching and research Molecular Ecology Notes 6:288-295 doi:10.1111/j.1471-8286.2005.01155.x

Qin L, Pan LL, Liu SS (2016) Further insight into reproductive incompatibility between putative cryptic species of the B*emisia tabaci* whitefly complex Insect Science 23:215-224 doi:10.1111/1744-7917.12296

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux Molecular Ecology Resources 8:103-106 doi:10.1111/j.1471-8286.2007.01931.x

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments Nature biotechnology 18:233-234

Stansly PA, Naranjo SE (eds) (2010) *Bemisia*: bionomics and management of a global pest vol 1st eition., vol Book, Whole. Springer, Dordrecht;London. doi:10.1007/978-90-481-2460-2

Sun DB, Xu J, Luan JB, Liu SS (2011) Reproductive incompatibility between the B and Q biotypes of the whitefly *Bemisia tabaci* in China: genetic and behavioural evidence Bulletin of Entomological Research 101:211-220 doi:10.1017/S0007485310000416

Syarifin Firdaus Ben Vosman Nurul Hidayati Ence Darmo Jaya Supena Richard GFVAWvH (2013) The *Bemisia tabaci* species complex = Additions from different parts of the world Insect Science 20:723-733 doi:10.1111/1744-7917.12001

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces Nucleic Acids Research 40:e115-e115 doi:10.1093/nar/gks596

White JA, Kelly SE, Perlman SJ, Hunter MS (2009) Cytoplasmic incompatibility in the parasitic wasp *Encarsia inaron*: disentangling the roles of Cardinium and Wolbachia symbionts Heredity 102:483-489

Xu J, De Barro PJ, Liu SS (2010) Reproductive incompatibility among genetic groups of *Bemisia tabaci* supports the proposition that the whitefly is a cryptic species complex Bulletin of Entomological Research 100:359-366 doi:10.1017/S0007485310000015

Zang L-S, Liu S-S (2007) A Comparative Study on Mating Behaviour between the B Biotype and a non-B Biotype of *Bemisia tabaci* (Hemiptera: Aleyrodidae) from Zhejiang, China Journal of Insect Behavior 20:157-171 doi:10.1007/s10905-006-9066-8

Appendix 4: Draft National Diagnostic Protocol

Draft Diagnostic Protocol for

Cryptic species of whitefly (*Bemisia tabaci*)

Introduction

Bemisia tabaci (Gennadius, 1889) Family: *Aleyrodidae,* is an agricultural pest which has been found on every continent except Antarctica and on more than 600 eudicot plant species (Oliveira et al., 2001, Quintela et al., 2016). These include crop, weed and ornamental species in the following dicot families; Fabaceae, Malvaceae, Asteraceae, Solanaceae, Cucurbitaceae, Euphorbiaceae, Begnoniaceae, Lythraceae, Zygophyllaceae, Cleomaceae, Rubiaceae, Sterculicaceae, Verbenaceae, Hypericaceae, Valerianaceae (References within Oliveira et al. (2001)). Recently B, tabaci has also been found utilising a monocot, Maize (Poaceae) as a reproductive host (Quintela et al., 2016).

Large infestations of *B. tabaci* can cause large crop losses. A sap sucking insect, feeding results in direct damage to crops, as well as honey dew secretions. These secretions can cause further damage to the plant due to sooty mildew growth. The honey dew secretions can also result in increased post-harvest costs.

B. tabaci are also the vectors of a number of plant viruses. These viruses can result in large scale crop losses and include members belonging to five genera: *Begomovirus* (*Geminiviridae*), *Carlavirus* (*Betaflexiviridae*), *Crinivirus* (*Closteroviridae*), *Ipomovirus* (*Potyviridae*) and *Torradovirus* (*Secoviridae*).

There are over 300 species within the *Begomovirus* genus, all of which are spread by *B. tabaci*. Three of the four species within the *Ipomoviruses* are also vectored by *B. tabaci*; *Cassava brown streak virus*, *Cucumber vein yellowing virus*, *Squash vein yellowing virus* (Webb et al., 2012, Maruthi et al., 2005, Mansour & Al-Musa, 1993). With the fourth species, *Sweet potato mild mottle virus* also thought to be vectored by *B. tabaci* (not yet fully confirmed).

Although a number of species within the genera Torradovirus (*Tomato torrado virus*, *Tomato marchitez virus*, Squash chlorotic leaf spot virus (Verbeek et al., 2014, Lecoq et al., 2016)), and Crinivirus (*Tomato chlorosis virus, Sweet potato chlorotic stunt virus, Cucurbit yellow stunting disorder virus, Bean yellow disorder virus, Lettuce infectious yellows virus (Duffus et al., 1986, Wisler et al., 1998, Martín et al., 2008, Célix et al., 1996, Schaefers & Terry, 1976)*) are vectored by *B. tabaci*, not all members in the genera are. With vectors of a number of Torradoviruses still unknown, these may also be vectored by *B. tabaci*. Within the Carlaviruses, only one member *Cow pea mild mottle virus* is known to be transmitted by *B. tabaci* (Iwaki et al., 1982).

B. tabaci is regarded as a cryptic species complex, consisting of at least 24 cryptic species which are not able to be reliably identified morphologically (Dinsdale et al., 2010, Liu et al., 2012). These cryptic species show differences in hosts, behaviour, invasiveness and insecticide resistance profiles. Different *B. tabaci* species have also been associated with, and experimentally shown, to have different begomovirus transmission capabilities (Guo et al., 2015, Weng et al., 2015)

An EPPO standard for the morphological differentiation of *B. tabaci* from other *Aleyrodidae* is available through the NDP website and online http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2338.2004.00729.x/full
Morphological methods however are unable to differentiate the cryptic species within *B. tabaci* therefore molecular methods are used. This diagnostic protocol covers the DNA barcoding of a region of the mitochondrial COI (mtCOI) gene which is commonly used to identify *B. tabaci* cryptic species.

Taxonomic information

Bemisia tabaci (Gennadius, 1889)
Kingdom: Animalia, Phylum Arthropoda, Class Insecta, Order Hemiptera, Family Aleyrodidae, Genus Bemisia

Cryptic species complexes: MEAM1, AUSI, AUSII, Asia I, Asia II 1-12, AsiaIII, Australia/Indonesia, China 1, China 2, Italy, Sub-saharan Africa 1-4, Uganda, New World, MED, Indian Ocean, MEAM2

Biotypes: Biotypes were first used to explain *B. tabaci* complex species, these are often still used. However, evidence suggests *B. tabaci* is a cryptic species rather than containing biotypes.

Synonyms: Bemisia argentifolii (Bellows et al., 1994), Aleurodes tabaci Gennadius, 1889, Aleurodes inconspicua Quaintance, 1900, Bemisia emiliae Corbett 1926, Bemisia bahiana Bondar, 1928, Bemisia costalimai Bondar, 1928, Bemisia signata Bondar, 1928, , Bemisia gossypiperda Misra & Lamba, 1929, , Bemisia achyranthes Singh, 1931, Bemisia hibisci Takahashi, 1933, Bemisia longispina Priesner & Hosny, 1934, Bemisia gossypiperda mosaicivectura Ghesquiere, 1934, Bemisia goldringi Corbett, 1935, Bemisia nigeriensis Corbett, 1935, Bemisia rhodesiaensis Corbett, 1936, Bemisia manihotis Frappa, 1938, Bemisia vayssierei Frappa, 1939, Bemisia lonicerae Takahashi 1957, Bemisia minima Danzig, 1964, Bemisia miniscula Danzig, 1964, Cortesiana restonicae Goux 1987

Common names include; silver leaf whitefly, sweet potato whitefly, cotton whitefly, tobacco whitefly, Cassava whitefly

Bemisia species in the Australia context
At the time of writing this report there have been five *B. tabaci* species identified in Australia. Individuals are predominantly identified as MEAM1, which have been found in Queensland, Northern Territory, South Australia, Western Australia and New South Wales. Three native species AUSI, AUSII and AUSIII. AUSI has been identified in Queensland and New South Wales and AUSII identified in Western Australia, Northern Territory, and the Torres Strait Islands (Queensland) and recently AUSIII has been identified in the Torres Strait Islands. A small number of individuals have also been identified in Northern Territory and Western Australia as an AsiaII species (currently unclassified, most similar to ASIAII-12). AUS/INDO species have previously been reported in Australia, however these have not been identified since the initial report.

A further small number of individuals were identified from archived samples (1996, 1997) collected in Queensland which are most similar to those of a Uganda species of Bemisia. These are yet to be classified and may not be *B. tabaci*.

MED is an invasive species of *B. tabaci* which is known to have high insecticide resistant profiles and is able to transmit a number of viruses. MED is not known to be present in Australia but is of high biosecurity concern.

Detection

Whiteflies are slightly smaller than fruit flies, ~1mm long with yellow bodies and white wings which they lay flat across their bodies when not in flight. Adults are often found resting and feeding on the underside of leaves. In heavily infested plants, whitefly clouds appear when the plant is disturbed.

Lifestages include egg, nymph, fourth nymph instar and adult. Eggs are laid in groups or as singles often on the leaf underside, where males (haploid) hatch from unfertilised eggs and females (diploid) from fertilised eggs.

Direct feeding damage

Symptoms of feeding damage differ depending on host and cryptic species. Stunting, defoliation and a reduction in yield is often seen

Silvering of leaves is common in some host plants including squash and cucumber. While the exact aetiology of the silvering is unknown, it is linked to the feeding of both the nymphs and adults (Yokomi et al., 1990, De Barro & Khan, 2007). A recent study identified the reduction in chlorophyll as one possible cause, as well as direct damage to the plant tissue (Zhang et al., 2017)

Indirect feeding damage

Virus damage

Figure 1: Tomato plant infected with *Tomato leaf curl virus* (Begomovirus; Geminiviridae). Symptoms include yellowing of leaves, leaf curling and plant stunting.
Australia. © Sharon van Brunscot

Figure 2: Cotton plant infected with *Cotton leaf curl Multan virus* (Begomovirus: Geminiviridae). Symptoms include leaf cupping, enlarged viens and leaf chlorosis.
South China. © Sharon van Brunscot

Visually similar species

Figure 3: *Trialeurodes vaporariorum*. Adults are slightly larger than *B.tabaci*, wings are held flat in a love heart shape. These are very difficult to distinguish in the field.
Australia. © Sharon van Brunscot ??

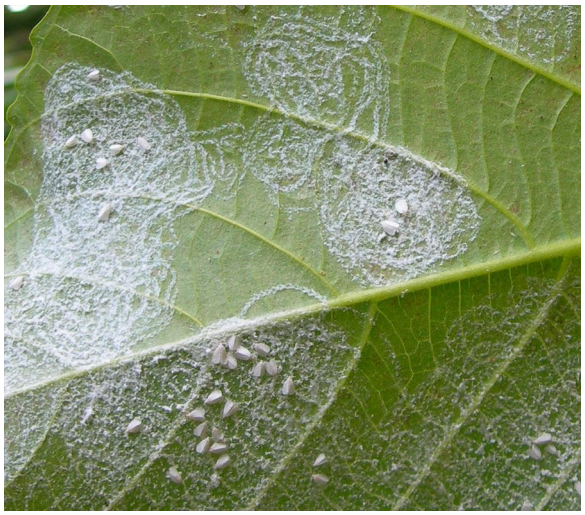


Figure 4: *Aleurodicus disperses* - spiralling whitefly. Adults are larger than *B. tabaci*. Wings are held in a triangular pyramid-like shape. They have tell-tale spiral egg laying.
Australia. © Sharon van Brunscot ??

Sampling

Adults or Nymphs can be used for molecular identification. Although not always successful, eggs have also been used previously for molecular identification.

A pooter or aspirator can be used to collect individuals directly off leaves. Gently turn over leaves to check for individuals on the underside.

Figure 5: 100s of whitefly individuals within the collection vial of a pooter. Collected from a lab colony.
UK © Sharon van Brunscot

Sweep nets can also be used to collect adults. These are useful if adults have been disturbed rather than waiting for the individuals to resettle. If possible use black sweep nets rather than the standard white nets as this helps to visualise the whiteflies for collection from the inside of the nets.

Storage
Samples can be stored in ≥70% ethanol in the freezer for many years. Samples stored for 21 years in ethanol (unknown percentage) at -20 have had the mtCOI region sequenced.

Identification

Morphological Identification

Morphological identification is not able to identify species within *B. tabaci* cryptic species complex.

Morphological identification of Aleyrodidae is based on the fourth larval instar. The EPPO diagnostic protocol (http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2338.2004.00729.x/full) provides in-depth information on the morphological detection of *Bemisia* sp. from other Aleyrodidae. The morphological standard above also outlines how to morphologically differentiate *B.tabaci* from *Trialeurodes vaporariorum*, glasshouse whitefly, which are commonly visually confused.

Molecular Detection

The 3' region of the mtCOI region is used for the molecular identification of the cryptic species complex *B. tabaci*. Simon et al primers TL2N3014 and C1J52195 (Table 1) are used to amplify an 800bp product from which a 654bp region is used to phylogenetically identify *B. tabaci* species (Dinsdale et al., 2010). Phylogenetic inference with reference sequences is required to identify the *B. tabaci* species.

A touch-up PCR protocol is used to incorporate degeneracy for the amplification of diverse *B. tabaci*. Other Bemisia species such as *B. atriplex* and *B. subdecipiens* will also amplify using mtCOI primers of Simon et al., therefore a positive PCR result does not definitively identify *B. tabaci*. It should be noted, that *B. atriplex* falls within the COI phylogenetic tree of the *B. tabaci* and may constitute a further *B. tabaci* species.

As mentioned earlier, *Trialeurodes vaporariorum,* Glasshouse whitefly is often confused with B.tabaci but can be differentiated through slight differences in morphology. However, once stored in ethanol it can become more difficult to morphologically differentiate these insects. Morphological differentiation also requires the fourth larval instar, whereas molecular identification can be carried out on adults, nymphs or the fourth larval instar (Eggs have been successfully tested previously, however, detection rate is usually lower).

While the mtCOI primers TL2N3014 and C1J52195 Simon et al. (1994) will amplify *B. tabaci*, they were found to not amplify *T. vaporariorum*. If *T. vaporariorum* need to be definitively identified other mtCOI primers such as those in Folmer et al. (1994) can be used (not covered here).

Protocol for molecular detection

Whitefly Chelex extraction

Equipment required
Micropipettes – 1-10ul, 10-100ul, 100-200ul, 200-1000ul (or variations to measure between 2-1000ul) and associated sterile tips
Dissecting microscope and fine tip brush are useful for separating individuals
Truncated tips (the ends of 100ul tips can be cut with scissors to produce truncated tips)
Magnetic stirrer and stir bar/flea
Heating block capable of 37°C and 96°C
Beadbeater, micropestle or heat blunted tips
If using a beadbeater; safelock eppendorf tubes and 0.5mm zirconium beads are required
Eppendorf tubes 1.5ml
Centrifuge

Reagents required
10mg/ul Protenase K

10% Chelex solution – 100ml
Can be stored with magnetic stir bar at room temperature for at least six months.

1M TRIS-HCL pH8	1ml
0.5M EDTA	200ul
Chelex®100 Resin (BioRad), 10g	
Water	make up to 100ml

Whitefly samples should be kept in ≥70% ethanol, 90% is ideal at -20°C for long term storage.

Individual whiteflies should be extracted separately. A dissecting microscope may be necessary to ensure only a single whitefly is transferred to each sterile 1.5ml Eppendorf tube. A fine paint brush, bristles can be trimmed if necessary, is useful for isolating and transferring an individual.

To ensure the ethanol has evaporated completely, the Eppendorf should be left uncapped for approximately 10min. Once the ethanol has evaporated add 50ul of 10% chelex solution and 6ul of 10mg/ul protenase K, to each sample. Use a truncated tip to transfer the chelex solution, which needs to be continuously stirred (magnetic stirrer) to keep the chelex beads in solution.

Once the chelex and proteinase K has been added, there are multiple methods to crush the whitefly

- A beadbeater is ideal to crush whitefly. A 1/3 scoop of 0.5mm zirconium beads is added to each sample and run at full speed for 5 minutes. If using a beadbeater 1.5ml safelock eppendorf tubes are recommended.
- Plastic micropestles also known as pellet pestles, fit directly into eppendorf tubes and can be used for hand grinding each sample. Smaller micropestles which fit into 0.5ml tubes can also be purchased. To avoid cross contamination a separate micropestle is required for each sample. However, a number of brands can be re-sterilised through bleach and autoclaving.
- Single use pestles can also be made by heating and blunt ending 1000ul tips. Due to the small size of the whitefly, the tip needs to be closed at the end to ensure the sample is not lost inside the tip while hand grinding.

Extraction of crushed whitefly

Once the whitefly is crushed, samples are spun for 1minute at full speed and then heated at 37°C for 1hr. Samples are then heated at 96°C for 8min to deactivate the proteinase K enzyme.

Due to the inhibitory nature of the chelex beads, before use in subsequent reactions the samples need to be spun in a centrifuge at full speed for 10 minutes. Samples can be stored at -20°C for later use, however, samples will need to be respun at full speed for 10 minutes after thawing.

PCR

Equipment required
PCR and band visualisation
Micropipettes – 1-10ul, 10-100ul, 100-200ul, 200-1000ul (or variations to measure between 2-1000ul) and associated sterile tips
PCR tubes
Thermocycler
Electrophoresis equipment
– tank, gel trays, gel combs, power pack, gloves
Gel documentation system or gel light box
Microwave

Reagents required
DNA polymerase and buffers – Protocol has been optimised for MyTaq™ Hot Start DNA polymerase (Bioline) and associated buffer
Molecular grade water
PCR primers (TL2N3014 and C1J52195)
Molecular ladder (to identify an ~800bp band)
Agarose gel
DNA binding stain eg Ethidium bromide or Sybr safe
Gel electrophoresis buffer eg TAE, TBE,
Loading dye
Note: Using ExoAp to clean PCR products (protocol used here) will not remove loading dye. Therefore, loading dye is only to be added to an aliquot of the sample to be run on the gel. Polymerases/buffers with loading dye incorporated can also not be used if cleaning by ExoAp. Column based PCR product cleaning kits will remove dye (not covered within this protocol)

Due to the inhibitory nature of the chelex beads, sample aliquots should be taken from the top half of the extraction, after spinning.
A master mix should be prepared from the single reaction amount, times by the number of samples, plus a positive (if available), a no template negative control, and a couple of extra reactions to account for pipetting error.
Primers TL2N3014 and C1J52195 (Simon et al., 1994) bind and amplify an 800bp product

PCR reaction mix	Amount per reaction
MyTaq Polymerase (5U/ul)	0.2ul
Primer F C1J-2195 (10uM)	0.5ul
Primer R TL2-N-3014 (10uM)	0.5ul
MyTaq Buffer HS 5x	6ul
Water	20.8ul
Template - Chelex extracted sample	2ul
Total	30

Other polymerases may be used in place of Bioline MyTaq hotstart, however the PCR protocol will need to be adjusted accordingly.

Touch-up thermocycle protocol
$95°C_{3min}(95°C_{30s},45°C_{30s},72°C_{1min})_{10\ cycles},(95°C_{30s},50°C_{30s},72°C_{1min})_{30cycles},\ 72°C_{2min},12°C_{2min}$

Visualisation of bands
Prepare a TAE 1% agarose gel with ethidium bromide or alternative DNA gel stain, with the appropriate number of wells, including a molecular ladder for each row. Add 5ul of sample and loading dye to each well, with an appropriate amount of molecular ladder added to one empty well per

row. At 110V, separation of the bands relative to the molecular ladder should be sufficient after approximately 40minutes. Using a UV gel documentation system or similar take a picture of the resulting gel. When compared to the molecular ladder, samples positive for COI will show a single clean band that is approximately 800bp in length.
The remaining 25ul of samples should be stored at 4°C.

Positive samples should be cleaned and sequenced. Samples can be cleaned using commercial PCR clean-up kits with the associated protocols or using an exonuclease enzyme clean protocol.

Enzyme PCR clean-up protocol

Equipment required
ExoAp cleanup of positive PCR products
Micropipettes – 1-10ul, 10-100ul, 100-200ul, 200-1000ul (or variations to measure between 2-1000ul) and associated sterile tips
Thermocycler or heating block capable of 37°C and 80°C

Reagents required
PCR reactions (without dye) can be cleaned using the enzymes exonuclease 1 and Antarctic phosphatase (ExoAp).

ExoAp formula for 25 reactions

Exonuclease 1 (20U/ul)	1.25ul
Antarctic phosphatase (5U/ul)	5ul
Exonuclease Buffer (10x)	2.5ul
Antarctic phosphatase buffer (10x)	2.5ul
Water	38.75
Total (25 reactions)	50ul
Per 25ul PCR reaction	2ul

To each positive 25ul sample, 2ul of ExoAp mix (1U/ul for each enzyme) is added. Samples are then heated at 37°C for 20 minutes followed by 80°C for 10 minutes. Samples can be stored in the fridge till sequencing.

Sequencing
The cleaned COI products are then able to be Sanger sequenced through a commercial facility such as Australian Genome Research facility or Macrogen Inc (Korea). A forward and reverse sequencing read for each sample is recommended if the researcher is unfamiliar with chromatograms and identifying clean sequence reads.

