UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

FCA
FACULDADE DE CIÊNCIAS AGRONÔMICAS
UNESP | BOTUCATU

BRUNO ROSSITTO DE MARCHI

ANÁLISE DE TRANSCRIPTOMAS DE MOSCA-BRANCA (*Bemisia tabaci)* E DIVERSIDADE GENÉTICA EM CLOROPLASTO DE MANDIOCAS (*Manihot esculenta*) INFECTADAS COM VÍRUS

Botucatu

2018

BRUNO ROSSITTO DE MARCHI

ANÁLISE DE TRANSCRIPTOMAS DE MOSCA-BRANCA (*Bemisia tabaci)* E DIVERSIDADE GENÉTICA EM CLOROPLASTO DE MANDIOCAS (*Manihot esculenta*) INFECTADAS COM VÍRUS

Tese apresentada à Faculdade de Ciências Agronômicas da Unesp Câmpus de Botucatu, para obtenção do título de Doutor em Proteção de Plantas.

Orientadora: Renate Krause-Sakate

Coorientadora: Laura M. Boykin

Botucatu

2018

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉCNICA DE AQUISIÇÃO E TRATAMENTO DA INFORMAÇÃO – DIRETORIA TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - UNESP – FCA – LAGEADO – BOTUCATU (SP)

De Marchi, Bruno Rossito, 1988-
D372a Análise de transcriptomas de mosca-branca (*Bemisia tabaci)* e diversidade genética em cloroplasto de mandiocas (*Manihot esculenta*) infectadas com vírus / Bruno Rossito. – Botucatu: [s.n.], 2018
96 p.: ils. color., tabs.

Tese (Doutorado)- Universidade Estadual Paulista Faculdade de Ciências Agronômicas, Botucatu, 2018
Orientadora: Renate Krause-Sakate
Coorientadora: Laura M. Boykin
Inclui bibliografia

1. Mandioca – Doenças e pragas. 2. Mosca branca. 3. Fitopatologia. 4. Vírus. 5. Plantas – Mitocôndria. I. Krause-Sakate, Renate. II. Boykin, Laura M. III. Universidade Estadual Paulista "Júlio de Mesquita Filho" (Câmpus de Botucatu). Faculdade de Ciências Agronômicas. IV. Título.

Elaborada por Ana Lucia G. Kempinas – CRB-8:7310

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Botucatu

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: ANÁLISE DE TRANSCRIPTOMAS DE MOSCA-BRANCA (Bemisia tabaci) E DIVERSIDADE GENÉTICA EM CLOROPLASTO DE MANDIOCAS (Manihot esculenta) INFECTADAS COM VÍRUS

AUTOR: BRUNO ROSSITTO DE MARCHI
ORIENTADORA: RENATE KRAUSE SAKATE
COORIENTADORA: LAURA M. BOYKIN

Aprovado como parte das exigências para obtenção do Título de Doutor em AGRONOMIA (PROTEÇÃO DE PLANTAS), pela Comissão Examinadora:

Profa. Dra. RENATE KRAUSE SAKATE
Depto de Proteção Vegetal / UNESP - Faculdade de Ciências Agronômicas de Botucatu

Profa. Dra. REGIANE CRISTINA OLIVEIRA DE FREITAS BUENO
Departamento de Proteção Vegetal / FCA - UNESP - Câmpus de Botucatu

Prof. Dr. JULIO MASSAHARU MARUBAYASHI
Insetos Vetores / Sementes Hortec - Bragança Paulista

DR. VALDIR ATSUSHI YUKI
Depto. Virologia Vegetal / Instituto Agronômico de Campinas

Dra. MÔNIKA FECURY MOURA
Departamento de Proteção Vegetal / FCA/Unesp Botucatu

Botucatu, 31 de julho de 2018

AGRADECIMENTOS

Aos meus pais Adriano F. De Marchi e Rosane A. R. De Marchi e ao meu irmão Adriano F. De Marchi Jr. pelo apoio e amor incondicional e por estarem sempre ao meu lado em todas as circunstâncias

A Prof. Dr. Renate Krause-Sakate, pela orientação, amizade, ensinamentos, seriedade e por me apoiar e auxiliar nas principais escolhas durante toda a minha formação acadêmica

To Dr. Laura Boykin, for the support, friendship, patience, for the opportunity to be part of her team and for all the shared knowledge during my year in Australia

To the Australian team, Tonny Kinene, James Wainaina, Ammar Ghalab, Anders Savil, Monika Kehoe for the friendship, support and the amazing moments we spent together in Perth

Aos meus amigos e companheiros de laboratório e de departamento, Luis F. M. Watanabe, Vinicius H. Bello, Felipe Barreto, Giovana Cruciol, Monika Fecury Moura, Marcos Roberto, João César, Tadeu, Marcelo Soman, Vanessa Carvalho, Vinicius Canassa, Eduardo Gorayeb, Leticia Aparecida, Kessia Pantoja, Denise Nozaki Nakada, Vinicius Suarez, Lucas Fusco, Isabela Vieira e Roseni pelo companheirismo, amizade, parceria e todos os bons momentos vividos

A minha namorada Rafaela Ruschel, pelo companheirismo, apoio em todas as decisões e por todos os momentos incríveis compartilhados. Sou muito grato por ter você ao meu lado

Aos amigos que compartilharam as alegrias e tristezas de um intercâmbio, Eduardo Medina, Lilian Sanglard, Danilo Camelo e Aida Carneiro. Serei eternamente grato por meu caminho ter cruzado o de vocês

Aos Professores e Pesquisadores que me guiaram ao longo da minha formação e sempre servirão como exemplo e inspiração para mim. Julio Marubayashi, Maria José De Marchi, Valdir A. Yuki, Marcelo Pavan, Edson L. Baldin, Carlos G. Raetano, Antonio Carlos Maringoni, Regiane C. F. Bueno, Jorge A. M. Rezende, Cristiane Muller e Jayme Augusto de Souza-Neto

À todos os funcionários da UNESP/FCA – Botucatu e The University of Western Australia (UWA)

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro, concedido por meio de bolsa de doutorado (Processo nº 131324/2015-5) e doutorado sanduíche na Austrália (Processo nº 200826/2015-8)

RESUMO

A mosca-branca, *Bemisia tabaci* (Gennadius) é uma praga de distribuição global que afeta centenas de diferentes plantas hospedeiras incluindo grandes culturas, olerícolas e ornamentais. *B. tabaci* causa danos principalmente através da transmissão de vírus de plantas como os begomovirus, crinivirus, ipomovirus, torradovirus e carlavirus. Atualmente, *B. tabaci* é considerada um complexo de pelo menos 40 espécies crípticas que apresentam diversidade genética, biológica e diferentes composições de bactérias endossimbiontes facultativas. No Brasil, tanto espécies nativas quanto exóticas de mosca-branca são encontradas e ainda há uma escassez de dados genômicos destas populações e das bactérias endossimbiontes. Na África Oriental, altas populações de moscas-brancas estão disseminando diferentes vírus de plantas que causam epidemias na cultura da mandioca (*Manihot esculenta*) e com perdas na produtividade. A principal forma de manejo desses vírus na África é através da utilização de variedades tolerantes. Portanto, é essencial a identificação de novos genes de resistência para o desenvolvimento de variedades e um manejo eficiente das doenças. O sequenciamento de transcriptomas é uma ferramenta que possibilita uma análise genômica da mosca-branca, dos vírus transmitidos por ela, das bactérias endossimbiontes e das plantas hospedeiras dessa praga. Portanto, os dados genômicos obtidos dão suporte para o desenvolvimento de novas técnicas que podem se tornar futuras alternativas de controle de mosca-branca e dos vírus associados. No Capítulo 1, dados de transcriptomas foram obtidos das diferentes espécies de *B. tabaci* encontradas no Brasil, tendo sido possível a obtenção de genomas mitocondriais completos de espécies exóticas e nativas de mosca-branca, além de genomas parciais do endossimbionte facultativo *Hamiltonella*. A análise filogenética revelou que as diferenças genéticas presentes no gene mtCOI entre as espécies nativas e as espécies exóticas se estendem ao genoma mitocondrial e ao endossimbionte facultativo *Hamiltonella*. Além disso, foi possível verificar uma deleção de aminoácidos somente no gene GroEL de *Hamiltonella* que está presente em populações de moscas-brancas nativas. Esse gene é conhecido por estar associado a transmissão de vírus de plantas. No Capítulo 2, foram sequenciados transcriptomas de plantas de mandioca naturalmente infectadas com vírus coletadas no campo em diferentes países da África. A partir desses dados, foi avaliada a diversidade de genes do cloroplasto e a relação com diferentes espécies de vírus. Há uma baixa diversidade dentre as cultivares de mandioca atualmente plantadas na África e não foi possível verificar nenhuma relação entre os genes de cloroplastos avaliados e as espécies de vírus detectadas ocorrendo naturalmente no campo. Os dados obtidos reforçam a necessidade da introdução de novos materiais genéticos para aumentar a diversidade genética das cultivares de mandioca plantadas nesses países, a fim de melhorar as alternativas de manejo das epidemias de vírus.

Palavras-chave: *Cassava mosaic virus*, *Cassava brown streak virus*, mitocôndria, GroEL, New world

ABSTRACT

The whitefly, *Bemisia tabaci* (Gennadius) is a global pest that affects hundreds of different plant hosts including vegetable, fiber and ornamental crops. *B. tabaci* causes damage mainly by the transmission of plant viruses such as begomoviruses, criniviruses, ipomoviruses, torradoviruses, and carlaviruses. Currently, *B. tabaci* is known as a complex of at least 40 putative cryptic species that shows genetic and biological diversity and a different composition of bacterial facultative endosymbionts. In Brazil, both exotic and indigenous species of whiteflies are found and there is still a lack of genomic data available among these populations and also their bacterial endosymbionts. In East Africa, high populations of whiteflies are transmitting different plant viruses that are causing epidemics in cassava crops (*Manihot esculenta*) and leading to great yield losses. Currently, the management of these viral diseases in Africa is carried out mainly by growing cassava tolerant varieties. Therefore, is essential to identify new target genes for the development of new varieties for an efficient management of viral diseases. Transcriptome sequencing is a tool that allows a genomic analysis of the whitefly, whitefly-transmitted viruses, bacterial endosymbionts and plant hosts. Therefore, the genomic data obtained gives support for the development of new technics that might aid for future management alternatives of whiteflies and their associated viruses. In Chapter 1, transcriptome data were obtained from different *B. tabaci* species found in Brazil which allowed to obtain complete mitochondrial genomes from different whitefly species and draft genomes of the facultative endosymbiont *Hamiltonella*. The phylogenetic analysis revealed that the genetic differences among exotic and native populations present in the mtCOI gene extends to the mitochondrial genome and to the facultative endosymbiont *Hamiltonella*. In addition, it was verified an amino acid deletion in the GroEL gene from *Hamiltonella* present only in native populations of whiteflies. This gene is known to be associated with the transmission of plant viruses. In Chapter 2, transcriptome data were obtained from virus-infected cassava plants collected in the field in different African countries. The data allowed to evaluate the diversity of chloroplast genes and their relationship with different virus species. The data revealed a low genetic diversity among cassava currently grow in East Africa. In addition, there was no direct relationship between the evaluated chloroplast genes and the virus species detected. The obtained data reinforce the need of introduction of new genetic accession to increase the genetic diversity of the currently grown cassava in Africa in order to improve the alternatives of management of viral diseases.

Keywords: *Cassava mosaic virus*, *Cassava brown streak virus*, mitochondrial, GroEL, New world

LISTA DE ILUSTRAÇÕES

CAPÍTULO 1

CAPÍTULO 2

LISTA DE TABELAS

CAPÍTULO 1

CAPÍTULO 2

LISTA DE ABREVIATURAS E SIGLAS

ACMV	African cassava mosaic disease
Bp	Base pair
CBSD	Cassava brown streak disease
CBSV	Cassava brown streak virus
CMD	Cassava mosaic disease
EACMV	East african cassava mosaic disease
IR	Inverted repeat
MEAM1	Middle East-Asia Minor 1
MED	Mediterranean
mtCOI	mitochondrial cytochrome oxidase I
NCBI	National Center for Biotechnology Information
NW1	New World 1
NW2	New World 2
ORF's	Open reading frame
rRNA	Ribosomal ribonucleic acid
SNP's	Single nucleotide polymorphism
tRNA	Transfer ribonucleic acid
UCBSV	Ugandan cassava brown streak virus

SUMÁRIO

INTRODUÇÃO GERAL

A mosca-branca, *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) é uma praga extremamente invasiva que causa danos em diversas culturas de importância econômica em todo o mundo (DE BARRO et al., 2011). O dano causado por *B. tabaci* pode ser direto pela sucção de seiva ou indireto através de alterações fisiológicas na planta hospedeira, desenvolvimento de fumagina e principalmente da transmissão de vírus de plantas pertencentes aos gêneros *Begomovirus, Carlavirus, Crinivirus, Ipomovirus* e *Torradovirus* (NAVAS-CASTILLO; FIALLO-OLIVÉ; SÁNCHEZ-CAMPOS, 2011). Atualmente, *B. tabaci* é classificada como um complexo de espécies crípticas composto de pelo menos 40 espécies com base no gene mitocondrial mtCOI (BOYKIN et al., 2011). As espécies do complexo *B. tabaci* são morfologicamente indistinguíveis mas diferem entre si em diversos aspectos incluindo a gama de hospedeiros, a capacidade em causar desordens fisiológicas nas plantas, atração a inimigos naturais, resposta à pesticidas e diferentes capacidades na transmissão de vírus de plantas (DE BARRO et al., 2011; XU; DE BARRO; LIU, 2010).

No Brasil, quatro diferentes espécies do complexo já foram relatadas até o momento. As espécies nativas das Américas, New World 1 (NW1) e New World 2 (NW2) que são pouco invasivas e atualmente encontradas com mais frequência em plantas daninhas (BARBOSA et al., 2014; MARUBAYASHI et al., 2013) e as espécies exóticas, Middle East-Asia Minor 1 (MEAM1 ou biótipo B) e Mediterranean (MED ou biótipo Q), que são distribuídas globalmente e de mais difícil controle comparadas as espécies nativas (GAUTHIER et al., 2014; HU et al., 2011). A espécie exótica, MEAM1 foi primeiramente relatada no Brasil na década de 1990 e implicou em surtos epidemiológicos de begomovirus em diferentes culturas, como o tomateiro (LOURENÇÃO; NAGAI, 1994). Mais recentemente, a espécie exótica MED também foi relatada no Brasil, primeiramente no Rio Grande do Sul (DA FONSECA BARBOSA et al., 2015) seguido de uma segunda introdução associada a plantas ornamentais (APARECIDA DE MORAES et al., 2017).

O Brasil é um país de grande extensão territorial e alta diversidade dos sistemas agrícolas, o que dificulta ainda mais a identificação e rastreamento das pragas. As diferentes introduções de pragas exóticas de *B. tabaci* no país destacou ainda mais a importância e necessidade de dados de referência, como os genomas mitocondriais

completos, para prevenir e rastrear a introdução destas pragas. Além disso, comparações de genomas mitocondriais completos fornecem um melhor entendimento das relações filogenéticas entre membros do complexo de espécies *B. tabaci* (TAY et al., 2014). Atualmente, existem 20 genomas mitocondriais completos de mosca-branca caracterizados e disponíveis no GenBank (TAY et al., 2014, 2017a; THAO; BAUMANN, 2004; WANG et al., 2013).

A complexidade de *B. tabaci* também depende de bactérias endossimbiontes as quais as funções ainda não são completamente compreendidas (PARRELLA et al., 2014). Os endossimbiontes obrigatórios são essenciais para os insetos sobreviverem em dietas pobres em nutrientes (BAUMANN, 2005; FERRARI; VAVRE, 2011). Para a mosca-branca, o endossimbionte obrigatório é *Portiera aleyrodidarum* (THAO; BAUMANN, 2004). Além disso, insetos também abrigam com frequência endossimbiontes facultativos que podem ter diferentes papeis no vetor, como alterar a suscetibilidade inseticidas (KONTSEDALOV et al., 2008, 2009), facilitar a transmissão de vírus (GOTTLIEB et al., 2010; RANA et al., 2012), conferir tolerância à altas temperaturas (BRUMIN; KONTSEDALOV; GHANIM, 2011) e resistência a inimigos naturais como vespas parasitas (MAHADAV et al., 2008). Em *B. tabaci*, diferentes endossimbiontes facultativos foram descritos, incluindo *Arsenophonus*, *Hamiltonella*, *Wolbachia*, *Cardinium*, *Fritschea* e *Rickettsia* (GOTTLIEB et al., 2010; HU et al., 2011). Dentre os endossimbiontes facultativos, destaca-se a bactéria *Hamiltonella defensa* que é encontrada esporadicamente em insetos sugadores incluindo afídios, psilídeos e moscas-brancas (RAO et al., 2012).

Estudos realizados anteriormente no Brasil demonstraram que diferentes espécies de *B. tabaci* como MEAM1, MED e NW2 abrigam *Hamiltonella* (APARECIDA DE MORAES et al., 2017; MARUBAYASHI et al., 2014). Mais recentemente, um levantamento relatou que *Hamiltonella* está amplamente distribuída pelo território brasileiro e foi encontrada em 89,5% dos espécimes de MEAM1 e 50% dos espécimes de MED analisados (DE MORAES APARECIDA et al., 2018). A alta incidência de *Hamiltonella* em populações de mosca-branca no Brasil pode ter implicações sérias para a transmissão de vírus. Foi evidenciado que a proteína GroEL produzida por *Hamiltonella* de *B. tabaci* MEAM1 estaria envolvida na transmissão do begomovirus *Tomato yellow leaf curl virus*, TYLCV (GOROVITS; CZOSNEK, 2013; GOTTLIEB et al., 2010), porém não tendo sido evidenciada interação da GroEL produzida por outros

endossimbiontes de *B. tabaci* (MEAM1 e MED) (GOTTLIEB et al., 2010). A *Hamiltonella* pode ter diferentes papéis dependendo do vetor. No pulgão-da-ervilha, *Acyrthosiphon pisum* (Harris), *Hamiltonella* pode bloquear o desenvolvimento larval de vespas parasitas, garantindo sua sobrevivência (SU et al., 2013). Devido ao fato da *Hamiltonella* ter um importante papel na transmissão de vírus de plantas e por ser encontrada com alta frequência nas diferentes espécies de *B. tabaci* presentes no Brasil, há uma necessidade de conhecer a variabilidade desse endossimbionte facultativo.

A importância de mosca-branca como uma praga que causa grandes danos econômicos e sociais se estende globalmente. O laboratório da Dra. Laura Boykin na "The University of Western Australia", é uma referência no mundo em estudos de *B. tabaci*. Particularmente neste laboratório, é realizado um estudo amplo na África Oriental com o objetivo de diminuir os danos causados por esta praga. A pesquisa desenvolvida visa melhorar a economia local e aumentar a segurança alimentar nesses países. Dessa forma, parte do trabalho foi realizado com amostras de mandioca coletadas em regiões afetadas pela mosca-banca na África Oriental, aonde essa praga tem sido responsável pela disseminação e epidemias de diversos vírus na cultura da mandioca (*Manihot esculenta* Crantz) (HILLOCKS; JENNINGS, 2003). Nos países da África, a mandioca é especialmente importante pois é altamente tolerante a seca e os tubérculos podem ser preservados no solo por alguns anos sendo considerada uma importante reserva de carboidratos para atenuar a fome global (RAHEEM; CHUKWUMA, 2001). As principais doenças de origem viral transmitidas por mosca-branca para a mandioca são CBSD (cassava brown streak disease), nas áreas costeiras do leste e sudoeste da Africa, e CMD (cassava mosaic disease) em todas as regiões africanas (CALVERT; THRESH, 2002). Há relatos que essas duas doenças juntas podem causar perdas de produção de mais de US$ 3 bilhões todos os anos na África (HILLOCKS; JENNINGS, 2003; LEGG et al., 2006). No Brasil, não há nenhum relato de vírus de plantas transmitidos por mosca-branca ocorrendo em mandioca até o momento.

Cassava mosaic disease, CMD ocorre por toda a África e diferentes espécies de geminivirus foram relatadas à doença, como o *African cassava mosaic virus*, *East African cassava mosaic virus* e *East African cassava mosaic virus* – Uganda (LEGG et al., 2006). Quanto ao CBSD, acredita-se que esta doença era causada por duas

espécies distintas de vírus pertencentes ao gênero *Ipomovirus*: *Cassava brown streak virus* (CBSV) e *Ugandan cassava brown streak virus* (UCBSV). Porém, estudos recentes baseados na análise de 470 amostras de folhas sintomáticas coletadas na Tanzânia sugerem que até quatro diferentes espécies de cassava brown streak viruses (CBSVs) podem estar associadas com o CBSD (NDUNGURU et al., 2015). Incidências do CBSD são particularmente altas em áreas costais da Tanzânia e Moçambique, onde a mandioca é alimento básico (HILLOCKS; JENNINGS, 2003).

O centro de origem da mandioca é a América do Sul e a maior parte da diversidade genética é encontrada em bancos de germoplasma nesse continente. Mais de 37% da diversidade de mandioca é mantida no banco de germoplasma do CIAT na Colômbia e no Brasil, aonde a mandioca foi domesticada há milhares de anos atrás, está mantido cerca de 24% da diversidade de mandioca (https://ciat.cgiar.org). Outros países da América do Sul (21%), América Central e Caribe (7%) e na Ásia (7%) também contribuem para a coleção. O germoplasma africano é conservado no banco de germoplasma do IITA na Nigéria (https://ciat.cgiar.org).

O manejo e controle de doenças virais em mandioca dependem principalmente de cultivares tolerantes. Assim, novos genes alvos para a resistência de vírus de plantas são essenciais para o desenvolvimento de mandioca resistentes a vírus por melhoramento convencional ou engenharia genética. Sequências completas de genomas de cloroplastos são valiosas para decifrar as relações filogenéticas entre diferentes grupos taxonômicos relacionados e para aprimorar o entendimento sobre evolução das espécies (DANIELL et al., 2016). Outra aplicação do genoma de cloroplasto na agricultura é a identificação de cultivares comerciais e determinação da pureza deles (DANIELL et al., 2016).

