African Cassava whitefly, *Bemisia tabaci*, systematics and patterns of molecular evolution

Tonny Kinene

BSc. Education, PGD IT, MSc. Mathematical modelling

THE UNIVERSITY OF WESTERN AUSTRALIA

This thesis is presented for the degree of Doctor of Philosophy of The University of Western Australia,

School of Molecular Sciences

March 2019

Thesis Declaration

This thesis is presented as a series of scientific manuscripts that include one published book chapter in chapter 2 and three other manuscripts. One of the manuscripts, chapter 3, has already been submitted to a journal for peer review. The rest of the manuscripts will be submitted to the relevant journals after thesis submission by the end of March 2019.

I declare that this thesis is my own original work and the results come from my own research while enrolled in the degree of Doctor of Philosophy at the University of Western Australia, School of Molecular Sciences. This work has not been submitted for any degree award in any other University.

The contributions of the co-authors in the manuscripts listed in this thesis are associated with sample collection, research direction, and editorial input in the manuscripts. I have permission from all the co-authors to include the listed manuscript in my thesis.

PhD Candidate

Tonny Kinene

_____

Abstract

The whitefly *Bemisia tabaci* is a species complex consisting of more than 34 morphologically indistinguishable species. *B. tabaci* is known as a global pest and is a vector of more than 200 plant viruses. In this thesis, I focused on whitefly species collected from cassavas plant from Uganda, Tanzania and Malawi in relation to other *B. tabaci* species in the world. Cassava (*Manihot esculenta*) is a major staple crop in Africa. Many smallholder farmers rely on cassava as a food security crop and a cash crop because of its tolerance to drought, poor soils, and low rainfall, as well as its low labour intensity and long harvesting window. The tuberous roots of cassava are eaten as food and are rich in carbohydrates, while the leaves are eaten as vegetables and are rich in protein. Unfortunately, cassava is highly vulnerable to both whiteflies and the viruses they transmit (Cassava mosaic *Geminiviruses*, Uganda cassava brown streak virus and cassava brown streak virus).

The research questions we seek to answer are:

(i) What role does the gene HSP90 play in *B. tabaci*'s ability to adapt to varying climate? Could it possible play a role in the superabundance phenomenon observed in east Africa?

(ii) What are the phylogenetic relationships of the various *B. tabaci* species utilising the common nuclear genes RNA polymerase II and Shaker cognate gene w?

(iii) What species tree of the *B. tabaci* species is inferred from thousands of nuclear genes generated from the transcriptomes?

Chapter 1 is an introduction to whitefly (*B. tabaci*) biology, systematics, economic importance, and genetic diversity, and to cassava viruses. It also describes the objectives of the study to prepare the background for the next five chapters. Chapter 2 deals with methods used to root phylogenetic trees, as phylogenetic trees are one of the tools we used in our further analysis. Therefore, understanding the importance of finding the correct root for the tree is advantageous, because incorrectly rooted phylogenetic trees might mislead evolutionary and taxonomic inferences. In chapter 3, the *B. tabaci* heat shock protein 90 (HSP90) was identified from next generation sequence data generated from 21 whitefly transcriptomes. The aim here was to identify genetic signatures for a superabundance of whiteflies. We detected recombination signals in the HSP90 gene for members of the *B. tabaci* species complex and we conclude that recombination might be driving the evolution of *B. tabaci* species. In chapter 4, genomic regions of interest were identified (RNA polymerase II and Shaker cognate gene w) and used to infer evolutionary relationships of *B. tabaci* species. Patterns of conflict were found between mitochondrial cytochrome oxidase 1 (mtCOI) and nuclear DNA segments. In chapter 5, the species tree for the *B. tabaci* complex was estimated using transcriptome data. Three thousand orthologous genes were identified using the Agalma pipeline and subsequently analysed in SVDQuartets and in RAxML to generate gene trees that were used in ASTRAL for species tree estimation. We compared the tree topologies from ASTRAL and SVDQuartets and also compared our results with the conventional mtCOI topology. We noted a difference in the grouping of the *B. tabaci* Uganda species in the mtCOI conventional tree compared to the species tree estimated from transcriptome data. Finally, chapter 6 presents a general discussion and the conclusions following from this research.

Table of Contents

Acknowledgements

Special thanks to the University of Western Australia, Scholarship for International Research Fees (SIRF), University Postgraduate Awards (UPA) and the University of Western Australia Safety Net top-up scholarship for the financial support offered to me to further my PhD studies. I thank the Natural Resources Institute, University of Greenwich for supporting the first two years of my PhD through a grant provided by the Bill and Melinda Gates Foundation (Investment ID OPP1058938).

Special thanks to my principle supervisor Dr Laura Boykin from the University of Western Australia (UWA) School of Molecular Sciences (SMS) and the Australian Research Council Centre of Excellence (ARC CoE) in Plant Energy Biology (PEB) for the continued support, guidance, mentorship, patience and proofreading of the numerous drafts of this thesis. I would also like to thank Prof. Ian Small from UWA, ARC CoE PEB, for the guidance and support during the course of my PhD.

My gratitude to Prof. Laura Kubatko from Ohio State University Department of Statistics, Evolution, Ecology and Organismal Biology for the support and guidance in reviewing the manuscripts and for making me feel at home as a mathematician.

Sincere thanks to my supervisor Prof. Livingstone S. Luboobi from Makerere University Department of Mathematics for the mentorship, invaluable comments and encouragement, and for making me a great research scientist.