Bioinformatic Analyses

Equipment required
Geneious – 14 day free trial available
https://www.geneious.com/free-trial/
Minimum computational requirements Intelx86/x86_64, 2048MB memory, 2GB free space, video 1024x768 resolution. (Standard laptops have been used successfully previously)

Downloading reference accession numbers

After positive PCR products have been sequenced, these sequences of interest need to be aligned trimmed and compared with reference sequences to determine their species.

All sequences analyses can be undertaken using the Bioinformatics programme Geneious (Kearse et al., 2012a). A free 14 day trial of Geneious is available (https://www.geneious.com/free-trial/). A number of Geneious tutorials are also freely available

Reference sequences are provided as a trimmed edited database (Geneious file). The accession numbers for sequences are provided for those deposited in GenBank. It should be noted that many of the GenBank sequences have been trimmed to the 654nt region used for species identification. After downloading the sequence file this file can be imported into a new Geneious folder

- Open Geneious >create new folder >name folder 'Reference sequences' >drag and drop Geneious file into new folder
- Extract reference sequences from the alignment file > Sequence > extract regions > extract region as list of sequences
- Extract sequences from list of sequences > Sequence > extract sequences from list
- Delete consensus sequence

Sequences of interest .abi files

The forward and reverse sequencing reads need to be combined into a single consensus sequence. First these sequences are assembled into a contig sequence. All .abi files can be assembled at the same time, with Geneious recognising the sequence names of the forward and reverse reads separated by an underscore, eg ID1_CIJ2195 and ID1_TL2N3014 will be assembled into the contig ID1 Assembly.

- Create new folder > name folder 'Sequences of interest' Drag and drop .ab1 file/s of sequences of interest into this folder
- Select all raw .abi sequences > Tools > align & assemble > De novo assemble -settings 'assemble by' '1st part of name' separate by '_underscore'
- Check contigs for quality, with base differences between the two ab1 files manually checked against the chromatogram.
- If chromatogram is not present, sequence tools on the right has a 'graphs' tool button with a check box for chromatograms

Once contigs have been checked for quality, the consensus sequence of the forward and reverse sequences can be exported.

- Export the consensus sequences > select all contigs > File > export > consensus sequence(s) > use default settings > keep sequences separate > save as Geneious format
- Create new consensus folder > drag and drop the consensus sequences into the new folder

Sequences of interest need to be trimmed to the 654bp region used for species identification

- Copy and paste 2-5 *B. tabaci* reference sequences from the reference sequence folder into the consensus sequences folder which contains the sequences of interest.
- Select all in folder > align/assemble > multiple align > Geneious alignment – with 'automatically determine direction (slower)' checked
- Once aligned, trim sequences of interest relative to the reference sequences

The sequences of interest can now be phylogenetically compared with the Reference sequence database.

- Copy the trimmed consensus sequences of interest and paste into the Reference sequences folder
- MAFFT alignment is faster with large datasets. Download plugin > Tools > Plugins> MAFFT > install
- Highlight all reference sequences and the sequences of interest

Draw phylogenetic tree from nucleotide alignment

- Select nucleotide alignment > Tools > tree > Parameters - Genetic distance model HKY, Tree build method – Neighbour-Joining, no outgroup, Resampling method – Bootstrap, Number of replicates – 1000, Support Threshold 60% > OK

Tree will take a while to run, once complete, the phylogenetic tree needs to be rooted with the 16 non-bemisia tabaci sequences which were included in the reference sequence database. Visualisation of the tree is easier in a new window.

- > View >open document in new window. Zoom and expansion tools on the right can be used to expand the tree for easier visualisation.
- The phylogenetic tree is currently unrooted. Identify the branch which contains the non-bemisia sequences (Table 2) – the search/filter button above the zoom and expansion tools can be used, search for subdecipiens or similar. Select whole branch by clicking on the first node of the branch. Once all nodes of this branch are highlighted, click on the root button, within the tree frame.
- Once tree is rooted, identify which clade/s the sequences of interest fall within.

Primer name	Sequence 5'-3'	Product size	Reference
CI-J-2195	TTGATTTTTTGGTCATCCAGAAGT	800bp	Simon et al 1994
TL2-N-3014	TCCAATGCACTAATCTGCCATATTA		

Table 1: Sequences of mtCOI primers used for the identification of species within *B. tabaci*

Reference sequences for outgroup
Bemisia_tuberculata_None_None_None_AY057220_None_1
Bemisia_afer_Australia_None_None_GU220055_None_1
Bemisia_afer_China_Beijing_None_GQ139515_None_1
Bemisia_berbericola_Usa_None_None_HQ457046_None_1
Bemisia_subdecipiens_None_None_None_GU220056_None_1
Aleurochiton_aceris_None_None_None_AY572538_None
Aleurocanthus_camelliae_None_None_None_AB536794_Camellia_Sinensis
Aleurocanthus_camelliae_None_None_None_AB536800_Camellia_Japonica
Aleurocanthus_spiniferus_None_None_None_AB536793_Citrus
Aleurocanthus_spiniferus_None_None_None_AB558172_Citrus
Aleurotrachelus_camelliae_Japan_Kurinodake_None_AB536801_Camellia_Japonica
Tetraleurodes_acaciae_China_HongKong_None_AY521262_None
Aleurodicus_dispersus_India_Karnataka_None_AJ748380_Cassava
Aleurodicus_dugesii_Usa_California_None_AY521251_None
Neomaskellia_andropogonis_None_None_None_AY572539_None
Vasdavidius_concursus_None_None_None_AY648941_None

Table 2: Reference sequences used for rooting mtCOI phylogenetic tree

References

Célix, A, López-Sesé, A, Almarza, N, Gómez-Guillamón, ML & Rodríguez-Cerezo, E. 1996 Characterization of cucurbit yellow stunting disorder virus, a Bemisia tabaci-transmitted Closterovirus. *Phytopathology* **86**, 1370-1376.

De Barro, P & Khan, S. 2007 Adult Bemisia tabaci biotype B can induce silverleafing in squash. *Bull Entomol Res* **97**, 433-436.

Dinsdale, A, Cook, L, Riginos, C, Buckley, YM & De Barro, P. 2010 Refined global analysis of Bemisia tabaci (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Annals of the Entomological Society of America* **103**, 196-208.

Duffus, JE, Larsen, RC & Liu, HY. 1986 Lettuce infectious yellows virus - a new type of whitefly-transmitted virus. *Phytopathology* **76**, 97-100.

Folmer, O, Black, M, Hoeh, W, Lutz, R & Vrijenhoek, R. 1994 DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology* **3**, 294-299.

Guo, T, Guo, Q, Cui, X-Y, Liu, Y-Q, Hu, J & Liu, S-S. 2015 Comparison of transmission of Papaya leaf curl China virus among four cryptic species of the whitefly Bemisia tabaci complex. **5**, 15432.

Iwaki, M, Thongmeearkom, P, Prommin, M, Honda, Y & Hibi, T. 1982 Whitefly transmission and some properties of cowpea mild mottle virus on soybean in Thailand. *Plant Disease* **66**, 365-368.

Kearse, M, Moir, R, Wilson, A, et al. 2012 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

Lecoq, H, Verdin, E, Tepfer, M, et al. 2016 Characterization and occurrence of squash chlorotic leaf spot virus, a tentative new torradovirus infecting cucurbits in Sudan. **161**, 1651-1655.

Liu, S-s, Colvin, J & De Barro, PJ. 2012 Species concepts as applied to the whitefly Bemisia tabaci systematics: how many species are there? *Journal of Integrative Agriculture* **11**, 176-186.

Mansour, A & Al-Musa, A. 1993 Cucumber Vein Yellowing Virus; Host Range and Virus Vector Relationships. *Journal of Phytopathology* **137**, 73-78.

Martín, G, Velasco, L, Segundo, E, Cuadrado, IM & Janssen, D. 2008 The complete nucleotide sequence and genome organization of bean yellow disorder virus, a new member of the genus Crinivirus. *Archives of Virology* **153**, 999-1001.

Maruthi, MN, Hillocks, RJ, Mtunda, K, et al. 2005 Transmission of Cassava brown streak virus by Bemisia tabaci (Gennadius). *Journal of Phytopathology* **153**, 307-312.

Oliveira, MRV, Henneberry, TJ & Anderson, P. 2001 History, current status, and collaborative research projects for Bemisia tabaci. *Crop Protection* **20**, 709-723.

Quintela, ED, Abreu, AG, Lima, JF, Mascarin, GM, Santos, JB & Brown, JK. 2016 Reproduction of the whitefly Bemisia tabaci (Hemiptera: Aleyrodidae) B biotype in maize fields (Zea mays L.) in Brazil. *Pest Manag Sci* **72**, 2181-2187.

Schaefers, G & Terry, E. 1976 Insect transmission of sweet potato disease agents in Nigeria. *Phytopathology* **66**, 642-645.

Simon, C, Frati, F, Beckenbach, A, Crespi, B, Liu, H & Flook, P. 1994 Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* **87**, 651-701.

Verbeek, M, van Bekkum, PJ, Dullemans, AM & van der Vlugt, RA. 2014 Torradoviruses are transmitted in a semi-persistent and stylet-borne manner by three whitefly vectors. *Virus Res* **186**, 55-60.

Webb, SE, Adkins, S & Reitz, SR. 2012 Semipersistent Whitefly Transmission of Squash vein yellowing virus, Causal Agent of Viral Watermelon Vine Decline. *Plant Disease* **96**, 839-844.

Weng, SH, Tsai, WS, Kenyon, L & Tsai, CW. 2015 Different transmission efficiencies may drive displacement of tomato begomoviruses in the fields in Taiwan. *Annals of Applied Biology* **166**, 321-330.

Wisler, G, Li, R, Liu, H-Y, Lowry, D & Duffus, J. 1998 Tomato chlorosis virus: a new whitefly-transmitted, phloem-limited, bipartite closterovirus of tomato. *Phytopathology* **88**, 402-409.

Yokomi, R, Hoelmer, K & Osborne, L. 1990 Relationships between the sweetpotato whitefly and the squash silverleaf disorder. *Phytopathology* **80**, 895-900.

Zhang, J, Wang, L, Zheng, Y, et al. 2017 Effects of Bemisia tabaci (Gennadius) infestation and squash silverleaf disorder on Cucurbita pepo L. leaf. *Scientia Horticulturae* **217**, 8-16.

Appendix 5: Peer-reviewed journal article (under review)

Malka O, Feldmesser E, Santos-Garcia D, Sharon E, Krause Sakate R, Delatte H, **van Brunschot S**, Patel M, Mugerwa H, Seal S, Colvin J, Morin S (2017). Species-complex evolution and host-plant associations in *Bemisia tabaci*: a plant-defense, detoxification perspective revealed by RNAseq analyses. Molecular Ecology.

Species-complex evolution and host-plant associations in *Bemisia tabaci*: a plant-defense, detoxification perspective revealed by RNAseq analyses

Osnat Malka[1], Ester Feldmesser[2], Diego Santos-Garcia[1], Elad Sharon[1], Renate Krause Sakate[3], Hélène Delatte[4], Sharon van Brunschot[5,6], Mitulkumar Patel[5], Habibu Mugerwa[5], Susan Seal[5], John Colvin[5] and Shai Morin[1]

[1]Department of Entomology, Faculty of Agriculture, the Hebrew University of Jerusalem, Rehovot 76100, Israel
[2]Department of Biological Services, Weizmann Institute of Science, Rehovot 76100, Israel.
[3]São Paulo State University, Departamento de Proteção Vegetal, São Paulo, Brazil
[4]French Agricultural Centre for International Development (CIRAD Réunion), Ligne Paradis 97410, Saint Pierre, La Réunion, France
[5]Agriculture, Health and Environment Department, Natural Resources Institute, University of Greenwich, Kent, UK
[6] School of Biological Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia

Abstract

Bemisia tabaci (Hemiptera: Aleyrodidae) is a group of more than 35 cryptic species. The mechanisms driving this unusual diversification are poorly understood, but explanations largely focus on allopatric forces and the assumption that most species share a similar and broad host-plant range. In contrast, the meta-analyses of field sampling data reported here, suggest that many *B. tabaci* species are more host-specific than thought, with only few species showing a truly broad host-plant range. We explored the possibility, therefore, that differences in plant-host utilization may have contributed to the speciation process in *B. tabaci*. We first compared the expression patterns of a set of 298 detoxification genes from six *B. tabaci* species, collected from different geographical regions and diet-breadth groups. Gene-expression data were then obtained from adults feeding on four plant hosts, considered to have different probabilities of being part of the common host range of *B. tabaci*. The six species' performance indicated that they can be divided into two groups; one group showing higher performance on the various host plants than the other, putatively reflecting differences in their diet breadth. In parallel, the same grouping pattern appeared when the species were clustered according to the expression patterns of their detoxification genes, despite the three species with the more extended diet-breadth being only distantly related phylogenetically. We also found that at least some of the common detoxification mechanisms present in those species had been lost by species placed in the lower performance group. The relationship between the expression patterns of the detoxification system and the species' feeding habits, supports the hypothesis that plant hosts have played an important role in the evolution of this diverse group of species. The relevance of our findings to the oscillation hypothesis of speciation is discussed.

Introduction

Species complexes are present in a wide range of taxonomic groups and are being discovered at an increasing rate (Bickford et al. 2006). Nevertheless, only a few studies have gone beyond the raw phylogenetic identification of the species complex composition. Yet, understanding of the evolutionary processes at work in complexes of radiating taxa that only experience infrequent events of gene flow is of great importance for both evolutionary and conservation biology (Lexer et al. 2016). This is because these processes occur at evolutionary time scales that are expected to have great impact not only on the establishment of reproductive barriers between the diverging species (Rundle and Nosil 2005), but also on functional changes occurring in their genomes (Simon et al. 2015).

Species complexes of herbivorous insects are of special interest, because host specialization is the favored and dominant evolutionary strategy in this insect group (Forister et al. 2015), allowing host-specialized herbivores to become adapted optimally to the nutritional and secondary defensive chemistry of their host plants (Cornell and Hawkins 2003). It is to be expected, therefore, that within species complexes of herbivorous insects, most species will be specialists or oligophages, with the exception of a few "true" generalists (Loxdale and Harvey 2016). Examination of such transitional, closely-related species groups, differing in their host range and behavioral, ecological and genomic host-utilization traits, is generally assumed to be the best experimental approach for studying the evolution of different degrees of generalism or specialism and their putative derived effects on divergence and speciation processes (Ali and Agrawal 2012). This is because species complexes tend to share a common genetic background, with biological differences mainly associated with their feeding ecology, thus avoiding the "phylogenetic noise" present when comparing phylogenetically distant taxa that have accumulated more diverse adaptations (Roy et al. 2016).

We focus here on the whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex, which is a cosmopolitan complex widely distributed throughout tropical and subtropical regions. Multiple phylogenetic analyses of mitochondrial cytochrome oxidase I (mtCOI) DNA sequence divergence and mating experiments (Boykin et al. 2007, Dinsdale et al. 2010, De Barro et al. 2011, Liu et al. 2012, Boykin and De Barro 2014, Hu et al. 2014, Qin et al. 2016), concluded that the group contains more than 35 distinct species, assigned to ~10 major clades. Although morphologically indistinguishable, the different species show variability in certain biological and ecological traits, including plant-virus transmission efficiency, insecticide resistance, fecundity, dispersal and mating behavior (Hadjistylli et al. 2016, Qin et al. 2016). The understanding of the mechanism/s driving the extreme and unusual diversification of the species complex evolution is limited and the explanations provided so far largely point towards geographic (allopatric) divergence as the key driving force, associated with the separation of continental landmasses, which overlapped with a period of global diversification across the plant and animal kingdoms (De Barro et al. 2005, Boykin et al. 2013). These arguments were based on the association of different species with particular geographic (continental) regions, the lack of (or minimal) gene flow between species and the assumption that most species share a similar and broad host range (De Barro 2005). However, our true knowledge of the host-plant range of the different species is patchy, with much of it assumed on the basis of sampling efforts for two invasive species in the complex, *B. tabaci* MED-Q1 and *B. tabaci* MEAM1. Moreover, the allopatric divergence model fails to explain prominent sympatric expansions in the species complex, as observed for example in the Asia II (~12 species) and Sub-Saharan Africa (~5 species) major genetic groups (Lee et al. 2013, Alemandri et al. 2015).

In this study, we explore the possibility that evolved differences in host utilization played a role in the evolution of the *B. tabaci* species complex. We pose the hypothesis that the species complex is a more host-specific taxon than commonly thought and, therefore, it is quite likely that during evolution, various species acquired genetic changes that relate to their ability or inability to utilize plant hosts. These plant-related differences might have contributed to the speciation process that led, for example, to the establishment of multiple species in the same geographic range. We test this hypothesis by focusing on one important component (in herbivory) of successful plant feeding, the ability to detoxify plant chemical defenses (Heckel 2014). More specifically, we ask if the expression profile of the insect's detoxification system is shaped more by the evolutionary history of the species complex or by the putative evolutionary history of the species-plant associations. A clear association with the species-complex phylogeny would suggest that independent selection forces shaped each species'

digestive/detoxification evolution within each geographical region. In this case, distinct host specificities may just result from local adaptation of species that have previously diverged in allopatry, allowing the coexistence of host-adapted insects as a consequence of reproductive isolation, rather than by causing it (Peccoud et al. 2010). In contrast, a pattern reflecting the putative evolutionary history of the species-plant associations, where gene expression more strongly recognizes the differences between less-restricted (more general or "extended") and more-restricted (less general) feeding habits, would support the possibility that variation in host-plant utilization might have played a significant role in the species complex divergence. This is because it provides a mechanistic platform for implementing theories on the evolutionary processes of speciation. Specifically, we raise the possibility that the *B. tabaci* species complex is alternating, in evolutionary time scales, between more "extended" and more restricted host-range phases, according to the oscillation hypothesis (Jans et al. 2006, Janz and Nylin 2008). In short, the oscillation hypothesis proposes that the increased diversity of herbivorous insects is largely a result of expansions in diet breadth that set the stage for subsequent local adaptation and specializations, leading to fragmentation and eventually speciation, even in sympatry or parapatry (Slove and Jans 2011). To provide a mechanism for the host expansion phases, it hypothesizes that even more specialized insects retain the essential "machinery" that allows the utilization of host plants that were lost from their repertoire.

As stated above, and without ignoring that plants species differ in many morphological and physiological traits, we decided to focus here on one essential part of the "machinery" known to be involved in the successful utilization of host plants, the ability to detoxify plant secondary (defensive) compounds.

Insect herbivores have evolved an arsenal of enzymatic mechanisms to counteract and adapt to the toxic effects of plant-chemical defenses, of which the use of detoxifying enzymes and xenobiotic transporters play a major role (Celorio-Mancera et al. 2012, 2013; Dermauw et al. 2013, Wybouw et al. 2015, Ragland et al. 2015, Koenig et al. 2015). Detoxification occurs in three phases: the first involves oxidation, hydrolysis and/or reduction by enzymes like P450 monooxygenases (P450s) and carboxylesterases (COEs) (Lindroth 1989, Feyereisen 2005); the second involves conjugation with hydrophilic groups such as glutathione, sulphate or sugars by glutathione S-transferases (GSTs), sulfotransferases or UDP-glucosyltransferases (UDPGTs) to increase polarity and facilitate excretion; the third phase involves active export of the conjugated toxins out of the cell by ATP-binding cassette transporters (ABC transporters) (Despres et al. 2007, Broehan et al. 2013).
Multiple recent transcriptomic analyses have improved our understanding on the operational mode of the detoxification system in generalist herbivores. These studies indicated that an extensive re-arrangement of detoxification gene expression (transcriptional plasticity) takes place shortly after a host-shift, clearly suggesting that the detoxification system of generalist herbivores plays an important role in the insects' survival, when they first encounter a novel or adverse host plant (Oppert et al. 2010, Grbić et al. 2011, Tao et al. 2012, Alon et al. 2012, Elbaz et al. 2012, Celorio-Mancera et al. 2013; Halon et al. 2015, Xu et al. 2015, Yu et al. 2016, Mathers et al. 2017). Given that transcriptional plasticity of these genes is expected to mitigate the stresses of novel host plants, however, selection of the mode of expression of these genes (constitutive versus plastic) in generalist species could be disproportionately strong and could lead to evolved differences in constitutive expression between closely-related species, differing in their host range (Ragland et al. 2015, Wybouw et al. 2015, Roy et al. 2016). Moreover, we and others have recently questioned the ability of generalist insects to tailor their detoxification profiles to a specific mixture of substrate toxins in their diets, as substantial evidence indicates that these species are probably incapable of responding to specific toxins with the "right" optimal gene/s (Li et al. 2002, Wen et al. 2009, Giraudo et al. 2010, Halon et al. 2015). Instead, we have proposed (Halon et al. 2015) that for "true" generalist herbivores, it might be more advantageous to move from one plant species to the other while constitutively expressing genes (coding for enzymes) capable of detoxifying various plant toxins. For example, the recent host-shift to tobacco (*Nicotiana tabacum*) by the peach–potato aphid, *Myzus persicae*, was related to constitutive overexpression of a P450 (CYP6CY3) that allows tobacco-adapted races of *M. persicae* to detoxify nicotine efficiently (Bass et al. 2013). Here, we hypothesized that species of *B. tabaci* with a more general feeding habit, will differ from those with a less-general feeding habit, by relying more on constitutive expression than on plastic expression of their detoxification genes.