O cloroplasto é um alvo comum para a replicação de vírus de plantas, e reciprocamente, os componentes do cloroplasto também tem um papel ativo na defesa contra vírus de plantas (DE TORRES ZABALA et al., 2015), o que faz desta organela uma potencial fonte para a descoberta de novos genes contra patógenos invasores como os vírus de plantas. De fato, mais e mais fatores do cloroplasto têm sido identificados interagindo com componentes virais. Esses fatores estão envolvidos na replicação viral, movimento, sintomas ou defesa da planta, sugerindo que os vírus evoluíram para interagir com o cloroplasto (ZHAO et al., 2016). Mais de 800

sequências de genomas completos de cloroplasto estão disponíveis na base de dados de genomas de organelas do National Center for Biotechnology Information (NCBI), incluindo os da mandioca. A sequência completa do genoma do cloroplasto da mandioca tem 161,453bp de comprimento e inclui um par de repetições invertidas (IR) de 26,954bp. O genoma inclui 128 genes, 96 são copias únicas e 16 são duplicados na IR. Há quatro genes de rRNA (RNA ribossomal) e 30 tRNAS (RNA transportador) distintos, sete deles são duplicados na IR (DANIELL et al., 2008).

O sequenciamento de transcriptomas é uma ferramenta essencial que possibilita uma análise genômica da mosca-branca, dos vírus transmitidos por ela, das bactérias endossimbiontes e das plantas hospedeiras desta praga. Dessa forma a tese foi dividida em dois capítulos sendo que no Capitulo 1, *"*Comparative transcriptome analysis reveals genetic diversity in the endosymbiont *Hamiltonella* between native and exotic populations of *Bemisia tabaci*", redigido em inglês conforme as normas da revista PLOS ONE, foram sequenciados dez transcriptomas de espécimes de diferentes populações brasileiras e foram caracterizados os genomas mitocondriais completos pertencentes a três espécies diferentes (MEAM1, MED e NW2) do complexo *B. tabaci* seguido de analise filogenética. Além disso, a diversidade do endossimbionte facultativo *Hamiltonella* foi inferida analisando-se 12 genes, com um foco especial para a região codificadora da proteína GroEL, envolvida na transmissão de begomovirus. No Capitulo 2, "Genetic diversity and SNP's from chloroplast coding regions of virus-infected cassava (*Manihot esculenta* Crantz)", redigido em inglês conforme as normas da revista PLOS ONE, o objetivo foi avaliar a diversidade genética de regiões codificantes do cloroplasto de mandioca de plantas naturalmente infectadas com vírus de diferentes variedades coletadas no campo na África Oriental.

CAPÍTULO 1

Comparative transcriptome analysis reveals genetic diversity in the endosymbiont *Hamiltonella* between native and exotic populations of *Bemisia tabaci* from Brazil

Bruno Rossitto De Marchi[1]; Tonny Kinene[2]; James Mbora Wainaina[2]; Renate Krause-Sakate[1]; Laura M. Boykin[2*]

[1]UNESP – Universidade Estadual Paulista, Faculdade de Ciências Agronomicas, Botucatu-SP, Brazil.

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

*Corresponding author

E-mail: laura.boykin@uwa.edu.au

Abstract

The whitefly, *Bemisia tabaci*, is a species complex of more than 40 cryptic species and a major agricultural pest. It causes extensive damage to plants mainly by transmitting plant viruses. There is still a lack of genomic data available for the different whitefly species found in Brazil and their bacterial endosymbionts. Understanding the genetic and transcriptomic composition of these insect pests, the viruses they transmit and the microbiota is crucial to sustainable solutions for farmers to control whiteflies. Illumina RNA-Seq was used to obtain the transcriptome of individual whiteflies from 10 different populations from Brazil including Middle East-Asia Minor 1 (MEAM1), Mediterranean (MED) and New World 2 (NW2). Raw reads were assembled using CLC Genomics Workbench and subsequently mapped to reference genomes. We obtained whitefly complete mitochondrial genomes and draft genomes from the facultative bacterial endosymbiont *Hamiltonella* for further phylogenetic analyses. In addition, nucleotide sequences of the GroEL chaperonin gene from *Hamiltonella* from different populations were obtained and analysed. There was concordance in the species clustering using the whitefly complete mitogenome and the mtCOI gene tree. On the other hand, the phylogenetic analysis using the 12 ORF's of *Hamiltonella* clustered the

native species NW2 apart from the exotics MEAM1 and MED. In addition, the amino acid analysis of GroEL chaperonin revealed a deletion only in *Hamiltonella* infecting NW2 among whiteflies populations analysed which was further confirmed by PCR and Sanger sequencing. The genomic data obtained in this study will aid understanding the functions that *Hamiltonella* may have in whitefly biology and serve as a reference for further studies regarding whiteflies in Brazil.

Keywords: mitochondrial; mitogenome; GroEL; chaperonin; new world

Introduction

The whitefly *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) is an insect pest of significant economic importance of a wide variety of agricultural crops such as tomatoes, beans, cassava, cotton and ornamentals [1]. *B. tabaci* causes damage directly through feeding and indirectly through the transmission of plant pathogenic viruses, belonging to the genera *Begomovirus*, *Carlavirus*, *Crinivirus*, *Ipomovirus* and *Torradovirus* [2]. Currently, *B. tabaci* is classified as a cryptic species complex composed of at least 40 morphologically indistinguishable species on the basis of the mtCOI gene [3]. In Brazil, four *B. tabaci* species have been reported to date, the natives species from the Americas, New World 1 (NW1) and New World 2 (NW2) and the exotic species Middle East-Asia Minor 1 (MEAM1) and Mediterranean (MED) [4–6]. The exotics, MEAM1 and MED are globally distributed and the most challenging to control [7,8]. Species of the complex *B. tabaci* differ in several aspects including the range of host plants utilized, the capacity to cause plant disorders, attraction of natural enemies, response to pesticides and plant virus transmission capabilities [1,9]. Brazil it's a country with extensive territory and high diversification of the agricultural systems which makes the identification and tracking of pests more complex. The first report of MED in Brazil [4] followed by a second incursion of this species associated to ornamentals plants [10] has highlighted the need of reference data, such as complete mitochondrial genome, to prevent and track the introduction of exotic pests to the country. In addition, full mitogenome comparisons provides a better understanding of evolutionary genetic relationships between members of *B. tabaci* [11]. Currently, there are 13 complete whitefly mitochondrial genomes characterized and available on GenBank [11–14].

The complexity of *B. tabaci* might also depend on the inherited bacterial endosymbionts whose functions are not fully understood [15]. The obligatory endosymbionts are essential for insects to live on nutritionally poor diets [16,17]. In whiteflies, the obligatory endosymbionts is *Portiera aleyrodidarum* [13]. In addition, insects often harbor facultative endosymbionts that can play different roles on the vector such as enhancing insecticide susceptibility [18,19], facilitating virus transmission [20,21], conferring tolerance to high-temperature [22] and resistance to natural enemies like parasitic wasps [23]. These bacteria have probably been acquired more recently than obligatory endosymbionts [16,24]. In *B. tabaci*, different facultative symbionts have been described, including *Arsenophonus*, *Hamiltonella*, *Wolbachia*, *Cardinium*, *Fritschea* and *Rickettsia* [8,20]. Among the facultative symbionts, *Hamiltonella defensa* is a maternally transmitted gamma-proteobacterium found sporadically in sap-feeding insects, including aphids, psyllids, and whiteflies [27–29]. Previous studies in Brazil have shown that different species of *B. tabaci,* such as MEAM1, MED and NW2 harbor *Hamiltonella* [10,30]. More recently, a survey has reported that *Hamiltonella* is highly distributed throughout the Brazilian territory and was detected in 89,5% of the MEAM1 specimens and approximately 50% of the MED specimens analyzed [31]. The high incidence of *Hamiltonella* in populations of whiteflies in Brazil may have serious implications to virus transmission. The GroEL proteins encoded by *Hamiltonella* has been found in Israeli *B. tabaci* (MEAM1) populations interacting with the coat protein of begomovirus and therefore facilitating virus transmission [20,24]. The GroEL produced by other symbionts of *B. tabaci* (MEAM1 and MED) did not interact with the virus and therefore were not involved in virus transmission [20]. *Hamiltonella* can play different roles depending on the vector species. In pea aphids, *Acyrthosiphon pisum* (Harris), *Hamiltonella* can block larval development of the solitary endoparasitoid wasps *Aphidius ervi* and *Aphidius eadyi*, rescuing the aphid host [32].

In this study, we sequenced ten transcriptomes of single whitefly specimens from different Brazilian populations and characterized the full mitochondrial genome belonging to three different species (MEAM1, MED and NW2) of the *B. tabaci* complex followed by phylogenetic analysis. In addition, the diversity of the facultative endosymbiont *Hamiltonella* was inferred analyzing 12 different ORF's. The GroEL amino acid sequences of *Hamiltonella* from different *B. tabaci* species were also

analyzed. Our goal was to add further details concerning phylogenetic diversity of mitochondrial genome and *Hamiltonella* among Brazilian populations of *B. tabaci*. In addition, a GroEL protein analysis was carried out that may give further insights about the functions that *Hamiltonella* may have in whitefly biology.

Methods

Whitefly sampling

Samples were obtained from pure colonies and straight from the field. Four different *B. tabaci* populations (153, 154, 156 and 320) were analysed, including two exotics species: Middle East-Asia Minor (MEAM1) and Mediterranean (MED); and a native species: New World 2 (NW2). Populations from colonies were previously identified by sequencing and analysis of the mtCOI gene using the primers C1-J-2195 and TL2-N-3014 [33].

RNA extraction

The RNA extraction of a single individual whitefly was carried out using the ARCTURUS PicoPure kit with modifications [34]. Extracted RNA was subjected to DNase treatment using the TURBO DNA free kit as described by the manufacturer (Ambion Life Technologies CA, USA). Subsequently, the RNA was concentrated using a vacuum centrifuge (Eppendorf, Germany) at 25^0C for one hour. The pellet was resuspended in 18 µl of RNase free water and stored at - 80 C waiting further analysis. Integrity of RNA was quantified by 2100 Bio-analyser (Aligent Technologies).

cDNA and Illumina library preparation

Total RNA from each individual whitefly sample was used for cDNA library preparation using the Illumina TruSeq Stranded Total RNA Preparation kit as described by the manufacturer (Illumina, San Diego, CA, USA). Later on, sequencing of 10 samples was carried out using the HiSeq2000 on a rapid run mode generating 2x50 bp paired end reads. Base calling, quality assessment and image analysis were conducted using the HiSeq control software v1.4.8 and Real Time Analysis v1.18.61 at the Macrogen Korea

Trimming and *de novo* sequence assembly

The raw transcriptome data was trimmed using the software CLC Genomics Workbench v8.5.1 (CLCGW) with quality scores limit set to 0.01, ambiguous limit set to 2. Trimmed reads were then assembled into contigs using *de novo* sequence assembly tool in CLCGW. The assembly parameters consisted of mismatch cost (2), insertion cost (3), deletion cost (3), length fraction (0.5), similarity fraction (0.9) and minimum contig length of either 500 bp or 1000 bp.

Obtaining and analysing complete mitochondrial genomes

The whiteflies complete mitochondrial genome was obtained by mapping of the assembled contigs to reference genomes from GenBank using the software Geneious v9.1.3 [35]. For each *B. tabaci* species of the complex, a different reference mitogenome was used: KU877168 for MEAM1, JQ906700 for MED and AY521259 for NW2. When the mapped contigs did not covering the full length of the mitochondrial genome of the reference sequence, we resorted to mapping trimmed reads to the reference sequence and thus the whole length of the reference was covered. Mapping was performed with the following setting in Geneious software; minimum overlap 10%, minimum overlap identity 80%, allow gaps 10%, fine tuning set to iterate up to 10 times at custom sensitivity. A consensus between the mapped trimmed reads and the reference was used to form new mitochondrial genome. Improvements on the draft mitochondrial genomes were carried out using the software Pilon, a tool for genome assembly improvement [36]. Subsequently, mitochondrial genomes were annotated by MITOS [37]. Other whiteflies mitogenomes sequences were downloaded from GenBank (KJ778614, KX714967, KY951451, KF734668, KR819174, KY951448, JQ906700, KY951447, KU877168, KY951449. KY951450, KY951452 and AY521259) [11,12,14,15,38–40] and added to the analysis. Sequences obtained were aligned using MAFFT v7.309 [41] followed by visualization and analysis in Geneious software. A total of 19 sequences were aligned and analysed.

Analysing *Hamiltonella* genetic diversity

Assembled contigs as well as trimmed reads from each sample were mapped to a *Hamiltonella* reference genome (CP016303) to obtain the facultative endosymbiont draft genome. Mapping was performed with the following setting in Geneious software; minimum overlap 10%, minimum overlap identity 80%, allow gaps 10%, fine tuning set to iterate up to 10 times at custom sensitivity. Afterward, draft

genomes were aligned to the reference using the whole genome alignment tool LASTZ version 1.02.00 [42] within Geneious v. 9.1.8. Genes with full coverage for all the samples were selected for further Bayesian phylogenetic analysis using the software ExaBayes version 1.4.1 [43].

Chaperonin GroEL gene analysis

Trimmed reads were mapped onto the reference sequence for chaperonin GroEL gene from *Hamiltonella* of *B. tabaci* (AF130421). As the mapped contigs did not cover the full length of the coding region of the reference sequence AF130421, we resorted to mapping trimmed reads to the reference sequence to get the full coverage of the whole length from the reference. A consensus between the mapped trimmed reads and the reference was used to form new chaperonin GroEL sequences, open reading frames (ORF) were predicted in Geneious. In addition, other chaperonin GroEL sequence from *A. pisum* was downloaded from Genbank (CP001277) and added to the analysis. Sequences obtained were aligned using MAFFT v7.309 [41] followed by visualization and analysis in Geneious software. A total of 9 sequences were aligned and analysed. In addition, primers were designed (GroEL 1,354 For-CCTC TGCG TCAG ATTG TGGT and GroEL 1,663 Rev – TCAT ACCA TTCA TTCC GCCC A) for a PCR reaction (95°C for 5min, 35 cycles at 95°C for 30 secs, 59.5°C for 30 secs, 72°C for 30 secs and 72°C for 10 min) followed by nucleotide sequencing to confirm the results obtained by NGS.

Bayesian phylogenetic analyses

All the phylogenetic analyses were run on 384 nodes on the Magnus supercomputer (Pawsey Centre, Western Australia). Mitochondrial genome phylogenetic analysis were performed by MrBayes 3.2.2. [44]. Analyses were run for 30 million generations with sampling every 1000 generations. Each analysis consisted of four independent runs, utilizing four coupled Markov chains. The run convergence was assessed by finding the plateau in the likelihood scores (standard deviation of split frequencies < 0.0015). In each of the runs, the first 25% trees were discarded as burn-in and the posterior probability is shown on each node.

In addition, the *Hamiltonella* phylogenetic analysis was performed on DNA sequences of 12 protein-coding genes for a dataset with 9 taxa using ExaBayes

version 1.4.1; [43]. Bayesian analysis was carried out for four independent runs for 1 million generations, with trees sampled every 500 generations. The run convergence was monitored by finding the plateau in the likelihood scores (standard deviation of split frequencies < 0.0015). In each of the runs, the first 25% trees were discarded as burn-in for the estimation of a majority rule consensus topology and posterior probability for each node. Bayesian run files are available from the authors upon request. Trees were visualized, edited and rooted using FigTree v1.4.3.

Results

The sequenced number of reads from the 10 samples ranged from 29,840,288 to 64,080,188 among the ten samples and the number of contigs assembled ranged from 7,730 to 41,165 (Table 1).

Table 1. Next generation sequencing data from single whitefly transcriptomes of *Bemisia tabaci* populations collected in Brazil.

Sample ID	Species	Colony / Open Field	Reference	Number of Reads	Number of Reads After Trimming	CLC minimum contig length	Number of contigs	Contig average length
153_1	MEAM1	Colony	KU877168	36,449,340	36,449,269	Trimmed reads	-	143.9
						500	25,640	1,240
						1000	10,079	2,087
153_2	MEAM1	Colony	N/A	44,093,422	41,041,633	Trimmed reads	-	146.0
						500	31,072	1,228
154_1	MED	Colony	JQ906700	37,206,396	37,206,313	Trimmed reads	-	144.7
						1000	9,737	1,983
154_2	MED	Colony	JQ906700	29,840,288	29,840,226	Trimmed reads	-	144.1
						1000	7,730	1,947
320_1	MED	Open Field	N/A	64,080,188	57,947,512	Trimmed reads	-	145.5

						500	41,165	1,240
320_3	MED	Open Field	JQ906700	39,894,296	39,894,200	Trimmed reads	-	145.1
						1000	9,719	1,969
156_2	NW2	Colony	AY521259	40,826,418	40,826,334	Trimmed reads	-	145.2
						1000	11,961	2,232
156_3	NW2	Colony	AY521259	43,194,264	42,194,175	Trimmed reads	-	145.0
						1000	12,802	2,244
156_4	NW2	Colony	N/A	48,800,318	45,102,719	Trimmed reads	-	145.9
						500	34,090	1,298
156_5	NW2	Colony	N/A	24,568,460	22,491,230	Trimmed reads	-	145.7
						500	21,597	1,310

Hamiltonella genetic diversity

The diversity of *Hamiltonella* in *B. tabaci* populations from Brazil was carried out analysing 12 ORF's from eight single whitefly transcriptomes totalizing an alignment of 6,378bp length. The analysed ORF's were: DNA transformation protein tfoX, porin OmpA, acyl carrier, 50S ribosomal protein L3, 50S ribosomal protein L23, 30S ribosomal protein S17, 50S ribosomal protein L24, 50S ribosomal protein L5, rpsN, nucleotide exchange factor GrpE, porin and DNA-binding protein. The phylogenetic analysis from the sequenced accessions separate into two deeply divergent clades representing the native species from the Americas (NW2) and the exotics (MEAM1 and MED) (Fig 1). Furthermore, the identity percentages among the 12 ORF's from *Hamiltonella* were obtained (S1 Table).

(INSERT FIG 1)

GroEL chaperonin analysis

The GroEL (Chaperonin 60) gene was obtained from six *B. tabaci* single whitefly transcriptomes of Brazil. In addition to the data obtained in this study, the

analysis also included *Hamiltonella* GroEL sequences downloaded from GenBank from the pea aphid *A. pisum* (CP001277) and from *B. tabaci* MEAM1 (CP016303). Analysis of the translated proteins revealed a three amino acids deletion present only in the *B. tabaci* NW2 species and in the pea aphid *A. pisum* (Fig 2). The deletion present in *Hamiltonella* from NW2 and pea aphid was a sequence of two glycine and one isoleucine. The nucleotide deletion was confirmed by PCR, which amplified an 300bp amplicon, followed by nucleotide sequencing.

(INSERT FIG 2)

Mitochondrial genome

Six complete mitochondrial genomes were obtained from three different species found in Brazil (NW2, MEAM1 and MED). Phylogenetic analysis of the complete mitochondrial genomes separated different species of the complex in distinct clades (Fig 3). Furthermore, the identity percentages among all the mitogenomes from *B. tabaci* were obtained (S2 Table).

(INSERT FIG 3)

Discussion

In this study we present the phylogenetic relationship of *B. tabaci* and its association with the facultative endosymbiont *Hamiltonella*. We show the evolutionary differences between native populations of *B. tabaci* species of the Americas and invasive *B. tabaci* species. From our data the genetic differences are not confined to the mtCOI gene, but extend to the rest of the mitogenome and to the facultative endosymbiont *Hamiltonella*. These findings will contribute to the understanding about the functions that *Hamiltonella* may have in *B. tabaci* biology and will aid the correct identification of *B. tabaci* specimens.

Six complete mitochondrial genomes were obtained from three different species of the complex *B. tabaci*: MEAM1, MED and NW2. Complete mitochondrial genomes comparisons are essential for better understanding the evolutionary genetic relationships between members of the *B. tabaci* species complex [3,11]. There are 13 complete whitefly mitochondrial genomes characterized and available on GenBank to date [11,12,14,15,38–40] which is a valuable contribution to a better understanding of

the taxonomy of this global pest. The addition of six more full mitogenomes from *B. tabaci* to the global database will contribute for more accurate identification and will serve as references for further full mitochondrial genome phylogenies studies. The phylogenetic analysis of the full mitogenome (Fig 3) conducted in this study separated the species in different clades in a similar pattern compared to phylogenetic trees of partial fragment of the mtCOI gene [1]. This reinforces that using a partial fragment of the mtCOI gene to infer phylogeny in *B. tabaci* is a reliable way to delimitate the species boundaries and it represents the whole mitochondrial genome.

A multilocus phylogenetic analysis was carried out for the facultative endosymbiont *Hamiltonella*. Previous studies in Brazil phylogenetically analyzed *Hamiltonella* based on 16S rRNA gene and found genetically homogeneous populations of *Hamiltonella* from MEAM1 and MED across Brazil [31]. Another phylogenetic study based on 16S rRNA gene, in Southeast Europe, have grouped *Hamiltonella* from both *B. tabaci* and *T. vaporariorum* in the same clades. Our phylogenetic analysis based on 12 ORF's clustered *Hamiltonella* from NW2 populations in a different clade from MEAM1 and MED populations (Fig 1) suggesting that native populations are infected with a genetically different *Hamiltonella* from invasive populations, reinforcing the inexistence of endosymbiont horizontal transmission between invasive and native species.

The differences of *Hamiltonella* from native and exotic species extends to the GroEL gene. Our analysis revealed an amino acid deletion of two glycine and one isoleucine in the GroEL gene present only in *Hamiltonella* from native populations (Fig 2). Glycine is a non-essential amino acid and isoleucine is an essential amino acid for *Hamiltonella* [45]. The absence of these amino acids could be affecting the confirmation of GroEL on populations of *Hamiltonella* presenting this deletion. Thus, this deletion in *Hamiltonella* could imply in biological changes in populations of whiteflies harbouring this facultative endosymbiont.

It's known that facultative bacterial endosymbionts are associated with viral transmission [20,21] and different members of the *B. tabaci* species complex may transmit the same viruses with different efficiencies [46]. The facultative endosymbiont *Hamiltonella* has been identified as a key driver in the transmission of begomoviruses [24]. *Hamiltonella* encodes a GroEL chaperonin homologue protein that is crucial in

safeguarding begomoviruses in the haemolymph [24]. Several studies have shown the interaction between the begomoviruses *Tomato yellow leaf curl virus* (TYLCV) coat protein (CP) and GroEL present in the haemolymph of *B. tabaci* [20,47]. Disturbing the association GroEL-TYLCV in vivo by feeding insects with an antibody raised against *Buchnera* GroEL leads to the degradation of the virus and to a markedly decrease in transmission efficiency of the virus [24,47,48].