I would like to thank my other supervisors, Dr Chris Abu Omongo and Dr Titus Alicai, of the National Research Organisation (NARO) and National Crop Resources Research

Institute (NaCRRI) in Uganda for their support, guidance and making the sample collection exercise a smooth experience.

Special thanks to Dr. Joseph Nduguru and Dr. Peter Sseruwagi of the Mikocheni Agricultural Research Organisation, Tanzania for the invaluable support and organisation of sample collection trips (field work). Sincere thanks to Anders Savill for setting up the server, and helping in installing Bioinformatics software. My gratitude to Dr. Monica Kehoe for her scientific guidance and assistance while collecting whiteflies in Tanzania, great thanks to the comrades with whom we started the project and those who helped me in building my laboratory skills: James Mbora Wainaina, Bruno Rossito De Marchi and all the PEB members.

Finally I would like to thank the PEB and SMS staff and support team, especially the purchasing team Nehal and Adam for making my purchase and deliveries on time. Proofreading of this thesis was done by Pia Smith (www.piacharlottesmith.com).

To my mother, sister Pauline, Joan, Dianah, Mary, Asha, uncle David K and my girl friend Shilla. Thank you for giving me all the moral support and for allowing me to be away whenever you needed me.

Statement of Contribution

This thesis contains work that has been published and/or prepared for publication.

Details of the work: **Kinene, T**., Wainaina, J., Maina, S., and Boykin, L.M. (2016) Rooting Trees, Methods for. In: Kliman, R.M. (ed.), Encyclopedia of Evolutionary Biology. vol. 3, pp. 489–493. Oxford: Academic Press
Location in thesis: Chapter 2
Student contribution to work: Phylogenetic analysis, literature search and writing the manuscript

Details of the work: **Tonny Kinene***, Bruno Rossito De Marchi, Titus Alicai, Livingstone S. Luboobi, Christopher Abu Omongo, Anders Savill and Laura M. Boykin. (2019). Recombination detected in the Heat Shock Protein 90 (HSP90) of the *Bemisia tabaci* species complex. *Submitted to BMC Evolutionary Biology*.
Location in thesis: Chapter 3
Student contribution to work: Sample collection and RNA extraction, experimental design, NGS data analysis, recombination analysis, phylogenetic analysis and writing of the manuscript.

Details of the work: **Tonny Kinene***, Laura Kubatko, Titus Alicai, Christopher Abu Omongo, Livingstone S. Luboobi, and Laura M. Boykin. (2019). Patterns of conflict between nuclear and mitochondrial cytochrome oxidase I (mtCOI) phylogenies in the *Bemisia tabaci* species complex. *To be submitted to PlosOne*
Location in thesis: Chapter 4
Student contribution to work: Experimental design, DNA extraction from individual whiteflies, PCR amplification with mtCoI and nuclear primers, phylogenetic analysis and writing

Details of the work: **Tonny Kinene***, Laura Kubatko, Titus Alicai, Bruno Rossito De Marchi, Livingstone S. Luboobi, Peter Sseruwagi, Joseph Ndunguru and Laura M. Boykin. (2019). Species tree estimation for the global *Bemisia tabaci* species complex using 3000 nuclear genes**.** *To be submitted to Molecular Phylogenetics and Evolution.*
Location in thesis: Chapter 5
Student contribution to work:

Sample collection from East Africa, experimental design, RNA extraction, cDNA library preparing, NGS data analysis, phylogenetic analysis and writing of the manuscript.

Student signature: [redacted]
Date: 25th March 2019

I, Laura Boykin certify that the student's statements regarding their contribution to each of the works listed above are correct.

As all co-authors' signatures could not be obtained, I hereby authorise inclusion of the co-authored work in the thesis.

Coordinating supervisor signature: [redacted]
Date: 27 March 2019

Chapter 1

1.1 General introduction

1.1.1 Whitefly Biology

The whiteflies *Bemisia tabaci* (Hemiptera; Aleyrodidae) are small insects that typically feed on the underside of plant leaves by sucking sap. They have wings that are less than 3 mm and a body length of at most 2 mm (Martin, 2004). These whiteflies are complicated to control both in the greenhouse and in the field because of their small sizes. They can easily go through the fine mesh of a greenhouse screen. Whiteflies have six life stages: eggs, four instars, and adult (Chapman, 1998). The eggs are ovoid, elongated-oval and are oviposited by an adult female either singly, scattered, grouped, or in a semi-circular, spiral or circle pattern (Dowell, 1981). The number of eggs oviposited by the adult female is temperature and season dependent but it has been reported to be at least 300 eggs from an individual female (Dittrich et al., 1986).

The eggs hatch into crawlers (first instar): these are elliptical in shape with an initially flattened dorsum that becomes convex on feeding. This is the only mobile stage and moves looking for feeding sites on the plant leaves. Therefore, the crawler possesses functional walking legs, short antennae and conspicuous eyes (Byrne and Bellows, 1991; Gill, 1990).

The next three instars are sessile and complete their life cycle on the same leaf (McAuslane, 2002). The second and third instars may be covered by white wax secretions (Gill, 1990). At the entry of the fourth instar, the whitefly feeds with inserted mouthparts that are later withdrawn at the end of the fourth instar; thus, the insect stops feeding until the adult emerges (Gelman et al., 2002). After development is completed

in the fourth instar, adult whiteflies emerge with an average length of 1 to 2 mm. The males are usually smaller than the females (Byrne and Bellows, 1991; Gill, 1990).