In order to test our assumptions and hypotheses, we obtained transcriptome profiles of six *B. tabaci* species (representing four different geographical and diet-breadth groups) experiencing host-shifts from a common ancestral host to three novel and putatively toxic hosts (justification of both insect and plant-species selection is provided below). We used a replicated RNA-seq approach combined with a manually curated dataset containing 298 *B. tabaci* detoxification genes. Rather than relying on identification and confirmation of individual candidate genes, we focused on recognizing patterns of expression of multiple genes within detoxification families in combination with enrichment analyses. Gene expression across species and plants was first tested to see whether variation was explained better by the host-plant utilized, or by the phylogenetic relatedness amongst the six *B. tabaci* species. Next, we tested whether or not the variation could be explained by the expression categories outlined above, predicting transcriptional differences between "more-general" host and "less-general" host species. We then compared these results with measures of adult survival to assess the concordance between the transcriptional and performance data. Taken together, our findings provide supporting evidence for the possible involvement of host-plant utilization differences in the divergence seen in the *B. tabaci* species complex.

Materials and Methods

Literature survey, clustering and ancestral host reconstruction analysis

We started our research by looking for evidence in the literature that support the general assumption that most species in the *B. tabaci* complex share a similar and broad host range (De Barro 2005). The literature survey was conducted by typing the terms (using Boolean operators): "*Bemisia tabaci*" AND botanical family name (for example "Acanthaceae") AND "Cytochrome oxidase I" OR "mtCOI" (both options) in Google Scholar (https://scholar.google.co.il), using the "Any time" option. Only reported insect collections in which the plant host was associated with a specific *mtCOI* barcode of *B. tabaci* were maintained. This dataset represents, to our best knowledge, all major field-sampling efforts published up to the end of 2016 (papers are listed in Table S1), as well as two unpublished field surveys provided by Mr Habibu Mugerwa and Dr. Hélène Delatte. Due to the very limited and sporadic sampling reports, data of species within the MED, NW, SSA, Italy, Asia II and China genetic groups were collapsed. The combined data set (absence/presence matrix) was used as input for the heatmap.2 function (Euclidean distance with a complete linkage method) of the R gplots package (R Core Team, 2017), which hierarchically clustered the different *B. tabaci* species into groups according to their botanical families/orders host range.

For ancestral host range reconstruction (order, family and genus levels), the maximum likelihood (hereafter ML) inferred tree was first produced, using selected sequences of all major genetic groups of *B. tabaci* and two related outgroups (*Bemisia afer* and *Dialeurodes citri*). *mtCOI* nucleotide sequences were downloaded from the GenBank database, clustered with CD-HIT (Fu et al. 2012) at 98% identity and a cluster representative for each major group was selected (in the case where more than one cluster was obtained, a single representative was selected). A codon-based alignment was performed with the RevTrans2.0 web-server (Wernersson and Pedersen 2003), and IQ-TREE (Nguyen et al. 2015) was used to calculate the best codon model (MGK+F3X4+R2) and ML tree (5000 ultrafast bootstraps and 5000 SH-aLRT). The resulting tree, together with the corresponding presence/absence matrix of order, family or genus host usage, were used for the ML (marginal) and maximum parsimony (MP) ancestral states reconstruction, using the ace (equal rate) and MPR functions of the ape package, respectively (R software) (Figs. S1A and S1B).

Bemisia tabaci species and host plants

From the literature survey analyses, six species of *B. tabaci*, representing different geographical and diet-breadth groups (see first paragraph of the '*Results'* section for more details), were selected for analyses: SSA1-SG3 (representing the sub-Saharan Africa (SSA) genetic group, collected from Bagamoyo Road, Tanzania in 2013), Asia II-1 (representing the Asia-II genetic group, collected from Lodhran, Pakistan in 2013), New-World 2 (hence after NW2) (representing the American genetic group, collected in Brazil in 2013), and MEAM1, MED-Q1 and Uganda-MED-ASL (representing the Africa/Middle east/Asia minor genetic group, where MEAM1 was collected from Peru in 2012; MED-Q1 from Montpellier, France in 2011 and Uganda-MED-ASL collected from Mukono, Uganda in 2012, respectively). The identity of the six species was verified using their *mtCOI* DNA sequences

(Boykin and De Barro 2014). At least two months before starting the experiments, all six colonies were transferred to eggplant, to allow them to establish a common baseline host plant. Colonies were reared under standard conditions of 28 ± 2 °C, 60% humidity and a 14:10 h light:dark cycle.

The selection of the experimental host plants was based on the results of the literature survey and the host reconstruction analysis, which identified a wide spectrum of insect species-botanical-family/order associations (Fig. 1 and first part of the *'Results'* section). Four host plants were selected: (i) eggplant (*Solanum melongena*, cv. Black Beauty) and (ii) pepper (*Capsicum annuum*, cv. California Wonder) (Solanaceae). The Solanaceae/Solanales seems to be one of the ancient host families/orders of *B. tabaci*, common to many species in the complex (Figs. 1, and S1C), but this observation relates mostly to the *Solanum* genus. Moreover, differences in the probability of being part of the ancestral host repertoire of *B. tabaci* were observed between the *Solanum* and *Capsicum* genera ($P = 0.99$ and $P = 0.5$, respectively) (Fig. S1D). The *Solanum – Capsicum* split occurred about 19 MYA (Särkinen et al. 2013) and toxic defensive secondary metabolites are known to be present in pepper, including flavonoids, phenols and capsaicinoids (Mokhtar et al. 2015); (iii) cassava (*Manihot esculenta*, cv. MCol22) (Euphorbiaceae). Like the Solanaceae/Solanales, the Euphorbiaceae/Malpighiales seem to be one of the ancient plant host families/orders of *B. tabaci*, common to many species in the complex (Figs. 1 and S1E). However, cassava is considered to be a well-defended plant and a suitable host only for some SSA and Asia II cryptic species of *B. tabaci* (Colvin et al. 2004, Ellango et al. 2015). Important defensive metabolites present in cassava include cyanogenic glucosides (Alves 2002) and flavonoids (Prawat et al. 1995); (iv) kale (*Brassica oleracea*, var. *sabellica*, cv. Dwarf Green Curled) (Brassicaceae, Brassicales). Unlike the Solanaceae/Solanales and the Euphorbiaceae/Malpighiales, Brassicaceae/Brassicales plants are not utilized by many species in the *B. tabaci* complex. Moreover, the probability of the family/order to be part of the ancestral host repertoire of *B. tabaci* was estimated to be only 0.5 (Fig. S1F and Table S2). According to the literature survey, species within the *B. tabaci* complex mainly utilize host plants from one tribe, the Brassiceae. This tribe emerged around 24 MYA from the Saharo-Sindian region, today a desert belt stretching from northwest Africa to India. It then expanded and colonized North Africa, while the Mediterranean was still forming (Arias et al. 2014). Important toxic metabolites present in kale include glucosinolates and flavonoids (Schmidt et al. 2010). All experimental plants were grown in rearing rooms maintained at 28 ± 2 °C, 60% humidity, and a 14:10 h light:dark cycle.

Performance assay

For the performance assay, groups of 50 *B. tabaci* adults, from each of the six species reared on eggplants, were transferred 1-3 days after emergence, to one of the four host plants (eggplant, pepper, kale and cassava), at the 5-8 leaf stage. The adults fed on the four host plants for 24 h, after which the proportion of survivors was recorded [(3 (biological replicates) X 4 (diets) X 6 (*B. tabaci* species)]. Proportional data were arcsin-square root transformed. Two-way ANOVA was carried out to compare the mean survival rate of the six species on the four plant hosts (plant and *B. tabaci* species as main effects). Sequential Bonferroni comparisons were performed on the 15 insect species pairs of means (k=6X5/2=15 comparisons) and the six plant species pairs of means (k=4X3/2=6 comparisons), using the conservative Dunn-Sidak method (Sokal and Rohlf 1995) at a specified experimental error rate of α=0.05 (α'=1-[1-0.05]$^{1/k}$). To avoid minute α' values obtained when carrying out unplanned comparisons amongst all possible pairs of means of the 'insect species' X 'plant species' interaction (k=24X23/2=276 comparisons), comparisons between the six insect species were made for each plant species separately. Again, Sequential Bonferroni comparisons were performed on the 15 insect species pairs of means (k=6X5/2=15 comparisons), using the conservative Dunn-Sidak method (Sokal and Rohlf 1995) at a specified experimental error rate of α=0.05 (α'=1-[1-0.05]$^{1/k}$).

Establishing a detoxification dataset for B. tabaci

Raw data from several *B. tabaci* transcriptomes were downloaded from the Sequence Read Archive (SRA) repository of NCBI. Transcriptomes were classified according to the *B. tabaci* species they belonged to; MEAM1: SRX022878 (Wang et al. 2011), SRA036954 (Xie et al. 2012), SRR835757 (Ye et al. 2014); MED-Q1: SRX018661 (Wang et al. 2010), SRR316271 (Su et al. 2012), SRR835756 (Ye et al. 2014) and Asia II-3: SRR062575 (Wang et al. 2012). Transcriptomes were assembled with Trinity v2.0.6 adapted to pair-end libraries (with the exception of SRA036954, which

was assembled as single-end) and with the following options “trimmomatic” and “normalize_reads” activated.
Re-assembled transcriptomes, plus the ones included in the bioproject PRJNA293094 (Ilias et al. 2015), were deposited in an in-house Galaxy server (Afgan et al. 2016) to perform the following steps. The sequences were clustered by CD-HIT (Fu et al. 2012) at 95 % similarity, considering the sequences belonging to the same cluster as allelic variants. For sequence annotation, a blastx similarity search against the NCBI protein database nr (e-value threshold 10^{-6}) was performed, keeping the accession number of the top hit in the insect model species: *Drosophila melanogaster*, *Helicoverpa armigera* and *Acyrthosiphon pisum* (Table S3). Due to the relatively low similarity to other model insects, sequences were only assigned to a specific detoxification family or sub-family, without the identification of their putative true orthologs. Transcriptomes belonging to several *B. tabaci* species were used and so this process allowed us to produce a non-redundant detoxification consensus gene dataset containing 104 P450s, 25 GSTs, 24 COEs, 71 UDPGTs, 20 sulfotransferases and 54 ABC transporters. Only the coding sequence (CDS) of each consensus gene was used in the differential gene expression analysis (see below).

RNA isolation and Illumina sequencing

RNA-seq experiments were carried out on colonies of the six analyzed *B. tabaci* species, reared for at least two months on eggplant. Groups of 200 newly emerged adults (1-3 days after emergence) from each species, were subjected to a feeding period of 72 h on 10% sucrose diet, to obtain a standardized detoxification genes expression pattern, which was host-plant independent. The groups were then transferred for a feeding period of 24 h, to the four experimental host plants: eggplant, pepper, kale and cassava. Next, the surviving adults (of both the 10% sucrose and plant feeding periods) were collected for RNA extraction. About 50 adults were pooled for each RNA sample to obtain sufficient RNA. Three feeding experiments per host plant were performed for each *B. tabaci* species [(3 (biological replicates) X 4 (diets) X 6 (*B. tabaci* species) = 72 RNA samples]. Total RNA was extracted using the Isolate II RNA Mini Kit (Bioline), according to the manufacturer's instructions. Library construction and sequencing was performed by the Centre for Genomic Technologies at the Hebrew University of Jerusalem. For mRNA enriched library preparation, the TruSeq RNA Library Prep Kit v2 (Illumina) was used. Libraries were sequenced using a NextSeq 500 desktop sequencer, which produced approximately 27 million 75 bp single-end reads per sample.

Gene expression analysis

The reads obtained were subjected to quality control using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Most of the sequenced bases had a quality score of > 36. No Illumina adapters were found in the sequences.
For mapping and expression analysis, a reference backbone dataset of 46898 genes, established for MED-Q1, was used (generously provided by Prof. Xiao-Wei Wang, Zhejiang University, China). The dataset was manually curated to include the consensus detoxification genes dataset described above and redundant detoxification sequences were further removed. The reads were mapped to the manually curated dataset, using RSEM (reads per kilobase of transcript per million mapped reads) v1.2.18 (Li and Dewey, 2011). The transcript reference was first prepared (rsem-prepare-reference), followed by rsem-calculate-expression with the parameter Bowtie2 (Langmead et al. 2009). The percentage of mapped reads ranged from 41 to 78.

Genes that had <10 reads in 4% of the samples were filtered out. The RSEM gene quantification for all the remaining genes was used as the input for the DESeq2 R package, version 1.10.1 (Love et al. 2014). The gene counts were normalized using DESeq2 defaults, taking into account the different read mapping percentage of the different samples. Differential expression analysis was performed using a full two-factorial model (*B. tabaci* and plant species as main effects and their interaction). All pairwise comparisons were applied with the parameter “cooksCutoff = FALSE”. False discovery rate (FDR) was corrected for all the 30012 genes that were not filtered out. Log_2 converted fold-change differences and corrected *P*-values were considered at 1 and 0.05, respectively, for the constitutive expression comparisons (between species feeding on eggplant) and 0.58 (1.5-fold expression change) and 0.05, respectively, for the plastic comparisons (within species after transfer from eggplant to cassava, kale or pepper).

Two tests were performed to show that DNA sequence differences between the six *B. tabaci* species did not bias our results due to differences in mapping efficiency. We first performed an analysis of sequence similarity, focusing on the set of 298 detoxification genes. We produced one assembled transcriptome for each of the six analyzed species, using RNA-seq data from all the species' RNA samples. Next, we used a "blast reciprocal best hit" approach to check the identity of each gene/contig (of the relevant detoxification gene) in the species' transcriptome to its putative orthologous gene/contig in the manually curated dataset (see above). All reported alignments include genes/contigs that had at least 70% of their sequence aligned, using a cutoff of at least 50% nt identity. As can be seen in Fig. S2, the mean identity for all six species was higher than 95%, meaning that the mean number of mismatches in the mapping process of reads of 75 bp long was up to ≈3 (1.22- 3.31), which is less than that allowed by the default option of Bowtie2. In addition, arcsin-square root transformed proportions of per cent identities showed only low correlations with estimated DESeq2 rld values of the detoxification genes, which were similar amongst all possible insect species and plant species combinations (Pearson's r = XX, P = XX). Rld stands for regularized log transformation of the original count data to a log2 scale by fitting a model with a term for each sample.

For visualizations by principal component analysis (PCA), ANOVA and hierarchical clustering, DESeq2 rld values were used as suggested by the authors of the package (Love et al. 2014). The PCA, ANOVA and hierarchical clustering were performed using the Partek® Genomics Suite® software, version 6.6 (v6.6; St. Louis, MO, 2014). The correlation method was applied to calculate the dispersion matrix of the PCA and the eigenvectors were normalized. ANOVA was used to calculate the mean sources of variation for all the genes (*B. tabaci* and plant species as main effects and their interaction). For the hierarchical clustering, Pearson's dissimilarity and complete linkage were applied.

Verification of differential expression by real-time polymerase chain reaction

The expression levels of nine detoxification genes that were differentially or equally expressed by the RNA-seq approach, were validated using quantitative reverse transcription PCR (qRT-PCR). Comparisons were made between the SSA1-SG3 and MEAM1 species, which showed the highest (98.33%) and lowest (95.58%) gene identity to the curated dataset. Primers were designed using the NCBI's Primer-Blast tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/), based on cDNA sequences retrieved from the RNA-seq. Specificity was improved using the Primersearch tool of EMBOSS (Rice et al. 2000) implemented in Galaxy (Blankenberg et al. 2007), allowing 30% mismatches. Only primer pairs with a sole target sequence were utilized (Table S4). The qRT-PCRs were performed using RNA extracted from the MEAM1 and SSA1-SG3 populations feeding on eggplant. An aliquot of 500 ng total RNA from each of the three biological replicates, for each species, served as template for cDNA synthesis using 5X All-In-One RT MasterMix (Applied Biological Materials), according to the manufacturer's instructions. For PCR, each18 µl reaction contained 9 µl Power SYBR Green PCR Master Mix (Applied Biosystems), 150 nM forward and reverse primers (Table S4) and 2 µl of cDNA. PCR cycling conditions were 95 °C for 1 min, followed by 40 cycles of 95 °C for 15 sec and 60 °C for 1 min. Melt curve analyses were performed to test the specificity of amplicons. A serial dilution of cDNA was used to generate standard curves for each gene in order to assess PCR efficiency and quantitative differences amongst samples. Quantification of transcript level was conducted according to the ΔCT method using the ribosomal protein L13a (RPL13A) as the reference gene (Collins et al. 2014). Comparisons of transcription levels were conducted using a one-way ANOVA model (species, MEAM1 versus SSA1-SG3, as the main effect) for each gene separately. Genes were considered significantly over- or under-transcribed when the ΔCT values of RNA samples from MEAM1 were different from ΔCT values of RNA samples from SSA1-SG3 at $P \leq 0.05$. All statistical analyses conducted were performed with JMP statistical software v13 (SAS Institute, USA).

A perfect match was observed between the RNA-seq and qRT-PCR analyses. Eight of the analyzed genes showed the same (and significant) differences in expression pattern between the MEAM1 and SSA1-SG3 species (two overexpressed in MEAM1 and six overexpressed in SSA1-SG3). The ninth gene (*GST_28*) showed non-significant differences in gene expression levels (between the two

species) both in the RNA-seq and qRT-PCR methods. All qRT-PCR data are summarized in Table S5.

Results

Do all the species of the Bemisia tabaci complex share a common host range?

The true extent of knowledge on the host range of the different *B. tabaci* species is patchy (Boykin et al. 2013). We first conducted a literature survey, therefore, in which we documented all of the major sampling efforts where a *mtCOI* barcode (of the collected *B. tabaci* sample) was linked with a specific plant-host species. Clustering analysis, at the botanical family level (Fig. 1A), indicated that one species in the complex (MEAM1) can be considered as a true "generalist", four species can be considered to be species with "extended" host ranges (Asia-I, Indian-Ocean, MED, SSA1), while the others can be roughly divided into two groups based on their characteristics of being with a more- or less-restricted host range. Mapping the data to the host plant orders (Fig. 1B), allowed the identification of nine orders (belonging to the Asterids and Rosids clades) that are commonly shared by most *B. tabaci* species: Asterales, Fabales, Rosales, Cucurbitales, Malvales, Malpighiales, Brassicales, Solanales and Lamiales. However, reconstruction of ancestral host range (using the ML algorithm) suggested that only plant species belonging to the Lamiales and Solanales orders can be considered to be ancestral hosts with high probability ($P \geq 0.91$), while the other seven orders presented a ≤ 0.5 probability (Table S2). Further analysis of these seven orders by a MP algorithm, showed that the Malpighiales may also be ancestral hosts of *B. tabaci*, while no decision could be made for the Malvales. The MP algorithm did not support the presence of the other five orders: Asterales, Fabales, Rosales, Cucurbitales and Brassicales in the ancestral host range of *B. tabaci* species (Fig. S3).

Based on these findings (and species availability), six *B. tabaci* species (representing different geographical and diet breadth major genetic groups) were selected for further analyses: MEAM1, MED-Q1, Uganda-MED-ASL, NW2, Asia II-1 and SSA1-SG3. In parallel, four plant species, representing different probabilities of being considered as a common/ancestral host plants, were included in the experimental setting: eggplant and pepper (representing the *Solanum* and *Capsicum* genera of the Solanaceae, Solanales), cassava (Euphorbiaceae, Malpighiales) and kale (Brassicaceae, Brassicales), (more details are provided in the "*Materials and Methods*" section).

Selected species performance on the various plant hosts

We first tested if the literature survey expectations "held" in our experimental system. Adult survival was monitored 24 h after subjecting newly emerged adults from the six selected *B. tabaci* species (all reared on eggplant for at least two months), to eggplant (control), cassava, kale and pepper plants. Both main treatments (*B. tabaci* species and plant host) were found to affect adult survival significantly ($P < 0.0001$ and $P = 0.0001$, respectively), but their interaction was found not to be significant ($P = 0.148$). As expected, eggplant was the most suitable host, significantly differing from pepper and cassava, which significantly differed from kale (Fig 2A). Survival over all four plants (Fig. 2B), indicated that the six *B. tabaci* species can be divided roughly into two performance groups with MEAM1, Asia II-1 and SSA1-SG3, showing higher performance on the various hosts compared to NW2, MED-Q1 and Uganda-MED-ASL ("high performance" and "low performance" groups, respectively). This result was in general agreement with our literature survey data (see above), which predicted MEAM1 and SSA1-SG3 to have an extended host repertoire compared to NW2 and Uganda-MED-ASL. The two exceptions were Asia II-1, which seemed to be capable of preforming well on more hosts than reported in the literature and MED-Q1, which did not tolerate host-plant switches well. Plant by plant examination indicated that the six species did not differ in their survival on eggplant (Fig. 2C). MEAM1 performed significantly better than NW2, MED-Q1 and Uganda-MED-ASL on kale and pepper (Figs 2D and 2E). SSA1-SG3 performed significantly better than NW, MED-Q1 and Uganda-MED-ASL on cassava, while MEAM1 and Asia II-1 differed significantly only from the latter two (Fig. 2F).