There is still a lack of knowledge regarding the interactions among vector, symbionts and whitefly-transmitted viruses within Brazilian populations. Previous surveys conducted in Brazil have found that field populations of MEAM1, MED and NW2 analyzed frequently harbor *Hamiltonella* [28, 41]. Therefore, it would be important to find out if there is a relationship between the diversity of *H. defensa* found in the current study and biological features of field populations of *B. tabaci*. The existence of biological and behavioral differences between the native species New World and the exotic MEAM1 have been reported already [5,49]. The New World *B. tabaci* are found more often colonizing weeds and wild plants [5]. In addition, previous transmission studies comparing NW2 and MEAM1 species, both harboring *Hamiltonella*, have found a different transmission efficiency of Brazilian whitefly-transmitted viruses between the species [49]. It was found that the begomovirus *Euphorbia yellow mosaic virus* (EuYMV) is transmitted more efficiently by NW2 compared to MEAM1 and that the crinivirus, *Tomato chlorosis virus* (ToCV) and the carlavirus, *Cowpea mild mottle virus* (CpMMV) are transmitted more efficiently by MEAM1 than NW2 [49]. Interestingly, each of these viruses has a different mode of transmission by the whitefly vector. The begomoviruses, EuYMV is transmitted in a persistent manner, the crinivirus, ToCV is transmitted in a semipersistent manner and the carlavirus, CpMMV is transmitted in a non-persistent mode [50,51], which is a very unusual mode of transmission among *B. tabaci* transmitted viruses.

The reasons for a difference in transmission efficiencies by NW2 and MEAM1 are still unknown but might be related to several factors related to the host, the viruses or other facultative endosymbiont found in the vector. However, the data found in this study of phylogenetic differences between insect symbiotic *Hamiltonella* from NW2 and MEAM1 added to the deletion of three amino acids present in the homologue GroEL Chaperonin protein might aid to explain the difference in the transmission efficiencies among native and exotic species present in Brazil if further studies were carried out.

The genomic data obtained in this study from the facultative endosymbionts, *Hamiltonella* and the complete mitochondrial of Brazilian *B. tabaci* populations is unprecedented and essential to serve as a reference for further studies regarding whiteflies in Brazil. The phylogenetic and amino acid analysis revealed genetic diversity between *Hamiltonella* from native and exotic populations that will aid for better understanding about the functions that *Hamiltonella* may have in whitefly biology.

Acknowledgements

The analyses were carried out using computational facilities at Pawsey Supercomputing Centre who are supported by the Australian Government and the Government of Western Australia.

References

1. De Barro PJ, Liu S-S, Boykin LM, Dinsdale AB. *Bemisia tabaci* : A Statement of Species Status. Annu Rev Entomol. Annual Reviews ; 2011;56: 1–19. doi:10.1146/annurev-ento-112408-085504

2. Navas-Castillo J, Fiallo-Olivé E, Sánchez-Campos S. Emerging Virus Diseases Transmitted by Whiteflies. Annu Rev Phytopathol. Annual Reviews ; 2011;49: 219–248. doi:10.1146/annurev-phyto-072910-095235

3. Boykin, Boykin, Karen Armstrong, Laura Kubatko, Paul De Barro. Species Delimitation and Global Biosecurity. Evol Bioinforma. 2011; 1. doi:10.4137/EBO.S8532

4. da Fonseca Barbosa L, Yuki VA, Marubayashi JM, De Marchi BR, Perini FL, Pavan MA, et al. First report of Bemisia tabaci Mediterranean (Q biotype) species in Brazil. Pest Manag Sci. 2015;71. doi:10.1002/ps.3909

5. Barbosa LFF, Marubayashi JM, De Marchi BR, Yuki VA, Pavan MAMA, Moriones E, et al. Indigenous American species of the Bemisia tabaci complex are still widespread in the Americas. Pest Manag Sci. 2014;70: 1440–1445. doi:10.1002/ps.3731

6. Marubayashi JM, Yuki VA, Rocha KCG, Mituti T, Pelegrinotti FM, Ferreira FZ, et al. At least two indigenous species of the *Bemisia tabaci* complex are

present in Brazil. J Appl Entomol. Blackwell Publishing Ltd; 2013;137: 113–121. doi:10.1111/j.1439-0418.2012.01714.x

7. Gauthier N, Clouet C, Perrakis A, Kapantaidaki D, Peterschmitt M, Tsagkarakou A. Genetic structure of *Bemisia tabaci* Med populations from home-range countries, inferred by nuclear and cytoplasmic markers: impact on the distribution of the insecticide resistance genes. Pest Manag Sci. 2014;70: 1477–1491. doi:10.1002/ps.3733

8. Hu J, De Barro P, Zhao H, Wang J, Nardi F, Liu S-S. An Extensive Field Survey Combined with a Phylogenetic Analysis Reveals Rapid and Widespread Invasion of Two Alien Whiteflies in China. Roberts RG, editor. PLoS One. Public Library of Science; 2011;6: e16061. doi:10.1371/journal.pone.0016061

9. Xu J, De Barro PJ, Liu SS. Reproductive incompatibility among genetic groups of Bemisia tabaci supports the proposition that the whitefly is a cryptic species complex. Bull Entomol Res. 2010;100: 359–366. doi:10.1017/S0007485310000015

10. Aparecida de Moraes L, Marubayashi Massaharu J, Yuki Atsushi V, Ghanim M, Bello VHVH, Rossitto De Marchi B, et al. New invasion of Bemisia tabaci Mediterranean species in Brazil associated to ornamental plants. Phytoparasitica. Springer Netherlands; 2017;45: 1–9. doi:10.1007/s12600-017-0607-9

11. Tay WT, Elfekih S, Court L, Gordon KH, De Barro PJ. Complete mitochondrial DNA genome of Bemisia tabaci cryptic pest species complex Asia I (Hemiptera: Aleyrodidae). Mitochondrial DNA. 2014;1736: 1–2. doi:10.3109/19401736.2014.926511

12. Wang H-L, Yang J, Boykin LM, Zhao Q-Y, Li Q, Wang X-W, et al. The characteristics and expression profiles of the mitochondrial genome for the Mediterranean species of the Bemisia tabaci complex. BMC Genomics. 2013;14: 401. doi:10.1186/1471-2164-14-401

13. Thao ML, Baumann P. Evolutionary relationships of primary prokaryotic

endosymbionts of whiteflies and their hosts. Appl Environ Microbiol. American Society for Microbiology; 2004;70: 3401–6. doi:10.1128/AEM.70.6.3401-3406.2004

14. Tay WT, Elfekih S, Polaszek A, Court LN, Evans GA, Gordon KHJ, et al. Novel molecular approach to define pest species status and tritrophic interactions from historical Bemisia specimens. Sci Rep. 2017;7: 429. doi:10.1038/s41598-017-00528-7

15. Parrella G, Nappo AG, Manco E, Greco B, Giorgini M. Invasion of the Q2 mitochondrial variant of Mediterranean *Bemisia tabaci* in southern Italy: possible role of bacterial endosymbionts. Pest Manag Sci. 2014;70: 1514–1523. doi:10.1002/ps.3686

16. Baumann P. Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. Annu Rev Microbiol. 2005;59: 155–189. doi:10.1146/annurev.micro.59.030804.121041

17. Ferrari J, Vavre F. Bacterial symbionts in insects or the story of communities affecting communities. Philos Trans R Soc Lond B Biol Sci. The Royal Society; 2011;366: 1389–400. doi:10.1098/rstb.2010.0226

18. Kontsedalov S, Zchori-Fein E, Chiel E, Gottlieb Y, Inbar M, Ghanim M. The presence of *Rickettsia* is associated with increased susceptibility of *Bemisia tabaci* (Homoptera: Aleyrodidae) to insecticides. Pest Manag Sci. 2008;64: 789–792. doi:10.1002/ps.1595

19. Kontsedalov S, Gottlieb Y, Ishaaya I, Nauen R, Horowitz R, Ghanim M. Toxicity of spiromesifen to the developmental stages of *Bemisia tabaci* biotype B. Pest Manag Sci. 2009;65: 5–13. doi:10.1002/ps.1636

20. Gottlieb Y, Zchori-Fein E, Mozes-Daube N, Kontsedalov S, Skaljac M, Brumin M, et al. 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. J Virol. American Society for Microbiology; 2010;84: 9310–7. doi:10.1128/JVI.00423-10

21. Rana VS, Singh ST, Priya NG, Kumar J, Rajagopal R. Arsenophonus GroEL

Interacts with CLCuV and Is Localized in Midgut and Salivary Gland of Whitefly B. tabaci. Oliveira PL, editor. PLoS One. Public Library of Science; 2012;7: e42168. doi:10.1371/journal.pone.0042168

22. Brumin M, Kontsedalov S, Ghanim M. Rickettsia influences thermotolerance in the whitefly Bemisia tabaci B biotype. Insect Sci. Blackwell Publishing Asia; 2011;18: 57–66. doi:10.1111/j.1744-7917.2010.01396.x

23. Mahadav A, Gerling D, Gottlieb Y, Czosnek H, Ghanim M. Parasitization by the wasp Eretmocerus mundus induces transcription of genes related to immune response and symbiotic bacteria proliferation in the whitefly Bemisia tabaci. BMC Genomics. 2008;9: 342. doi:10.1186/1471-2164-9-342

24. Gorovits R, Czosnek H. Insect Symbiotic Bacterial GroEL (Chaperonin 60) and Plant Virus Transmission. In: Mande SC, Kumar CMS, Sharma A, editors. Moonlighting Cell Stress Proteins in Microbial Infections. 2013. pp. 173–187. doi:10.1007/978-94-007-6787-4

25. Kanakala S, Ghanim M. Advances in the Genomics of the Whitefly Bemisia tabaci: An Insect Pest and a Virus Vector. Springer International Publishing; 2015. pp. 19–40. doi:10.1007/978-3-319-24235-4_2

26. Bing X-L, Yang J, Zchori-Fein E, Wang X-W, Liu S-S. Characterization of a newly discovered symbiont of the whitefly Bemisia tabaci (Hemiptera: Aleyrodidae). Appl Environ Microbiol. American Society for Microbiology; 2013;79: 569–75. doi:10.1128/AEM.03030-12

27. Rao Q, Wang S, Su YL, Bing XL, Liu SS, Wang XW. Draft genome sequence of “Candidatus Hamiltonella defensa,” an endosymbiont of the whitefly Bemisia tabaci [Internet]. Journal of Bacteriology. 2012. p. 3558. doi:10.1128/JB.00069-12

28. Luan J-B, Shan H-W, Isermann P, Huang J-H, Lammerding J, Liu S-S, et al. Cellular and molecular remodelling of a host cell for vertical transmission of bacterial symbionts. Proc R Soc B Biol Sci. 2016;283: 20160580. doi:10.1098/rspb.2016.0580

29. Luan J, Sun X, Fei Z, Douglas AE. Maternal Inheritance of a Single Somatic

Animal Cell Displayed by the Bacteriocyte in the Whitefly Bemisia tabaci. Curr Biol. 2018;28: 459–465.e3. doi:10.1016/j.cub.2017.12.041

30. Marubayashi JM, Kliot A, Yuki VA, Rezende JAM, Krause-Sakate R, Pavan MA, et al. Diversity and Localization of Bacterial Endosymbionts from Whitefly Species Collected in Brazil. Bourtzis K, editor. PLoS One. Public Library of Science; 2014;9: e108363. doi:10.1371/journal.pone.0108363

31. De Moraes Aparecida L, Muller C, Oliveira de Freitas Bueno RC, Santos A, Bello VH, Rossitto De Marchi B, et al. Distribution and phylogenetics of whiteflies in Brazil and their endosymbiont relationships.

32. Su Q, Oliver KM, Pan H, Jiao X, Liu B, Xie W, et al. Facultative Symbiont <I>Hamiltonella</I> Confers Benefits to <I>Bemisia tabaci</I> (Hemiptera: Aleyrodidae), an Invasive Agricultural Pest Worldwide. Environ Entomol. 2013;42: 1265–1271. doi:10.1603/EN13182

33. Frohlich DR, Torres-Jerez I, Bedford ID, Markham PG, Brown JK. A phylogeographical analysis of the Bemisia tabaci species complex based on mitochondrial DNA markers. Mol Ecol. 1999;8: 1683–1691. doi:10.1046/j.1365-294x.1999.00754.x

34. Sseruwagi P, Wainaina J, Ndunguru J, Tumuhimbise R, Tairo F, Guo J-Y, et al. The first transcriptomes from field-collected individual whiteflies (Bemisia tabaci, Hemiptera: Aleyrodidae). Gates Open Res. 2017;1: 16. doi:10.12688/gatesopenres.12783.1

35. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28: 1647–1649. doi:10.1093/bioinformatics/bts199

36. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. Wang J, editor. PLoS One. Public Library of Science; 2014;9: e112963. doi:10.1371/journal.pone.0112963

37. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al.

MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013;69: 313–319. doi:10.1016/j.ympev.2012.08.023

38. Tay WT, Elfekih S, Court LN, Gordon KHJ, Delatte H, De Barro PJ. The Trouble with MEAM2: Implications of Pseudogenes on Species Delimitation in the Globally Invasive Bemisia tabaci (Hemiptera: Aleyrodidae) Cryptic Species Complex. Genome Biol Evol. Oxford University Press; 2017;9: 2732–2738. doi:10.1093/gbe/evx173

39. Wang H-L, Zhang Z, Bing X-L, Liu Y-Q, Liu S-S, Wang X-W. A complete mitochondrial DNA genome derived from a Chinese population of the *Bemisia afer* species complex (Hemiptera: Aleyrodidae). Mitochondrial DNA Part A. 2016;27: 3500–3501. doi:10.3109/19401736.2015.1066367

40. Wang H-L, Xiao N, Yang J, Wang X-W, Colvin J, Liu S-S. The complete mitochondrial genome of *Bemisia afer* (Hemiptera: Aleyrodidae). Mitochondrial DNA. 2016;27: 98–99. doi:10.3109/19401736.2013.873921

41. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. Oxford University Press; 2002;30: 3059–66. Available: http://www.ncbi.nlm.nih.gov/pubmed/12136088

42. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, et al. Human – Mouse Alignments with BLASTZ. Genome Res. 2003;13: 103–107. doi:10.1101/gr.809403.

43. Aberer AJ, Kobert K, Stamatakis A. ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era. Mol Biol Evol. Oxford University Press; 2014;31: 2553–2556. doi:10.1093/molbev/msu236

44. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, H?hna S, et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol. 2012;61: 539–542. doi:10.1093/sysbio/sys029

45. Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA. Hamiltonella defensa, genome evolution of protective bacterial endosymbiont from pathogenic

ancestors. Proc Natl Acad Sci. 2009;106: 9063–9068. doi:10.1073/pnas.0900194106

46. Polston JE, De Barro P, Boykin LM. Transmission specificities of plant viruses with the newly identified species of the *Bemisia tabaci* species complex. Pest Manag Sci. John Wiley & Sons, Ltd; 2014;70: 1547–1552. doi:10.1002/ps.3738

47. Morin S, Ghanim M, Sobol I, Czosnek H. The GroEL Protein of the Whitefly Bemisia tabaci Interacts with the Coat Protein of Transmissible and Nontransmissible Begomoviruses in the Yeast Two-Hybrid System. Virology. 2000;276: 404–416. doi:10.1006/viro.2000.0549

48. Morin S, Ghanim M, Zeidan M, Czosnek H, Verbeek M, van den Heuvel JFJM. A GroEL Homologue from Endosymbiotic Bacteria of the WhiteflyBemisia tabaciIs Implicated in the Circulative Transmission of Tomato Yellow Leaf Curl Virus. Virology. 1999;256: 75–84. doi:10.1006/viro.1999.9631

49. De Marchi BR, Marubayashi JM, Favara GM, Yuki VA, Watanabe LFLFM, Barbosa LFLF, et al. Comparative transmission of five viruses by Bemisia tabaci NW2 and MEAM1. Trop Plant Pathol. Tropical Plant Pathology; 2017;1. doi:10.1007/s40858-017-0186-9

50. Marubayashi JM, Yuki VA, Wutke EB. Transmission of the Cowpea mild mottle virus by whitefly Bemisia tabaci biotype B for plants of beans and soy. Summa Phytopathol. Grupo Paulista de Fitopatologia; 2010;36: 158–160. doi:10.1590/S0100-54052010000200009

51. Zanardo LG, Carvalho CM. Cowpea mild mottle virus (Carlavirus, Betaflexiviridae): a review. Trop Plant Pathol. Tropical Plant Pathology; 2017; 417–430. doi:10.1007/s40858-017-0168-y

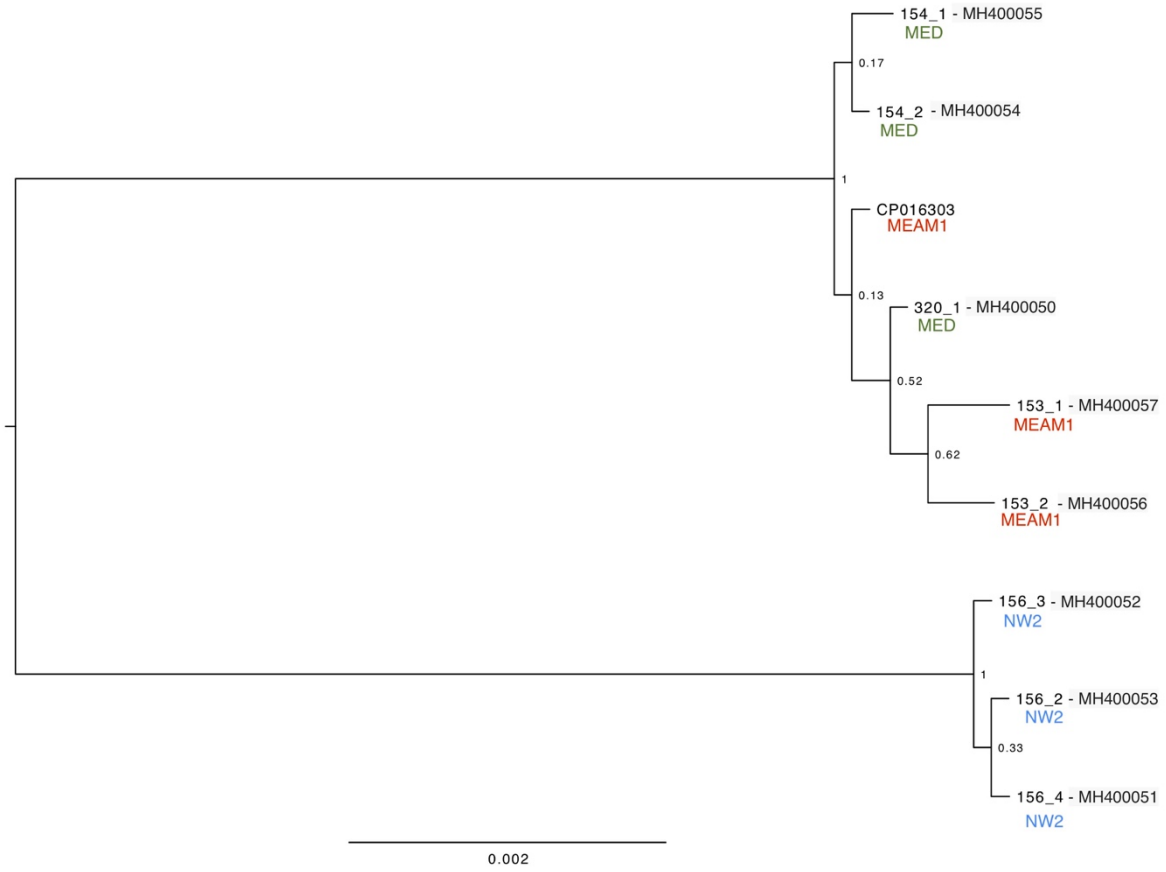


Fig 1. Phylogenetic analysis from 12 ORF's of *Hamiltonella* from Brazilian populations of *Bemisia tabaci.* Analysis was carried out using the software ExaBayes version 1.4.1. MEAM1, Middle East-Asia Minor-1; MED, Mediterranean; NW2, New World 2.

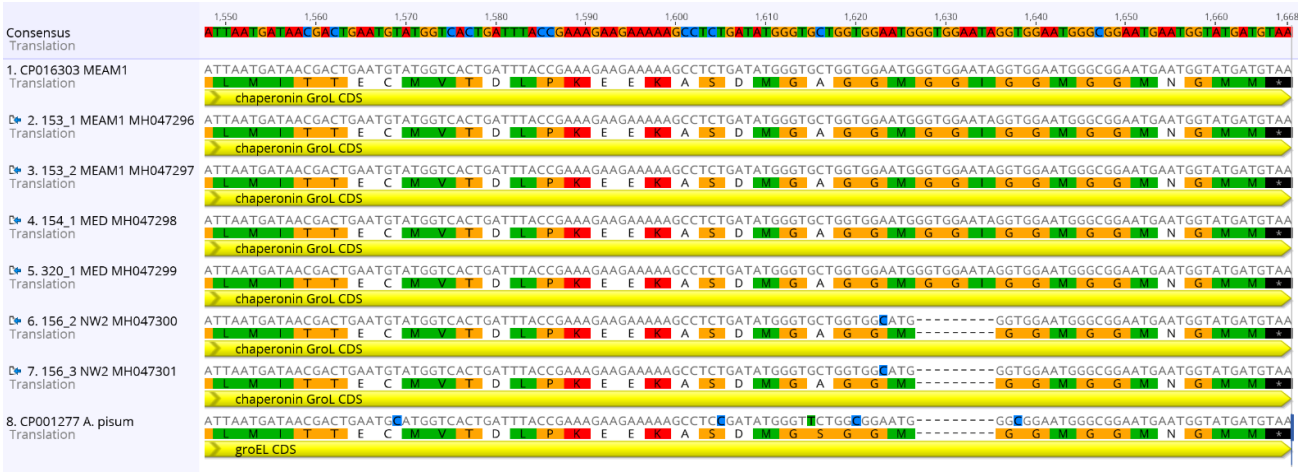


Fig 2. The chaperonin GroEL protein analysis of the *Hamiltonella* endosymbiont. Analysis was visualized on Geneious v9.1.3 and revealed a three amino acids deletion only in the *Bemisia tabaci* New World 2 species and the pea aphid *Acyrthosiphon pisum*.