Whiteflies have an arrhenotokous parthenogenesis reproductive system where the fertilised eggs develop into diploid females and the unfertilised eggs into haploid males (Byrne and Bellows, 1991). Overall the life cycle of whiteflies depends on the temperature, season, host plant and species. The ratio of males to females is usually 1:2 under field conditions (Sharaf and Batta, 1985). The lifespan of female whiteflies differs from that of males, with females living up to 35 days while male live 20 days (Butler et al., 1983).

1.1.2 Systematics of *B. tabaci*

The whiteflies *B. tabaci* are a species complex composed of at least 34 species that are morphologically indistinguishable (Boykin et al., 2012; Boykin et al., 2007; Liu et al., 2012). The *B. tabaci* species complex are known as serious pests for agriculture (Gill and Brown, 2009) and vector for more than 200 plant viruses (Polston et al., 2014). They cause extensive damage to major food staples, such as cassava, through direct feeding by both the juveniles and adult whiteflies and indirect damage through transmission of viruses (Brown et al., 1995a; Hoddle, 1999). Apart from transmission of viruses, the whiteflies also excrete honeydew, which covers the leaf surfaces and develops into black sooty mold that interferes with the plant's photosynthesis (Omongo et al., 2012).

The genetic diversity of the *B. tabaci* species complex was first realised in the 1950s when morphologically indistinguishable populations of *B. tabaci* were reported to differ

in host range and capacity to transmit *Begomoviruses* (Costa and Russell, 1975; Bird, 1957). This led to the coining of the terms 'biotypes', 'strain' or 'races', which were used to differentiate between the different genetic groups of *B. tabaci*. The term 'host race' was used to describe the feeding patterns of the *B. tabaci* species (Bird, 1957). In the 1980s the term 'biotype' was substituted for race (Costa and Brown, 1990) but the meaning of the concept did not fundamentally change from the observed phenotypic differences between the *B. tabaci* species.

In 1986, the outbreak of *B. tabaci* on ornamental plants in Florida, and in 1988 in Arizona (Byrne and Miller, 1990; Price et al., 1987), led to the understanding of the difference between *B. tabaci* biotypes A and B. The term biotype was used based on phenotypic differences: host ranges, preferences, and virus transmitted. For example B biotype was distinguished with its ability to induce silvering on leaves of pumpkin plants (Costa and Brown, 1991). A number of phenotypic differences were noticed between *B. tabaci* biotypes and this led to the proposition that *B. tabaci* should be described as a species complex (Brown et al., 1995b).

Originally, the morphological similarities within *B. tabaci* species made people think that it was a single species (Macfadyen et al., 2018). However, based on the genetic diversity revealed by molecular techniques (partial fragment of mitochondrial cytochrome oxidase I gene) (Boykin et al., 2007; Sseruwagi et al., 2005; Boykin et al., 2013) and mating experiments (Colvin et al., 2004; Liu and Pearl, 2007; Xu et al., 2010), *B. tabaci* is now recognised as a species complex composed of at least 34 species (Boykin et al., 2012).

The rapid identification of species within the *B. tabaci* species complex led to many nomenclature changes, causing confusion in the literature over the last decade (Boykin et al., 2018). Many names were used for the same species and this made it very difficult to compare studies of ecological importance across time (Boykin et al., 2018). For example, in 2004 Sub-Saharan Africa (SSA) samples were classified as Sub-Saharan I–V (Berry et al., 2004). However, Sseruwagi et al. (2005) described eight genetic clusters as Ug 1–8 of SSA species, leading to more confusion in the literature (Boykin et al., 2018). Further confusion occurred when Ug 1 of Sseruwagi et al. (2005) was subdivided into SSA1 subclade 1 (Mugerwa et al., 2012) and SSA1 subgroup 1 (Legg et al., 2014b). Adding to the confusion in the nomenclature was the description of Ug 2 by Sseruwagi et al. (2005) as SSA2 (Mugerwa et al., 2012; Legg et al., 2014b), Banks et al. (1999) made it worse by calling the same species the S biotype.

Boykin et al. (2018) recommends using a unified nomenclature across all the *B. tabaci* species and further points out avoiding subclades/groups until confirmation from mating studies and molecular techniques. Inadequate sampling caused the confusion in the nomenclature of *B. tabaci* species and this was due to the fact that earlier studies failed to include all reported mitochondrial cytochrome oxidase I (mtCOI) sequences in their phylogenetic analysis (Boykin et al., 2018; Macfadyen et al., 2018). In order to manage whitefly outbreaks in East Africa, we need to get an accurate naming system of the members of the *B. tabaci* species complex.

1.1.3 Cassava and cassava viruses transmitted by *B. tabaci*

Cassava (*Manihot esculenta*) is a staple food crop for million of families in Africa. It originated from South America and was introduced to sub-Saharan African by the Portuguese in the 16th Century (Fauquet and Fargette, 1990). Cassava is tolerant to drought and can be planted in different agro-ecological zones where other crops may fail. This makes it the best crop to tackle hunger in uncertain climatic conditions (Bennett, 2015; Alabi et al., 2011), and many smallholder famers heavily rely on it as a food security crop and a cash crop. The tuberous roots of cassava are a source of dietary starch and are eaten fresh, cooked or processed depending on the cassava variety. In most cases, cassava is intercropped with vegetables, plantains, groundnuts and other legumes. Cassava can be harvested between 6 months and 3 years after planting; therefore, it can be harvested whenever there is demand. This qualifies it as the best food security crop.