General expression profiles of detoxification genes/contigs on the different host plants

As indicated above, we focused this study on the detoxification system, one of the major gene groups involved in host utilization in herbivorous insects. We mapped the sequencing reads obtained from the individual samples to the 298 detoxification genes dataset of *B. tabaci*. We first used ordination and statistical methods to analyze co-expression between the candidate genes, comparing gene expression values from the six analyzed species on the four host plants. Principal component analysis (PCA) showed that the samples group together mainly according to their species association (Fig. S4). This finding was backed up by two-way analysis of variance (ANOVA) on rld values, which indicated that the majority of variance in the experimental system was associated with differences between species (70.79%), while only 3.66% and 8.77% were associated, respectively, with the effects of 'plants' and the 'plants' X 'species' interaction. Hierarchical clustering analysis was then conducted, using Pearson's dissimilarity for distance measure and complete agglomeration method for clustering (Fig. 3). Similar to the PCA and ANOVA findings, the four diet samples of each species clustered together at the first hierarchical level. In addition, there was no clear grouping of samples by host-plant within each species. An important result was obtained at the next clustering level (above species), where the detoxification gene expression pattern of the species clustered according to their host-performance groups (putatively reflecting their diet breadth) and not according to their geographical association: MEAM1, Asia II-1 and SSA1-SG3 in one group and NW2, MED-Q1 and Uganda-MED-ASL in the other. As argued above, an identified link between the species-plant associations and their gene expression pattern, best supports an evolutionary scenario in which differences in host-plant utilization contribute to the species-complex divergence. If, as commonly argued, only geography played/s a role in the species divergence and all species indeed share a similar host range, we would have expected MEAM1 to group with its own geographical group (MED-Q1 and Uganda-MED-ASL), and Asia II-1 to share a clade with the NW2 species (see the phylogenetic tree presented in Fig. S1A).

Constitutive expression differences between B. tabaci species

The transcriptomic profile of each species on eggplant (an ancestral suitable host plant for all six species, Fig. 2C) was used as a baseline and the constitutive expression differences in detoxification genes were compared between the six *B. tabaci* species. As explained earlier, constitutive differences (also referred by some authors as selected differences) might reflect targets of selection and adaptation processes in the different species, imposed by different suites of plant-chemical defenses. These differences, alternatively, may also simply result from independent responses to genetic drift. To avoid non-conclusive differences in expression patterns, only genes significantly overexpressed or underexpressed in one species compared to all others were considered in the first instance.

Overall, from the 298 genes analyzed, 107 were significantly constitutively overexpressed or underexpressed in one species (compared to all others), with a slightly higher percentage of overexpressed ones (61%) (Fig. 4). SSA1-SG3 showed the highest number of constitutively overexpressed genes (29), followed by Asia II-1 (15) and NW2 (13). On the other hand, MEAM1, MED-Q1 and Uganda-MED-ASL had the lowest number of constitutively overexpressed genes (7, 4 and 2, respectively). A different pattern appeared when genes significantly underexpressed in one species (compared to all others) were considered. Here, there was a remarkable difference between the three species in the "low performance" group, NW2, MED-Q1 and Uganda-MED-ASL, with 9, 14 and 18, underexpressed genes, respectively, to the three "high performance" species, MEAM1, Asia II-1 and SSA1-SG3, with 2, 2 and 0, underexpressed genes, respectively. Assuming again that the two performance groups reflect differences in the species diet breadth, these results highlight the possibility that the evolution of reduced host range in *B. tabaci*, involves the suppression of unnecessary constitutive detoxification activity (to avoid increased toxicity of the modified compound or to reduce energetic costs).

We then attempted to identify detoxification genes that were constitutively overexpressed in all three "high performance" species (MEAM1, Asia II-1 and SSA1-SG3), relative to each of the "low performance" species. Thirty-eight genes that met this criterion, when compared to Uganda-MED-ASL (in comparison, only four genes showed the opposite expression pattern, $\chi^2_{(1)} = 31.81$, $P \leq 0.0001$), 28 genes when compared to MED-Q1 (again, only four genes were found to show the opposite expression pattern, $\chi^2_{(1)} = 20.25$, $P \leq 0.0001$) and 22 genes when compared to NW2 (this time 12 genes showed the opposite expression pattern, $\chi^2_{(1)} = 2.99$, $P = 0.084$). A set of eight

detoxification genes were overexpressed in all three "high performance" species relative to all three "low performance" species. This gene set might represent an ancestral detoxification tool-kit in the *B. tabaci* species complex, associated with a more general feeding habit (see Discussion).

Plastic expression differences within and between B. tabaci species

Plastic gene-expression responses to host switching in generalist insect herbivores are relatively well documented. It is commonly argued that these insects initially survive a host plant transfer by rearranging the expression profile of genes coding for detoxification enzymes (Celorio-Mancera et al. 2012, Dermauw et al. 2013, 2015). As both our literature survey and performance assays suggested, host-range differences between the analyzed *B. tabaci* species focused on detoxification plastic responses (within each species) after a transfer from a common ancestral host, eggplant, to the novel hosts, cassava, pepper and kale.

Our gene expression data indicated that 151 from the 298 (51%) detoxification genes analyzed were differentially-expressed. Of these, 95 were differentially-expressed in only one species, while 56 were differentially-expressed in more than one species. Interestingly, while the relative proportion of the six detoxification gene families in the 95 list (differentially-expressed in only one species) did not differ from that of the complete list of the 298 detoxification genes analyzed, ($\chi^2_{(5)} = 8.95$, $P = 0.11$), a significant enrichment of genes belonging to the P450 and UDPGT families was observed in the list of 56 genes differentially-expressed in more than one species ($\chi^2_{(5)} = 14.52$, $P = 0.012$), suggesting an important role to these two families in detoxification of plant defenses by *B. tabaci*.

The six species largely differed in the number of detoxification genes showing plastic responses to host transfer (Fig. 5) and three main expression patterns were revealed: (I) High plastic response to more than one plant host was seen in Asia II-1 and NW2 (74 and 84 genes, respectively), which modified their detoxification expression profile largely when transferred to cassava and to a lesser extent to pepper. The majority of genes responding to the transfer to pepper responded also to the transfer to cassava (23 from 24 or 96% in Asia II-1 and 26 from 37 or 70% in NW2). (II) High plastic response to just one plant host by the MED-Q1 and SSA1-SG3 species (41 and 27 genes, respectively) which modified their detoxification expression profile when transferred from eggplant to cassava or kale, respectively. (III) Low level of plastic response to host transfer by MEAM1 and Uganda-MED-ASL after transfer from eggplant to cassava (5 and 9 genes, respectively). In contrast to all other species, SSA1-SG3 (which is performing well on cassava) showed no plastic response (only one gene) when transferred from eggplant to cassava although the colony had been reared on eggplant for multiple generations, suggesting an optimally tuned constitutive (selected) expression of detoxification genes in this species for feeding and utilizing cassava plants. Following this observation, it is interesting to note that out of the 29 genes uniquely constitutively overexpressed in the SSA1-SG3 (Fig. 4), 12 were plastically overexpressed by Asia II-1, NW2 and MED-Q1 when transferred from eggplant to cassava.

Do B. tabaci species share a common "essential detoxification machinery" that allows the utilization of non-optimal host plants?

The oscillation hypothesis predicts that even species with restricted host range retain an essential "machinery" that allows some level of utilization of host plants that have been lost from the species repertoire (Janz and Nylin 2008). We asked, therefore, if a common and essential detoxification mechanism can be identified in *B. tabaci*. To avoid considering non-consistent changes from which a clear pattern cannot be obtained, only genes differentially-expressed in more than one species and showing the same expression pattern (upregulation or downregulation) in at least two species were considered. This reduced the original list of 56 genes, differentially-expressed in more than one species, to 45.

The list of 45 genes (Table 1) was significantly enriched in genes belonging to the P450 (24 genes from which 14 belong to sub-family 6 and 8 to sub-family 4) and UDPGT (14 genes) detoxification families ($\chi^2_{(5)} = 15.26$, $P = 0.0093$). Six genes were upregulated in response to host switch in four species, 12 were upregulated and two were downregulated in three species and 16 were upregulated and 9 were downregulated in two species (Table 1). Two interesting observations were made: (i) in nearly all genes (42 from 45), the plastic changes in expression in all species and on all host plants

occurred in the same direction (upregulation or downregulation). In the three exceptions, it was only the SSA1-SG3 species which did not follow the expression direction shared amongst the other species on the different plant hosts. (ii) With five exceptions (out of 30 cases), the detoxification genes which plastically responded to host transfer from eggplant to pepper and kale were the same ones that responded to the transfer to cassava, possibly suggesting the existence of an "essential detoxification mechanism". This machinery is assumed to participate in the insects' efforts to neutralize a wide and unrelated range of plant defensive chemistry shortly after host switches occur.

Discussion

We present here, two independent lines of evidence to suggest that species within the *B. tabaci* complex differ in their ability to accept or utilize multiple plant hosts. First, our literature survey of field-collection data clearly indicated that only a few species in the complex can be found on hosts from multiple botanical families, while the majority of species were limited to only few families (Fig. 1A).

Second, our adult survival assays clearly separated the six species into two performance groups, showing higher (MEAM1, Asia II-1, SSA1-SG3) or lower (NW2, MED-Q1, Uganda-MED-ASL) ability to survive on novel and toxic plant hosts (Fig. 2B), likely reflecting less-restricted and more-restricted feeding habits, respectively. At the same time, a consistent linkage between the species feeding habit and the mode of their detoxification gene expression (constitutive versus plastic) could not be established. Moreover, our hypothesis that species with a more general feeding habit will depend more on constitutive expression than on plastic expression of their detoxification system was only partially supported. While MEAM1 and SSA1-SG3 (less-restricted species) seemed to be considerably less plastic than MED-Q1 and NW2 (more-restricted species), Asia II-1 (a less-restricted species) was highly plastic while Uganda-MED-ASL (a more-restricted species) showed very low plasticity (Fig. 5). The same inconsistency was also observed when genes significantly overexpressed in one species compared to all others (constitutive overexpression). Whilst SSA1-SG3 showed the highest number of constitutively overexpressed genes (29) followed by Asia II-1 (15), MEAM1, the most generalist species in the *B. tabaci* complex, uniquely overexpress only seven genes (Fig. 4). To place these non-conclusive results into a broader perspective, two important issues are highlighted: (i) that the majority of studies that raise the possibility that plastic-regulation of detoxification genes are important for generalist insect survival, when first encountering an adverse host plant, have focused on non-phloem-feeding Orders such as Lepidoptera and spider-mites (Grbić et al. 2011, Schuler 2011, Dermauw et al. 2012, Tao et al. 2012, Celorio-Mancera et al. 2013, Vogel et al. 2014, Wybouw et al. 2015); (ii) that large-scale gene expression analyses clarifying the molecular mechanisms of host alternation and adaptation in major generalist phloem-feeding species are extremely limited. A recent study on the generalist aphid species, *M. persicae*, for example, compared colonies that were reared in parallel for one year on *Brassica rapa* or *Nicotiana benthamiana*. Comparison of the colony's transcriptomes identified only 171 differentially expressed genes putatively involved in host adjustment from a total of >18,000 genes (<1%). Moreover, the *B. rapa* clone successfully colonized *N. benthamiana* with no significant differences observed in survival, reproduction rates, weight, development time and longevity (Mathers et al. 2017). Although the outlined experimental system differed from ours, the common findings in the two systems highlights the possibility that successful short- or long-term host shifts of generalist phloem-feeders do not necessarily require significant plastic or constitutive changes in gene expression.

At the same time, two characteristics of our detoxification gene expression comparison support an evolutionary scenario in which differences in plant-host adaptations (or lack of adaptation) played a role in the *B. tabaci* species-complex evolution. The hierarchical clustering analysis clustered the six species according to their aforementioned host-performance groups and not according to their phylogenetic relationship (Fig. 3). This suggests the existence of a common detoxification gene expression pattern that can be associated with a more general feeding nature of some species, regardless of the estimated time of their separation. Moreover, it allows us to assume logically that some version of this common detoxification gene expression pattern was already present in the common ancestor of the *B. tabaci* species complex, which likely displayed a feeding habit toward the less-restricted side of the *B. tabaci* spectrum. We also show that at least some of these detoxification capabilities were lost by species with a more restricted host range (Fig. 4), suggesting that selection pressures might be particularly strong on some detoxification genes that are not required when adapting to a less-general diet. Moreover, their production (or the production of their coded enzymes)

might be energetically costly or capable of endangering the organism by producing modified/bio-activated deleterious molecules (Feyereisen 1999, Zeng et al. 2009, Halon et al. 2015). These two complementing findings best support an evolutionary model of species expansion (see below), in which generalist stages persist over long evolutionary time spans and can produce multiple descendants with a more restricted host range (Stireman 2005, Nylin et al. 2013).

The hypothesis that the different species in the *B. tabaci* complex differ in their host range is definitely not new (Brown et al. 1995, 1996, Perring 2001), although it was rightly argued that this is more an assumption than a "solid" fact, as experimental studies comparing performance across different hosts are few and restricted not only in the numbers of host plant species included, but also primarily to comparisons between MEAM1 and other non-MEAM1 species (Zhang et al. 2006, Xu et al. 2011). What is new here is the postulated link we make between the speciation process in this group and the observed differences in the species' host-plant ranges (especially less-restricted to more-restricted transitions). The very few studies that previously considered this issue did not provide a clear statement, but argued that there is only limited evidence to support this possibility (De Barro 2005, De Barro et al. 2005). Alternatively, it was assumed that the host range of each of the various species identified in Asia or Africa are fairly equivalent, and that it is possible that in small spatial-scales, density-dependent competitive interactions have operated to exclude an invader belonging to a different species (De Barro 2005, De Barro 2011).

If we take a broader perspective, however, it is clear that much of what we know about host-associated genetic differentiation involves insect species with narrow host-plant ranges. For example, the pea aphid, *Acyrthosiphon pisum*, encompasses at least eight conspecific host races specialized on different legume species (Peccoud et al. 2009). A question arises, therefore, regarding the feasibility and importance of host-associated differentiation in the speciation process of generalist insects such as *B. tabaci*. Evidence of host-associated differentiation in generalist species is accumulating. For example, the polyphagous aphid *M. persicae* comprises a specialized form on tobacco (*N. tabacum*) as a summer host that is formally designated as the subspecies *M. persicae nicotianae*. In Greece, sexual populations in tobacco-growing areas are genetically differentiated from those in other regions (Margaritopoulos et al. 2007). Other examples include grasshoppers and green mirids, which feed on multiple hosts from different families, yet were shown to exhibit host-associated differentiation (Antwi et al. 2015). Moreover, retrospective reading of the *B. tabaci* literature highlights multiple occasions in which the presence of host specialization was detected. Bird (1957) reported in Puerto Rico the existence of the narrowly specific 'Jatropha' race of *B. tabaci* occupying a distinct niche of Euphorbiaceae plants (*Jatropha gossypifolia* and *Croton lobatus*) and the more polyphagous 'Sida' race which could colonize numerous plant species belonging to the Malvaceae, Euphorbiaceae, and Leguminosae families. Burban et al. (1992) observed in the Ivory Coast a race considered largely confined to cassava (Euphorbiaceae) and a polyphagous race that did not utilize cassava as a host. More recently, two genetic variants with restricted host ranges were recorded in Italy. Variant T had only been collected from natural areas of Italy, where it was found to be monophagous on *Euphorbia characias* (Euphorbiaceae) (Simon et al. 2003). Variant Ru was collected almost exclusively from *Rubus ulmifolius* (Rosaceae) and once from *Vitis vinifera* (Vitaceae) (Parrella et al. 2012). Sseruwagi et al. (2005) reported a new *B. tabaci* genotype cluster, Ug3, which occurred exclusively on *Ocimum gratissimum* (Lamiaceae) and this was not found to colonize any other plant species.

We agree that all of these reported cases of host-associated specialization and speciation in *B. tabaci* might still be considered as the "exception" and not the "rule". Still, we would like to argue here that recognizing the *B. tabaci* complex as a group of separate species with large/limited ranges of host plants, may give us an opportunity to develop a new working hypothesis for providing an evolutionary mechanism/s for explaining the species extreme expansion, especially in some sympatric occasions. As already stated above, the oscillation hypothesis has drawn attention to the possibility that lineages could alternate between more generalized and more specialized phases (Jans et al. 2006, Janz and Nylin, 2008). We find it quite appealing to apply the major principles of the oscillation hypothesis for explaining the extensive (and not common in the genus) speciation process that has occurred in the *B. tabaci* group. This is not only because we found expression similarity in species with a more general feeding nature, but also because we found evidence for the maintenance of putative essential host-utilization "machinery", common to many of the *B. tabaci* species analyzed (Table 1). We hypothesize that, on an evolutionary time scale, multiple reorganizations of the ancestral "machinery" occurred, through a process termed genetic accommodation (West-Eberhard

2005). This happened to enable processes in which females were evolving to prefer the plants that maximize their fitness, given the local patterns of seasonality, the local abundances of potential host plants, the life-history consequences of feeding on one plant or the other, and the fitness consequences of these particular life histories under the local patterns of seasonality and climate, leading eventually to speciation (Schluter 2009).

We conclude by presenting a new evolutionary scenario for *B. tabaci* diversification. This scenario takes into account and tries to resolve contradictions that currently appear in the literature reporting on the molecular dating of the species-complex diversification. While Boykin et al. (2013) provided detailed analysis of the diversification of all major clades of the *B. tabaci* species complex and came to the conclusion that the major lineages within the complex arose approximately 60-30 MYA, Santos-Garcia et al. (2015) estimated the divergence time of the *B. tabaci* complex to be 18.43 (28.5-9.85) MYA, and Hsieh et al. (2014) estimated that the lineage is even younger and originated 6.47 MYA. A consideration of these data with our literature survey and host reconstruction analysis, allows us to suggest that *B. tabaci* diversification is a combination of allopatric and sympatric forces. The allopatric forces probably include two migration waves associated with the break-up of Gondwanaland and the collision between the Indian and Asian plates (50-35 MYA), and the collision of the African (Arabian) and Eurasian plates during the Mid-Miocene climatic optimum (20–14 MYA) (Zachos et al. 2001). The collision of the Arabian and Eurasian plates allowed overland migration from Africa to Asia through Arabia (Yu et al. 2014) and the Beringia Land Bridge allowed migration between Asia and North America (Tiffney 1985). In the late Miocene through the Pliocene, regional intensification of aridification occurred on all continents and is thought to have advanced the rapid development of modern arid ecosystems and tropical grasslands (Horn et al. 2014). Many angiosperm clades are thought to have diversified rapidly and dispersed due to Miocene aridification, including major botanical families that are common hosts of *B. tabaci* (Fig. 1A). Taking into account the logical assumption that herbivorous insects disperse with their hosts, we argue here that the clustering together of the three species showing a less-restricted feeding habit (MEAM1, Asia II-1 and SSA1-SG3), not only allows us to raise the possibility that the common ancestor of the *B. tabaci* species complex displayed a relatively less-restricted feeding habit, but also that its increased ability to utilize many host species might actually have been the cause of subsequent speciation and not its consequence, by enabling geographic expansion and therefore an opportunity for spatial isolation and genetic divergence to occur within each continent (Futuyma and Agrawal 2009).

Appendix 6: Conference presentation

Stainton D, van Brunschot S, Hereward J, Walter G. (2017) Invasive and native whiteflies in Australia: cryptic species complex, endosymbionts and begomoviruses. Proceedings of the International Fuscom Meeting, 8th-9th August, Toowoomba, Queensland, Australia.

Link to presentation:
https://www.dropbox.com/s/1zg9stqdjhrm6fq/2017_Fuscom_Stainton_UQ1305.pptx?dl=0

Abstract

Invasive and native whiteflies in Australia: cryptic species complex, endosymbionts and begomoviruses.

Daisy Stainton[1], Sharon van Brunschot[2], James Hereward[1], Gimme Walter[1]

[1] *School of Biological Sciences, The University of Queensland, Brisbane, Australia.*
[2] *Natural Resources Institute, University of Greenwich, Kent, UK.*

Cotton leaf curl disease (CLCuD) is caused by a number of viruses, which are spread by the silverleaf whitefly *Bemisia tabaci*, a major pest of cotton in its own right. Epidemics of CLCuD overseas have reduced cotton yield and resulted in large financial losses. Symptoms include upward leaf curling, vein thickening and leaf-like growths on the underside of the leaves (enations). None of the viruses which cause CLCuD are present in Australia. However, a number of whitefly species are present here, including native Australian species, and these may have the potential to transmit CLCuD should an incursion occur.

Many insects, including *B. tabaci,* contain bacteria within their cells. These endosymbionts can influence the biology of the insects and may potentially play a role in their ability to transmit plant viruses. Little is known about the endosymbionts that are present in Australian whiteflies. This project is investigating the identity of the whiteflies present in Australia, the endosymbionts they harbour, and the ability of these whiteflies to transmit the viruses that cause CLCuD.
Genetic characterisation of the species present in Australia has identified a number of *B. tabaci* in Australia, including a species not previously known to be present here. Next-generation sequencing tools have been used to gain further insight into the whiteflies present in Australia and the endosymbionts that are hosted by these insects. The endosymbiont profiles harboured by the different species are being determined. Overseas experiments associated with this project, and within special containment facilities, are investigating which whitefly species can transmit the viruses that cause CLCuD.

Although the viruses that cause CLCuD are not present in Australia, epidemics of CLCuD overseas have reduced cotton yield and resulted in large financial losses. Within Australia there are a number of *B. tabaci* species, so understanding the ability of CLCuD to establish in Australia requires knowledge of the potential vectors that are present here. At the completion of this project, we will not only have gained in depth knowledge of the *B. tabaci* whitefly present in Australia, but will also have enhanced our understanding of the threat of CLCuD establishing in Australia should an incursion occur.