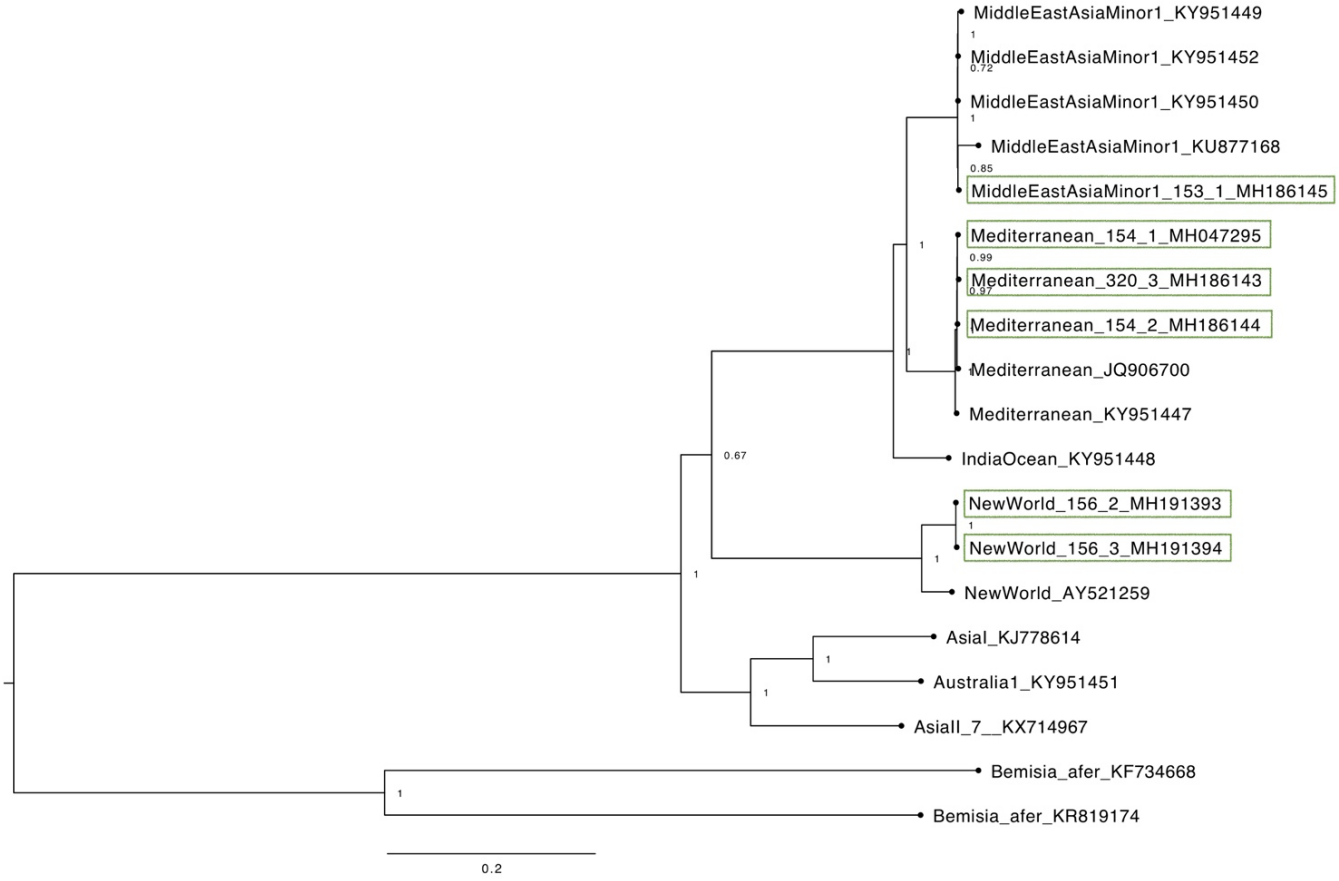


Fig 3. Phylogenetic analysis of the complete mitochondrial genome of whiteflies populations from Brazil obtained from single whitefly transcriptomes. The alignment totalized 19 samples and 16073bp length including references from GenBank. Brazilian samples obtained in this study were highlighted in green.

CAPÍTULO 2

Genetic diversity and SNP's from the chloroplast coding regions of virus-infected cassava (*Manihot esculenta* Crantz)

Bruno Rossitto De Marchi[1-2], Tonny Kinene[2], James Mbora Wainaina[2], Renate Krause-Sakate[1], Laura M. Boykin[2*]; Joseph Ndunguru[3], Titus Alicai, Monica Kehoe[2], Elijah Ateka[4], Fred Tairo[3], Jamisse Amisse, Peter Sseruwagi[3]

[1] UNESP – Universidade Estadual Paulista, Faculdade de Ciências Agronômicas, Botucatu-SP, Brazil

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

[3] Mikocheni Agricultural Research Institute, Sam Nujoma Road, Box 6226, Dar es Salaam, Tanzania

[4] Department of Horticulture, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

*Corresponding author

E-mail: laura.boykin@uwa.edu.au

Abstract

Cassava is a staple food crop in sub-Saharan Africa, as is a rich source of carbohydrates and proteins which currently supports livelihoods of more than 800 million people worldwide. However, its continued abundance is at stake due to vector-transmitted diseases such as Cassava mosaic disease (CMD) and Cassava brown streak disease (CBSD). Currently, the management and control of viral diseases in cassava relies mainly on tolerant cultivars. Thus, new target genes for plant virus resistance are essential for the development of virus-resistant cassava by either conventional breeding or genetic engineering. The chloroplast is a common target of plant viruses for viral pathogenesis or propagation; and conversely, chloroplast and its components also can play active roles in plant defense against viruses. Non-infected and infected cassava samples were obtained from different locations in Tanzania,

Kenya and Mozambique. RNA extraction followed by cDNA library preparation and Illumina sequencing was performed. Assembling and mapping of the reads was carried out and 33 cassava partial chloroplast genomes were obtained. Bayesian phylogenetic analysis of 62 chloroplast protein-coding genes for a dataset with 39 taxa was performed and SNPs identified for the chloroplast dataset. Phylogenetic analysis revealed a genetic diversity present in cassava chloroplast partial genome among cassava varieties of East Africa. The results should be incorporated into identifying resistances genes to different viruses for breeding and obtaining new cassava varieties resistant to viruses.

Introduction

Cassava, *Manihot esculenta* Crantz subsp. *esculenta* (Euphorbiaceae) is a perennial woody shrub also known as manioc or tapioca. Cassava is cultivated for its starchy storage roots throughout tropical and subtropical regions of the world and provides staple food for over 800 million people in Africa (51%), Asia (29%) and South America (20%) (http://faostat.fao.org). Africa now produces more cassava than the rest of the world and it is one of the continent's staple food crops (HILLOCKS; THRESH, 2002). It is highly tolerant to drought and its storage roots can be preserved in soil for a few years, cassava is considered to be an important reserve of carbohydrates to relieve global famine (RAHEEM; CHUKWUMA, 2001). The use of stem cuttings as the usual means of propagating cassava exposes the crop to the risk of proliferating virus disease problems. As a result of the widespread use of unselected planting material, cassava's continued abundance is threatened in all major areas of production in Africa by virus diseases such as Cassava mosaic disease (CMD) and Cassava brown streak disease (CBSD), which together were reported to cause production losses of more than US$3 billion every year in Africa (HILLOCKS; JENNINGS, 2003; LEGG et al., 2006). CMD occurs throughout Africa and several different geminiviruses have been shown to be responsible (e.g. *African cassava mosaic virus*, *East African cassava mosaic virus*, *East African cassava mosaic virus* – Uganda) (LEGG et al., 2006). CBSD was thought to be caused by two distinct virus species of Ipomoviruses; *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV). However, recent findings based on analysis of 470 symptomatic leaf samples collected from Tanzania seem to suggest that up to four different species of cassava brown streak viruses (CBSVs) could be associated with CBSD (NDUNGURU et al., 2015).

Incidences of CBSD are particularly high in the coastal areas of Tanzania and Mozambique, where cassava is the main staple food (HILLOCKS; JENNINGS, 2003). These viruses are transmitted by the whitefly, *Bemisia tabaci* (Gennadius) which is one of the most serious pests of vegetable, fibre and ornamental crops worldwide and has been listed as one of the world's 100 worst invasive alien species (LOWE S., BROWNE M., BOUDJELAS S., 2000). In East Africa, high populations of whiteflies are visible in cassava crops and are causing tremendous yield losses (HILLOCKS; JENNINGS, 2003). However, a recent study suggests that aphids may play a role in the transmission of CBSV as well, due to an occurrence of a highly-conserved DAG motif which is associated to the aphid vector (ATEKA et al., 2017). Currently, the management and control of viral diseases in cassava relies mainly on tolerant cultivars. The main goal of the breeders is to obtain cultivars resistant to both diseases. Although high resistance for CMD has been found, only limited success has been reported for CBSD (TUMWEGAMIRE et al., 2018).

Thus, new target genes for plant virus resistance are essential for the development of virus-resistant cassava by either conventional breeding or genetic engineering. Chloroplast is a common target of plant viruses for viral pathogenesis or propagation; and conversely, chloroplast and its components also can play active roles in plant defense against viruses (DE TORRES ZABALA et al., 2015). Consequently, this organelle is a potential source for discovering new genes against invading pathogens such as plant viruses. A series of typical changes followed by chlorotic symptoms imply the occurrence of chloroplast-virus interactions. (DARDICK, 2007; MOCHIZUKI et al., 2014; ZHAO et al., 2016). Many plant viruses have as a signature of infection the attachment of the viral particle to the chloroplast membrane which affects large numbers of chloroplast and photosynthesis related genes (BHATTACHARYYA; CHAKRABORTY, 2018; WEI et al., 2010). In other words, damage to the chloroplast is a fundamental step in successful infection for plant viruses. Indeed, more and more chloroplast factors have been identified to interact with viral components. These factors are involved in virus replication, movement, symptoms or plant defense, suggesting that viruses have evolved to interact with chloroplast (ZHAO et al., 2016). Currently, over 800 complete chloroplast genome sequences have been made available in the National Center for Biotechnology Information (NCBI) organelle genome database, including cassava. The complete sequence of the

chloroplast genome of cassava is 161,453 bp in length and includes a pair of inverted repeats (IR) of 26,954 bp. The genome includes 128 genes; 96 are single copy and 16 are duplicated in the IR. There are four rRNA genes and 30 distinct tRNAs, seven of which are duplicated in the IR (DANIELL et al., 2008). Complete chloroplast genome sequences are valuable for deciphering phylogenetic relationships between closely related taxa and for improving our understanding of the evolution of plant species (DANIELL et al., 2016). A key application of the chloroplast genome in agriculture is the identification of commercial cultivars and the determination of their purity (DANIELL et al., 2016). The goal of this study was to evaluate the genetic diversity of cassava chloroplast coding regions from cassava plants naturally infected with viruses from different varieties collected from the field in East Africa. Other sequences from GenBank, including cassava wild relatives, were added to the phylogenetic analysis. The results might help to correctly identify varieties commonly used in East Africa and aid breeders to obtain new cassava varieties resistant to viruses.

Methods

Cassava sampling

The dataset consisted of 33 cassava plants including Tanzanian, Kenyan and Mozambican plants. In addition, six cassava and cassava relatives chloroplast reference genomes from GenBank have been added to the final dataset (EU117376, SRR2847471, SRR2847474, SRR2847450, SRR2847419 and SRR2847403) (DANIELL et al., 2008; WANG et al., 2014). Field samples collection was performed differently for each batch of samples. Tanzanian samples were collected in cassava fields (3–6 months old) in the major cassava growing zones of the country (Coast, Lake, Southern, Zanzibar and Mafia Islands) inspected for CBSD symptoms (NDUNGURU et al., 2015). Kenyan samples were collected in four regions of Kenya, namely, Nyanza, Coast, Western and Eastern, between May and October 2013. The crops were 6–12 months old, a growth stage of cassava when CBSD above-ground symptoms are clearly visible (ATEKA et al., 2017).

RNA extraction and Next Generation Sequencing

RNA extraction was performed from approximately 100mg of cassava leaf as described by the CTAB (cetyltrimethyl ammonium bromide) protocol (LODHI et al., 1994).

Total RNA quality check was performed followed by cDNA libraries preparation using the IlluminaTruSeq Stranded Total RNA Sample Preparation kit (Illumina, San Diego, California). Paired end reads were generated using the Illumina MiSeq System. The methods used for Next Generation Sequencing were previously described (ATEKA et al., 2017; NDUNGURU et al., 2015).

De novo Sequence Assembly and mapping to chloroplast genome

The raw transcriptome data was trimmed and assembled using the software CLC Genomics Workbench v9.5.2, as previously described (ATEKA et al., 2017; NDUNGURU et al., 2015). Assembled contigs were mapped to an existent cassava chloroplast genome from GenBank (EU117376) (DANIELL et al., 2008) cultivar TME3 and merged to produce a single draft genome using Geneious v9.1.3 (KEARSE et al., 2012). Assemblies were refined by repeatedly mapping trimmed raw reads to the draft sequencing and adjusting as necessary. Mapping was performed with the following setting in Geneious software; minimum overlap 10%, minimum overlap identity 80%, allow gaps 10%, fine tuning set to iterate up to 10 times at custom sensitivity. A consensus between the mapped trimmed reads and the reference was used to form new draft chloroplast genomes.

Draft chloroplast genomes were annotated using CpGAVAS (LIU et al., 2012). Subsequently, pairwise chloroplast alignment among the 39 cassava chloroplast genomes was performed using MAFFT v7.222 (KATOH et al., 2002) within the Geneious software. Further annotations adjustment was performed by direct comparison with the reference genome (EU117376). Finally, the alignment was refined removing regions and genes with gaps or artefacts for further phylogeny reconstruction.

Virus identification

For each sample, assembled contigs were sorted by length and the longest subjected to a BLAST search (blastn and blastx) (ALTSCHUL et al., 1990). In addition, reads were also imported into Geneious v9.1.3 (KEARSE et al., 2012) and provided with reference sequences obtained from Genbank (NC012698 for CBSV, GQ329864 for CBSV-T and NC014791 for UCBSV). Mapping was performed with minimum

overlap 10%, minimum overlap identity 80%, allow gaps 10% and fine tuning set to iterate up to 10 times.

Bayesian phylogenetic analyses

Phylogenetic Bayesian analyses were performed on DNA sequences of 62 protein-coding genes for a dataset with 39 taxa using ExaBayes version 1.4.1; (ABERER; KOBERT; STAMATAKIS, 2014) and were run in parallel across 384 nodes on the Magnus supercomputer (located at the Pawsey Centre, Western Australia). Analyses were run for 1 million generations with sampling every 500 generations. Each analysis consisted of four independent runs, each utilizing four coupled Markov chains. The run convergence was monitored by finding the plateau in the likelihood scores (standard deviation of split frequencies < 0.0015). The first 25% of each run was discarded as burn-in for the estimation of a majority rule consensus topology and posterior probability for each node. Bayesian run files are available from the authors upon request.

Results and discussion

In this study, we obtained draft chloroplast genomes from 33 cultivated cassava and wild relative species belonging to three different countries in East Africa. Most of the plants analysed were naturally infected by the two main viral diseases affecting cassava production in Africa: CBSD and CMD. It was verified a low genetic diversity of the chloroplast coding regions analysed among cultivated cassava indicating the need of enhancing the diversity of germplasm available. In addition, there was no direct relationship between chloroplast coding regions and determinate virus species.

Illumina sequencing of libraries prepared from total DNA produced between 2,071,164 and 23,427,360 paired-end reads with a maximum sequence length of 100 nucleotides and minimum of 30 nucleotides for Kenyan samples, maximum sequence length of 300 nucleotides and minimum of 100 nucleotides for Tanzanian samples and 100 nucleotide sequence length for the healthy sample.

Table 1. Next generation sequencing data and additional information from cassava samples collected in Tanzania, Kenya, Mozambique and accessions obtained from GenBank.

Country/ Sample ID	Species/ Cultivar name	District/ Region	Virus species	No. of reads obtained	No. of reads after trimming	GenBank Accession	Reference	Additional Information
Tanzania 03 (Tz: Ser-6)	Rumara	N/A	CBSV	4,988,502	3,936,531	N/A	This study	N/A
Tanzania 04 (Tz: Tan 19-1)	Kibandamen o	N/A	CBSV	2,735,840	2,279,998	N/A	This study	N/A
Tanzania 05 (Tz: Tan 23)	Mkunungu	N/A	UCBSV	4,978,272	4,115,933	N/A	This study	N/A
Tanzania 06 (Tz: Tan 26)	Kibandamen o	N/A	CBSV	2,732,618	2,249,619	N/A	This study	N/A
Tanzania 07	Unknown	N/A	Not detected	3,500,480	3,031,625	N/A	This study	N/A
Tanzania 08	Unknown	N/A	Not detected	5,719,724	4,387,729	N/A	This study	N/A
Tanzania 09 (Tz: Nya-38)	Mkangawan du	N/A	CBSV	2,560,544	2,126,102	N/A	This study	N/A
Tanzania 10 (Tz Mf-49)	Kibembe	N/A	CBSV	2,088,040	1,752,853	N/A	This study	N/A
Tanzania 11 (Tz: Maf-51)	Kibembe	N/A	UCBSV	2,071,164	1,852,435	N/A	This study	N/A
Tanzania 12 (Tz: Maf-58)	Mwarabu	N/A	UCBSV	2,548,594	2,030,993	N/A	This study	N/A
Kenya 01	Local	Bumula (Western)	CBSV & UCBSV	15,573,022	15,282,670	N/A	This study	N/A
Kenya 02	Magana	Bumula (Western)	CBSV-Tz	18,381,178	17,703,620	N/A	This study	N/A

Kenya 03	Tereka	Bumula (Western)	N/A	17,711,474	17,024,115	N/A	This study	N/A
Kenya 04	Magana	Teso (Western)	UCBSV	22,509,994	21,799,542	N/A	This study	N/A
Kenya 05	Magana	Busia (Western)	UCBSV & CBSV	22,191,388	21,848,109	N/A	This study	N/A
Kenya 07	Local	Bondo (Nyanza)	EACMV - Ug	21,488,968	21,240,329	N/A	This study	N/A
Kenya 10	Local	Bondo (Nyanza)	EACMV - Ug	23,427,360	23,129,332	N/A	This study	N/A
Kenya 11	Kibandamen o	Malindi (Coast)	EACMZV, EACMV-Ke, CBSV, UCBSV	19,856,998	19,163,125	N/A	This study	N/A
Kenya 12	Local	Malindi (Coast)	UCBSV & EACMV - Ke	19,187,524	18,865,588	N/A	This study	N/A
Kenya 13	Kibandamen o	Msambwe ni (Coast)	UCBSV & CBSV - Kilifi (Kenya)	21,512,336	21,261,827	N/A	This study	N/A
Kenya 14	Kibandamen o	Msambwe ni (Coast)	CBSV & EACMV - Ke, UCBSV	22,057,126	21,755,991	N/A	This study	N/A
Kenya 15	Kibandamen o	Msambwe ni (Coast)	CBSV	22,124,706	21,851,119	N/A	This study	N/A
Tanzania_DRJL030	Unknown	N/A	Healthy	7,329,128	6,584,936	N/A	This study	Tall cassava - Healthy
Mozambique 4	Ezalamalithi	Namapa	CBSV	N/A	N/A	N/A	This study	N/A
Mozambique 5	Calamidade	Namapa	CBSV	N/A	N/A	N/A	This study	N/A
Mozambique 8	Calamidade	Nampula	CBSV	N/A	N/A	N/A	This study	N/A
Mozambique 10	Buana	Gile District	Not detected	N/A	N/A	N/A	This study	N/A

Mozambique 11	Fernando	Alto Molocue District	Not detected	N/A	N/A	N/A	This study	N/A
Mozambique 16	Bwana	Alto Molocue	CBSV	N/A	N/A	N/A	This study	N/A
Mozambique 17	Mulaleia	Mocuba District	Not detected	N/A	N/A	N/A	This study	N/A
Mozambique 20	Cadri	Mocuba	CBSV	N/A	N/A	N/A	This study	N/A
Mozambique 23	Mulaleia	Quelimane	CBSV	N/A	N/A	N/A	This study	N/A
Mozambique R1*	Relative plant	Namapa	CBSV	N/A	N/A	N/A	This study	N/A
Unknown_TME3	TME3	N/A	N/A	N/A	N/A	EU117376	Daniell et al., 2008	Derived from the original cassava germplasm transferred to Africa or recent African breeding improvement
Tanzania_Tree Cassava	Tree Cassava	N/A	N/A	N/A	N/A	SRR2847471	Bredeson et al., 2016	Presumed to be an *M. glaziovii–M. esculenta* hybrid
Australia_TMS I50395-Unk	TMS I50395/Unk	N/A	N/A	N/A	N/A	SRR2847474	Bredeson et al., 2016	N/A
Unknown_Namikonga	Namikonga	N/A	N/A	N/A	N/A	SRR2847450	Bredeson et al., 2016	Arose by introgression of *M. glaziovii* into *M. esculenta*, derived from the Amani program. It's CBSD-resistant but CMD-susceptible
Tanzania_Manihot_glaziovii	*Manihot glaziovii*	N/A	N/A	N/A	N/A	SRR2847419	Bredeson et al., 2016	N/A

Brazil_FLA444-1	FLA 444-1	N/A	N/A	N/A	N/A	SRR2847403	Bredeson et al., 2016	*M. esculenta* ssp. *flabellifolia*. Brazilian accession of the wild progenitor species of cassava

Phylogeny reconstruction included 62 protein-coding genes for a dataset with 39 taxa. The single-copy genes analyzed (54 genes) were psbA, atpA, atpF, atpH, atpI, rps2, rpoC1, psbM, psbD, psbC, psbZ, rps14, psaB, psaA, ycf3, rps4, ndhJ, ndhK, ndhC, atpE, atpB, rbcL, psaI, ycf4, cemA, petA, psbJ, psbL, psbF, psbE, petG, psaJ, rpl33, rpl20, clpP, psbB, psbT, psbN, psbH, petB, petD, rps11, rpl36, rps8, rpl14, rpl16, rps3, rpl22, psaC, ndhE, ndhG, ndhI, ndhA and ndhH. The genes duplicated in the IR analyzed (4 genes) were rpl2, rps7, rpl23 and rps12.

Single nucleotide polymorphisms (SNP's) were evaluated by the number of differences between each of the samples, calculated by Geneious v9.1.3 (Supplementary Figure 1). The nucleotide phylogenetic analysis clustered samples in three main clades, herein named as clades A, B and C (Fig 1). The clade "A" consisted from the wild relative species *Manihot glaziovii* and a relative plant from Mozambique. *M. glaziovii* (also named as India rubber tree species) was domesticated in South America and imported to East Africa in the early twentieth century. Since then, it has been used in cassava breeding programs to select the natural resistance of *M. glaziovii* to cassava pathogens, such as plant viruses (BREDESON et al., 2016; NICHOLS, 1947).