About 277 million tonnes of cassava are produced annually in the world, of which 158 million tonnes come from Africa, and about 54 million tonnes come from Nigeria – this makes Nigeria the largest producer of cassava in Africa. Surprisingly, most of the cassava produced in Africa is not traded but rather consumed domestically (Bennett, 2015). Despite the importance of cassava as a food security crop and a cash crop its production is constrained by whiteflies and the viruses they transmit. In East Africa the whitefly is a vector for two major cassava viruses: cassava mosaic *Geminiviruses* (CMGs) and cassava brown streak viruses (CBSV). These viruses have led to two ongoing pandemics of cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) (Alicai et al., 2016; Alicai et al., 2007; Colvin et al., 2004; Legg et al., 2006; Maruthi et al., 2005). Annual losses due to the diseases are estimated at US$1 billion [IITA bulletin 2215, 2014].

Cassava mosaic disease is caused by CMGs that belong to the genus *Begomovirus* in the family *Geminiviridae* (Sseruwagi et al., 2004; Maruthi et al., 2002; Legg et al., 2006). The CMG genome consists of two circular single-stranded DNA molecules, referred to as DNA – A and DNA – B, contained in geminate particles (Maruthi et al., 2002). DNA – A and B are required for successful infectivity (Stanley and Gay, 1983) because the DNA – A component encodes the viral coat proteins required for replication and encapsidation of both DNA – A and B, while the DNA – B component encodes transport proteins responsible for movement and symptom expression of the virus in the cassava plant (Chikoti et al., 2013). CMGs are transmitted by *B. tabaci* and also through planting infected stem cuttings. Nine species of CMGs are currently recognised in Africa (Fauquet et al., 2003, Fauquet et al., 2008, Sseruwagi et al., 2004).

Cassava brown streak disease is caused by cassava brown streak virus (CBSV) and Uganda cassava brown streak virus, (UCBSV); both viruses have positive single-stranded RNA ((+) ssRNA) genomes and belong to the genus *Ipomovirus* in the family of *Potyviridae*. These viruses are also transmitted by *B. tabaci* and through the planting of infected stem cuttings. Symptoms of these viruses in infected cassava plants are generally similar (Legg et al., 2011; Winter et al., 2010, Ndunguru et al., 2015).

1.1.4 Management of whiteflies

Previous studies focused on the viruses transmitted by the whitefly but not the whitefly itself, yet understanding and controlling the vector is very important in managing virus spread (Legg et al., 2014a). Pesticide control strategies have been implemented in the past on large cotton fields to manage the destruction caused by whiteflies but this came with a disadvantage as whiteflies eventually developed a high resistance to pesticides;

thus, this approach became significantly unsustainable (Olsen and Schaal, 1999). It is important to note that a few farmers in East Africa try to use pesticides but the majority of farmers cannot because of the cost. The majority of them are smallholder farmers who don't hold large fields (De Bon et al., 2014).

Legg et al. (2003) pointed out the need to enhance naturally occurring predators and parasitic wasps to control the population of *B. tabaci* as part of the integrated pest management strategy; unfortunately, the ecology and impact of parasitoids and predators of the *B. tabaci* species in East Africa is not clearly known, although they exist (Fishpool and Burban, 1994). It's also important to note that parasitoids alone might not be sufficient to control *B. tabaci* without other control approaches (Hoelmer, 1996).

Recently, breeding cassava for resistance in East Africa received attention among researchers (Mukiibi et al., 2019) but this faced hindrances such as the whiteflies' ability to quickly adapt to the new genotypes. This made cassava resistance to whiteflies a temporary strategy for whitefly management. However, breeding for host plant resistance against viruses has been successful, for example to CMGs (Dixon et al., 2003), and this might be the reason for the limited focus on cassava resistance to whiteflies research (Legg et al., 2014a). Omongo et al. (2012) pointed out that researchers had identified cassava genotypes that showed resistance to whiteflies; however, the resistant cassava varieties were likely to be CBSD susceptible. Therefore, more research is needed for further breeding necessary to develop cassava varieties that are resistant to both the viruses and the whitefly (Thresh and Cooter, 2005).

Intercropping and other planting strategies, such as crop rotation, have been practised by smallholder farmers in East Africa. Fauquet and Fargette (1990) suggested that high-density intercropping may be effective in controlling whitefly populations but they weren't sure of the strategy when replicated in the field. Cultural practices like crop rotation, intercropping and host-free periods have been used in horticultural crops and they have managed to control the population of whiteflies to some extent (Hilje et al., 2001). The advantage of intercropping cassava with legumes is that legumes don't affect cassava yield, yet they manage to reduce whitefly egg densities (Islami et al., 2011). Gold et al. (1990) observed that intercropping cowpeas with cassava is more effective in reducing the whitefly egg density than intercropping with maize.

1.1.5 Next generation sequencing (NGS)

Next generation sequencing has reshaped biology as a quantitative and computationally intense science; the advancement in technology has enabled researchers to collect enormous amounts of data. Recently the MEAM1 genome was published from an isogenic colony of whiteflies developed at the United States Department of Agriculture (USDA) Agricultural Research Service (ARS) in Charleston, South Carolina (SC) and it contains a total of 15,664 protein coding genes (Chen et al., 2016). This genome was sequenced on both Illumina HiSeq 2500 and PacBio. More genomes of whiteflies will be published in the future, and they could be used as reference for resolving *B. tabaci* systematics as well as help in the development of sustainable management strategies. The numbers of genes created will be used to answer biological and phylogenetic questions in the whitefly community.