Appendix 7: Peer-reviewed journal article

Mabvakure B, Martin DP, Kraberger S, Cloete L, **van Brunschot S,** Geering AD, Thomas JE, Bananej K, Lett J-M, Lefeuvre P (2016) Ongoing geographical spread of *Tomato yellow leaf curl virus*. Virology 498:257-264

Virology 498 (2016) 257–264

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Ongoing geographical spread of *Tomato yellow leaf curl virus* ☆

Batsirai Mabvakure [a], Darren P. Martin [b], Simona Kraberger [c,d], Leendert Cloete [a], Sharon van Brunschot [e,f], Andrew D.W. Geering [f], John E. Thomas [f], Kaveh Bananej [g], Jean-Michel Lett [h], Pierre Lefeuvre [h], Arvind Varsani [b,c,i,*], Gordon W. Harkins [a,**]

[a] South African National Bioinformatics Institute, University of the Western Cape, Bellville 7535, South Africa
[b] Department of Clinical Laboratory Sciences, University of Cape Town, Cape Town, South Africa
[c] School of Biological Sciences and Biomolecular Interaction Centre, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
[d] Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO, USA
[e] School of Biological Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia
[f] The University of Queensland, Queensland Alliance for Agriculture and Food Innovation, Ecosciences Precinct, GPO Box 267, Brisbane, QLD 4001, Australia
[g] Department of Plant Virus Research, Iranian Research Institute of Plant Protection (IRIPP), Tehran 19395-1454, Iran
[h] CIRAD, UMR PVBMT, Pôle de Protection des Plantes, 7 chemin de l'IRAT, Saint-Pierre, Ile de la Réunion 97410, France
[i] The Center for Functional Microbiomics, The Biodesign Institute and School of Life sciences, Arizona State University, Tempe, AZ 85287, USA

ARTICLE INFO

Article history:
Received 11 July 2016
Returned to author for revisions
29 August 2016
Accepted 30 August 2016

Keywords:
TYLCV
Phylogeography
Geminivirus
Tomato

ABSTRACT

Tomato yellow leaf curl virus (TYLCV) seriously impacts tomato production throughout tropical and sub-tropical regions of the world. It has a broad geographical distribution and continues to spread to new regions in the Indian and Pacific Oceans including Australia, New Caledonia and Mauritius. We undertook a temporally-scaled, phylogeographic analysis of all publicly available, full genome sequences of TYLCV, together with 70 new genome sequences from Australia, Iran and Mauritius. This revealed that whereas epidemics in Australia and China likely originated through multiple independent viral introductions from the East-Asian region around Japan and Korea, the New Caledonian epidemic was seeded by a variant from the Western Mediterranean region and the Mauritian epidemic by a variant from the neighbouring island of Reunion. Finally, we show that inter-continental scale movements of TYLCV to East Asia have, at least temporarily, ceased, whereas long-distance movements to the Americas and Australia are probably still ongoing.

1. Introduction

Tomato yellow leaf curl virus (TYLCV) is a monopartite begomovirus in the family *Geminiviridae* and is one of many closely related viruses that cause tomato yellow leaf curl disease (TYLCD) (Abhary et al., 2007; Navot et al., 1991). TYLCD was initially recognised in the Jordan Valley, Israel, in the 1930s, but it was not until the early 1960s that TYLCV was identified (Cohen and Nitzany, 1960, 1966). Subsequently, the virus has spread unabated into the Mediterranean basin and into most tropical and sub-tropical regions of the world and is recognised as one of the world's most devastating pathogens of tomato (Abhary et al., 2007; Delatte et al., 2007; Delatte et al., 2005; Diaz-Pendon et al., 2010; Duffy and Holmes, 2007; Kenyon et al., 2014; Lefeuvre et al., 2010; Moriones and Navas-Castillo, 2000; Péréfarres et al., 2012; Picó et al., 1996; Polston and Anderson, 1997; Stonor et al., 2003; Van Brunschot et al., 2010).

Although there are seven recognised strains of TYLCV (Brown et al., 2015), only two, the mild (Mld) and Israel (IL) strains, have ever been found outside of Iran. The global dissemination of TYLCV-Mld and TYLCV-IL from the Middle East or the Eastern Mediterranean (Duffy and Holmes, 2007; Lefeuvre et al., 2010) is attributed to the movement of infected planting material (Seal et al., 2006), together with spread of the Middle East-Asia Minor (MEAM1 formally referred to as the B biotype) and the Mediterranean (MED formally referred to as the Q biotype) cryptic species of its whitefly vector, *Bemisia tabaci* (Czosnek et al., 2002; Diaz-Pendon et al., 2010; Horowitz et al., 2007; Seal et al., 2006). Recent reports suggest that TYLCV is possibly unique amongst begomoviruses in that it is capable of both replicating within *B. tabaci* (Pakkianathan et al. (2015), as well as being seed

☆GenBank Accession #s: KX347094–KX347145, KX347155–KX347172.

* Corresponding author at: The Center for Functional Microbiomics, The Biodesign Institute and School of Life sciences, Arizona State University, Tempe, AZ 85287, USA.
** Corresponding author.
E-mail addresses: arvind.varsani@asu.edu (A. Varsani), gordon@sanbi.ac.za (G.W. Harkins).

http://dx.doi.org/10.1016/j.virol.2016.08.033
0042-6822/

transmitted in tomato (Kil et al., 2016). These characteristics may have contributed to it achieving a geographical range that is far broader than those of almost all other begomovirus species.

As with other begomoviruses, TYLCV is able to rapidly adapt to new environments as a consequence of its high rates of mutation and recombination (Delatte et al., 2005; Duffy and Holmes, 2007, 2008; Lefeuvre et al., 2010; Monci et al., 2002). For example, TYLCV-IL is a recombinant of TYLCV-Mld and *Tomato leaf curl Karnataka virus* (another tomato-infecting begomovirus), while other begomoviruses from the Mediterranean basin are recombinants of TYLCV-IL and TYLCV-Mld (Navas-Castillo et al., 2000).

The global spread of TYLCV began in the 1980s, after the emergence of the Mld and IL strains (Duffy and Holmes, 2008; Lefeuvre et al., 2010). The region centred on Iran harbours the highest diversity of TYLCV, although there has been little obvious movement of viruses out of this region since before the early 1980s (Lefeuvre et al., 2010). A previous phylogeographic study by Lefeuvre et al. (2010) included 91 coat protein and 82 full genome sequences of TYLCV, which had been generated over 22 years. However, this study was limited in geographical scope, as the virus isolates were primarily from the Mediterranean basin, the Middle East and the Americas, with Southeast Asia, the Pacific and Indian Ocean island nations/territories and Australia being greatly underrepresented. Furthermore, the analytical tools to account for the potentially confounding influences of recombination were not then available.

Here we analyse a much larger TYLCV sequence dataset comprising 414 full-genome sequences (70 of which are published here for the first time) sampled over 26 years from 33 countries to infer the historic global movement dynamics of TYLCV. Using fully probabilistic Bayesian modelling methods and accounting for recombination, we specifically focus on the contributions of south-western Pacific (Australia and New Caledonia) and south-eastern Indian Ocean (Mauritius and Reunion) states to the spread of TYLCV. Also, because of the intensified sampling for TYLCV over the past five years in various other parts of the world, we are also able to provide much more clarity on TYLCV movements into and across Asia, the Americas and the Caribbean.

2. Methods and materials

2.1. Sampling, TYLCV genome recovery and sequencing

Total DNA was extracted from tomato samples displaying leaf curl symptoms from Australia (n=52), Iran (n=12) and Mauritius (n=6). Circular DNA was enriched by rolling circle amplification (RCA) using Templiphi (GE Healthcare, USA). Unit length TYLCV genomes were recovered from the RCA concatemers using *Xmn*I, *Nco*I, *Bam*HI or *Sal*I restriction enzymes, and cloned into pJET 1.2 plasmid vector (ThermoFisher, USA) for *Xmn*I digested genomes and into pBluescript SK (Stratagen, USA) for *Nco*I, *Bam*HI or *Sal*I digested genomes. The recombinant plasmids were Sanger sequenced by primer walking at Macrogen Inc. (South Korea). Complete genome sequences were assembled using DNA Baser V4 (Heracle Biosoft S.R.L., Romania).

2.2. Construction of a recombination free dataset

A dataset of 435 full TYLCV genomes was assembled, which contained sequences of isolates sampled from 33 countries between 1988 and 2014 (Supplementary Table 1) including 356 full genome sequences retrieved from GenBank.

A preliminary multiple sequence alignment was generated using the slow, iterative refinement method (FFT-NS-I) implemented in MAFFT version 7 (Katoh and Standley, 2014). This alignment was then manually edited using IMPALE (available from http://web.cbio.uct.ac.za/~arjun/).

The resulting alignment was used for recombination analyses using the seven detection methods implemented in RDP version 4.36 (Martin et al., 2015) with default settings and a Bonferroni corrected *p*-value cut-off of 0.05. Events detected with three or more methods coupled with significant phylogenetic support were considered credible evidence of recombination. The breakpoint positions and recombinant sequence(s) inferred for every detected potential recombination event were manually checked and adjusted where necessary using the extensive phylogenetic and recombination signal analysis features available in RDP4.56 (Martin et al., 2015).

The final TYLCV recombination-free dataset (RF-dataset) comprised 414 TYLCV sequences, all generated following recombination analysis and the removal of (i) all tracts of sequence from the alignment that were detected to have been acquired through recombination (replaced in the alignment with gap characters), and (ii) 21 sequences from the TYLCV dataset that were inferred to have acquired >30% (or >810 nucleotides) of their genomes via recombination with non-TYLCV parental viruses.

2.3. Geographical clustering

Geographical clustering was done as described by Lefeuvre et al. (2010), using the centroid hierarchical clustering method (Rokach and Maimon, 2005) implemented in R (R Core Team, 2013) to determine the most appropriate regional grouping scheme for the phylogeographic analyses.

2.4. Identification of best-fit evolutionary models

The best-fit nucleotide substitution model was inferred using jModelTest (Posada, 2008) implemented in MEGA6 (Tamura et al., 2013) and the best-fit molecular clock, and demographic models were inferred using Path Sampling and Stepping stone methods with 100 path steps and a chain length of one million (Baele et al., 2012; Baele et al., 2013) using BEAST v1.8.1. (Drummond and Rambaut, 2007) and the BEAGLE high-performance library v2.1.2 (Ayres et al., 2012).

We used linear regression techniques available in TempEst (Rambaut et al., 2016) to visually examine the degree of divergence accumulation that had occurred over the sampling time interval as a proxy for temporal signal. This method explores the root-to-tip distances of the branches in the maximum likelihood tree as a function of sampling time. In this analysis, TempEst outputs the correlation coefficient and the coefficient of determination, for which higher values indicate strong temporal signal in the data, and improved fit of the data to the strict clock nucleotide substitution model, respectively.

2.5. Phylogeographic analyses

A discrete reversible diffusion model with the Bayesian stochastic search variable selection (BSSVS) procedure (Lemey et al., 2009), implemented in BEAST v1.8.1. (Drummond and Rambaut, 2007), was used to conduct Bayes factor (BF) tests that identified the statistically supported epidemiological links between the geographical regions considered (Lemey et al., 2009). Statistically supported links between locations were identified as those with an associated BF test statistic >5: where BF scores >100 were taken as representing decisive support for one or more movements between locations, BF scores >10 were taken as indicating strong support for movement(s), and BF scores <5 were taken as indicating negligible support (Kass and Raftery, 1995).

To determine whether the inference of the most probable root location (interpreted as the geographical origin of TYLCV) was biased towards locations/countries with the largest sample sizes, we compared the root probability results obtained with and without applying a tip swap location randomization procedure in BEAST v1.8.1 (Drummond and Rambaut, 2007).

The length of the Markov chains that were explored during these analyses had between 1×10^8 and 3×10^8 steps for the ten replicate runs of carried out with each model. When similar results were obtained with independent replicate runs of the chain for a particular model, the log and tree files were combined using LogCombiner (a computer program available in the BEAST v1.8.1 package; (Drummond and Rambaut, 2007). For all models, the runs were continued until effective sample size (ESS) values for all individual model parameters exceeded 200.

TreeAnnotator, which is also available as a component of the BEAST v1.8.1 package, was used to produce an annotated maximum clade credibility tree from the posterior distribution of trees produced during the MCMC analyses. This tree was visualized in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). The program SPREAD version 1.06 (Bielejec et al., 2011) was used to calculate BFs (Kass and Raftery, 1995) for potential TYLCV movements inferred using the discrete reversible diffusion model. SPREAD was also used to generate a key markup language (kml) file for the visualization of TYLCV movements using Google-Earth (Supplementary Data 1).

3. Results and discussion

3.1. Classification of geographical data into regions

Prior to analysing the movement dynamics of TYLCV, it was necessary to classify the 414 sequences in the RF-dataset based on their geographical origins. A hierarchical clustering method based on the geographical distances separating all these sequences indicated that they fell into twelve reasonably distinct geographical clusters that we named Africa (n=2), North & Central America (n=15), Australia (n=60), China (n=149), East Asia (n=64), Eastern Mediterranean (n=13), Western Mediterranean (n=17), Mauritius (n=6), Middle East (n=71), New Caledonia (n=6), Reunion Island (n=2) and Caribbean (n=9) (Fig. 1; Supplementary Table 1).

3.2. Estimation of best-fit evolutionary models and TYLCV nucleotide substitution rate estimates

The nucleotide substitution model that best fit the RF-dataset

	Country	Number
Africa	Egypt	2
Australia	Australia	60
China	China	149
East Asia	Korea	45
	Japan	19
Eastern Mediterranean	Israel	2
	Jordan	8
	Lebanon	2
	Turkey	1
Mauritius	Mauritius	6
Middle East	Iran	53
	Iraq	1
	Kuwait	2
	Oman	15
	Saudi Arabia	2
New Caledonia	New Caledonia	6
North & central America	Guatemala	1
	Mexico	6
	USA	8
Reunion Island	Reunion Island	2
Caribbean	Costa Rica	3
	Cuba	1
	Dominican Republic	1
	Grenada	2
	Puerto Rico	1
	Venezuela	1
Western Mediterranean	Italy	1
	Morocco	10
	Netherlands	1
	Portugal	1
	Spain	3
	Tunisia	1

	Root location probability	
Location	Normal	Randomised
Africa	0.0010	0.0010
Australia	0	0.0502
China	0	0.3883
East Asia	0	0.2968
Eastern Mediterranean	0.5500	0.2125
Mauritius	0.0020	0.0021
Middle East	0.4230	0.0171
New Caledonia	0.0010	0.0067
North & Central America	0	0.0093
Reunion	0.0010	0.0057
Caribbean	0.0010	0.0041
Western Mediterranean	0.0210	0.0062

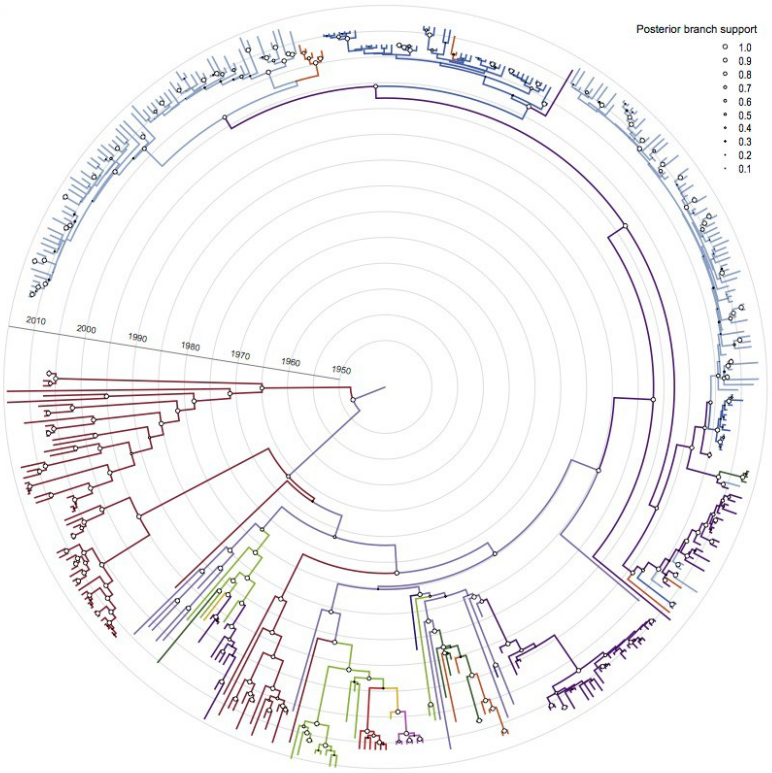


Fig. 1. Temporally scaled maximum clade credibility (MCC) tree constructed using 414 recombination-free TYLCV sequences. Branches are coloured according to where out of the twelve considered geographic regions the ancestral sequences represented by these branches most likely occurred. The posterior probability support of branches is represented by circles at the tip nodes of branches, with circle sizes being proportional to degrees of branch support. Also indicated are both the numbers of TYLCV samples from each region, and the probabilities of each region being the origin of the most recent common ancestor of all the sampled TYLCVs.

was the generalized time-reversible model with four gamma rate categories and a proportion of invariant sites (GTR+G4+I). The best fit molecular clock and demographic model combination, identified using marginal likelihood estimation (MLE) by the path sampling and stepping stone methods (Baele et al., 2012), was the uncorrelated lognormal relaxed molecular clock model with the Bayesian Gaussian Markov random field (GMRF) skygrid coalescent tree prior (Gill et al., 2013; Minin et al., 2008).

The correlation between root-to-tip divergence and sampling time for the RF-dataset, inferred using TempEst, yielded an r value of 0.2727, with a residual r^2 of 7.4374×10^{-2}, which indicated that although a strict molecular clock model was unlikely to fit our data very well, the data likely contained a detectable signal of sequence divergence throughout the sampling interval.

The mean TYLCV nucleotide substitution rate for the RF-dataset was determined to be 8.8929×10^{-4} (95% HPD 7.7457×10^{-4} to 9.9679×10^{-4}) subs/site/year, which is faster than previously reported rates for complete TYLCV genomes with recombination included (Duffy and Holmes, 2008; Lefeuvre et al., 2010; Yang et al., 2014), but is similar to those reported for a largely-recombination free TYLCV coat protein dataset (Lefeuvre et al., 2010).

The maximum clade credibility (MCC) tree for the RF-dataset (Fig. 1; Supplementary Fig. 1) indicated that the most recent common ancestor (MRCA) of the TYLCV isolates examined here occurred in the either the Eastern Mediterranean (p =0.55) or Middle East (p=0.42) around 1946 (95% HPD= 1914–1971) (Fig. 1). It is unlikely that this inference is attributable to uneven sampling density among the locations, since the most probable locations of the MRCA inferred from MCC trees with randomized sampling locations were China, followed by East-Asia (Fig. 1). The modal location state estimates, indicated by the branch colours in the MCC tree, also reveal a reasonably strong spatial structure for this virus across its geographic range.

It must be stressed, however, that our conclusions regarding a Eastern Mediterranean or Middle-Eastern origin of the TYLCV MRCA have two associated caveats. First, discrete phylogeographic analyses such as we have performed are incapable of inferring any origin location outside of the regions from which the analysed sequences were sampled. Second, besides biases caused by uneven spatial sampling density, these analyses could have also been biased by uneven temporal sampling density across different locations. For example, it is possible that the fact that the oldest TYLCV sequences that we analysed were sampled from the Eastern Mediterranean may have unduly influenced the identification of this region as the most-probable location of the MRCA. However, it is noteworthy that there seemed to be no obvious over-all association between the temporal depth of sampling at particular locations and the inferred probability of those locations being the MRCA. Specifically, while the second most probable MRCA location, the Middle-East (p=0.423) had only a single sequence sampled prior to 2000, the two regions with the most samples collected prior to 2000, East Asia (n=5) and the Western Mediterranean (n=4), had associated posterior probabilities of being the MRCA location of 0.00 and 0.02, respectively.

3.3. Geographical dissemination of TYLCV

A total of 18 statistically supported epidemiological links were inferred across the twelve geographical regions that were

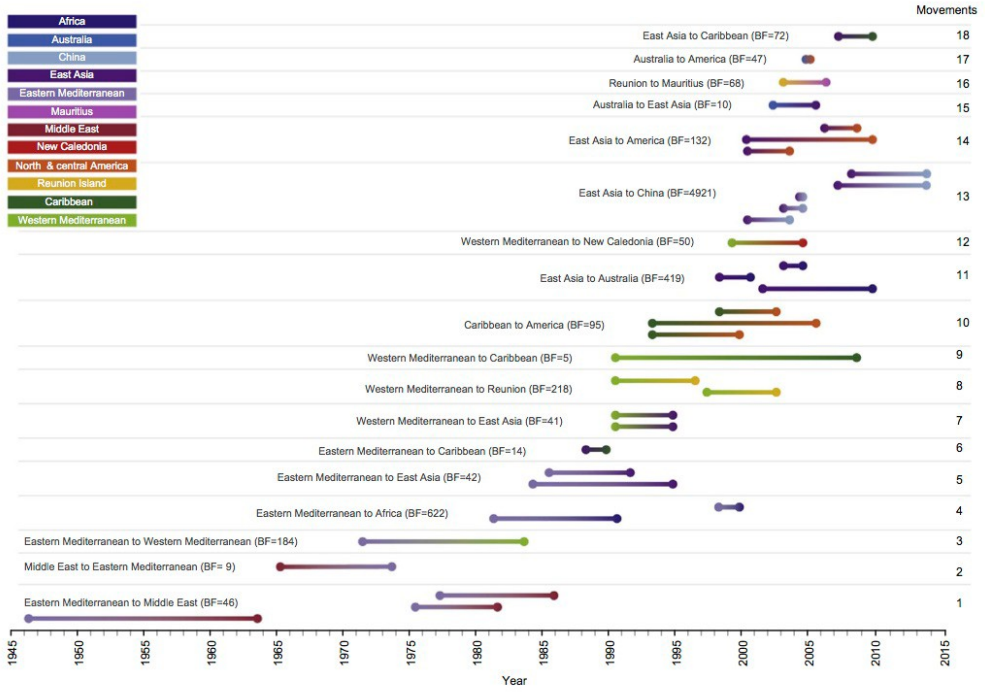


Fig. 2. Graph showing the timing of historic TYLCV movement events between the twelve analysed regions. Colours gradients across lines indicate potential movements of ancestral TYLCV variants from the region represented by the colour at the left of the line to the region represented by the colour on the right of the line.