The clade "B" of the Phylogenetic Tree (Fig 1) was composed by two sub-clades. The upper sub-clade containing Brazil_FLA_444-1, that is an accession from Brazil of a wild progenitor species of cassava, *M. esculenta* ssp. *flabellifolia*. The bottom sub-clade was composed by Namikonga, TME3 and three cassava plants from Kenya. The SNP's analysis revealed that Namikonga, Kenya_02_Megana and Kenya_04_Megana are 100% identical based on the chloroplast coding-regions analysed. The variety Namikonga is CBSD-resistant but CMD-susceptible. It has emerged by introgression of the wild *M. glaziovii* into *M. esculenta*, (BREDESON et al., 2016). Curiously, both Megana cassava plants, which are 100% identical to

Namikonga based on the chloroplast coding regions analysed, were naturally infected by CBSD, that Namikonga was presumed to be resistant.

The clade “C” consisted mainly of cassava samples, excepted for Tanzania_Tree_Cassava (SRR2847471) which is supposed to be a *M. glaziovii–M. esculenta* hybrid (BREDESON et al., 2016). Other three samples (Kenya_11_Kibandameno, Unknown_Healthy and Australia_TMS_I50395) showed 100% identity with Tanzania_Tree_Cassava on the SNP’s analysis (Supplementary Figure 1). It is known that DNA barcodes derived from the chloroplast genome can be used to identify varieties (DANIELL et al., 2016). The 100% identity among the chloroplast regions from these four cassava plants may indicate that they are all an *M. glaziovii–M. esculenta* hybrid. Curiously, Kenya_11_Kibandameno was naturally infected with EACMZV, EACMV-K, UCBSV and CBSV. It suggests that these *M. glaziovii–M. esculenta* related varieties are not conferring resistance to EACMZV, EACMV-K, UCBSV and CBSV isolates present in Kenya.

The variety Kibandameno seems to either present high variability or to be misidentified in Kenya and Tanzania. The cassava plant identified as Kenya_11_Kibandameno showed 46 SNP’s with Kenya_15_Kibandameno (Supplementary Figure 1) suggesting that the identification of varieties might be incorrect. This also happened to the variety Megana in Kenya. Samples Kenya02_Megana and Kenya_04_Megana showed no SNP’s between them. However, 20 SNP’s were present between Kenya02_Megana and Kenya_05_Megana, indicating another possible variety misidentification.

Our phylogenetic analysis suggests that chloroplast coding-regions from East-African cassava varieties presents low genetic diversity (Fig 1). Cassava plants from Tanzania were more genetically similar to each other compared to plants collected in Mozambique and Kenya. Cassava plants from Kenya showed the most genetic diversity and are more closely related to Mozambique varieties. Previous studies reported an extraordinarily limited global chloroplast diversity of cultivated cassava (BREDESON et al., 2016). They suggested that this is a consequence of a limited chloroplast diversity relative to the wild progenitor species, in the beginning of cassava domestication (BREDESON et al., 2016). Apparently, there was no direct relationship between cassava chloroplast coding regions diversity and the natural infection of a

determined virus species. This may indicate that there is a need of the introgression of new wild relative's plants to increase the genetic diversity of the cultivated cassava and find new target genes resistant to CBSD and CMD in East Africa.

Acknowledgements

Computational resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia supported this work.

References

ABERER, A. J.; KOBERT, K.; STAMATAKIS, A. ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era. **Molecular Biology and Evolution**, v. 31, n. 10, p. 2553–2556, out. 2014.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 5 out. 1990.

APARECIDA DE MORAES, L. et al. New invasion of Bemisia tabaci Mediterranean species in Brazil associated to ornamental plants. **Phytoparasitica**, v. 45, n. 4, p. 1–9, 10 ago. 2017.

ATEKA, E. et al. Unusual occurrence of a DAG motif in the Ipomovirus Cassava brown streak virus and implications for its vector transmission. **PLOS ONE**, v. 12, n. 11, p. e0187883, 20 nov. 2017.

BARBOSA, L. F. F. et al. Indigenous American species of the Bemisia tabaci complex are still widespread in the Americas. **Pest Management Science**, v. 70, n. 10, p. 1440–1445, out. 2014.

BAUMANN, P. Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. **Annual Review of Microbiology**, v. 59, n. 1, p. 155–189, out. 2005.

BERNT, M. et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. **Molecular Phylogenetics and Evolution**, v. 69, n. 2, p. 313–319, nov. 2013.

BHATTACHARYYA, D.; CHAKRABORTY, S. Chloroplast: the Trojan horse in plant-virus interaction. **Molecular Plant Pathology**, v. 19, n. 2, p. 504–518, fev. 2018.

BING, X.-L. et al. Characterization of a newly discovered symbiont of the whitefly Bemisia tabaci (Hemiptera: Aleyrodidae). **Applied and environmental microbiology**, v. 79, n. 2, p. 569–75, 15 jan. 2013.

BOYKIN, L. M. et al. Species delimitation and global biosecurity. **Evolutionary Bioinformatics**, v. 2011, n. 7, p. 1–37, dez. 2011.

BREDESON, J. V et al. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. **Nature biotechnology**, v. 34, n. 5, p. 562–570, 2016.

BRUMIN, M.; KONTSEDALOV, S.; GHANIM, M. Rickettsia influences thermotolerance in the whitefly Bemisia tabaci B biotype. **Insect Science**, v. 18, n. 1, p. 57–66, 1 fev. 2011.

CALVERT, L. A.; THRESH, J. M. The Viruses and Virus diseases of Cassava. **Cassava: biology, production and utilization**, p. 237–260, 2002.

DA FONSECA BARBOSA, L. et al. First report of Bemisia tabaci Mediterranean (Q biotype) species in Brazil. **Pest Management Science**, v. 71, n. 4, 2015.

DANIELL, H. et al. The complete nucleotide sequence of the cassava (Manihot esculenta) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron. **Theoretical and Applied Genetics**, v. 116, n. 5, p. 723–737, mar. 2008.

DANIELL, H. et al. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. **Genome Biology**, v. 17, n. 1, p. 134, 23 dez. 2016.

DARDICK, C. Comparative expression profiling of Nicotiana benthamiana leaves systemically infected with three fruit tree viruses. **Molecular Plant-Microbe Interactions MPMI**, v. 20, n. 8, p. 1004–1017, ago. 2007.

DE BARRO, P. J. et al. *Bemisia tabaci* : A Statement of Species Status. **Annual Review of Entomology**, v. 56, n. 1, p. 1–19, 7 jan. 2011.

DE MARCHI, B. R. et al. Comparative transmission of five viruses by Bemisia tabaci NW2 and MEAM1. **Tropical Plant Pathology**, v. 1, n. 6, 2017.

DE MORAES APARECIDA, L. et al. Distribution and phylogenetics of whiteflies in Brazil and their endosymbiont relationships. **unpublised**, 2018.

DE TORRES ZABALA, M. et al. Chloroplasts play a central role in plant defence and are targeted by pathogen effectors. **Nature Plants**, v. 1, n. 6, p. 15074, 2015.

DEGNAN, P. H. et al. Hamiltonella defensa, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. **Proceedings of the National Academy of Sciences**, v. 106, n. 22, p. 9063–9068, 2 jun. 2009.

FERRARI, J.; VAVRE, F. Bacterial symbionts in insects or the story of communities affecting communities. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, v. 366, n. 1569, p. 1389–400, 12 maio 2011.

FROHLICH, D. R. et al. A phylogeographical analysis of the Bemisia tabaci species complex based on mitochondrial DNA markers. **Molecular Ecology**, v. 8, n. 10, p. 1683–1691, out. 1999.

GAUTHIER, N. et al. Genetic structure of *Bemisia tabaci* Med populations from home-range countries, inferred by nuclear and cytoplasmic markers: impact on the distribution of the insecticide resistance genes. **Pest Management Science**, v. 70, n. 10, p. 1477–1491, out. 2014.

GOROVITS, R.; CZOSNEK, H. Insect Symbiotic Bacterial GroEL (Chaperonin 60) and Plant Virus Transmission. In: MANDE, S. C.; KUMAR, C. M. S.; SHARMA, A. (Eds.). . **Moonlighting Cell Stress Proteins in Microbial Infections**. [s.l: s.n.]. v. 7p. 173–187.

GOTTLIEB, Y. et al. 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**, v. 84, n. 18, p. 9310–7, 15 set. 2010.

HILLOCKS, R. J.; THRESH, J. M. (EDS.). **Cassava: biology, production and utilization**. Wallingford: CABI, 2002.

HILLOCKS, R.; JENNINGS, D. Cassava brown streak disease: A review of present knowledge and research needs. **International Journal of Pest Management**, v. 49, n. 3, p. 225–234, jul. 2003.

HU, J. et al. An Extensive Field Survey Combined with a Phylogenetic Analysis Reveals Rapid and Widespread Invasion of Two Alien Whiteflies in China. **PLoS ONE**, v. 6, n. 1, p. e16061, 21 jan. 2011.

KANAKALA, S.; GHANIM, M. Advances in the Genomics of the Whitefly Bemisia tabaci: An Insect Pest and a Virus Vector. In: [s.l.] Springer International Publishing, 2015. p. 19–40.

KATOH, K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic acids research**, v. 30, n. 14, p. 3059–66, 15 jul. 2002.

KEARSE, M. et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, n. 12, p. 1647–1649, 15 jun. 2012.

KONTSEDALOV, S. et al. The presence of *Rickettsia* is associated with increased susceptibility of *Bemisia tabaci* (Homoptera: Aleyrodidae) to insecticides. **Pest Management Science**, v. 64, n. 8, p. 789–792, ago. 2008.

KONTSEDALOV, S. et al. Toxicity of spiromesifen to the developmental stages of *Bemisia tabaci* biotype B. **Pest Management Science**, v. 65, n. 1, p. 5–13, jan. 2009.

LEGG, J. P. et al. Cassava Mosaic Virus Disease in East and Central Africa: Epidemiology and Management of A Regional Pandemic. In: **Plant Virus Epidemiology**. [s.l.] Academic Press, 2006. v. 67p. 355–418.

LIU, C. et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. **BMC genomics**, v. 13, n. 1, p. 715, 2012.

LODHI, M. A. et al. A simple and efficient method for DNA extraction from grapevine cultivars andVitis species. **Plant Molecular Biology Reporter**, v. 12, n. 1, p. 6–13,

mar. 1994.

LOURENÇÃO, A. L.; NAGAI, H. Outbreaks of Bemisia tabaci in the São Paulo State, Brazil. **Bragantia**, v. 53, n. 1, p. 53–59, 1994.

LOWE S., BROWNE M., BOUDJELAS S., D. P. M. **100 of the World's Worst Invasive Alien Species A selection from the Global Invasive Species Database**. Auckland, New Zealand: The Invasive Species Specialist Group (ISSG) a specialist group of the Species Survival Commission (SSC) of the World Conservation Union (IUCN), 2000.

LUAN, J.-B. et al. Cellular and molecular remodelling of a host cell for vertical transmission of bacterial symbionts. **Proceedings of the Royal Society B: Biological Sciences**, v. 283, n. 1833, p. 20160580, 29 jun. 2016.

LUAN, J. et al. Maternal Inheritance of a Single Somatic Animal Cell Displayed by the Bacteriocyte in the Whitefly Bemisia tabaci. **Current Biology**, v. 28, n. 3, p. 459–465.e3, 5 fev. 2018.

MAHADAV, A. et al. Parasitization by the wasp Eretmocerus mundus induces transcription of genes related to immune response and symbiotic bacteria proliferation in the whitefly Bemisia tabaci. **BMC Genomics**, v. 9, n. 1, p. 342, 18 jul. 2008.

MARUBAYASHI, J. M. et al. At least two indigenous species of the *Bemisia tabaci* complex are present in Brazil. **Journal of Applied Entomology**, v. 137, n. 1–2, p. 113–121, 1 fev. 2013.

MARUBAYASHI, J. M. et al. Diversity and Localization of Bacterial Endosymbionts from Whitefly Species Collected in Brazil. **PLoS ONE**, v. 9, n. 9, p. e108363, 26 set. 2014.

MARUBAYASHI, J. M.; YUKI, V. A.; WUTKE, E. B. Transmission of the Cowpea mild mottle virus by whitefly Bemisia tabaci biotype B for plants of beans and soy. **Summa Phytopathologica**, v. 36, n. 2, p. 158–160, jun. 2010.

MOCHIZUKI, T. et al. Quantitative transcriptional changes associated with chlorosis severity in mosaic leaves of tobacco plants infected with cucumber mosaic virus.

Molecular Plant Pathology, v. 15, n. 3, p. 242–254, abr. 2014.

MORIN, S. et al. A GroEL Homologue from Endosymbiotic Bacteria of the WhiteflyBemisia tabaciIs Implicated in the Circulative Transmission of Tomato Yellow Leaf Curl Virus. **Virology**, v. 256, n. 1, p. 75–84, 1999.

MORIN, S. et al. The GroEL Protein of the Whitefly Bemisia tabaci Interacts with the Coat Protein of Transmissible and Nontransmissible Begomoviruses in the Yeast Two-Hybrid System. **Virology**, v. 276, n. 2, p. 404–416, 2000.

NAVAS-CASTILLO, J.; FIALLO-OLIVÉ, E.; SÁNCHEZ-CAMPOS, S. Emerging Virus Diseases Transmitted by Whiteflies. **Annual Review of Phytopathology**, v. 49, n. 1, p. 219–248, 8 set. 2011.

NDUNGURU, J. et al. Analyses of twelve new whole genome sequences of cassava brown streak viruses and ugandan cassava brown streak viruses from East Africa: Diversity, supercomputing and evidence for further speciation. **PLoS ONE**, v. 10, n. 10, p. 1–18, 6 out. 2015.

NICHOLS, R. F. W. Breeding Cassava for Virus Resistance. **The East African Agricultural Journal**, v. 12, n. 3, p. 184–194, 1 jan. 1947.

PARRELLA, G. et al. Invasion of the Q2 mitochondrial variant of Mediterranean *Bemisia tabaci* in southern Italy: possible role of bacterial endosymbionts. **Pest Management Science**, v. 70, n. 10, p. 1514–1523, out. 2014.

POLSTON, J. E.; DE BARRO, P.; BOYKIN, L. M. Transmission specificities of plant viruses with the newly identified species of the *Bemisia tabaci* species complex. **Pest Management Science**, v. 70, n. 10, p. 1547–1552, 1 out. 2014.

RAHEEM, D.; CHUKWUMA, C. Foods from cassava and their relevance to Nigeria and other African countries. **Agriculture and Human Values**, v. 18, n. 4, p. 383–390, 2001.

RANA, V. S. et al. Arsenophonus GroEL Interacts with CLCuV and Is Localized in Midgut and Salivary Gland of Whitefly B. tabaci. **PLoS ONE**, v. 7, n. 8, p. e42168, 10 ago. 2012.

RAO, Q. et al. **Draft genome sequence of “Candidatus Hamiltonella defensa,” an**

endosymbiont of the whitefly Bemisia tabaciJournal of Bacteriology, 1 jul. 2012. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/22689243>. Acesso em: 22 nov. 2017

RONQUIST, F. et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. **Systematic Biology**, v. 61, n. 3, p. 539–542, maio 2012.

SCHWARTZ, S. et al. Human – Mouse Alignments with BLASTZ. **Genome Research**, v. 13, n. 1, p. 103–107, 2003.

SSERUWAGI, P. et al. The first transcriptomes from field-collected individual whiteflies (Bemisia tabaci, Hemiptera: Aleyrodidae). **Gates Open Research**, v. 1, p. 16, 28 dez. 2017.

SU, Q. et al. Facultative Symbiont <I>Hamiltonella</I> Confers Benefits to <I>Bemisia tabaci</I> (Hemiptera: Aleyrodidae), an Invasive Agricultural Pest Worldwide. **Environmental Entomology**, v. 42, n. 6, p. 1265–1271, 1 dez. 2013.

TAY, W. T. et al. Complete mitochondrial DNA genome of Bemisia tabaci cryptic pest species complex Asia I (Hemiptera: Aleyrodidae). **Mitochondrial DNA**, v. 1736, n. ii, p. 1–2, 2014.

TAY, W. T. et al. Novel molecular approach to define pest species status and tritrophic interactions from historical Bemisia specimens. **Scientific Reports**, v. 7, n. 1, p. 429, 2017a.

TAY, W. T. et al. The Trouble with MEAM2: Implications of Pseudogenes on Species Delimitation in the Globally Invasive Bemisia tabaci (Hemiptera: Aleyrodidae) Cryptic Species Complex. **Genome biology and evolution**, v. 9, n. 10, p. 2732–2738, 1 out. 2017b.

THAO, M. L.; BAUMANN, P. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. **Applied and environmental microbiology**, v. 70, n. 6, p. 3401–6, 1 jun. 2004.

TUMWEGAMIRE, S. et al. Exchanging and managing in-vitro elite germplasm to

combat Cassava Brown Streak Disease (CBSD) and Cassava Mosaic Disease (CMD) in Eastern and Southern Africa. **Food Security**, v. 10, n. 2, p. 351–368, 17 abr. 2018.

WALKER, B. J. et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. **PLoS ONE**, v. 9, n. 11, p. e112963, 19 nov. 2014.

WANG, H.-L. et al. The characteristics and expression profiles of the mitochondrial genome for the Mediterranean species of the Bemisia tabaci complex. **BMC genomics**, v. 14, p. 401, 2013.

WANG, H.-L. et al. The complete mitochondrial genome of *Bemisia afer* (Hemiptera: Aleyrodidae). **Mitochondrial DNA**, v. 27, n. 1, p. 98–99, 2 jan. 2016a.

WANG, H.-L. et al. A complete mitochondrial DNA genome derived from a Chinese population of the *Bemisia afer* species complex (Hemiptera: Aleyrodidae). **Mitochondrial DNA Part A**, v. 27, n. 5, p. 3500–3501, 2 set. 2016b.

WANG, W. et al. Cassava genome from a wild ancestor to cultivated varieties. **Nature communications**, v. 5, p. 5110, 2014.

WEI, T. et al. Sequential Recruitment of the Endoplasmic Reticulum and Chloroplasts for Plant Potyvirus Replication. **Journal of Virology**, v. 84, n. 2, p. 799–809, 15 jan. 2010.

XU, J.; DE BARRO, P. J.; LIU, S. S. Reproductive incompatibility among genetic groups of Bemisia tabaci supports the proposition that the whitefly is a cryptic species complex. **Bulletin of Entomological Research**, v. 100, n. 03, p. 359–366, 24 jun. 2010.

ZANARDO, L. G.; CARVALHO, C. M. Cowpea mild mottle virus (Carlavirus, Betaflexiviridae): a review. **Tropical Plant Pathology**, p. 417–430, 2017.

ZHAO, J. et al. **Chloroplast in plant-virus interactionFrontiers in Microbiology**, 2016. Disponível em: <https://trello-attachments.s3.amazonaws.com/57e4a3c4e8e5ee2f49659dc7/5840df4a0c532da34475dfd4/ffe302e7b366836984de30b7c06e541f/Chloroplast_in_Plant_Virus_interaction

_2016.pdf>. Acesso em: 15 mar. 2017

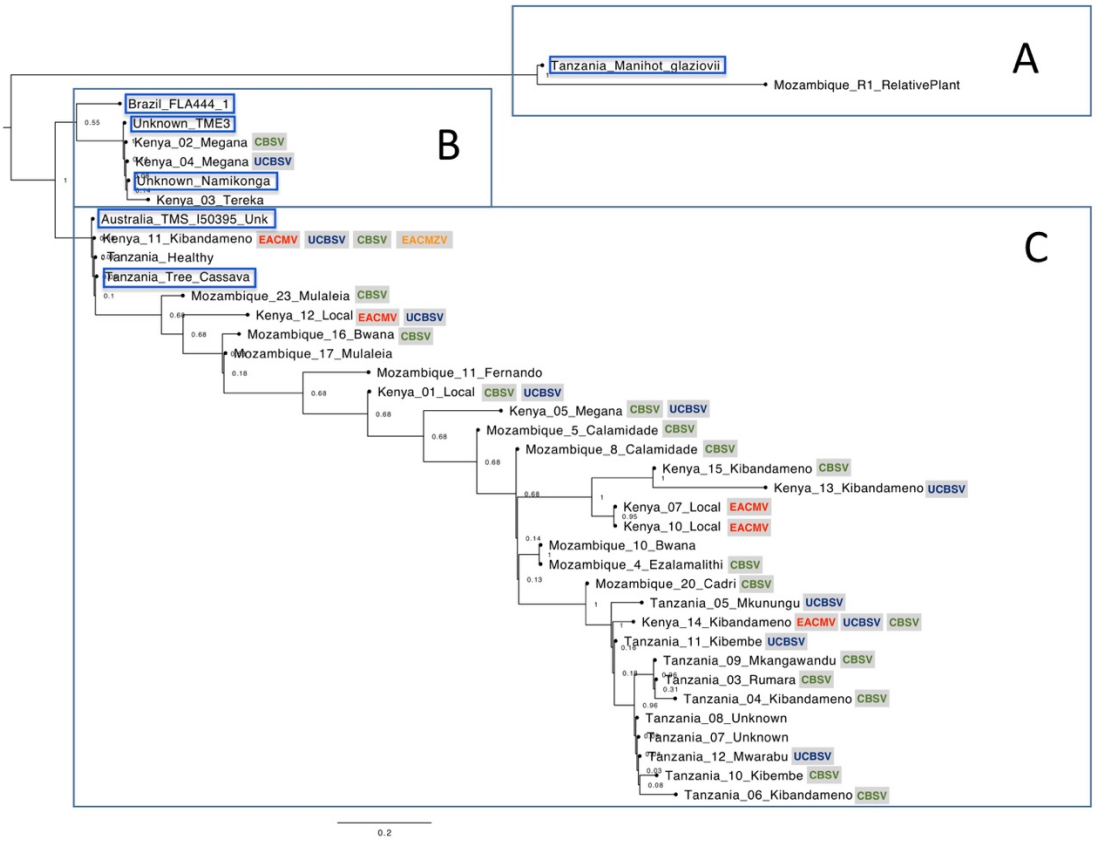
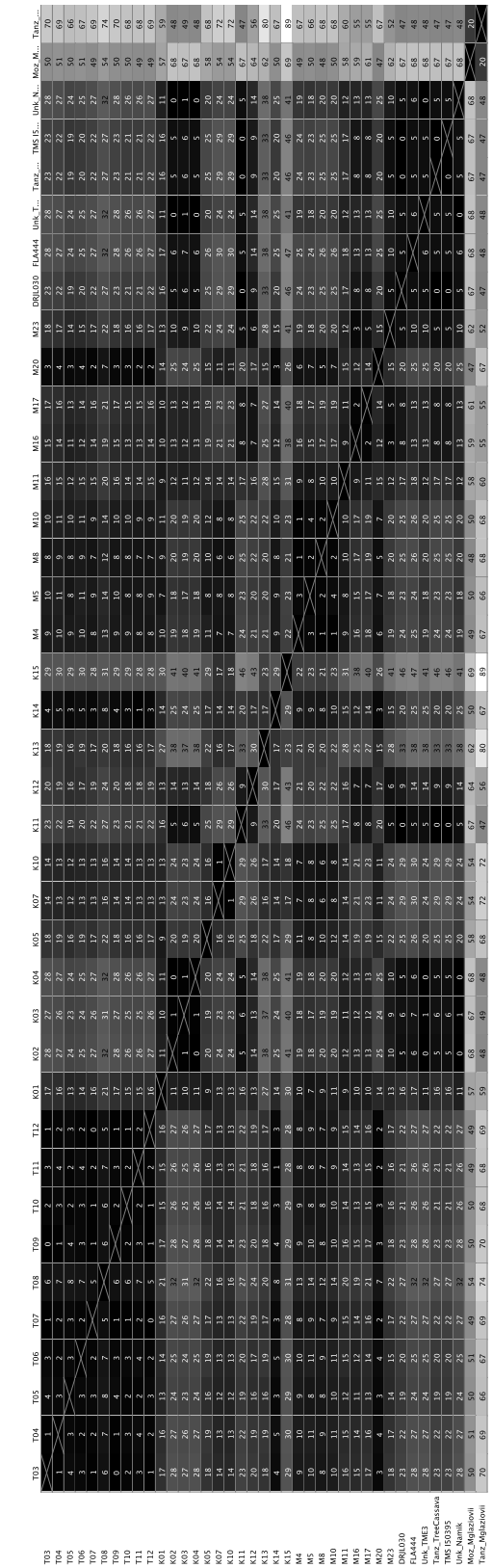


Fig 1. Phylogenetic relationships among cassava chloroplast coding regions collected in East Africa clustered cassava plants in three main clades (A, B and C). The viruses detected on the samples are indicated in colourful letters inside grey squares: *Ugandan cassava brown streak virus* (UCBSV), *Cassava brown streak virus* (CBSV), *East African cassava mosaic virus* (EACMV-Ug) and *East African cassava mosaic Zanzibar virus* (EACMZV). Samples obtained from GenBank are highlighted in blue.