Various studies have carried out transcriptome analysis to study whiteflies; for example, the study of the adaptation of whiteflies species to tobacco (Xia et al., 2017). In

addition, Sseruwagi et al. (2017) carried out a transcriptome analysis on field-collected whiteflies in East Africa and found an eleven amino acid residue deletion in the NusG gene. The comparative transcriptome analysis study by Hasegawa et al. (2018) revealed a network of genes activated in *B. tabaci* when fed on tomato plants infected with tomato yellow leaf curl virus. Lastly, Kaur et al. (2017) identified a temporal shift in gene expression and differential regulation of novel orphan genes when they carried out a transcriptome analysis on the MEAM1 *B. tabaci* species. In this study we carried out transcriptome analysis on 49 individual whiteflies, and generated thousands of genes that we used in the estimation of the *B. tabaci* species tree.

1.1.6 Phylogenetic analysis

A phylogeny represents evolutionary relationships or histories between species. This is one of the widely used methods in whitefly research to understand whitefly species relationships (Boykin et al., 2012; Boykin et al., 2007). It enriches our understanding of how genes, genomes and species evolve. In this thesis I reconstructed phylogenetic trees using the latest software, such as MrBayes 3.3.2 (Ronquist et al., 2012), and species tree estimation methods such as SVDQuartets (Chifman and Kubatko, 2014), ASTRAL (Mirarab et al., 2014; Zhang et al., 2018), and RAxML (Stamatakis, 2014). We also described how and why it is important to accurately root the phylogenetic tree (Kinene et al., 2016).

1.2 Statement of the problem

Whiteflies are a threat to one of the major food crops in Africa – Cassava (*Manihot esculenta*). This crop is tolerant to drought, poor soils, and low rainfall intensity and many smallholder farmers rely on it as a food security crop and a cash crop. Despite its

importance as a food security crop, its existence is threatened by high populations of whiteflies in East Africa termed as superabundant whitefly populations. The superabundant whitefly populations cause both direct damage through feeding and indirect damage through transmission of CMGs, UCBSV and CBSV (Maruthi et al., 2005; Legg et al., 2006). These viruses have caused two ongoing pandemics of cassava mosaic disease (CMD) and cassava brown streak disease (CBSD). Yield losses due to viral diseases have been estimated at 47% (Legg et al., 2006, Omongo et al., 2012). This has resulted in recurrent hunger and famine in East Africa.

Accurate identification of the whitefly species spreading viruses in East Africa is important in order to advise on control strategies. However, the focus on developing virus-resistant cassava varieties and lack of research attention on whitefly control has resulted in an increase (superabundance) of whitefly populations in smallholder farmers' fields. Apart from the viral disease spread by whiteflies in the farmers' fields, whiteflies also cause direct damage through feeding and the excretion of honeydew, which turns into black mould and affects plant photosynthesis. Yield losses due to direct damage are estimated at 40% (Omongo et al., 2012).

Smallholder farmers don't have enough knowledge and training about the vector; some are not even aware of its existence and its role in virus spread. Therefore, it becomes hard to convince them to carry out *B. tabaci* management strategies; this is why cassava virus pandemics have remained elusive. We use thousands of nuclear genes to explore the systematics and patterns of evolution of *B. tabaci* species in East Africa and also hypothesise as to what could be driving the high populations of *B. tabaci* in East Africa.

1.3 Objective of the study

1.3.1 General objective

To investigate the genetic variability of the *B. tabaci* species in E. Africa, generate multi-gene phylogeny that will help clarify *B. tabaci* species boundaries and provide an improved and robust systematics framework for the members of the *B. tabaci* species complex.

1.3.2 Specific objectives

1. To establish the role of Heat shock protein 90 (HSP90) in *B. tabaci*'s ability to adapt to the varying climate in East Africa and its relation to the superabundance phenomenon.
2. To identify phylogenetic relationships of various *B. tabaci* species utilising common nuclear genes (RNA polymerase II and Shaker cognate gene W)
3. To estimate the *B. tabaci* species tree using thousands of nuclear genes generated from the transcriptome dataset of individual whiteflies.

1.4 Significance of the study

This research provides a framework for exploration of the genetic variability of the African cassava whitefly (*B. tabaci* species complex) present in East Africa, and the superabundant phenomenon. To date, the majority of research has been done on whiteflies in Asia, China and America but little has been done on species in Africa. In addition, even with the research done in all parts of the world the whitefly community still faces the same problem of incomplete systematics and confusing nomenclature. Previous studies have used inconsistent nomenclature and overlapping names, causing

repetition in studies and hence the failure of control measures against these invasive species. Common nuclear genes (RNA polymerase II a, Shaker cognate gene W) were used to infer species relationships and investigate whether they were more informative than the mtCOI gene. Identifying the role of HSP90 in thermal adaption (Aim 1) is important for finding the genetic signature associated with the superabundant phenomenon of the whiteflies. Estimating a robust species tree from gene trees (Aim 3) will be important in establishing a systematics framework for Africa *B. tabaci* species.