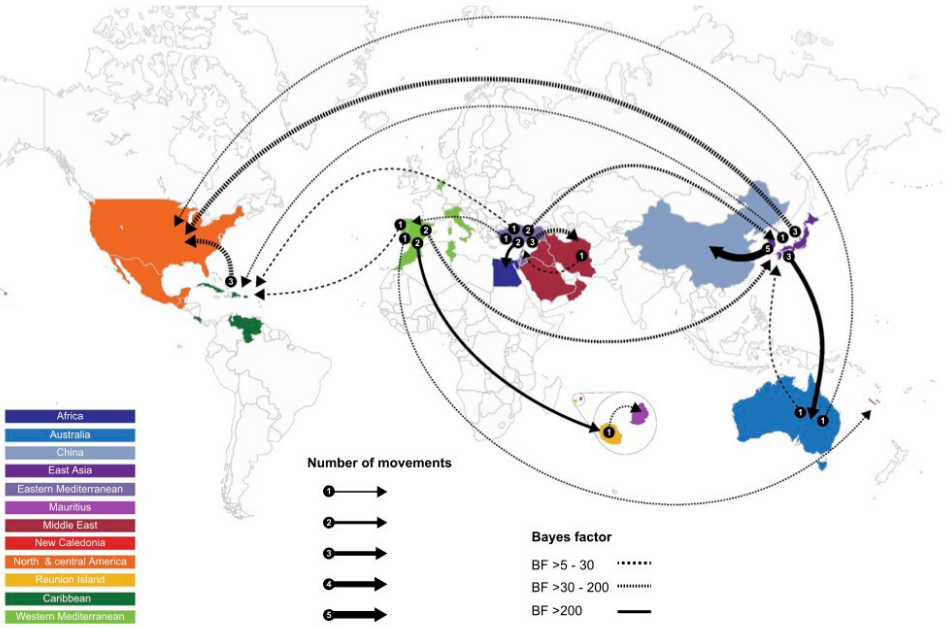


Fig. 3. A summary of the 18 statistically supported epidemiological links between the twelve geographic regions considered here. Dashed and solid lines represent the degree of Bayes factor support and the thickness of the lines represent the numbers of independent movements that were inferred to have occurred between locations.

considered (summarised in Figs. 2 and 3). The earliest detectable TYLCV dispersal event involved an unknown TYLCV strain moving from the Eastern Mediterranean to the Middle East between 1946 and 1964 (Figs. 1–3). This was followed by at least three more movements in the opposite direction between 1965 and 1974, 1971 and 1986 and 1988–2007. The estimated date of the existence of the MCRA of these isolates coincided with the first suspected cases of TYLCV in the Jordan Valley during the late 1950s to early 1960s (Cohen and Nitzany, 1960, 1966). This outbreak in the Eastern Mediterranean (Almusa, 1982; Makkouk, 1978; Makkouk et al., 1979) is thought to have represented the first opportunity for TYLCV to spread to the rest of the world (Lefeuvre et al., 2010).

Further movements from the Eastern Mediterranean included the dispersal of a TYLCV-Mld variant to the Western Mediterranean between 1971 and 1984 and two subsequent movements of the TYLCV-IL strain in the same direction between 1981 and 2001 and 1990 and 1995 (Figs. 1–3). While the first of these inferred movements is very close to the time of the first reports of TYLCV in the Western-Mediterranean around 1983, the timings of the other movements correspond to outbreaks in Italy and Spain between 1988 and 1992 (Credi et al., 1989; Gallitelli et al., 1991; Kheyr-Pour et al., 1991; Louro et al., 1996; Navas-Castillo et al., 1997; Noris et al., 1994).

Other inferred dispersal events of TYLCV-IL also correspond with actual observations, for example the arrival of the virus in Egypt during the early 1980s (Nakhla et al., 1993), in Japan in 1996, in China in 2006 and in South Korea in 2008 (Kenyon et al., 2014; Kim et al., 2011; Lee et al., 2010; Ueda et al., 2012).

Our analysis indicates that the Caribbean has possibly been a major portal for the introduction of TYLCV into the Western Hemisphere. TYLCV likely first entered the Caribbean from the Eastern Mediterranean between 1988 and 1991, a timeframe coinciding with serious TYLCV outbreaks in the Dominican Republic in the early 1990s (Czosnek and Laterrot, 1997; Polston et al., 1999; Salati et al., 2002). It was previously suggested that this introduction was via the Dominican Republic or Cuba, possibly in a shipment of tomato seedlings from Israel (Polston et al., 1999). In 1993, TYLCD decimated tomato production in the Dominican Republic (Polston et al., 1999) and from this focal point, the virus moved into Jamaica and Cuba (Mcglashan et al., 1994; Polston et al., 1999; Roye et al., 1999; Zubiaur et al., 2004).

From there, TYLCV quickly spread to the USA, where it was identified in Virginia, Tennessee and South Carolina in the mid-1990s (Polston et al., 1995), then in Florida (Polston and Anderson, 1997), Georgia (Momol et al., 1999), Mississippi (Ingram and Henn, 2001) and in Central America (Mexico) (Ascencio-Ibáñez et al., 1999; Banuelos-Hernandez et al., 2012; Brown and Idris, 2006). Other independent introductions of TYLCV into the Caribbean likely occurred when TYLCV-IL moved from East-Asia between 2006 and 2011 (Figs. 1–3), and TYLCV-Mld arrived there from the Western Mediterranean between 1990 and 2009 (Bird et al., 2001; Mcglashan et al., 1994; Roye et al., 1999; Zambrano et al., 2007).

Besides the introduction of TYLCV-IL into the Caribbean, other long distance movements of TYLCV from the Western Mediterranean were also detected in our analyses. One of these was to East Asia (between 1990 and 1995), preceding the first reports of TYLCV in Japan in 1996 (Kato et al., 1998; Kenyon et al., 2014), one to New Caledonia (between 1998 and 2006) (Péréfarres et al., 2012) and two to Reunion Island: a TYLCV-Mld movement between 1987 and 1997, coinciding with the first report of TYLCV there in 1997; (Peterschmitt et al., 1999) and a TYLCV-IL movement between 1998 and 2005 (Delatte et al., 2007). Although we were unable to obtain TYLCV samples from France (TYLCV is presently thought to have been eradicated there), it is plausible

that France was the Western Mediterranean country from which TYLCV entered Reunion Island and New Caledonia, since Reunion is a French overseas department and New Caledonia, a French territory (with both islands maintaining highly connected trade and transport links with metropolitan France). The importance of such links is highlighted by the fact that we also detected the movement of TYLCV-Mld from Reunion to the neighbouring island of Mauritius between 2003 and 2007 (Lobin et al., 2010).

TYLCV-IL was first reported in Australia in the peri-urban areas of Brisbane in 2006 and subsequently in the production areas of Bundaberg and Gatton (Van Brunschot et al., 2010). Consistent with this infection history, our analyses indicated that there were at least three separate introductions of TYLCV-IL from East Asia into Australia, two of which, occurred before this date: one between 1998 and 2001 from Japan to Brisbane, a second between 2003 and 2004 from Japan to Bundaberg, and a third between 2001 and 2009, which forms a monophyletic clade on the tree that is distinct from the two other monophyletic clades comprising the remaining Australian sequences from Brisbane and Bundaberg (Fig. 1).

We also inferred one movement of a Brisbane group TYLCV-IL virus from Australia to East Asia between 2000 and 2011 (most likely to Japan) and another to the USA (to either Hawaii or California) (Melzer et al., 2010; Rojas and Kon, 2007) between 2004 and 2005. These movements coincided with outbreaks of TYLCV in Arizona (Idris et al., 2007) and Texas (Isakeit et al., 2007), which were previously thought to have an independent, possibly East-Asian, origin to the TYLCV-IL variants found in the Caribbean (Duffy and Holmes, 2008; Lefeuvre et al., 2010).

We also detected two movements of TYLCV-IL from East Asia to the North American region (but also possibly directly to Hawaii), between 2006 and 2009 and between 2000 and 2010, which is consistent with the first reports of TYLCV-IL in Hawaii in 2009 (Melzer et al., 2010). An additional movement from East Asia, which further underlines the importance of this region as a major hub of global TYLCV dissemination, was to the Caribbean between 2006 and 2011.

Probably due to the proximity of Japan, South Korea and China, we inferred seven independent short range movements of TYLCV-IL from Japan/Korea into China between 2000 and 2012. These movements coincide with the first reported cases of TYLCV in China (around Shanghai) in 2006 (Yongping et al., 2008). Despite the rapid spread of TYLCV within China (Kenyon et al., 2014), there were no statistically supported movements of TYLCV from China to any of the other eleven regions studied. However, because the sequences from China make up ~36% of the total sample, and of these, 80% were collected between 2010 and 2014, it is possible that the absence of statistically supported movements out of China is in fact an artefact of the sampling scheme. Specifically, the relatively large proportion of samples originating from China increases the probability of detecting rare and infrequent relatively recent movements into this region.

4. Concluding remarks

While the phylogeographic analyses that we have performed broadly confirm the findings of similar analyses with smaller datasets, they substantially clarify the movement dynamics of TYLCV in the Western Hemisphere and Far East regions. We conclude that introductions of TYLCV into Australia and China have likely been from the East-Asia region, whereas the introduction to New Caledonia was likely from the Western Mediterranean. The fact that this association between New Caledonia and the Western Mediterranean is mirrored by that between the Western Mediterranean and Reunion Island, another region that is politically tied to France, suggests that metropolitan France might be the actual origin of TYLCV on these islands.

Regarding other inter-continental scale movements, we conclude that there have been at least nine independent introductions of TYLCV to the American/Caribbean region from the East-Asian, Australian, and Western Mediterranean regions, and at least five introductions of TYLCV to the East-Asian region from the Eastern Mediterranean and Western Mediterranean regions. Although there is no evidence of any movements to East Asia from the Mediterranean basin since 1995, at least five movements of TYLCV into the American/Caribbean region have occurred since the year 2000, suggesting that the flow of TYLCV variants into the Americas from elsewhere in the world is likely ongoing. The recent discovery that at least some TYLCV-IL variants are likely seed-transmissible (Kil et al., 2016) should be seriously investigated as a potential contributor to these movements. In this regard it would be of great interest to compare the seed transmission potential of all the main TYLCV lineages so as to determine whether a potentially causal association exists between the seed-transmissibility of particular lineages and their geographical ranges.

Crucially, the large numbers of inferred TYLCV movements over the past two decades, with multiple independent movements into and out of many of the regions analysed, strongly suggests that not enough is presently being done to control the ongoing spread of this major crop pathogen. Despite its already near cosmopolitan distribution, it is important that new containment strategies are implemented that account for the seed-transmissibility of TYLCV. Besides containing the current geographical range of TYLCV, such strategies will be crucial for impeding movements across this range of arising pathogenic and/or resistance-breaking variants of the virus.

Acknowledgements

We thank Michel Peterschmitt for sharing his TYLCV sequences from tomato samples collected in Morocco. BM, DPM, AV and GWH are supported by the National Research Foundation of South Africa (Grant No. TTK1207122745). SvB was supported by the Cooperative Research Centre for National Plant Biosecurity (CRC60070) and currently by the Cotton Research and Development Corporation, Australia (Grant No UQ1305). JML and PL are supported by the Région Réunion and the European Union (FEDER; Action # 1.03).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2016.08.033. These data include Supplementary Figure 1, Supplementary Table 1 and Google earth animated kml showing the dispersal dynamics in real time (years).

References

Abhary, M., Patil, B., Fauquet, C.M., 2007. Molecular biodiversity, taxonomy, and nomenclature of tomato yellow leaf curl-like viruses. In: Czosnek, H. (Ed.), Tomato Yellow Leaf Curl Virus Disease, pp. 85–118.

Almusa, A., 1982. Incidence, economic importance, and control of tomato yellow leaf curl in Jordan. Plant Dis. 66, 561–563.

Ascencio-Ibáñez, J.T., Diaz-Plaza, R., Méndez-Lozano, J., Monsalve-Fonnegra, Z.I., Argüello-Astorga, G.R., Rivera-Bustamante, R., 1999. First report of tomato yellow leaf curl virus in Yucatan, Mexico. Plant Dis. 83, 1178.

Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O., Huelsenbeck, J.P., Ronquist, F., Swofford, D.L., Cummings, M.P., Rambaut, A., Suchard, M.A., 2012. BEAGLE: an application programming interface and high-performance

computing library for statistical phylogenetics. Syst. Biol. 61, 170–173.
Baele, G., Li, W.L., Drummond, A.J., Suchard, M.A., Lemey, P., 2013. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. Mol. Biol. Evol. 30, 239–243.
Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A., Alekseyenko, A.V., 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29, 2157–2167.
Banuelos-Hernandez, B., Mauricio-Castillo, J.A., Cardenas-Conejo, Y., Guevara-Gonzalez, R.G., Arguello-Astorga, G.R., 2012. A new strain of tomato severe leaf curl virus and a unique variant of tomato yellow leaf curl virus from Mexico. Arch. Virol. 157, 1835–1841.
Bielejec, F., Rambaut, A., Suchard, M.A., Lemey, P., 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27, 2910–2912.
Bird, J., Idris, A.M., Rogan, D., Brown, J.K., 2001. Introduction of the Exotic Tomato yellow leaf curl virus-Israel in Tomato to Puerto Rico. Plant Dis. 85, 1028.
Brown, J.K., Idris, A.M., 2006. Introduction of the exotic monopartite tomato yellow leaf curl virus into west coast Mexico. Plant Dis. 90, 1360.
Brown, J.K., Zerbini, F.M., Navas-Castillo, J., Moriones, E., Ramos-Sobrinho, R., Silva, J. C., Fiallo-Olive, E., Briddon, R.W., Hernandez-Zepeda, C., Idris, A., Malathi, V.G., Martin, D.P., Rivera-Bustamante, R., Ueda, S., Varsani, A., 2015. Revision of Begomovirus taxonomy based on pairwise sequence comparisons. Arch. Virol. 160, 1593–1619.
Cohen, S., Nitzany, F.E., 1966. Transmission and host range of tomato yellow leaf curl virus. Phytopathology 56, 1127–1131.
Cohen, S., Nitzany, F.E., 1960. Curly Top Virus of Tomatoes: Its Identification and Mode of Transmission, Israeli Plant Protection and Inspection Services.
Credi, R., Betti, L., Canova, A., 1989. Association of a geminivirus with a severe disease of tomato in Sicily. Phytopathol. Mediterr. 25, 223–226.
Czosnek, H., Laterrot, H., 1997. A worldwide survey of tomato yellow leaf curl viruses. Arch. Virol. 142, 1391–1406.
Czosnek, H., Ghanim, M., Ghanim, M., 2002. The circulative pathway of begomoviruses in the whitefly vector Bemisia tabaci - insights from studies with Tomato yellow leaf curl virus. Ann. Appl. Biol. 140, 215–231.
Delatte, H., Holota, H., Naze, F., Peterschmitt, M., Reynaud, B., Lett, J.M., 2005. The presence of both recombinant and nonrecombinant strains of Tomato yellow leaf curl virus on tomato in Réunion Island. Plant Pathol. 54, 262.
Delatte, H., Holota, H., Moury, B., Reynaud, B., Lett, J.M., Peterschmitt, M., 2007. Evidence for a founder effect after introduction of Tomato yellow leaf curl virus-mild in an insular environment. J. Mol. Evol. 65, 112–118.
Diaz-Pendon, J.A., Canizares Mc Fau - Moriones, E., Moriones E Fau - Bejarano, E.R., Bejarano Er Fau - Czosnek, H., Czosnek H Fau - Navas-Castillo, J., Navas-Castillo, J., 2010. Tomato yellow leaf curl viruses: menage a trois between the virus complex, the plant and the whitefly vector. Mol. Plant Pathol. 11, 441–450.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC. Evolut. Biol. 7, 214.
Duffy, S., Holmes, E.C., 2007. Multiple introductions of the Old World begomovirus Tomato yellow leaf curl virus into the New World. Appl. Environ. Microbiol. 73, 7114–7117.
Duffy, S., Holmes, E.C., 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. J. Virol. 82, 957–965.
Gallitelli, D., Luisoni, E., Martinelli, G.P., Caciagli, P., Milne, R.G., Accotto, G.P., Antignus, Y., 1991. Tomato yellow leaf curl disease in Sardinia. Inf. Fitopatol. 41, 42–46.
Gill, M.S., Lemey, P., Faria, N.R., Rambaut, A., Shapiro, B., Suchard, M.A., 2013. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol. Biol. Evol. 30, 713–724.
Horowitz, R., Denholm, I., Morin, S., 2007. Resistance to Insecticides in the TYLCV vector, Bemisia tabaci. In: Czosnek, H. (Ed.), Tomato Yellow Leaf Curl Virus Disease: Management, Molecular Biology, Breeding for Resistance. Springer, Netherlands, Dordrecht, pp. 305–325.
Idris, A.M., Guerrero, J.C., Brown, J.K., 2007. Two distinct isolates of tomato yellow leaf curl virus threaten tomato production in Arizona and Sonora, Mexico. Plant Dis. 91, 910.
Ingram, D.M., Henn, A., 2001. First report of tomato yellow leaf curl virus in Mississippi. Plant Dis., 85.
Isakeit, T., Idris, A.M., Sunter, G., Black, M.C., Brown, J.K., 2007. Tomato yellow leaf curl virus in tomato in texas, originating from transplant facilities. Plant Dis. 91, 466.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.
Kato, K., Onuki, M., Fuji, S., Hanada, K., 1998. The first occurrence of Tomato yellow leaf curl virus in tomato (Lycopersicon esculentum Mill.) in Japan. Annu. Phyopathol. Soc. Jpn. 64, 552–559.
Katoh, K., Standley, D.M., 2014. MAFFT: iterative refinement and additional methods. Methods Mol. Biol. 1079, 131–146.
Kenyon, L., Tsai, W.S., Shih, S.L., Lee, L.M., 2014. Emergence and diversity of begomoviruses infecting solanaceous crops in East and Southeast Asia. Virus Res 186, 104–113.
Kheyr-Pour, A., Bendahmane, M., Matzeit, V., Accotto, G.P., Crespi, S., Gronenborn, B., 1991. Tomato yellow leaf curl virus from Sardinia is a whitefly-transmitted monopartite geminivirus. Nucleic Acids Res. 19, 6763–6769.
Kil, E.J., Kim, S., Lee, Y.J., Byun, H.S., Park, J., Seo, H., Kim, C.S., Shim, J.K., Lee, J.H., Kim, J.K., Lee, K.Y., Choi, H.S., Lee, S., 2016. Tomato yellow leaf curl virus (TYLCV-IL): a seed-transmissible geminivirus in tomatoes. Sci. Rep. 6, 19013.
Kim, S.H., Oh, S., Oh, T.K., Park, J.S., Kim, S.C., Kim, S.H., Kim, Y.S., Hong, J.K., Sim, S.Y., Park, K.S., Lee, H.G., Kim, K.J., Choi, C.W., 2011. Genetic diversity of tomato-infecting Tomato yellow leaf curl virus (TYLCV) isolates in Korea. Virus Genes 42, 117–127.
Lee, H., Song, W., Kwak, H.R., Kim, J.D., Park, J., Auh, C.K., Kim, D.H., Lee, K.Y., Lee, S., Choi, H.S., 2010. Phylogenetic analysis and inflow route of Tomato yellow leaf curl virus (TYLCV) and Bemisia tabaci in Korea. Mol. Cells 30, 467–476.
Lefeuvre, P., Martin, D.P., Harkins, G., Lemey, P., Gray, A.J., Meredith, S., Lakay, F., Monjane, A., Lett, J.M., Varsani, A., Heydarnejad, J., 2010. The spread of tomato yellow leaf curl virus from the Middle East to the world. PLoS Pathog. 6, e1001164.
Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A., 2009. Bayesian phylogeography finds its roots. PLoS Comput. Biol. 5, e1000520.
Lobin, K., Druffel, K.L., Pappu, H.R., Benimadhu, S.P., 2010. First report of tomato yellow leaf curl virus in tomato in Mauritius. Plant Dis. 94, 1261.
Louro, D., Noris, E., Veratti, F., Accotto, G.P., 1996. First report of tomato yellow leaf curl virus in Portugal. Plant Dis. 80, 1079.
Makkouk, K.M., 1978. Study on tomato viruses in Jordan valley with special emphasis on tomato yellow leaf curl. Plant Dis. Rep. 62, 259–262.
Makkouk, K.M., Shehab, S., Majdalani, S.E., 1979. Tomato yellow leaf curl – incidence, yield losses and transmission in Lebanon. Phytopathol. Z. 96, 263–267.
Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1, vev003.
Mcglashan, D., Polston, J.E., Bois, D., 1994. Tomato yellow leaf curl geminivirus in Jamaica. Plant Dis. 78, 1219.
Melzer, M.J., Ogata, D.Y., Fukuda, S.K., Shimabuku, R., Borth, W.B., Sether, D.M., Hu, J. S., 2010. First Report of Tomato yellow leaf curl virus in Hawaii. Plant Dis. 94, 641.
Minin, V.N., Bloomquist, E.W., Suchard, M.A., 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol. 25, 1459–1471.
Momol, M.T., Simone, G.W., Dankers, W., Sprenkel, R.K., Olson, S.M., Momol, E.A., Polston, J.E., Hiebert, E., 1999. First report of tomato yellow leaf curl virus in tomato in South Georgia. Plant Dis. 83, 487.
Monci, F., Sanchez-Campos, S., Navas-Castillo, J., Moriones, E., 2002. A natural recombinant between the geminiviruses Tomato yellow leaf curl Sardinia virus and Tomato yellow leaf curl virus exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. Virology 303, 317–326.
Moriones, E., Navas-Castillo, J., 2000. Tomato yellow leaf curl virus, an emerging virus complex causing epidemics worldwide. Virus Res 71, 123–134.
Nakhla, M.K., Mazyad, H.M., Maxwell, D.P., 1993. Molecular characterization of four tomato yellow leaf curl virus isolates from Egypt and development of detection methods. Phytopathol. Mediterr. 32, 163–173.
Navas-Castillo, J., Sanchez-Campos, S., Diaz, J.A., Saez-Alonso, E., Moriones, E., 1997. First report of Tomato yellow leaf curl virus is in Spain: Coexistence of two different geminiviruses in the same epidemic outbreak. Plant Dis. 81, 1461.
Navas-Castillo, J., Sanchez-Campos, S., Noris, E., Louro, D., Accotto, G.P., Moriones, E., 2000. Natural recombination between Tomato yellow leaf curl virus-Is and Tomato leaf curl virus. J. Gen. Virol. 81, 2797–2801.
Navot, N., Pichersky, E., Zeidan, M., Zamir, D., Czosnek, H., 1991. Tomato yellow leaf curl virus: a whitefly-transmitted geminivirus with a single genomic component. Virology 185, 151–161.
Noris, E., Hidalgo, E., Accotto, G.P., Moriones, E., 1994. High Similarity among the Tomato yellow leaf curl virus isolates from the west mediterranean basin – the nucleotide-sequence of an infectious clone from Spain. Arch. Virol. 135, 165–170.
Pakkianathan, B.C., Kontsedalov, S., Lebedev, G., Mahadav, A., Zeidan, M., Czosnek, H., Ghanim, M., 2015. Replication of Tomato yellow leaf curl virus in its whitefly vector, Bemisia tabaci. J. Virol. 89, 9791–9803.
Péréfarres, F., De Bruyn, A., Kraberger, S., Hoareau, M., Barjon, F., Lefeuvre, P., Pellegrin, F., Caplong, P., Varsani, A., Lett, J.M., 2012. Occurrence of the Israel strain of Tomato yellow leaf curl virus in New Caledonia and Loyalty Islands. New Dis. Rep. 25, 6.
Peterschmitt, M., Granier, M., Mekdoud, R., Dalmon, A., Gambin, O., Vayssieres, J.F., Reynaud, B., 1999. First report of tomato yellow leaf curl virus in Réunion Island. Plant Dis. 83, 303.
Picó, B., Díez, M.J., F., N, 1996. Viral diseases causing the greatest economic losses to the Tomato crop. II. The Tomato yellow leaf curl virus—a review. Sci. Hortic. 67, 151–196.
Polston, J.E., Anderson, P.K., 1997. The emergence of whitefly-transmitted geminiviruses in tomato in the western hemisphere. Plant Dis. 81, 1358–1369.
Polston, J.E., Keinath, A.P., Chellemi, D.O., 1995. Occurrence of Tomato Mottle Geminivirus in South-Carolina, Tennessee, and Virginia. Plant Dis. 79, 539.
Polston, J.E., McGovern, R.J., Brown, L.G., 1999. Introduction of tomato yellow leaf curl virus in Florida and implications for the spread of this and other geminiviruses of tomato. Plant Dis. 83, 984–988.
Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.
R Core Team, 2013. R: A Language and Environment for Statistical Computing.
Rambaut, A., Lam, T.T., Max Carvalho, L., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2.
Rojas, M.R., Kon, T., 2007. First report of tomato yellow leaf curl virus associated with tomato yellow leaf curl disease in California. Plant Dis. 91, 1056.
Rokach, L., Maimon, O., 2005. Clustering methods. In: Maimon, O., Rokach, L. (Eds.), Data Mining and Knowledge Discovery Handbook. Springer, US, Boston, MA, pp. 321–352.
Roye, M.E., Wernecke, M.E., McLaughlin, W.A., Nakhla, M.K., Maxwell, D.P., 1999.