Supplementary Figure 1. Number of single-nucleotide polymorphisms (SNPs) present among cassava (*Manihot esculenta*) chloroplast protein-coding genes from 39 different varieties. Samples were labelled with countries initials or varieties names: T (Tanzania), K (Kenya) and M (Mozambique).

CONSIDERAÇÕES FINAIS

- As diferenças genéticas entre as espécies de *Bemisia tabaci* se estendem do gene mtCOI para o genoma mitocondrial completo, bem como quanto ao endossimbionte facultativo *Hamiltonella* indicando que não há transmissão horizontal desse endossimbionte entre essas populações;
- Foi encontrada uma deleção no gene GroEL somente em *Hamiltonella* de populações nativas de mosca-branca que pode ter influência na síntese da proteína e, consequentemente, implicações biológicas no vetor;
- A análise do cloroplasto indicou baixa diversidade genética dos genótipos de mandioca da África Oriental, reforçando a necessidade de introdução de novos genótipos visando resistência aos principais vírus;
- Não houve uma relação direta entre a diversidade dos genes de cloroplastos avaliados e as espécies de vírus detectadas no campo.

REFERÊNCIAS

ABERER, A. J.; KOBERT, K.; STAMATAKIS, A. ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era. **Molecular Biology and Evolution**, v. 31, n. 10, p. 2553–2556, out. 2014.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 5 out. 1990.

APARECIDA DE MORAES, L. et al. New invasion of Bemisia tabaci Mediterranean species in Brazil associated to ornamental plants. **Phytoparasitica**, v. 45, n. 4, p. 1–9, 10 ago. 2017.

ATEKA, E. et al. Unusual occurrence of a DAG motif in the Ipomovirus Cassava brown streak virus and implications for its vector transmission. **PLOS ONE**, v. 12, n. 11, p. e0187883, 20 nov. 2017.

BARBOSA, L. F. F. et al. Indigenous American species of the Bemisia tabaci complex are still widespread in the Americas. **Pest Management Science**, v. 70, n. 10, p. 1440–1445, out. 2014.

BAUMANN, P. Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. **Annual Review of Microbiology**, v. 59, n. 1, p. 155–189, out. 2005.

BERNT, M. et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. **Molecular Phylogenetics and Evolution**, v. 69, n. 2, p. 313–319, nov. 2013.

BHATTACHARYYA, D.; CHAKRABORTY, S. Chloroplast: the Trojan horse in plant-virus interaction. **Molecular Plant Pathology**, v. 19, n. 2, p. 504–518, fev. 2018.

BING, X.-L. et al. Characterization of a newly discovered symbiont of the whitefly Bemisia tabaci (Hemiptera: Aleyrodidae). **Applied and environmental microbiology**, v. 79, n. 2, p. 569–75, 15 jan. 2013.

BOYKIN, L. M. et al. Species delimitation and global biosecurity. **Evolutionary Bioinformatics**, v. 2011, n. 7, p. 1–37, dez. 2011.

BREDESON, J. V et al. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. **Nature biotechnology**, v. 34, n. 5, p. 562–570, 2016.

BRUMIN, M.; KONTSEDALOV, S.; GHANIM, M. Rickettsia influences thermotolerance in the whitefly Bemisia tabaci B biotype. **Insect Science**, v. 18, n. 1, p. 57–66, 1 fev. 2011.

CALVERT, L. A.; THRESH, J. M. The Viruses and Virus diseases of Cassava. **Cassava: biology, production and utilization**, p. 237–260, 2002.

DA FONSECA BARBOSA, L. et al. First report of Bemisia tabaci Mediterranean (Q biotype) species in Brazil. **Pest Management Science**, v. 71, n. 4, 2015.

DANIELL, H. et al. The complete nucleotide sequence of the cassava (Manihot esculenta) chloroplast genome and the evolution of atpF in Malpighiales: RNA editing and multiple losses of a group II intron. **Theoretical and Applied Genetics**, v. 116, n. 5, p. 723–737, mar. 2008.

DANIELL, H. et al. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. **Genome Biology**, v. 17, n. 1, p. 134, 23 dez. 2016.

DARDICK, C. Comparative expression profiling of Nicotiana benthamiana leaves systemically infected with three fruit tree viruses. **Molecular Plant-Microbe Interactions MPMI**, v. 20, n. 8, p. 1004–1017, ago. 2007.

DE BARRO, P. J. et al. *Bemisia tabaci* : A Statement of Species Status. **Annual Review of Entomology**, v. 56, n. 1, p. 1–19, 7 jan. 2011.

DE MARCHI, B. R. et al. Comparative transmission of five viruses by Bemisia tabaci NW2 and MEAM1. **Tropical Plant Pathology**, v. 1, n. 6, 2017.

DE MORAES APARECIDA, L. et al. Distribution and phylogenetics of whiteflies in Brazil and their endosymbiont relationships. **unpublised**, 2018.

DE TORRES ZABALA, M. et al. Chloroplasts play a central role in plant defence and are targeted by pathogen effectors. **Nature Plants**, v. 1, n. 6, p. 15074, 2015.

DEGNAN, P. H. et al. Hamiltonella defensa, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. **Proceedings of the National Academy of Sciences**, v. 106, n. 22, p. 9063–9068, 2 jun. 2009.

FERRARI, J.; VAVRE, F. Bacterial symbionts in insects or the story of communities affecting communities. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, v. 366, n. 1569, p. 1389–400, 12 maio 2011.

FROHLICH, D. R. et al. A phylogeographical analysis of the Bemisia tabaci species complex based on mitochondrial DNA markers. **Molecular Ecology**, v. 8, n. 10, p. 1683–1691, out. 1999.

GAUTHIER, N. et al. Genetic structure of *Bemisia tabaci* Med populations from home-range countries, inferred by nuclear and cytoplasmic markers: impact on the distribution of the insecticide resistance genes. **Pest Management Science**, v. 70, n. 10, p. 1477–1491, out. 2014.

GOROVITS, R.; CZOSNEK, H. Insect Symbiotic Bacterial GroEL (Chaperonin 60) and Plant Virus Transmission. In: MANDE, S. C.; KUMAR, C. M. S.; SHARMA, A. (Eds.). . **Moonlighting Cell Stress Proteins in Microbial Infections**. [s.l: s.n.]. v. 7p. 173–187.

GOTTLIEB, Y. et al. 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**, v. 84, n. 18, p. 9310–7, 15 set. 2010.

HILLOCKS, R. J.; THRESH, J. M. (EDS.). **Cassava: biology, production and utilization**. Wallingford: CABI, 2002.

HILLOCKS, R.; JENNINGS, D. Cassava brown streak disease: A review of present knowledge and research needs. **International Journal of Pest Management**, v. 49, n. 3, p. 225–234, jul. 2003.

HU, J. et al. An Extensive Field Survey Combined with a Phylogenetic Analysis Reveals Rapid and Widespread Invasion of Two Alien Whiteflies in China. **PLoS ONE**, v. 6, n. 1, p. e16061, 21 jan. 2011.

KANAKALA, S.; GHANIM, M. Advances in the Genomics of the Whitefly Bemisia tabaci: An Insect Pest and a Virus Vector. In: [s.l.] Springer International Publishing, 2015. p. 19–40.

KATOH, K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. **Nucleic acids research**, v. 30, n. 14, p. 3059–66, 15 jul. 2002.

KEARSE, M. et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, n. 12, p. 1647–1649, 15 jun. 2012.

KONTSEDALOV, S. et al. The presence of *Rickettsia* is associated with increased susceptibility of *Bemisia tabaci* (Homoptera: Aleyrodidae) to insecticides. **Pest Management Science**, v. 64, n. 8, p. 789–792, ago. 2008.

KONTSEDALOV, S. et al. Toxicity of spiromesifen to the developmental stages of *Bemisia tabaci* biotype B. **Pest Management Science**, v. 65, n. 1, p. 5–13, jan. 2009.

LEGG, J. P. et al. Cassava Mosaic Virus Disease in East and Central Africa: Epidemiology and Management of A Regional Pandemic. In: **Plant Virus Epidemiology**. [s.l.] Academic Press, 2006. v. 67p. 355–418.

LIU, C. et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. **BMC genomics**, v. 13, n. 1, p. 715, 2012.

LODHI, M. A. et al. A simple and efficient method for DNA extraction from grapevine cultivars andVitis species. **Plant Molecular Biology Reporter**, v. 12, n. 1, p. 6–13, mar. 1994.

LOURENÇÃO, A. L.; NAGAI, H. Outbreaks of Bemisia tabaci in the São Paulo State, Brazil. **Bragantia**, v. 53, n. 1, p. 53–59, 1994.

LOWE S., BROWNE M., BOUDJELAS S., D. P. M. **100 of the World's Worst Invasive Alien Species A selection from the Global Invasive Species Database**. Auckland, New Zealand: The Invasive Species Specialist Group (ISSG) a specialist group of the Species Survival Commission (SSC) of the World Conservation Union (IUCN), 2000.

LUAN, J.-B. et al. Cellular and molecular remodelling of a host cell for vertical transmission of bacterial symbionts. **Proceedings of the Royal Society B: Biological Sciences**, v. 283, n. 1833, p. 20160580, 29 jun. 2016.

LUAN, J. et al. Maternal Inheritance of a Single Somatic Animal Cell Displayed by the

Bacteriocyte in the Whitefly Bemisia tabaci. **Current Biology**, v. 28, n. 3, p. 459–465.e3, 5 fev. 2018.

MAHADAV, A. et al. Parasitization by the wasp Eretmocerus mundus induces transcription of genes related to immune response and symbiotic bacteria proliferation in the whitefly Bemisia tabaci. **BMC Genomics**, v. 9, n. 1, p. 342, 18 jul. 2008.

MARUBAYASHI, J. M. et al. At least two indigenous species of the *Bemisia tabaci* complex are present in Brazil. **Journal of Applied Entomology**, v. 137, n. 1–2, p. 113–121, 1 fev. 2013.

MARUBAYASHI, J. M. et al. Diversity and Localization of Bacterial Endosymbionts from Whitefly Species Collected in Brazil. **PLoS ONE**, v. 9, n. 9, p. e108363, 26 set. 2014.

MARUBAYASHI, J. M.; YUKI, V. A.; WUTKE, E. B. Transmission of the Cowpea mild mottle virus by whitefly Bemisia tabaci biotype B for plants of beans and soy. **Summa Phytopathologica**, v. 36, n. 2, p. 158–160, jun. 2010.

MOCHIZUKI, T. et al. Quantitative transcriptional changes associated with chlorosis severity in mosaic leaves of tobacco plants infected with cucumber mosaic virus. **Molecular Plant Pathology**, v. 15, n. 3, p. 242–254, abr. 2014.

MORIN, S. et al. A GroEL Homologue from Endosymbiotic Bacteria of the WhiteflyBemisia tabaciIs Implicated in the Circulative Transmission of Tomato Yellow Leaf Curl Virus. **Virology**, v. 256, n. 1, p. 75–84, 1999.

MORIN, S. et al. The GroEL Protein of the Whitefly Bemisia tabaci Interacts with the Coat Protein of Transmissible and Nontransmissible Begomoviruses in the Yeast Two-Hybrid System. **Virology**, v. 276, n. 2, p. 404–416, 2000.

NAVAS-CASTILLO, J.; FIALLO-OLIVÉ, E.; SÁNCHEZ-CAMPOS, S. Emerging Virus Diseases Transmitted by Whiteflies. **Annual Review of Phytopathology**, v. 49, n. 1, p. 219–248, 8 set. 2011.

NDUNGURU, J. et al. Analyses of twelve new whole genome sequences of cassava brown streak viruses and ugandan cassava brown streak viruses from East Africa: Diversity, supercomputing and evidence for further speciation. **PLoS ONE**, v. 10, n. 10, p. 1–18, 6 out. 2015.

NICHOLS, R. F. W. Breeding Cassava for Virus Resistance. **The East African Agricultural Journal**, v. 12, n. 3, p. 184–194, 1 jan. 1947.

PARRELLA, G. et al. Invasion of the Q2 mitochondrial variant of Mediterranean *Bemisia tabaci* in southern Italy: possible role of bacterial endosymbionts. **Pest Management Science**, v. 70, n. 10, p. 1514–1523, out. 2014.

POLSTON, J. E.; DE BARRO, P.; BOYKIN, L. M. Transmission specificities of plant viruses with the newly identified species of the *Bemisia tabaci* species complex. **Pest Management Science**, v. 70, n. 10, p. 1547–1552, 1 out. 2014.

RAHEEM, D.; CHUKWUMA, C. Foods from cassava and their relevance to Nigeria and other African countries. **Agriculture and Human Values**, v. 18, n. 4, p. 383–

390, 2001.

RANA, V. S. et al. Arsenophonus GroEL Interacts with CLCuV and Is Localized in Midgut and Salivary Gland of Whitefly B. tabaci. **PLoS ONE**, v. 7, n. 8, p. e42168, 10 ago. 2012.

RAO, Q. et al. **Draft genome sequence of "Candidatus Hamiltonella defensa," an endosymbiont of the whitefly Bemisia tabaciJournal of Bacteriology**, 1 jul. 2012. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/22689243>. Acesso em: 22 nov. 2017

RONQUIST, F. et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. **Systematic Biology**, v. 61, n. 3, p. 539–542, maio 2012.

SCHWARTZ, S. et al. Human – Mouse Alignments with BLASTZ. **Genome Research**, v. 13, n. 1, p. 103–107, 2003.

SSERUWAGI, P. et al. The first transcriptomes from field-collected individual whiteflies (Bemisia tabaci, Hemiptera: Aleyrodidae). **Gates Open Research**, v. 1, p. 16, 28 dez. 2017.

SU, Q. et al. Facultative Symbiont <I>Hamiltonella</I> Confers Benefits to <I>Bemisia tabaci</I> (Hemiptera: Aleyrodidae), an Invasive Agricultural Pest Worldwide. **Environmental Entomology**, v. 42, n. 6, p. 1265–1271, 1 dez. 2013.

TAY, W. T. et al. Complete mitochondrial DNA genome of Bemisia tabaci cryptic pest species complex Asia I (Hemiptera: Aleyrodidae). **Mitochondrial DNA**, v. 1736, n. ii, p. 1–2, 2014.

TAY, W. T. et al. Novel molecular approach to define pest species status and tritrophic interactions from historical Bemisia specimens. **Scientific Reports**, v. 7, n. 1, p. 429, 2017a.

TAY, W. T. et al. The Trouble with MEAM2: Implications of Pseudogenes on Species Delimitation in the Globally Invasive Bemisia tabaci (Hemiptera: Aleyrodidae) Cryptic Species Complex. **Genome biology and evolution**, v. 9, n. 10, p. 2732–2738, 1 out. 2017b.

THAO, M. L.; BAUMANN, P. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. **Applied and environmental microbiology**, v. 70, n. 6, p. 3401–6, 1 jun. 2004.

TUMWEGAMIRE, S. et al. Exchanging and managing in-vitro elite germplasm to combat Cassava Brown Streak Disease (CBSD) and Cassava Mosaic Disease (CMD) in Eastern and Southern Africa. **Food Security**, v. 10, n. 2, p. 351–368, 17 abr. 2018.

WALKER, B. J. et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. **PLoS ONE**, v. 9, n. 11, p. e112963, 19 nov. 2014.

WANG, H.-L. et al. The characteristics and expression profiles of the mitochondrial

genome for the Mediterranean species of the Bemisia tabaci complex. **BMC genomics**, v. 14, p. 401, 2013.

WANG, H.-L. et al. The complete mitochondrial genome of *Bemisia afer* (Hemiptera: Aleyrodidae). **Mitochondrial DNA**, v. 27, n. 1, p. 98–99, 2 jan. 2016a.

WANG, H.-L. et al. A complete mitochondrial DNA genome derived from a Chinese population of the *Bemisia afer* species complex (Hemiptera: Aleyrodidae). **Mitochondrial DNA Part A**, v. 27, n. 5, p. 3500–3501, 2 set. 2016b.

WANG, W. et al. Cassava genome from a wild ancestor to cultivated varieties. **Nature communications**, v. 5, p. 5110, 2014.

WEI, T. et al. Sequential Recruitment of the Endoplasmic Reticulum and Chloroplasts for Plant Potyvirus Replication. **Journal of Virology**, v. 84, n. 2, p. 799–809, 15 jan. 2010.

XU, J.; DE BARRO, P. J.; LIU, S. S. Reproductive incompatibility among genetic groups of Bemisia tabaci supports the proposition that the whitefly is a cryptic species complex. **Bulletin of Entomological Research**, v. 100, n. 03, p. 359–366, 24 jun. 2010.

ZANARDO, L. G.; CARVALHO, C. M. Cowpea mild mottle virus (Carlavirus, Betaflexiviridae): a review. **Tropical Plant Pathology**, p. 417–430, 2017.

ZHAO, J. et al. **Chloroplast in plant-virus interactionFrontiers in Microbiology**, 2016. Disponível em: <https://trello-attachments.s3.amazonaws.com/57e4a3c4e8e5ee2f49659dc7/5840df4a0c532da34475dfd4/ffe302e7b366836984de30b7c06e541f/Chloroplast_in_Plant_Virus_interaction_2016.pdf>. Acesso em: 15 mar. 2017

Apêndice A: Artigo publicado na revista Gates Open Research em colaboração com a Dra. Laura M. Boykin durante o doutorado sanduíche na The University of Western Australia, UWA, Perth, WA.

RESEARCH ARTICLE - Gates Open Research

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

Peter Sseruwagi[1*], James Wainaina[2*], Joseph Ndunguru[1], Robooni Tumuhimbise [3], Fred Tairo[1], Jian-Yang Guo[4,5], Alice Vrielink[2], Amanda Blythe[2], Tonny Kinene [2], Bruno De Marchi[2,6], Monica A. Kehoe[7], Sandra Tanz[2], Laura M. Boykin[2]

[1]Mikocheni Agriculture Research Institute (MARI), Dar es Salaam, P.O. Box 6226, Tanzania [2]School of Molecular Sciences and Australian Research Council Centre of Excellence in Plant Energy Biology, University of Western Australia, Perth, WA, 6009, Australia [3]National Agricultural Research Laboratories, P.O. Box 7065, Kampala Kawanda - Senge Rd, Kampala, Uganda [4]Ministry of Agriculture Key Laboratory of Agricultural Entomology, Institute of Insect Sciences, Zhejiang University, Hangzhou, 310058, China [5]State Key Laboratory for the Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, 100193, China [6]Faculdade de Ciências Agronômicas, Universidade Estadual Paulista , Botucatu, Brazil [7]Department of Primary Industries and Regional Development, DPIRD Diagnostic Laboratory Services, South Perth, WA, Australia

[*] Equal contributors

Abstract

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

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

assembly of trimmed reads yielded between 65,000-162,000 Contigs across *B. tabaci* transcriptomes.

Results: Bayesian phylogenetic analysis of mitochondrion cytochrome I oxidase (mtCOI) grouped the four whiteflies within the SSA1 clade. BLASTn searches on the four transcriptomes identified five endosymbionts; the primary endosymbiont *Portiera aleyrodidarum* and four secondary endosymbionts: *Arsenophonus, Wolbachia, Rickettsia,* and *Cardinium spp.* that were predominant across all four SSA1 B. *tabaci* samples with prevalence levels of between 54.1 to 75%. Amino acid alignments of the *Nus*G gene of *P. aleyrodidarum* for the SSA1 *B. tabaci* transcriptomes of samples WF2 and WF2b revealed an eleven amino acid residue deletion that was absent in samples WF1 and WF2a. Comparison of the protein structure of the *Nus*G protein from *P. aleyrodidarum* in SSA1 with known *Nus*G structures showed the deletion resulted in a shorter D loop. **Conclusions:** The use of field-collected specimens means time and money will be saved in future studies using single whitefly transcriptomes in monitoring vector and viral interactions. Our method is applicable to any small organism where RNA quantity has limited transcriptome studies.

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

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

Introduction

Members of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex are classified as the world's most devastating insect pests. There are 34 species globally[1] and the various species in the complex are morphologically identical. They transmit over 100 plant viruses[2,3], become insecticide resistant[4], and ultimately cause billions of dollars in damage annually for farmers[5]. The adult whiteflies are promiscuous feeders, and will move between viral infected crops and native weeds that act as viral inoculum 'sources', and deposit viruses to alternative crops that act as viral 'sinks' while feeding.