1.5 Scope of the study

The study was designed to investigate the genetic variability of the *B. tabaci* species in E. Africa, generate multi-gene phylogeny to clarify *B. tabaci* species boundaries and provide an improved and robust systematics framework for the members of the *B. tabaci* species complex. The whitefly samples were collected from Uganda, Tanzania and Malawi under the arrangement of regional agricultural research institutes; for example, the National Crop Resources Research Institute (Uganda), the Mikocheni Agricultural Research Institute (Tanzania) and the Department of Agriculture research services (Malawi). The global samples used in this study were sent by collaborators from Brazil, Thailand, and the United Kingdom. Sample preparation and all the downstream analysis were carried out at the University of Western Australia. The results of this research will inform researchers of the species of whiteflies available in East Africa and draws conclusions on the drivers of the high population (superabundance) of whiteflies seen in the field. We hope cassava breeders will use this study to breed cassava for resistance to the right whitefly species and virus.

1.6 Thesis structure

This PhD thesis is in accordance with graduate research school regulations of the University of Western Australia and is presented as a series of scientific manuscripts that are under review in peer-reviewed journals or about to be submitted for review. The thesis consists of an introductory chapter on the research (Chapter 1), a general discussion (chapter 6) and four research chapters, each presenting relevant scientific papers addressing the study objectives. The four research chapters can be read as a part of the whole thesis or as separate entities. Each of these research chapters entails an independent introduction, literature review, methods, results, and discussion sections. Each chapter is independently referenced, the references for chapter 1 (Introduction) and chapter 6 (Discussion) are in the bibliography section following chapter 6.

Chapter 1 is an introduction to the whitefly (*B. tabaci*) biology, systematics, economic importance, genetic diversity, cassava viruses, and lastly the objectives of the study. This prepares the background for the next four chapters. Chapter 2 deals with methods used to root phylogenetic trees, as phylogenetic trees are one of the tools we used in our further analysis; therefore, understanding the importance of finding the correct root for the tree is advantageous because incorrectly rooted phylogenetic trees might mislead evolutionary and taxonomic inferences. In chapter 3, the *B. tabaci* heat shock protein 90 (HSP90) was identified from next generation sequence data generated from 21 whitefly transcriptomes. The aim here was to identify genetic signatures for superabundance of whiteflies. We detected recombination signals in the HSP90 gene for members of the *B. tabaci* species complex and we conclude that recombination might be driving the evolution of *B. tabaci* species. In chapter 4, genomic regions of interest were identified (RNA polymerase II and Shaker cognate gene w) and used to infer evolutionary relationships of *B. tabaci* species. Patterns of conflict were found between

mitochondrial cytochrome oxidase 1 (mtCOI) and nuclear DNA segments. In chapter 5, the species tree for the *B. tabaci* complex was estimated using transcriptome data. Three thousand orthologous genes were identified using the Agalma pipeline and subsequently analysed in SVDQuartets, and in RAxML to generate gene trees that were used in ASTRAL for species tree estimation. We compared the tree topologies from ASTRAL and SVDQuartets and also compared our results with the conventional mtCOI topology. In conclusion we detected recombination in the conserved HSP90 gene, identified patterns of conflict between nuclear and mtCOI tree topologies and finally we estimated the *B. tabaci* species tree using 49 transcriptomes. We noted a difference in the groupings of the *B. tabaci* Uganda species in the mtCOI conventional tree compared to the species tree from transcriptome data. Finally, chapter 6 presents the general discussion and conclusions based on the research results. The Bibliography (for chapter 1 and chapter 6) and appendix follow chapter 6.

Chapter 2

2.1 Methods for rooting phylogenetic trees

This book chapter was published in the *Encyclopedia of Evolutionary Biology* published by Elsevier.

Kinene, T., Wainaina, J., Maina, S., and Boykin, L.M. (2016) Rooting Trees, Methods for. In: Kliman, R.M. (ed.), Encyclopedia of Evolutionary Biology. vol. 3, pp. 489–493. Oxford: Academic Press.

Rooting Trees, Methods for

T Kinene, J Wainaina, S Maina, and LM Boykin, The University of Western Australia, Perth, WA, Australia

Rooted versus Unrooted

Phylogenetic trees are either rooted or unrooted, depending on the research questions being addressed. The root of the phylogenetic tree is inferred to be the oldest point in the tree and corresponds to the theoretical last common ancestor of all taxonomic units included in the tree. The root gives directionality to evolution within the tree (Baldauf, 2003). Accurate rooting of a phylogenetic tree is important for directionality of evolution and increases the power of interpreting genetic changes between sequences (Pearson *et al.*, 2013).

Many techniques such as molecular clock, Bayesian molecular clock, outgroup rooting, or midpoint rooting methods tend to estimate the root of a tree using data and assumptions (Boykin *et al.*, 2010). However, Steel (2012) discusses root location in random trees and points out that information in the prior distribution of the topology alone can convey the location of the root of the tree. These results show that the tree models that treat all taxa equally and are sampling consistently convey information about the location of the ancestral root in unrooted trees (Steel, 2012).

Why Do We Need a Rooted Tree?

We are interested in rooting a phylogenetic tree in order to show the path of evolution of biological species. Therefore most users of phylogenetic trees want rooted trees because they give an indication of the directionality of evolutionary change. The root of phylogenetic tree is crucial in evolutionary interpretation of the tree (Williams, 2014), because an unrooted tree species shows only the relationships among the taxa and does not define the evolutionary path (**Figure 1(a)**).

When Do We Need an Unrooted Tree?