Tomato dwarf leaf curl virus, a new bipartite geminivirus associated with tomatoes and peppers in Jamaica and mixed infection with tomato yellow leaf curl virus. Plant Pathol. 48, 370–378.
Salati, R., Nahkla, M.K., Rojas, M.R., Guzman, P., Jaquez, J., Maxwell, D.P., Gilbertson, R.L., 2002. Tomato yellow leaf curl virus in the Dominican Republic: characterization of an infectious clone, virus monitoring in whiteflies, and identification of reservoir hosts. Phytopathology 92, 487–496.
Seal, S.E., vandenBosch, F., Jeger, M.J., 2006. Factors influencing begomovirus evolution and their increasing global significance: Implications for sustainable control. Crit. Rev. Plant Sci. 25, 23–46.
Stonor, J., Hart, P., Gunther, M., DeBarro, P., Rezaian, M.A., 2003. Tomato leaf curl geminivirus in Australia: occurrence, detection, sequence diversity and host range. Plant Pathol. 52, 379–388.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
Ueda, S., Onuki, M., Yamashita, M., Yamato, Y., 2012. Pathogenicity and insect transmission of a begomovirus complex between tomato yellow leaf curl virus and Ageratum yellow vein betasatellite. Virus Genes 44, 338–344.
Van Brunschot, S.L., Persley, D.M., Geering, A.D.W., Campbell, P.R., Thomas, J.E., 2010. Tomato yellow leaf curl virus in Australia: distribution, detection and discovery of naturally occurring defective DNA molecules. Australas. Plant Pathol. 39, 412–423.
Yang, X.L., Zhou, M.N., Qian, Y.J., Xie, Y., Zhou, X.P., 2014. Molecular variability and evolution of a natural population of tomato yellow leaf curl virus in Shanghai. China J. Zhejiang Univ. Sci. B 15, 133–142.
Yongping, Z., Weimin, Z., Huimei, C., Yang, Q., Kun, S., Yanhui, W., Longying, Z., Li, Y., Zhang, H., 2008. Molecular identification and the complete nucleotide sequence of TYLCV isolate from Shanghai of China. Virus Genes 36, 547–551.
Zambrano, K., Carballo, O., Geraud, F., Chirinos, D., Fernandez, C., Marys, E., 2007. First report of Tomato yellow leaf curl virus in Venezuela. Plant Dis. 91, 768.
Zubiaur, Y.M., Fonseca, D., Quinones, M., Palenzuela, I., 2004. Presence of Tomato yellow leaf curl virus infecting squash (Curcubita pepo) in Cuba. Plant Dis. 88, 572.

Appendix 8: Conference presentation

Gambley C, Campbell P, Grundy P, Sharman M and **van Brunschot S**. (2016) Cotton leaf curl disease: how prepared are we? Proceedings of the International Fuscom Meeting, 12th-13th April, Toowoomba, Queensland, Australia.

Link to presentation:
https://www.dropbox.com/s/w8jmk337a08slk9/2016_Fuscom_Gambley.pptx?dl=0

Appendix 9: Conference presentation

van Brunschot S, Walter G, Hereward J, Gambley C, De Barro P, Seal S, and Colvin J (2016). Whiteflies and begomoviruses: an Australian perspective. Plant Health Australia Workshop on Whitefly and Whitefly-Transmitted Viruses, 1st September, Brisbane, Queensland, Australia.

Link to presentation:
https://www.dropbox.com/s/anzkabeu3hu1qla/2016_PHAWorkshop_UQ1305.pptx?dl=0

Appendix 10: Spotlight article

Maas S, van Brunschot SL (2015) Can whiteflies spread exotic disease. Spotlight Magazine – Autumn Edition. Australian Cotton Research and Development Corporation

MATRIX HELPS SIMPLIFY SLW SPRAY DECISIONS

The CottASSIST Silverleaf Whitefly (SLW) Matrix web tool allows users to enter regular sampling information to track the development of SLW populations over time. The tool then compares these populations with the control thresholds, which are based on the pest population size, day degrees and crop stage.

CottonInfo's Sandra Williams says the tool is easy to use and navigate.

"Once you have set-up farms and crops, the only required information is SLW sample data," she said.

"The recommendation page presents the industry's threshold matrix with your whitefly population data overlayed. In addition to this is a full explanation of the threshold that applies to your current SLW population and crop stage."

SLW management can be complicated as they have a very large host range, quick reproduction rate and can rapidly develop resistance to insecticides. Therefore to prevent risk of honey dew and resistance developing to the small number of registered products, it is fundamental to follow industry thresholds developed from research.

DAF QLD SLW researcher Dr Richard Sequeira says the CottASSIST SLW Matrix Tool reduces a multi-dimensional whitefly spray decision and takes the process down to a simple and practical exercise that any crop manager will be able to use with confidence and ease.

The industry's SLW threshold matrix is contained in the *Cotton Pest Management Guide*, however the CottASSIST SLW Matrix tool also automatically calculates the day degrees needed to use the matrix.

The tool can be accessed through www.cottassist.com.au

More information

Sandra Williams
02 6799 1500
sandra.williams@csiro.au

email us

RESEARCH IS ONGOING TO DETERMINE, IN THE EVENT OF AN INCURSION, IF AUSTRALIAN WHITEFLIES TRANSMIT COTTON LEAF CURL DISEASE.

CAN WHITEFLIES SPREAD EXOTIC DISEASE?

One of the subtleties of whiteflies (*Bemisia tabaci*) is that they are not one species, but a complex of at least 28 different species that look identical. On closer inspection however, they differ in their biology, genetics, and even their relative abilities to transmit plant viruses.

CRDC is supporting ongoing, cutting-edge whitefly research by Dr Sharon van Brunschot, a Research Fellow at The University of Queensland (UQ), to assess this pest's potential to vector the group of viruses that cause cotton leaf curl disease, in the event of an incursion. Sharon is part of an established group of cotton researchers at UQ led by Professor Gimme Walter, who are working together to better understand the ecology of various cotton pests.

"There are underlying complexities in the virus-vector interaction that require testing and research in order to understand and estimate the risk cotton leaf curl disease poses to the Australian cotton industry," Sharon said.

"While the viruses that cause cotton leaf curl have not been found in Australia, the disease continues to be a major constraint to cotton production across Asia and Africa. We have at least three different whitefly species in Australia, but we do not know if they would be competent vectors of these viruses.

"From a biosecurity perspective, this research will allow us to determine the actual risk this disease poses to Australian cotton."

Sharon will travel to the UK in July to undertake transmission experiments using virus isolates sourced from the Indian subcontinent, Australian whitefly species and cotton cultivars to determine if the whitefly species we have in Australia are efficient vectors of these viruses

"It would be a very high-risk exercise to do this research in Australia, so it is fortunate that we have strong collaborations with world experts at The University of Greenwich, via our collaborator at CSIRO Dr Paul De Barro," Sharon said.

Improving diagnostic capability

Sharon is also in the process of developing diagnostic tools for the quick and reliable detection of the virus in infected plants as well as in whiteflies. These diagnostic tests will improve the industry's capacity to prevent and respond to any incursion of exotic cotton leaf curl disease that may eventuate.

Furthermore, on a research expedition to China last year, Sharon identified that a likely incursion pathway for cotton leaf curl disease to Australia is via the importation of ornamental hibiscus. Utilising this information, CRDC researcher Dr Cherie Gambley (DAF QLD) then successfully negotiated the tightening of import conditions for hibiscus entering Australia. This change will strengthen the protection of the Australian cotton industry from this devastating disease.

Sharon Van Brunschot will travel to the UK to test the ability of whiteflies to vector the cotton leaf curl disease as part of a CRDC-funded project.

New management potential

Further subtleties exist in the interaction of these viruses and their whitefly vectors, including the potential involvement of endosymbiotic bacteria. Whitefly harbor an array of microbes in specific cells and parts of the insect's body, which may influence important processes including reproductive capacity, lifespan, susceptibility to insecticides and plant virus transmission.

"I have discovered that there are significant differences in endosymbionts harboured by the native and introduced whitefly species that are present in Australia," Sharon said.

"I am currently examining the influence these microbes have on processes including reproductive capacity and virus transmission.

"Knowing more about these relationships may inform new methods for whitefly management in the future."

More information

Sharon van Brunschot
s.vanbrunschot@uq.edu.au

email us

Appendix 11: Conference seminar

van Brunschot SL, Walter G, Hereward J, Gambley C, Campbell P, De Barro P, Howie L (2014) Investigating the interactions of viruses, whiteflies and endosymbionts. Proceedings of the International Fuscom Meeting, 3rd-5th November, Toowoomba, Queensland, Australia.

Link to presentation:
https://www.dropbox.com/s/mhr9noed2zmbxus/2014_Fuscom_UQ1305.pptx?dl=0

Appendix 12: Conference seminar

van Brunschot SL, Hereward J, Gambley C, De Barro P, Walter G (2014) Investigating the interactions of viruses, vectors and endosymbionts. Proceedings of the 11th Australasian Plant Virology Workshop, 13th-15th August, Brisbane, Queensland, Australia.

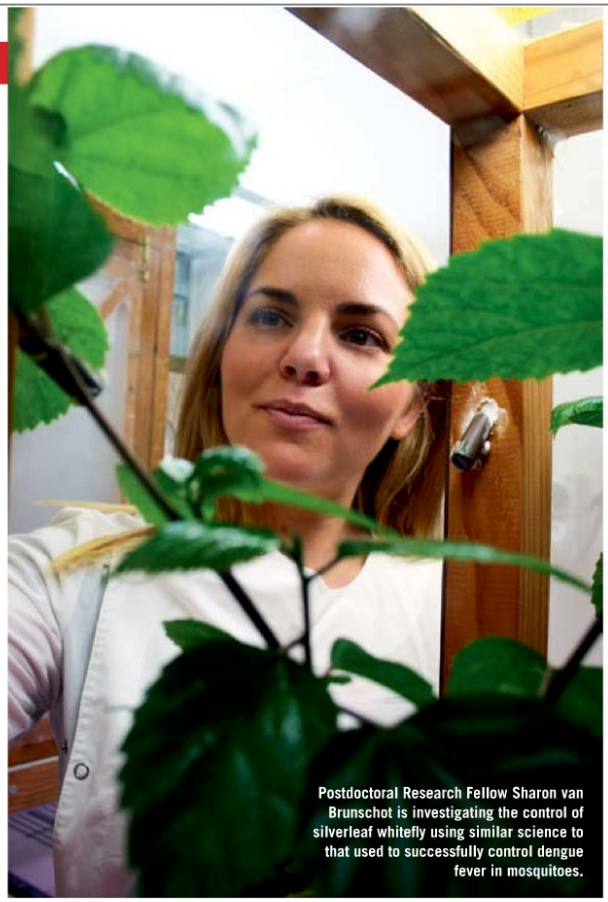
Link to presentation:
https://www.dropbox.com/s/ci8lsivuqaw57oq/2014_APVW_UQ1305.pptx?dl=0

Appendix 13: Spotlight article

Maas S, van Brunschot SL (2014) Cutting edge research underway for whitefly control. Spotlight Magazine – Winter Edition. Australian Cotton Research and Development Corporation

INVESTIGATING THE INTRICATE RELATIONSHIPS BETWEEN SILVERLEAF WHITEFLY AND ITS 'SYMBIOTIC' BACTERIA COULD OFFER NEW WAYS TO MANAGE THIS DAMAGING PEST.

CUTTING EDGE RESEARCH UNDER WAY FOR WHITEFLY CONTROL



Postdoctoral Research Fellow Sharon van Brunschot is investigating the control of silverleaf whitefly using similar science to that used to successfully control dengue fever in mosquitoes.

- Research hopes to close a knowledge gap that exists in our understanding of the complex interactions that occur between whiteflies and the bacterial endosymbionts that reside within the whitefly's body and cells. These endosymbionts can influence diverse processes in their insect hosts including reproductive capacity, thermal tolerance, lifespan, susceptibility to insecticides, and plant virus transmission.
- Cotton leaf curl disease (CLCuD) can decimate production and is spread by the whitefly (Bemisia tabaci). CLCuD exists across Africa and Southern Asia.
- Currently, we have no knowledge of the ability of whitefly in Australian cotton to transmit viruses of the CLCuD complex, however research is now underway.

Manipulation of symbiotic bacteria (endosymbionts) is cutting-edge science that has previously been used by researchers at Monash University to formulate ways to control dengue fever in mosquitoes.

Silverleaf whitefly (SLW) (*Bemisia tabaci*) is a major pest of Australian cotton systems because it contaminates cotton lint with honeydew. The adult insects (which are actually tiny bugs with piercing mouthparts and not flies at all) are highly mobile, can develop insecticide resistance quickly and populations can expand rapidly when natural enemies are reduced by insecticides. SLW also vector the exotic plant virus complex that causes cotton leaf curl disease, a significant biosecurity threat to Australian cotton.

This new CRDC-funded research is focused on examining the symbiotic relationship that SLW shares with bacteria (endosymbionts) that are harbored in the insects' body and cells. Clarifying the influence that these endosymbionts have on the biology of SLW may reveal opportunities for new pest management approaches.

The project is being undertaken by Sharon van Brunschot, a Postdoctoral Research Fellow at The University of Queensland (UQ) in association with Dr James Hereward (UQ), Dr Cherie Gambley (DAFF QLD) and Dr Paul De Barro (CSIRO), under the supervision of Associate Professor Gimme Walter (UQ).

"Bacterial endosymbionts can influence a diverse range of processes in their insect hosts, including reproductive capacity, thermal tolerance, lifespan, susceptibility to insecticides, and plant virus transmission," Sharon said.

"This study of endosymbiont-whitefly interactions aims to determine the influence of these microbes on key processes such as fitness, insecticide resistance, and plant virus transmission, to explore possibilities for manipulating these processes to control SLW populations in the future."

Cotton leaf curl

Another component of Sharon's research is aimed at improving the industry's capacity to prevent and respond to any incursion that may eventuate of the exotic Cotton leaf curl disease (CLCuD).

CLCuD is a major biosecurity threat to Australia as it causes serious economic losses to cotton production where it occurs overseas, particularly on the Indian Subcontinent.

Sharon will be developing new diagnostic tools to enable the quick and reliable detection of the virus in infected plants and also in SLW vectors. She will also be examining the capacity of SLW populations in Australian cotton to transmit viruses that cause CLCuD.

"We do not know if the SLW in Australia are competent vectors of the viruses in the CLCuD complex," Sharon said.

"We need to test these specific virus-vector interactions to understand more accurately the risks of introduction and establishment of this disease.

"This important aspect of my research is timely as CLCuD has recently been identified in China, which represents a movement of this virus out of its previously limited home range of South Asia/Africa, a move that brings the disease even closer to Australia."

Sharon will perform virus transmission experiments with virus sourced from the Indian Subcontinent, using Australian whiteflies and cotton cultivars. These experiments will be performed in the secure quarantine facilities of The University of Greenwich (London), in collaboration with Professor John Colvin and Dr Susan Seal.