The crop of importance for this study was cassava (*Manihot esculenta*). Cassava supports approximately 800 million people in over 105 countries as a source of food and nutritional secu- rity, especially within rural smallholder farming communities[6]. Cassava production in Sub-Saharan Africa (SSA), especially the East African region, is hampered by whitefly-transmitted DNA and RNA viruses.

Among the DNA viruses, nine cassava-infecting cassava mosaic begomoviruses (CMBs) including: *African cassava mosaic virus* (ACMV), *African cassava mosaic Bukina faso virus* (ACMBV), *African cassava mosaic Madagascar virus* (ACMMV), *East African cassava mosaic virus* (EACMV), *East African cassava mosaic Cameroon virus* (EACMCV), *East African cassava mosaic Kenya virus* (EACMKV), *East African*

cassava mosaic Malawi virus (EACMMV), *East African cassava mosaic Zanzibar virus* (EACMZV) and *South African cassava mosaic virus* (SACMV) have been reported in SSA[7]. The DNA viruses cause cassava mosaic disease (CMD) leading to 28–40% crop losses with estimated economic losses of up to $2.7 billion dollars per year in SSA[8]. The CMD pandemics in East Africa, and across other cassava producing areas in SSA, were correlated with *B. tabaci* outbreaks[9] Relevant to this study are two RNA ipomoviruses, in the family *Potyviridae*: *Cassava brown streak virus* (CBSV) and the *Ugandan cassava brown streak virus* (UCBSV), both devastating cassava in East Africa. *Bemisia tabaci* species have been hypothesized to transmit these RNA viruses with limited transmission efficiency[10,11]. Recent studies have shown that there are multiple species of these viruses[12], which further strengthens the need to obtain data from individual whiteflies as pooled samples could contain different species with different virus composition and transmission efficiency. In addition, CBSV has been shown to have a higher rate of evolution than UCBSV[13] increasing the urgency of understanding the role played by the different whitefly species in the system.

Endosymbionts and their role in *B. tabaci* Viral-vector interactions within *B. tabaci* are further influenced by bacterial endosymbionts forming a tripartite interaction. *B. tabaci* has one of the highest numbers of endosymbiont bacterial infections with eight different vertically transmitted bacteria reported[14–17]. They are classified into two categories; primary (P) and secondary (S) endosymbionts, many of which are in specialised cells called bacteriocytes, while a few are also found scattered throughout the whitefly body. A single obligate *P-endosymbiont P. aleyrodidarum* is systematically found in all *B. tabaci* individuals. *Portiera* has a long co-evolutionary history with all members of the *Aleyrodinae* subfamily[16]. In this study, we further explore genes within the *P. aleyrodidarum* retrieved from individual whitefly transcriptomes, including the transcription termination/antitermination protein *Nus*G. *Nus*G is a highly conserved protein regulator that suppresses RNA polymerase, pausing and increasing the elongation rate[18,19]. However, its importance within gene regulation is species specific; in *Staphylococcus aureus* it is dispensable[20,21].

The S-endosymbionts are not systematically associated with hosts, and their contribution is not essential to the survival and reproduction. Seven facultative S-endosymbionts, *Wolbachia, Cardinium, Rickettsia, Arsenophonus, Hamiltonella defensa* and *Fritschea bemisae* and *Orientia-like organism* have been detected in various *B. tabaci* populations[14,22–25]. The presence of S-endosymbionts can influence key biological parameters of the host. *Hamiltonella* and *Rickettsia* facilitate plant virus transmission with increased acquisition and retention by whiteflies[25]. This is done by protection and safe transit of virions in the haemolymph of insects through chaperonins (*GroEL*) and protein complexes that aid in protein folding and repair mechanisms[20].

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

The advent of next generation sequencing (NGS) and specifically transcriptome sequencing has allowed the unmasking of this tripartite relationship of vector-viral-

microbiota within insects[24,26–28]. Furthermore, NGS provides an opportunity to better understand the co-evolution of *B. tabaci* and its bacterial endosymbionts[26]. The endosymbionts have been implicated in influencing species complex formation in *B. tabaci* through conducting sweeps on the mitochondrial genome[27]. Applying transcriptome sequencing is essential to reveal the endosymbionts and their effects on the mitogenome of *B. tabaci,* and predict potential hot spots for changes that are endosymbiont induced.

Several studies have explored the interaction between whitefly and endosymbionts[29,30] and have resulted in the identification of candidate genes that maintain the relationship[31,32]. This has been explored as a source of potential RNAi pesticide control targets[32–34]. RNAi-based pest control measures also provide opportunities to identify species-specific genes for target gene sequences for knock-down. However, to date all transcriptome sequencing has involved pooled samples, obtained through rearing several generations of isolines of a single species to ensure high quantities of RNA for subsequent sequencing. This remains a major bottle neck in particular within arthropoda, where collected samples are limited due to small morphological sizes[32]. In addition, the development of isolines is time consuming and often has colonies dying off mainly due to inbreeding depression[34].

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

Methods

Whitefly sample collection and study design

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

Extraction of total RNA from single whitefly

RNA extraction was carried out using the ARCTURUS® PicoPure® kit (Arcturus, CA, USA), which is designed for fixed paraffin-embedded (FFPE) tissue samples. Briefly, 30 µl of extraction buffer were added to an RNase-free micro centrifuge tube containing a single whitefly and ground using a sterile plastic pestle. To the cell extract an equal volume of 70% ethanol was added. To bind the RNA onto the column, the RNA purification columns were spun for two minutes at 100 x g and immediately followed by centrifugation at 16,000 x g for 30 seconds. The purification columns were then subjected to two washing steps using wash buffer 1 and 2 (ethyl alcohol). The purification column was transferred to a fresh RNase-free 0.5 ml micro centrifuge tube, with 30 µl of RNAse-free water added to elute the RNA. The column was incubated at room temperature for five minutes, and subsequently spun for one minute at 1,000 x g, followed by 16,000 x g for one minute. The eluted RNA was returned into the column and re-extracted to increase the concentration. Extracted RNA was treated with DNase using the TURBO DNA free kit, as described by the manufacturer (Ambion, Life Technologies, CA, USA). Concentration of RNA was done in a vacuum centrifuge (Eppendorf, Germany) at room temperature for 1 hour, the pellet was suspended in 15 µl of RNase-free water and stored at -80°C awaiting analysis. RNA was quantified, and the quality and integrity assessed using the 2100 Bioanalyzer (Agilent Technologies, CA, USA). Dilutions of up to x10 were made for each sample prior to analysis in the bioanalyzer.

cDNA and Illumina library preparation

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

Analysis of NGS data using the supercomputer

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

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

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

Sequence alignment and phylogenetic analysis of *Nus*G gene in *P. aleyrodidarum* across *B. tabaci* species: Sequence alignment of the *Nus*G gene from the P-endosymbiont *P. aleyrodidarum* from the SSA1 *B. tabaci* in this study was compared with another *B. tabaci* species, *Trialeurodes vaporariorum* and *Alerodicus dispersus* using MAFFT (version 7.017)[38]. The Jmodel version 2[37] was used to search for phylogenetic models with the Akaike information criterion selecting the optimal that was to be implemented in MrBayes 3.2.2. MrBayes run was carried out using the command: "lset nst=6 rates=gamma" for 50 million generations, with trees sampled every 1000 generations. In each of the runs, the first 25% (2,500) trees were discarded as burn in.

Analysis and modelling the structure of the *Nus*G gene The structures for *Portiera aleyrodidarum BT* and *B. tabaci* SSA1 whitefly were predicted using Phyre2[39] with 100% confidence and compared to known structures of *Nus*G from other bacterial species. All models were prepared using Pymol (The PyMOL Molecular Graphics System, Version 1.5.0.4).

Results

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

In this study, we sampled four individual adult *B. tabaci* from cassava fields in Uganda (WF2) and Tanzania (WF1, WF2a, WF2b). Total RNA from single whitefly yielded high quality RNA with concentrations ranging from 69 ng to 244 ng that were used for library preparation and subsequent sequencing with Illumina Hiseq 2000 on a rapid run mode. The number of raw reads generated from each single whitefly ranged between 39,343,141 and 42,928,131 (Table 1). After trimming, the reads were assembled using Trinity resulting in 65,550 to 162,487 transcripts across the four SSA1 *B. tabaci* individuals (Table 1).

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

Comparison of the diversity of bacterial endosymbionts across individual whitefly transcripts was conducted with BLASTn searches on the non-redundant nucleotide database and by identifying the number of genes from each bacterial endosymbi- ont (Supplementary Table 1). We identified five main endosym- bionts including: *P. aleyrodidarum*, the primary endosymbionts and four secondary endosymbionts: *Arsenophonus, Wolbachia, Rickettsia sp, and Cardinium* spp (Table 2). *P. aleyrodidarum* predominate across all four SSA1 *B. tabaci* study samples based on number of core gene families identified for WF1 (74.82%), WF2 (72.18%), WF2a (53.17%) and WF2b (71.70%). This was followed by *Arsenophonus*, *Wolbachia, Rickettsia sp,* and *Cardinium* spp, which occurred at an average of 18.0%, 5.9%, 1.6%

and <1%, respectively across all four study samples (Table 2). The secondary endosymbiont *Hamiltonella* had the least number of core genes (n=1) and was detected in only one of the SSA1 *B. tabaci* (WF2a) samples.

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

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

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

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

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

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

Sequence alignment and Bayesian phylogenetic analysis of *Nus*G gene

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

Structure analysis of Portiera *Nus*G genes

Structures of the *Nus*G protein sequence of the primary endosymbiont *P. aleyrodidarum* in the four SSA1 *B. tabaci* samples were predicated using Phyre2 with 100% confidence, and compared to known structures of *Nus*G from other bacterial species including (*Shigella flexneri, Thermus thermophilus* and *Aquifex aeolicus*; (PDB entries 2KO6, 1NZ8 and 1M1H, respectively) and Spt4/5 from yeast (*Saccharomyces cerevisiae;* PDB entry 2EXU)[18,40,41]. The 11-residue deletion was found in a loop region that is variable in length and structure across bacterial species, but is absent from archaeal and eukaryotic species (Figure 3 and Figure 4A). The effect of the deletion appears to shorten the loop in *Nus*G from the African whiteflies (WF2 and WF2b). A model of bacterial RNA polymerase (orange surface representation; PDB entry 2O5I) bound to the N-terminal domain of the *Thermus thermophilus Nus*G shows that the loop region is not involved in the interaction between *Nus*G and RNA polymerase (Figure 4B).

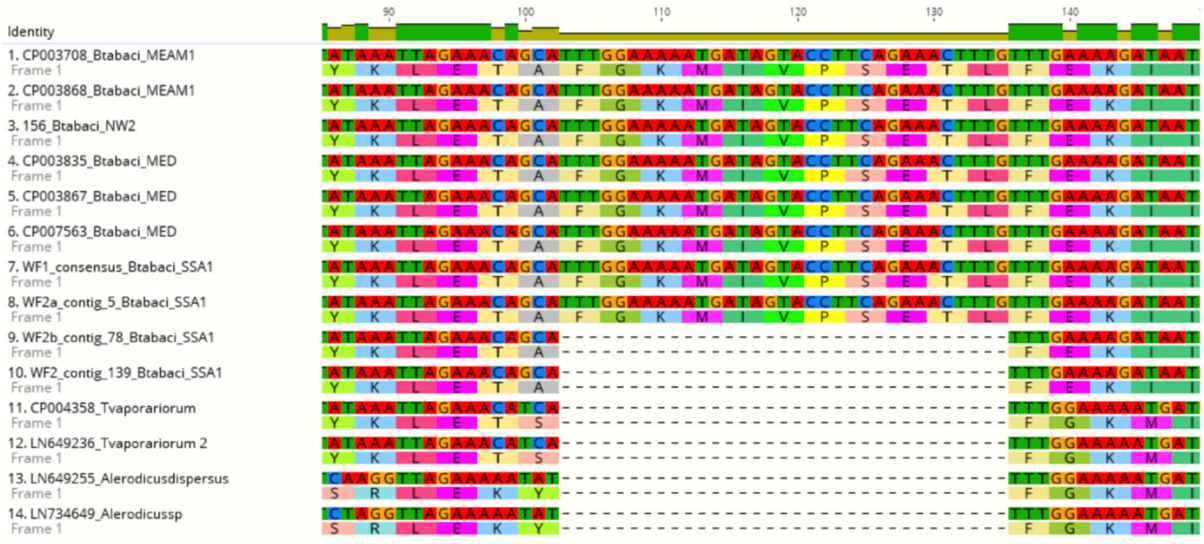


Figure 1. **Sequence alignment of nucleotide sequences of *Nus*G gene in *P. aleyrodidarum* across whitefly species sequences using MAFFT v 7.017.**

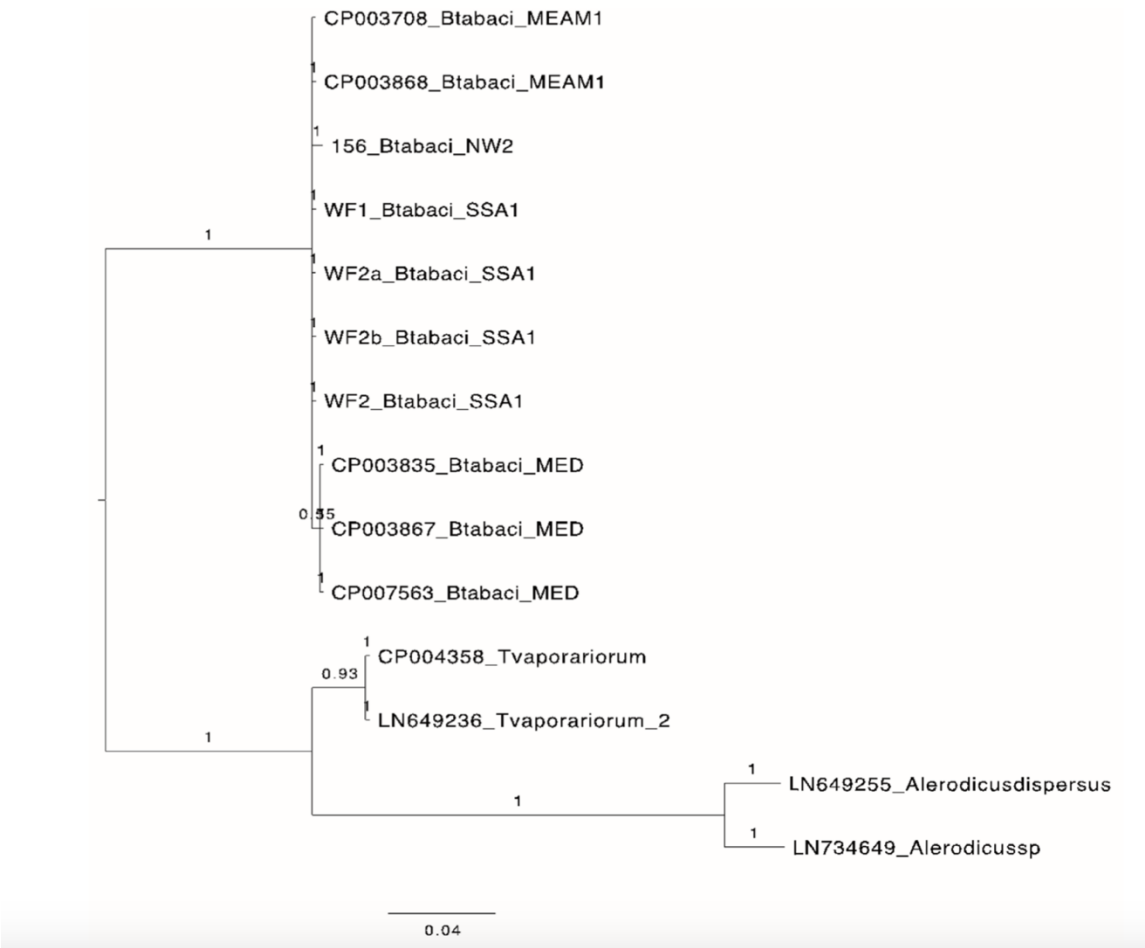


Figure 2. **Bayesian phylogenetic tree of *Nus*G gene of *P. aleyrodidarum* across whitefly species using MrBayes -3.2.2.**

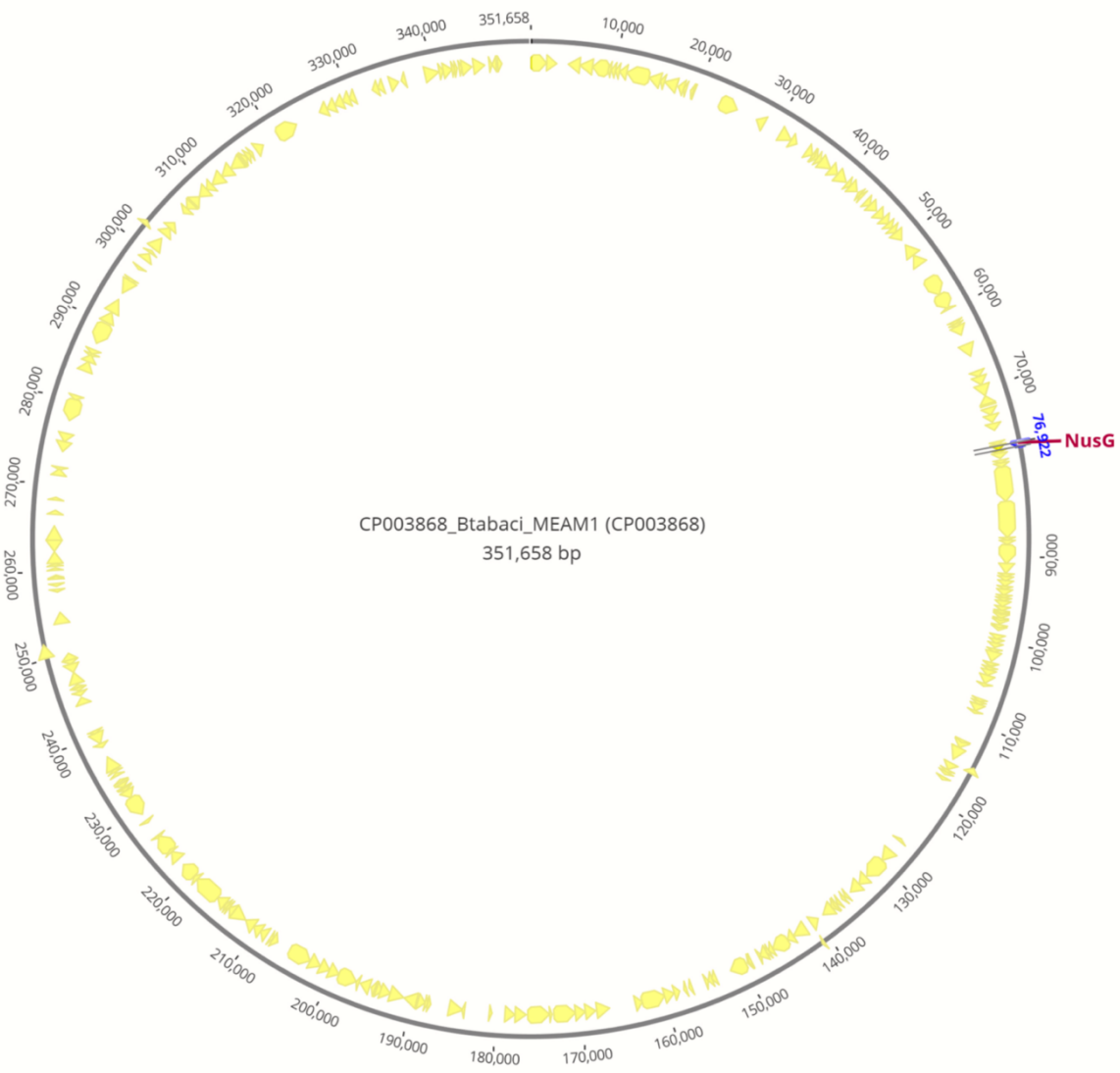


Figure 3. Primary endosymbiont *Portiera aleyrodidarum* whole genome from GenBank CP003868 showing the section of the NusG gene included in the analyses (position 76,922).

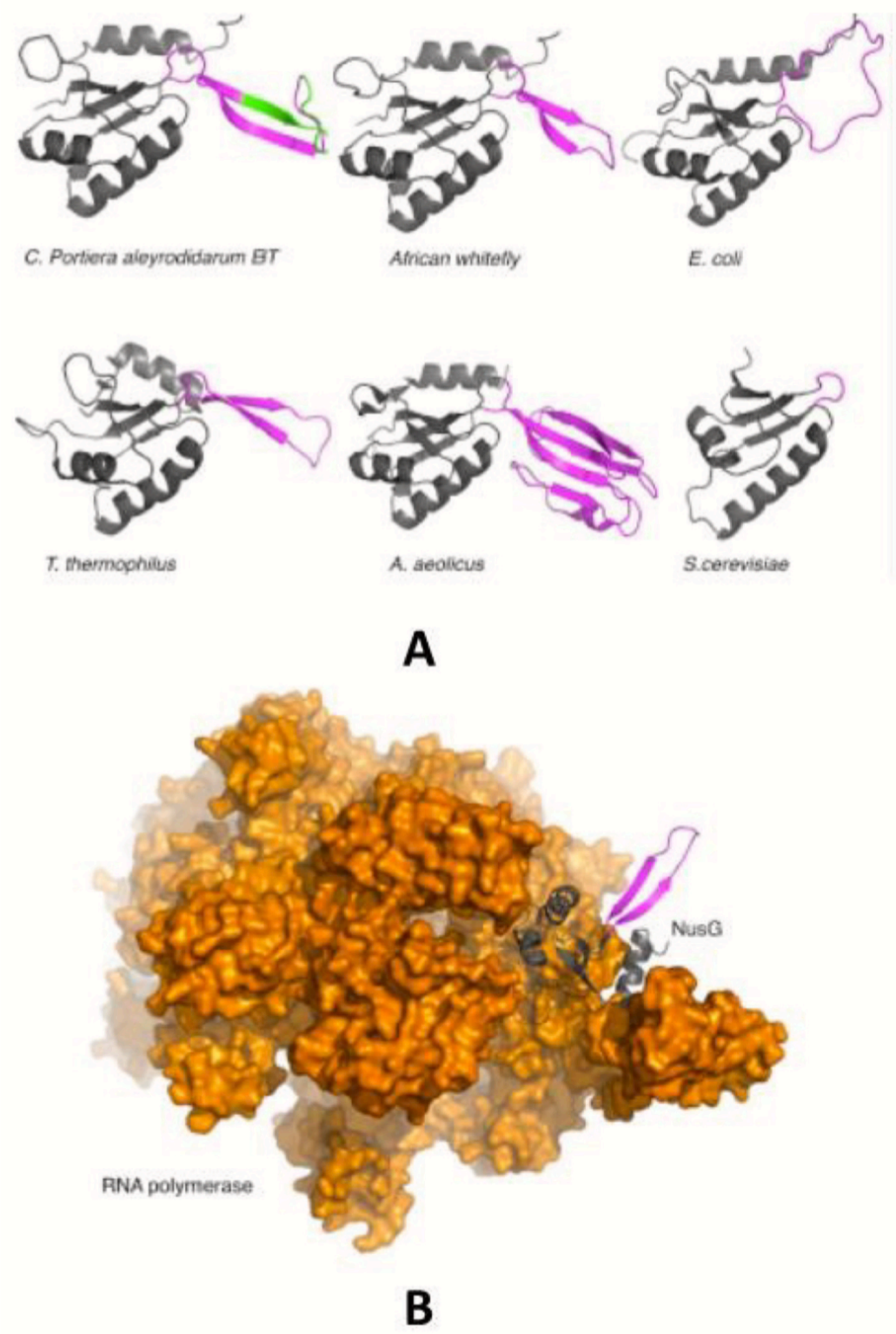


Figure 4. Structure analysis of *Nus*G from *P. aleyrodidarum* in *B. tabaci* and other endosymbionts. A. Phyre2 based structure prediction of *Nus*G from *Candidatus Portiera aleyrodidarum* in *B. tabaci* SSAI whitefly and comparisons to the structures of *Nus*G from other bacterial species as indicated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

Discussion

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

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

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

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

Why the single whitefly transcriptome approach?