An unrooted tree is desired when we do not have a distantly related group (sequence) for comparison or when primary interest is focused only on relationships among the taxa rather than on the directionality of evolutionary change. Unrooted trees are beneficial in depicting clusters of related sequences. Unrooted gene trees have also become more prevalent within the multispecies coalescent phylogenetic framework, leading

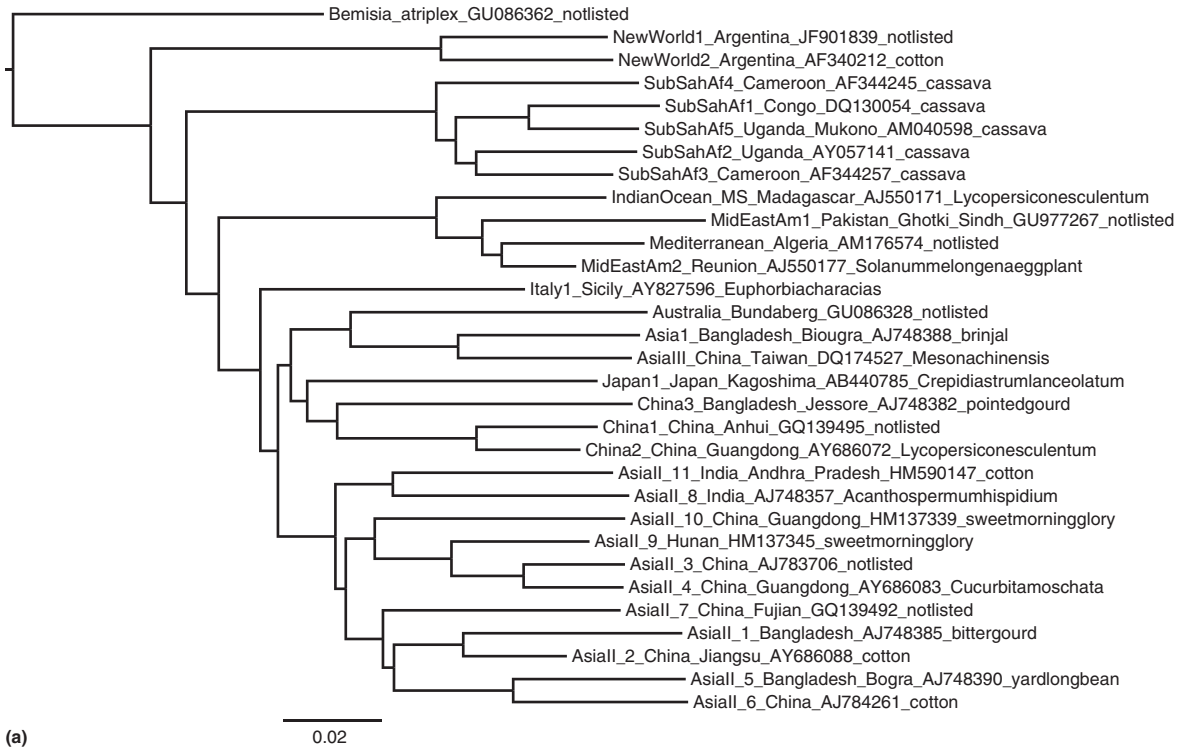


Figure 1 (a) Outgroup rooted phylogenetic tree of the *Bemisia tabaci* species complex (whiteflies) from a modified dataset (Boykin *et al.*, 2013). Tip labels correspond to geographic location_sublocation_GenBank accession number_host. (b) Unrooted phylogenetic tree of the *Bemisia tabaci* species complex (whiteflies) from a modified dataset (Boykin *et al.*, 2013). Tip labels correspond to geographic location_sublocation_GenBank accession number_host. (c) Unrooted star phylogenetic tree of the *Bemisia tabaci* species complex (whiteflies) from a modified dataset (Boykin *et al.*, 2013). Tip labels correspond to geographic location_sublocation_GenBank accession number_host.

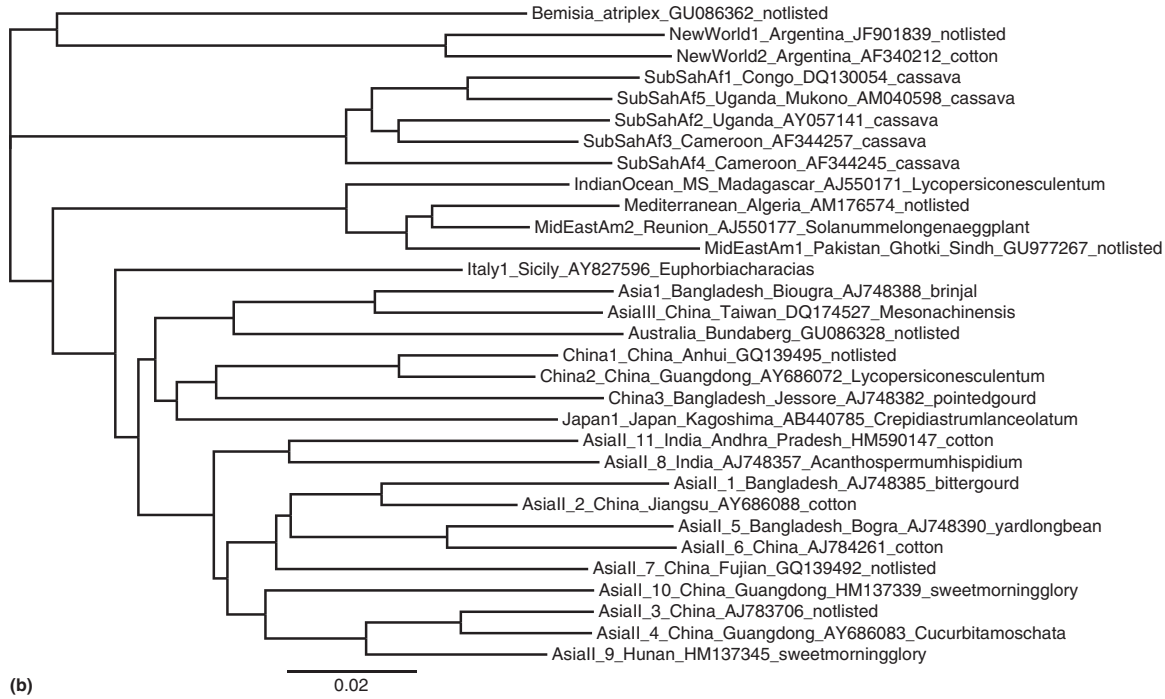


Figure 1 Continued.