> "THE STUDY MAY REVEAL OPPORTUNITIES FOR NEW PEST MANAGEMENT APPROACHES"

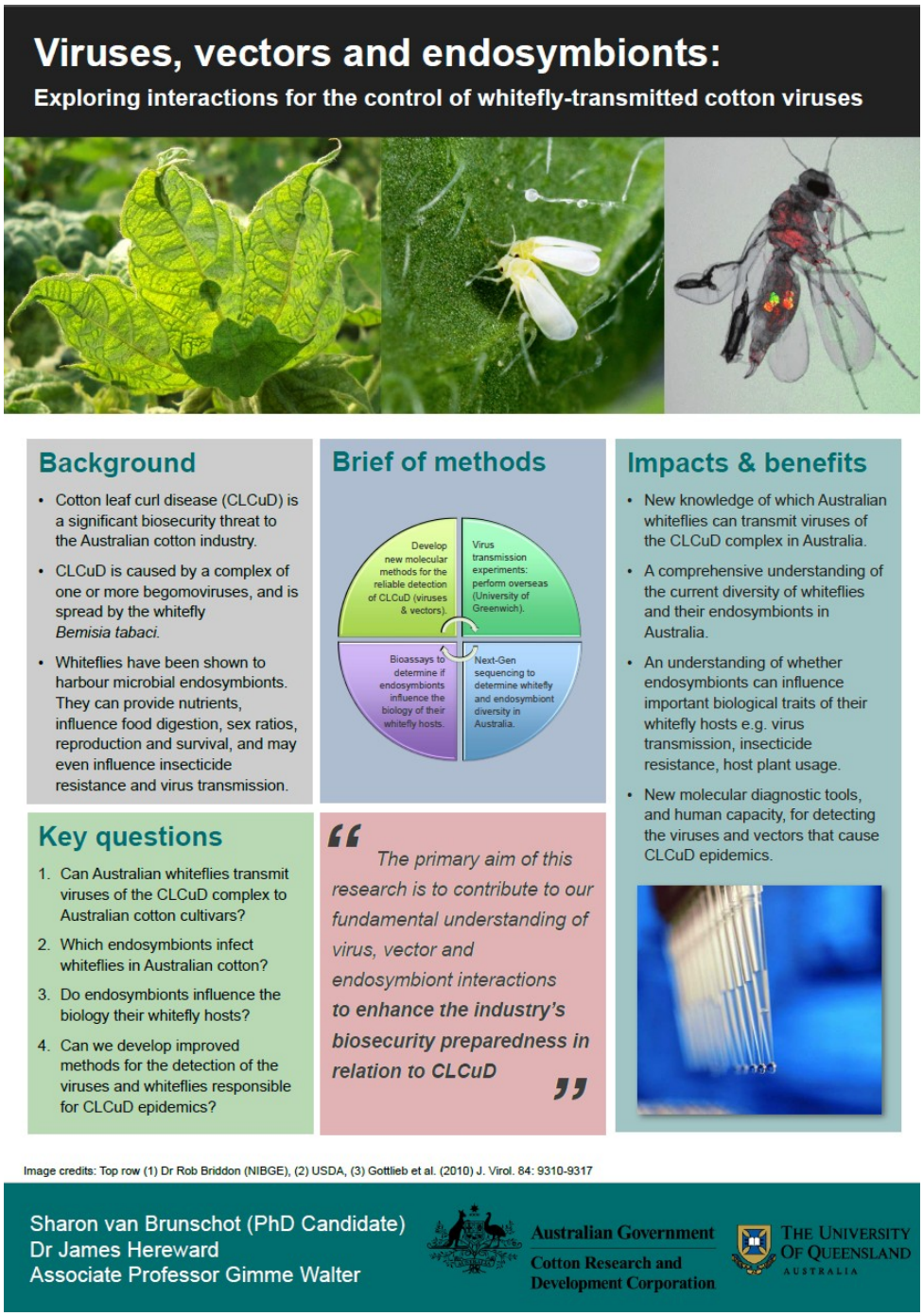
More information

susan.maas@crdc.com.au

email us

Appendix 14: Conference poster

van Brunschot SL, Hereward J, Gambley C, De Barro P, Walter G (2013) Viruses, vectors and endosymbionts: exploring interactions for the control of cotton leaf curl disease. Cotton Science Conference, 8th-11th September, Narrabri, New South Wales, Australia.



Appendix 15: Conference presentation

van Brunschot SL, Hereward J, Gambley C, De Barro P, Walter G (2013) Exploring the interactions of whiteflies, endosymbionts and Cotton leaf curl virus. Northern Farming Systems IPM Researchers Forum, 30th July, Toowoomba, Queensland, Australia.

Link to presentation:
https://www.dropbox.com/s/uzealhspsk6qzez/2013_IPM%20Forum_UQ1305.pptx?dl=0

12. References

Ahmed MZ, De Barro PJ, Greeff JM, Ren S-X, Naveed M, Qiu B-L, 2011. Genetic identity of the Bemisia tabaci species complex and association with high cotton leaf curl disease (CLCuD) incidence in Pakistan. *Pest Management Science* **67**, 307-17.

Allie F, Pierce EJ, Okoniewski MJ, Rey C, 2014. Transcriptional analysis of South African cassava mosaic virus-infected susceptible and tolerant landraces of cassava highlights differences in resistance, basal defense and cell wall associated genes during infection. *BMC Genomics* **15**, 1-30.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-10.

Ashfaq M, Asif M, Anjum ZI, Zafar Y, 2013. Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (Gossypium: Malvaceae). *Molecular Ecology Resources* **13**, 573-82.

Ashfaq M, Asif, M., Anjum, Z. I., Zafar, Y., 2013. Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (Gossypium: Malvaceae). *Molecular Ecology Resources* **13**, 573-82.

Ballard JWO, Whitlock MC, 2004. The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729-44.

Braig HR, Zhou W, Dobson SL, O'neill SL, 1998. Cloning and Characterization of a Gene Encoding the Major Surface Protein of the Bacterial EndosymbiontWolbachia pipientis. *Journal of Bacteriology* **180**, 2373-8.

Briddon RW, Markham PG, 2000. Cotton leaf curl virus disease. *Virus Research* **71**, 151-9.

Brownstein MJ, Carpten JD, Smith JR, 1996. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* **20**, 1004-6, 8-10.

Célix A, López-Sesé A, Almarza N, Gómez-Guillamón ML, Rodríguez-Cerezo E, 1996. Characterization of cucurbit yellow stunting disorder virus, a Bemisia tabaci-transmitted Closterovirus. *Phytopathology* **86**, 1370-6.

Chan KM, Levin SA, 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution* **59**, 720-9.

Chapuis M-P, Estoup A, 2007. Microsatellite Null Alleles and Estimation of Population Differentiation. *Molecular biology and evolution* **24**, 621-31.

Chen S, Yao H, Han J*, et al.*, 2010a. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* **5**, e8613.

Chen S, Yao H, Han J*, et al.*, 2010b. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. *PLoS ONE* **5**, e8613.

De Barro P, Khan S, 2007. Adult Bemisia tabaci biotype B can induce silverleafing in squash. *Bull Entomol Res* **97**, 433-6.

De Barro P, Scott K, Graham G, Lange C, Schutze M, 2003. Isolation and characterization of microsatellite loci in Bemisia tabaci. *Molecular Ecology Resources* **3**, 40-3.

De Barro PJ, 2012. The Bemisia tabaci Species Complex: Questions to Guide Future Research. *Journal of Integrative Agriculture* **11**, 187-96.

De Barro PJ, Hart PJ, 2000. Mating interactions between two biotypes of the whitefly, Bemisia tabaci (Hemiptera: Aleyrodidae) in Australia. *Bulletin of Entomological Research* **90**, 103-12.

De Barro PJ, Liu S-S, Boykin LM, Dinsdale AB, 2011. Bemisia tabaci: A Statement of Species Status. *Annual Review of Entomology* **56**, 1-19.

Dempster AP, Laird NM, Rubin DB, 1977. Maximum Likelihood from incomplete data via the EM algorithm *Journal of the Royal Statistical Society Series B-Methodological* **39**, 1-38.

Deng D, Mcgrath PF, Robinson DJ, Harrison BD, 1994. Detection and differentiation of whitefly-transmitted geminiviruses in plants and vector insects by the polymerase chain reaction with degenerate primers. *Annals of Applied Biology* **125**, 327-36.

Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P, 2010. Refined global analysis of Bemisia tabaci (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Annals of the Entomological Society of America* **103**, 196-208.

Dolores LM, Gonzales, M. C., Magdalita, P. M., 2014. Isolation and Identification of a Begomovirus Associated With the Leaf Curl Disease of Gumamela (Hibiscus rosa-sinensis) in the Philippines. *Asian Journal of Agriculture and Food Sciences* **02**.

Doyle J, Doyle J, 1987. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull* **19**, 11-5.

Duffus JE, Larsen RC, Liu HY, 1986. Lettuce infectious yellows virus - a new type of whitefly-transmitted virus. *Phytopathology* **76**, 97-100.

Firdaus S, Vosman B, Hidayati N, Jaya Supena ED, G.F. Visser R, Van Heusden AW, 2013. The Bemisia tabaci species complex: Additions from different parts of the world. *Insect Science* **20**, 723-33.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R, 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology* **3**, 294-9.

Fondong V, Pita J, Rey M, De Kochko A, Beachy R, Fauquet C, 2000. Evidence of synergism between African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. *Journal of General Virology* **81**, 287-97.

Gottlieb Y, Ghanim M, Chiel E*, et al.*, 2006. Identification and Localization of a Rickettsia sp. in Bemisia tabaci (Homoptera: Aleyrodidae). *Applied and Environmental Microbiology* **72**, 3646-52.

Gottlieb Y, Zchori-Fein E, Mozes-Daube N*, et al.*, 2010. The Transmission Efficiency of Tomato Yellow Leaf Curl Virus by the Whitefly Bemisia tabaci Is Correlated with the Presence of a Specific Symbiotic Bacterium Species. *Journal of Virology* **84**, 9310-7.

Gunning RV, Byrne FJ, Conde BD, Connelly MI, Hergstrom K, Devonshire AL, 1995. First Report of B-Biotype *Bemisia tabaci* (Gennadius)(Hemiptera: Aleyrodidae) in Australia. *Australian Journal of Entomology* **34**, 116-.

Guo T, Guo Q, Cui X-Y, Liu Y-Q, Hu J, Liu S-S, 2015. Comparison of transmission of Papaya leaf curl China virus among four cryptic species of the whitefly Bemisia tabaci complex. **5**, 15432.

Hadjistylli M, Schwartz SA, Brown JK, Roderick GK, 2014. Isolation and Characterization of Nine Microsatellite Loci from Bemisia tabaci (Hemiptera: Aleyrodidae) Biotype B. *Journal of Insect Science (Tucson)* **14**, 1-3.

Hsieh C-H, Ko C-C, Chung C-H, Wang H-Y, 2014a. Multilocus approach to clarify species status and the divergence history of the Bemisia tabaci (Hemiptera: Aleyrodidae) species complex. *Molecular Phylogenetics and Evolution* **76**, 172-80.

Hsieh CH, Ko CC, Chung CH, Wang HY, 2014b. Multi locus approach to clarify species status and the divergence history of the Bemisia tabaci (Hemiptera: Aleyrodidae) species complex. *MOLECULAR PHYLOGENETICS AND EVOLUTION* **76**, 172-80.

Hu J, Chen Y-D, Jiang Z-L*, et al.*, 2015. Global haplotype analysis of the whitefly Bemisia tabaci cryptic species Asia I in Asia. *Mitochondrial DNA* **26**, 232-41.

Idris A, Al-Saleh M, Amer M, Abdalla O, Brown J, 2014. Introduction of Cotton leaf curl Gezira virus into the United Arab Emirates. *Plant Disease* **98**, 1593-.

Iwaki M, Thongmeearkom P, Prommin M, Honda Y, Hibi T, 1982. Whitefly transmission and some properties of cowpea mild mottle virus on soybean in Thailand. *Plant Disease* **66**, 365-8.

Jiang Y, De Blas C, Bedford I, Nombela G, Muniz M, 2004. Effect of Bemisia tabaci biotype in the transmission of tomato yellow leaf curl sardinia virus (TYLCSV-ES) between tomato and common weeds. *Spanish Journal of Agricultural Research* **2**, 115-9.

Katoh K, Standley DM, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-80.

Kearse M, Moir R, Wilson A*, et al.*, 2012a. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-9.

Kearse M, Moir R, Wilson A*, et al.*, 2012b. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-9.

Kearse M, Moir R, Wilson A*, et al.*, 2010. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-9.

Kontsedalov S, Zchori-Fein E, Chiel E, Gottlieb Y, Inbar M, Ghanim M, 2008. The presence of Rickettsia is associated with increased susceptibility of Bemisia tabaci (Homoptera: Aleyrodidae) to insecticides. *Pest Management Science* **64**, 789-92.

Kumar J, Kumar J, Singh SP, Tuli R, 2014. Association of satellites with a mastrevirus in natural infection: Complexity of Wheat dwarf India virus disease. *Journal of Virology* **88**, 7093-104.

Lecoq H, Verdin E, Tepfer M*, et al.*, 2016. Characterization and occurrence of squash chlorotic leaf spot virus, a tentative new torradovirus infecting cucurbits in Sudan. **161**, 1651-5.

Lee W, Park J, Lee G-S, Lee S, Akimoto S-I, 2013. Taxonomic Status of the Bemisia tabaci Complex (Hemiptera: Aleyrodidae) and Reassessment of the Number of Its Constituent Species. *PLoS ONE* **8**, e63817.

Li R, Salih S, Hurtt S, 2004. Detection of Geminiviruses in Sweetpotato by Polymerase Chain Reaction. *Plant Disease* **88**, 1347-51.

Liu S-S, Colvin J, De Barro PJ, 2012. Species concepts as applied to the whitefly Bemisia tabaci systematics: how many species are there? *Journal of Integrative Agriculture* **11**, 176-86.

Mansour A, Al-Musa A, 1993. Cucumber Vein Yellowing Virus; Host Range and Virus Vector Relationships. *Journal of Phytopathology* **137**, 73-8.

Martín G, Velasco L, Segundo E, Cuadrado IM, Janssen D, 2008. The complete nucleotide sequence and genome organization of bean yellow disorder virus, a new member of the genus Crinivirus. *Archives of Virology* **153**, 999-1001.

Maruthi MN, Colvin J, Seal S, 2001. Mating compatibility, life-history traits, and RAPD-PCR variation in Bemisia tabaci associated with the cassava mosaic disease pandemic in East Africa. *Entomologia Experimentalis et Applicata* **99**, 13-23.

Maruthi MN, Colvin J, Thwaites RM, Banks GK, Gibson G, Seal SE, 2004. Reproductive incompatibility and cytochrome oxidase I gene sequence variability amongst host-adapted and geographically separate Bemisia tabaci populations (Hemiptera: Aleyrodidae). *Systematic Entomology* **29**, 560-8.

Maruthi MN, Hillocks RJ, Mtunda K*, et al.*, 2005. Transmission of Cassava brown streak virus by Bemisia tabaci (Gennadius). *Journal of Phytopathology* **153**, 307-12.

Meglecz E, Costedoat C, Dubut V*, et al.*, 2010. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* **26**, 403-4.

Meglecz E, Pech N, Gilles A*, et al.*, 2014. QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate. *Mol Ecol Resour* **14**, 1302-13.

Monsen KJ, Blouin MS, 2003. Genetic structure in a montane ranid frog: restricted gene flow and nuclear–mitochondrial discordance. *Molecular Ecology* **12**, 3275-86.

Mouton L, Thierry M, Henri H*, et al.*, 2012. Evidence of diversity and recombination in Arsenophonus symbionts of the Bemisia tabaci species complex. *BMC Microbiology* **12**, S10.

Muñiz M, 2000. Host suitability of two biotypes of Bemisia tabaci on some common weeds. *Entomologia Experimentalis et Applicata* **95**, 63-70.

Muñiz M, Nombela G, 2001. Bemisia tabaci: A new clip-cage for biological studies. In. United Kingdom: European Whitefly Studies Network (EWSN). (2016.)

Oliveira MRV, Henneberry TJ, Anderson P, 2001. History, current status, and collaborative research projects for Bemisia tabaci. *Crop Protection* **20**, 709-23.

Omondi BA, Sseruwagi P, Obeng-Ofori D, Danquah EY, Kyerematen RA, 2005. Mating interactions between okra and cassava biotypes of Bemisia tabaci (Homoptera: Aleyrodidae) on eggplant. *International Journal of Tropical Insect Science* **25**, 159-67.

Peakall R, Smouse PE, 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537-9.

Peakall ROD, Smouse PE, 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-95.

Piertney SB, Maccoll AD, Bacon PJ, Racey PA, Lambin X, Dallas JF, 2000. Matrilineal genetic structure and female-mediated gene flow in red grouse (Lagopus lagopus scoticus): an analysis using mitochondrial DNA. *Evolution* **54**, 279-89.

Pinho C, Harris DJ, Ferrand N, 2007. Comparing patterns of nuclear and mitochondrial divergence in a cryptic species complex: the case of Iberian and North African wall lizards (Podarcis, Lacertidae). *Biological journal of the Linnean Society* **91**, 121-33.

Posada D, 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-6.

Qin L, Pan LL, Liu SS, 2016. Further insight into reproductive incompatibility between putative cryptic species of the Bemisia tabaci whitefly complex. *Insect Science* **23**, 215-24.

Quintela ED, Abreu AG, Lima JF, Mascarin GM, Santos JB, Brown JK, 2016. Reproduction of the whitefly Bemisia tabaci (Hemiptera: Aleyrodidae) B biotype in maize fields (Zea mays L.) in Brazil. *Pest Manag Sci* **72**, 2181-7.

Rambaut A, Drummond A, 2003. Tracer: MCMC trace analysis tool. *University of Oxford*.

Rojas M, Gilbertson R, Maxwell D, 1993. Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses. *Plant Disease* **77**, 340-7.

Ronquist F, Huelsenbeck JP, 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-4.

Rousset F, 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-6.

Schaefers G, Terry E, 1976. Insect transmission of sweet potato disease agents in Nigeria. *Phytopathology* **66**, 642-5.

Schuelke M, 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature biotechnology* **18**, 233-4.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P, 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* **87**, 651-701.

Stansly PA, Naranjo SE, eds, 2010. *Bemisia: bionomics and management of a global pest*. Dordrecht;London: Springer.

Sun DB, Xu J, Luan JB, Liu SS, 2011. Reproductive incompatibility between the B and Q biotypes of the whitefly Bemisia tabaci in China: genetic and behavioural evidence. *Bulletin of Entomological Research* **101**, 211-20.

Syarifin Firdaus Ben Vosman Nurul Hidayati Ence Darmo Jaya Supena Richard GFVaWVH, 2013. The Bemisia tabaci species complex – Additions from different parts of the world. *Insect Science* **20**, 723-33.

Thao ML, Baumann P, 2004. Evidence for Multiple Acquisition of Arsenophonus by Whitefly Species (Sternorrhyncha: Aleyrodidae). *Current Microbiology* **48**, 140-4.

Untergasser A, Cutcutache I, Koressaar T*, et al.*, 2012. Primer3—new capabilities and interfaces. *Nucleic acids research* **40**, e115-e.

Van Brunschot SL, Bergervoet JHW, Pagendam DE*, et al.*, 2014. A bead-based suspension array for the multiplexed detection of begomoviruses and their whitefly vectors. *Journal of Virological Methods* **198**, 86-94.

Verbeek M, Van Bekkum PJ, Dullemans AM, Van Der Vlugt RA, 2014. Torradoviruses are transmitted in a semi-persistent and stylet-borne manner by three whitefly vectors. *Virus Res* **186**, 55-60.

Webb SE, Adkins S, Reitz SR, 2012. Semipersistent Whitefly Transmission of Squash vein yellowing virus, Causal Agent of Viral Watermelon Vine Decline. *Plant Disease* **96**, 839-44.

Weeks AR, Breeuwer JA, 2003. A new bacterium from the Cytophaga-Flavobacterium-Bacteroides phylum that causes sex ratio distortion. *Insect symbiosis*, 165-76.

Weng SH, Tsai WS, Kenyon L, Tsai CW, 2015. Different transmission efficiencies may drive displacement of tomato begomoviruses in the fields in Taiwan. *Annals of Applied Biology* **166**, 321-30.

White JA, Kelly SE, Perlman SJ, Hunter MS, 2009. Cytoplasmic incompatibility in the parasitic wasp Encarsia inaron: disentangling the roles of Cardinium and Wolbachia symbionts. *Heredity* **102**, 483-9.

White TJ, Bruns T, Lee S, Taylor J, 1990a. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In. *PCR Protocols.* San Diego: Academic Press, 315-22.

White TJ, Bruns T, Lee S, Taylor J, 1990b. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications* **18**, 315-22.

Wisler G, Li R, Liu H-Y, Lowry D, Duffus J, 1998. Tomato chlorosis virus: a new whitefly-transmitted, phloem-limited, bipartite closterovirus of tomato. *Phytopathology* **88**, 402-9.

Wyatt S, Brown JK, 1996a. Detection of subgroup III geminivirus isolates in leaf extracts by degenerate primers and polymerase chain reaction. *Phytopathology* **86**, 1288-93.

Wyatt SD, Brown JK, 1996b. Detection of subgroup III geminivirus isolates in leaf extracts by degenerate primers and polymerase chain reaction. *Phytopathology.* **86**.

Xu J, De Barro PJ, Liu SS, 2010. Reproductive incompatibility among genetic groups of Bemisia tabaci supports the proposition that the whitefly is a cryptic species complex. *Bulletin of Entomological Research* **100**, 359-66.

Yokomi R, Hoelmer K, Osborne L, 1990. Relationships between the sweetpotato whitefly and the squash silverleaf disorder. *Phytopathology* **80**, 895-900.

Zang L-S, Liu S-S, 2007. A Comparative Study on Mating Behaviour between the B Biotype and a non-B Biotype of Bemisia tabaci (Hemiptera: Aleyrodidae) from Zhejiang, China. *Journal of Insect Behavior* **20**, 157-71.

Zchori-Fein E, Brown JK, 2002. Diversity of Prokaryotes Associated with Bemisia tabaci (Gennadius) (Hemiptera: Aleyrodidae). *Annals of the Entomological Society of America* **95**, 711-8.

Zhang J, Wang L, Zheng Y*, et al.*, 2017. Effects of Bemisia tabaci (Gennadius) infestation and squash silverleaf disorder on Cucurbita pepo L. leaf. *Scientia Horticulturae* **217**, 8-16.