The sequencing of the whitefly transcriptome is crucial in understanding whitefly-microbiota-viral dynamics and thus circumventing the bottlenecks posed in sequencing the whitefly genome. The genome of whitefly is highly heterozygous[42]. Assembling of heterozygous genomes is complex due to the de Bruijn graph structures predominantly used[43]. To deal with the heterozygosity, previous studies have employed inbred lines, obtained from rearing a high number of whitefly isolines[35,44]. However, rearing whitefly isolines is time consuming and often colonies may suffer contaminations, leading to collapse and failure to raise the high numbers required for transcriptome sequencing.

We optimised the ARCTURUS® PicoPure® kit (Arcturus, CA, USA) protocol for individual whitefly RNA extraction with the dual aim of determining if we could obtain sufficient quantities of RNA from a single whitefly for transcriptome analysis and secondly, determine whether the optimised method would reveal whitefly microbiota as proof of concept. Using our method, the quantities of RNA obtained from field-

collected single whitefly samples were sufficient for library preparation and subsequent transcriptome sequencing. Across all transcriptomes over 30M reads were obtained. The amount of transcripts were comparable to those reported in other arthropoda studies from field collections[32]. However, we did not observe any difference in assembly qualities[32]; probably due to the fact that our field- collected samples had degraded RNA based on RIN, and thus direct comparison with 32 was inappropriate.

Degraded insect specimen have been used successfully in previous studies[45]. This is significant, considering that the majority of insect specimens are usually collected under field conditions and stored in ethanol with different concentrations ranging from 70 to 100%[46–48] rendering the samples liable to degradation. However, to ensure good keeping of insect speci- men to be used for mRNA and total RNA isolation in molecular studies, and other downstream applications such as histology and immunocytochemistry, it is advisable to collect the samples in an RNA stabilizing solution such as RNAlater. The solution stabilizes and protects cellular RNA in intact, unfrozen tissue, and cell samples without jeopardizing the quality, or quantity of RNA obtained after subsequent RNA isolation. The success of the method provided an opportunity to unmask vector-microbiota-viral dynamics in individual whiteflies in our study, and will be useful for similar studies on other small organisms.

Endosymbionts diversity across individual SSA1 *B. tabaci* transcriptomes

In this study, we identified bacterial endosymbionts (Table 2) that were comparable to those previously reported in *B. tabaci*[49] and more specifically SSA1 on cassava[23,37]. Secondary endosym- bionts have been implicated with different roles within *B. tabaci*. *Rickettsia* has been adversely reported across putative *B. tabaci* species, including the Eastern African region[23,50,51]. This endo- symbiont has been associated with influencing thermo tolerance in *B. tabaci* species[49]. *Rickettsia* has also been associated with altering the reproductive system of *B. tabaci,* and within the females. This has been attributed to increasing fecundity, greater survival, host reproduction manipulation and the production of a higher proportion of daughters all of which increase the impact of virus[49]. In addition, *Rickettsia* and *Hamiltonella* play a role in plant virus transmission in whiteflies[25] by protecting the safe transit of virions in the haemolymph of insects through chaper- onins (*GroEL*) and protein complexes that aid in protein folding and repair mechanisms[20]. However, *Hamiltonella* was reported to be absent in the indigenous whitefly populations studied elsewhere[15,50,52–54] and in Malawi, Nigeria, Tanzania and Uganda[50,55] as also confirmed in our study. *Arsenophonus, Wolbachia, Arsenophonus* and *Cardinium spp* have been detected within MED and MEAM1 *Bemisia* species[14,50]. In addition, 50 and 22 reported *Arsenophonus* within SSA1 *B. tabaci* in Eastern Africa that were collected on cassava. These endosymbionts have been associated with several deleterious functions within *B. tabaci* that include manipulating female-male host ratio through feminizing genetic males, coupled with male killing[56,57].

Within the context of SSA agricultural systems, the role of endo- symbionts in influencing *B. tabaci* viral transmission is important. Losses attributed to *B. tabaci* transmitted viruses within different crops are estimated to be in billions of US dollars[46]. Bacterial endosymbionts have been associated with influencing viral acquisition, transmission and retention, such as in tomato leaf curl virus[58,59]. Thus, better understanding of the diversity of the endosymbionts provides additional evidence on

which members of *B. tabaci* species complex more proficiently trans- mit viruses, and thus the need for concerted efforts towards the whitefly eradication.

Conclusions

Our study provides a proof of concept that single whitefly RNA extraction and RNA sequencing is possible and the method could be optimised and applicable to a range of small insect transcriptome studies. It is particularly useful in studies that wish to explore vector-microbiota-viral dynamics at individual insect level rather than pooling of insects. It is useful where genetic material is both limited, as well as of low quality, which is applicable to most agriculture field collections. In addition, the single whitefly RNA sequencing technique described in this study offers new opportunities to understand the biology, and relative economic importance of the several whitefly species occurring in ecosystems within which food is produced in Sub- Saharan Africa, and will enable the efficient development and deployment of sustainable pest and disease management strategies to ensure food security in the developing countries. However, this method still requires further optimisation to recover viral reads, especially in cases with very low viral titre as observed in this study. Finally, future studies could use freshly collected whiteflies on CBSD-affected plants to increase the detection of the causal viruses.

Data availability

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

SRR5110306, SRR5110307, SRR5109958, KY548924, MG680297.

Competing interests

No competing interests were disclosed.

Grant information

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

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

Acknowledgements

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

References

1. De Barro PJ, Liu SS, Boykin LM, *et al.*: ***Bemisia tabaci*: a statement of species status.** *Annu Rev Entomol.* 2011; **56**: 1–19. PubMed Abstract | Publisher Full Text

2. Polston JE, Capobianco H: **Transmitting plant viruses using whiteflies.** *J Vis Exp.* 2013; **8**(81): e4332. PubMed Abstract | Publisher Full Text | Free Full Text

3. Jones DR: **Plant viruses transmitted by whiteflies.** *Eur J Plant Pathol.* 2003; **109**: 195–219. Publisher Full Text

4. Vassylyev DG, Vassylyeva MN, Perederina A, *et al.*: **Structural basis for transcription elongation by bacterial RNA polymerase.** *Nature.* 2007; **448**(7150): 157–162. PubMed Abstract | Publisher Full Text

5. Vassiliou V, Emmanouilidou M, Perrakis A, *et al.*: **Insecticide resistance in *Bemisia tabaci* from Cyprus.** *Insect Sci.* 2011; **18**(1): 30–39. Publisher Full Text

6. FAO: **Save and Grow: Cassava. A Guide to Sustainable Production Intensification.** 2013. Reference Source

7. Fauquet CM, Briddon RW, Brown JK, *et al.*: **Geminivirus strain demarcation and nomenclature.** *Arch Virol.* 2008; **153**(4); 738–821. PubMed Abstract | Publisher Full Text

8. Patil BL, Fauquet CM: **Cassava mosaic geminiviruses: actual knowledge and perspectives.** *Mol Plant Pathol.* 2009; **10**: 685–701. PubMed Abstract | Publisher Full Text

9. Legg JP, Owor B, Sseruwagi P, *et al.*: **Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic.** *Adv Virus Res.* 2006; **67**(6): 355–418. PubMed Abstract | Publisher Full Text

10. Maruthi MN, Hillocks RJ, Mtunda K, *et al.*: **Transmission of *Cassava brown streak virus* by *Bemisia tabaci* (Gennadius).** *J Phytopathol.* 2005; **153**(5): 307–312. Publisher Full Text

11. Mware B, Narla R, Amata R, *et al.*: **Efficiency of cassava brown streak virus transmission by two whitefly species in coastal Kenya.** *J Gen Mol Virol.* 2009; **1**(4): 40–45. Reference Source

12. Ndunguru J, Sseruwagi P, Tairo F, *et al.*: **Analyses of Twelve New Whole Genome Sequences of Cassava Brown Streak Viruses and Ugandan Cassava Brown Streak Viruses from East Africa: Diversity, Supercomputing and Evidence for Further Speciation.** *PLoS One.* 2015; **10**(10): 1–18. PubMed Abstract | Publisher Full Text | Free Full Text

13. Alicai T, Ndunguru J, Sseruwagi P, *et al.*: ***Cassava brown streak virus* has a rapidly evolving genome: implications for virus speciation, variability, diagnosis and host resistance.** *Sci Rep.* 2016; **6**: 36164. PubMed Abstract | Publisher Full Text | Free Full Text

14. Gueguen G, Vavre F, Gnankine O, *et al.*: **Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex.** *Mol Ecol.* 2010; **19**(19): 4365–4378. PubMed Abstract | Publisher Full Text

15. Marubayashi JM, Kliot A, Yuki VA, *et al.*: **Diversity and Localization of Bacterial Endosymbionts from Whitefly Species Collected in Brazil.** *PLoS One.* 2014; **9**(9): e108363. PubMed Abstract | Publisher Full Text | Free Full Text

16. Thao ML, Baumann P: **Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts.** *Appl Environ Microbiol.* 2004; **70**(6): 3401–3406. PubMed Abstract | Publisher Full Text | Free Full Text

17. Mooney RA, Schweimer K, Rösch P, *et al.*: **Two structurally independent domains of *E. coli* NusG create regulatory plasticity *via* distinct interactions with RNA polymerase and regulators.** *J Mol Biol.* 2009; **391**(2): 341–358. PubMed Abstract | Publisher Full Text | Free Full Text

18. Yakhnin AV, Murakami KS, Babitzke P: **NusG Is a Sequence-specific RNA Polymerase Pause Factor That Binds to the Non-template DNA within the Paused Transcription Bubble.** *J Biol Chem.* 2016; **291**(10): 5299–5308. PubMed Abstract | Publisher Full Text | Free Full Text

19. Zchori-Fein E, Brown JK: **Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae).** *Ann Entomol Soc Am.* 2002; **95**(6): 711–718. Publisher Full Text

20. Gottlieb Y, Ghanim M, Chiel E, *et al.*: **Identification and localization of a *Rickettsia* sp. in *Bemisia tabaci* (Homoptera: Aleyrodidae).** *Appl Environ Microbiol.* 2006; **72**(5): 3646–3652. PubMed Abstract | Publisher Full Text | Free Full Text

21. Marubayashi JM, Yuki VA, Rocha KC, *et al.*: **At least two indigenous species of the *Bemisia tabaci* complex are present in Brazil.** *J Appl Entomol.* 2013; **137**(1–2): 113–121. Publisher Full Text

22. Ghosh S, Mitra PS, Loffredo CA, *et al.*: **Transcriptional profiling and biological pathway analysis of human equivalence PCB exposure *in vitro*: Indicator of disease and disorder development in humans.** *Environ Res.* 2015; **138**: 202–216. PubMed Abstract | Publisher Full Text | Free Full Text

23. Rosario K, Marr C, Varsani A, *et al.*: **Begomovirus-Associated Satellite DNA Diversity Captured Through Vector-Enabled Metagenomic (VEM) Surveys Using Whiteflies (Aleyrodidae).** *Viruses.* 2016; **8**(2): pii: E36. PubMed Abstract | Publisher Full Text | Free Full Text

24. Rosario K, Seah YM, Marr C, *et al.*: **Vector-Enabled Metagenomic (VEM) Surveys Using Whiteflies (Aleyrodidae) Reveal Novel Begomovirus Species in the New and Old Worlds.** *Viruses.* 2015; **7**(10): 5553–5570. PubMed Abstract | Publisher Full Text | Free Full Text

25. Kliot A, Cilia M, Czosnek H, *et al.*: **Implication of the bacterial endosymbiont *Rickettsia* spp. in interactions of the whitefly *Bemisia tabaci* with *tomato yellow leaf curl virus*.** *J Virol.* 2014; **88**(10): 5652–5660. PubMed Abstract | Publisher Full Text | Free Full Text

26. Poelchau MF, Coates BS, Childers CP, *et al.*: **Agricultural applications of insect ecological genomics.** *Curr Opin Insect Sci.* 2016; **13**: 61–69. PubMed Abstract | Publisher Full Text

27. Kapantaidaki DE, Ovčarenko I, Fytrou N, *et al.*: **Low levels of mitochondrial DNA and symbiont diversity in the worldwide agricultural pest, the greenhouse whitefly *Trialeurodes vaporariorum* (Hemiptera: Aleyrodidae).** *J Hered.* 2015; **106**(1): 80–92. PubMed Abstract | Publisher Full Text 28. Morin S, Ghanim M, Sobol I, *et al.*: **The GroEL protein of the whitefly *Bemisia tabaci* interacts with the coat protein of transmissible and nontransmissible begomoviruses in the yeast two-hybrid system.** *Virology.* 2000; **276**(2): 404–416. PubMed Abstract | Publisher Full Text

29. Xue J, Zhou X, Zhang CX, *et al.*: **Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation.** *Genome Biol.* 2014; **15**(12): 521. PubMed Abstract | Publisher Full Text | Free Full Text

30. Kim JK, Won YJ, Nikoh N, *et al.*: **Polyester synthesis genes associated with stress resistance are involved in an insect-bacterium symbiosis.** *Proc Natl Acad Sci U S A.* 2013; **110**(26): E2381–9. PubMed Abstract | Publisher Full Text | Free Full Text

31. Wang XW, Zhao QY, Luan JB, *et al.*: **Analysis of a native whitefly transcriptome and its sequence divergence with two invasive whitefly species.** *BMC Genomics.* 2012; **13**(1): 529. PubMed Abstract | Publisher Full Text | Free Full Text

32. Kono N, Nakamura H, Ito Y, *et al.*: **Evaluation of the impact of RNA preservation methods of spiders for *de novo* transcriptome assembly.** *Mol Ecol Resour.* 2016; **16**(3): 662–672. PubMed Abstract | Publisher Full Text

33. Xue X, Li SJ, Ahmed MZ, *et al.*: **Inactivation of *Wolbachia* reveals its biological roles in whitefly host.** *PLoS One.* 2012; **7**(10): e48148. PubMed Abstract | Publisher Full Text | Free Full Text

34. Charlesworth D, Willis JH: **The genetics of inbreeding depression.** *Nat Rev Genet.* 2009; **10**(11): 783–96. PubMed Abstract | Publisher Full Text

35. Grabherr MG, Haas BJ, Yassour M, *et al.*: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nat Biotechnol.* 2011; **29**(7): 644–52. PubMed Abstract | Publisher Full Text | Free Full Text

36. Huelsenbeck JP, Andolfatto P, Huelsenbeck ET: **Structurama: bayesian**

inference of population structure. *Evol Bioinform Online.* 2011; **7**: 55–9. PubMed Abstract | Publisher Full Text | Free Full Text

37. Darriba D, Taboada GL, Doallo R, *et al.*: **jModelTest 2: more models, new heuristics and parallel computing.** *Nat Methods.* 2012; **9**(8): 772. PubMed Abstract | Publisher Full Text | Free Full Text

38. Katoh K, Misawa K, Kuma K, *et al.*: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids Res.* 2002; **30**(14): 3059–3066. PubMed Abstract | Publisher Full Text | Free Full Text

39. Kelley LA, Mezulis S, Yates CM, *et al.*: **The Phyre2 web portal for protein modeling, prediction and analysis.** *Nat Protoc.* 2015; **10**(6): 845–858. PubMed Abstract | Publisher Full Text | Free Full Text

40. Boykin LM, Shatters RG Jr, Rosell RC, *et al.*: **Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences.** *Mol Phylogenet Evol.* 2007; **44**(3): 1306–1319. PubMed Abstract | Publisher Full Text

41. Hsieh CH, Ko CC, Chung CH, *et al.*: **Multilocus approach to clarify species status and the divergence history of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex.** *Mol Phylogenet Evol.* 2014; **76**: 172–180. PubMed Abstract | Publisher Full Text

42. Xie W, Meng QS, Wu QJ, *et al.*: **Pyrosequencing the *Bemisia tabaci* transcriptome reveals a highly diverse bacterial community and a robust system for insecticide resistance.** *PLoS One.* 2012; **7**(4): e35181. PubMed Abstract | Publisher Full Text | Free Full Text

43. Kajitani R, Toshimoto K, Noguchi H, *et al.*: **Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads.** *Genome Res.* 2014; **24**(8): 1384–1395. PubMed Abstract | Publisher Full Text | Free Full Text

44. Leshkowitz D, Gazit S, Reuveni E, *et al.*: **Whitefly (*Bemisia tabaci*) genome project: analysis of sequenced clones from egg, instar, and adult (viruliferous and non-viruliferous) cDNA libraries.** *BMC Genomics.* 2006; **7**: 79. PubMed Abstract | Publisher Full Text | Free Full Text

45. Gallego Romero I, Pai AA, Tung J, *et al.*: **RNA-seq: impact of RNA degradation on transcript quantification.** *BMC Biol.* 2014; **12**(1): 42. PubMed Abstract | Publisher Full Text | Free Full Text

46. Legg JP, Sseruwagi P, Boniface S, *et al.*: **Spatio-temporal patterns of genetic change amongst populations of cassava *Bemisia tabaci* whiteflies driving virus pandemics in East and Central Africa.** *Virus Res.* 2014; **186**: 61–75. PubMed Abstract | Publisher Full Text

47. Wainaina JM, De Barro P, Kubatko L, *et al.*: **Global phylogenetic relationships, population structure and gene flow estimation of *Trialeurodes vaporariorum* (Greenhouse whitefly).** *Bull Entomol Res.* 2018; **108**(1): 5–13. PubMed Abstract | Publisher Full Text

48. Tajebe LS, Boni SB, Guastella D, *et al.*: **Abundance, diversity and geographic distribution of cassava mosaic disease pandemic-associated *Bemisia tabaci* in Tanzania.** *J Appl Entomol.* 2015; **139**(8): 627–637. Publisher Full Text

49. Rao Q, Luo C, Zhang H, *et al.*: **Distribution and dynamics of *Bemisia tabaci* invasive biotypes in central China.** *Bull Entomol Res.* 2011; **101**(1): 81–88. PubMed Abstract | Publisher Full Text

50. Tajebe LS, Guastella D, Cavalieri V, *et al.*: **Diversity of symbiotic bacteria associated with *Bemisia tabaci* (Homoptera: Aleyrodidae) in cassava mosaic disease pandemic areas of Tanzania.** *Ann Appl Biol.* 2015; **166**(2): 297–310. Publisher Full Text

51. Rao Q, Rollat-Farnier PA, Zhu DT, *et al.*: **Genome reduction and potential metabolic complementation of the dual endosymbionts in the whitefly *Bemisia tabaci*.** *BMC Genomics.* 2015; **16**(1): 226. PubMed Abstract | Publisher Full Text | Free Full Text

52. Bing XL, Yang J, Zchori-fein E: **Characterization of a newly discovered symbiont of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae).** *Appl Environ Microbiol.* 2013; **79**(2): 569–75. PubMed Abstract | Publisher Full Text | Free Full Text

53. Singh ST, Priya NG, Kumar J: **Diversity and phylogenetic analysis of endosymbiotic bacteria from field caught *Bemisia tabaci* from different locations of North India based on 16S rDNA library screening.** *Infect Genet Evol.* 2012; **12**(2): 411–9. PubMed Abstract | Publisher Full Text

54. Gnankine O, Mouton L, Henri H, *et al.*: **Distribution of *Bemisia tabaci* (Homoptera: Aleyrodidae) biotypes and their associated symbiotic bacteria on host plants in West Africa.** *Insect Conserv Divers.* 2013; **6**(3): 411–421. Publisher Full Text

55. Ghosh S, Bouvaine S, Maruthi MN: **Prevalence and genetic diversity of endosymbiotic bacteria infecting cassava whiteflies in Africa.** *BMC Microbiol.* 2015; **15**(1): 93. PubMed Abstract | Publisher Full Text | Free Full Text

56. Brumin M, Kontsedalov S, Ghanim M: ***Rickettsia* influences thermotolerance in the whitefly *Bemisia tabaci* B biotype.** *Insect Sci.* 2011; **18**(1): 57–66. Publisher Full Text

57. Brumin M, Levy M, Ghanim M: **Transovarial transmission of *Rickettsia* spp. and organ-specific infection of the whitefly *Bemisia tabaci*.** *Appl Environ Microbiol.* 2012; **78**(16): 5565–5574. PubMed Abstract | Publisher Full Text | Free Full Text

58. Rosario K, Capobianco H, Ng TF, *et al.*: **RNA viral metagenome of whiteflies leads to the discovery and characterization of a whitefly-transmitted carlavirus in North America.** *PLoS One.* 2014; **9**(1): e86748. PubMed Abstract | Publisher Full Text | Free Full Text

59. Himler AG, Adachi-Hagimori T, Bergen JE: **Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias.** *Science.* 2011; **332**(6026): 254–6. PubMed Abstract | Publisher Full Text