to systematic approaches for inferring unrooted species trees from unrooted gene tree topologies (Liu and Yu, 2011). Gene trees and species trees can have similar topologies but often there is considerable discordance between gene tree and species trees (Degnan, 2013). For this reason, understanding how to root gene trees will have implications for accurate species tree inference (**Figures 1(b)** and **1(c)**).

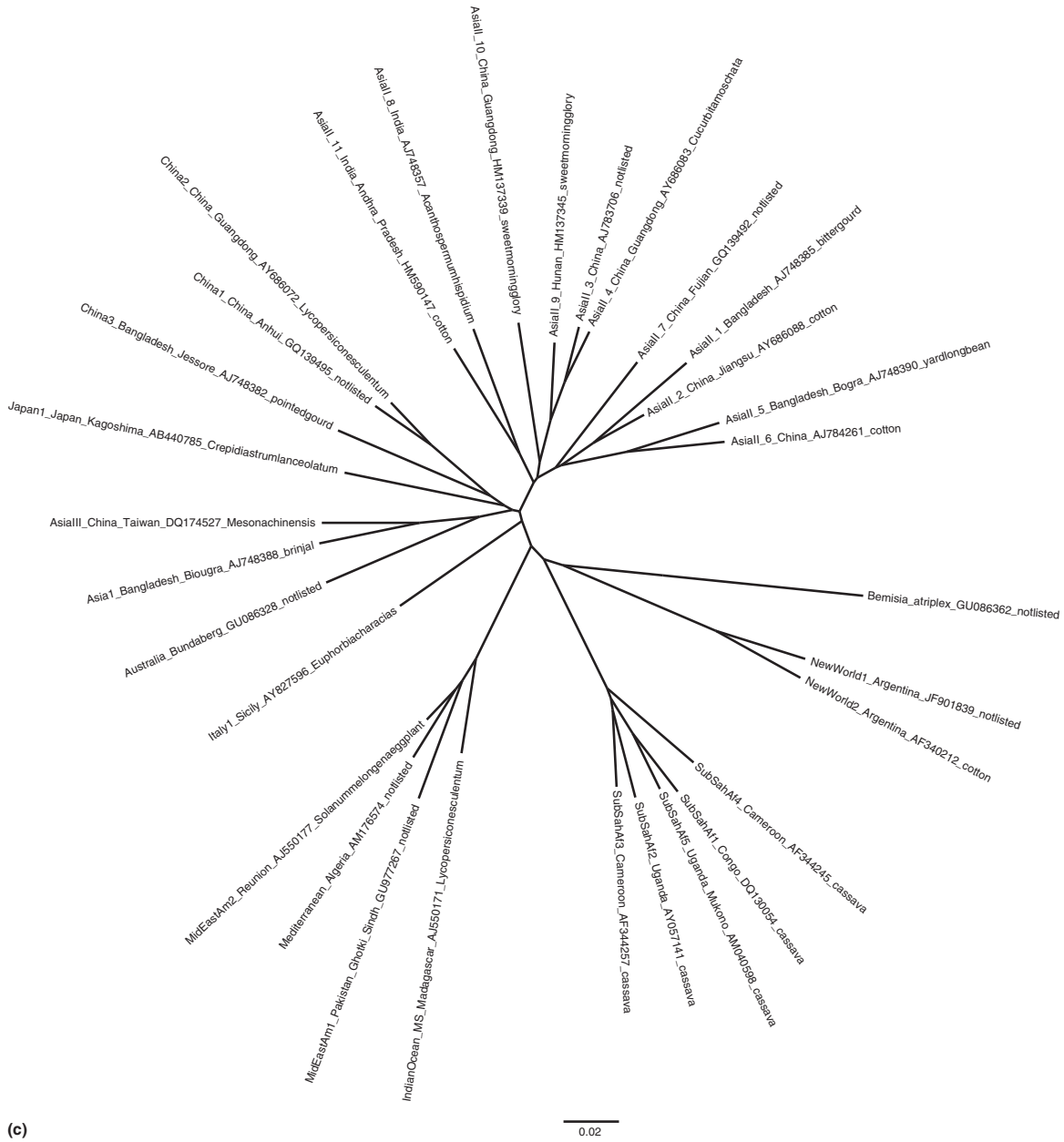
How Do You Root a Phylogenetic Tree?

Outgroup Rooting

There are several rooting methods (**Table 1**), however the most popular and widely used is the outgroup method (Wheeler, 1990; Tarrío *et al.*, 2000; Hess and De Moraes Russo, 2007; Boykin *et al.*, 2010). Outgroup method assumes that one or more of the taxa are divergent from the rest of the taxa (ingroup). The branch linking the ingroup and outgroup becomes the starting point, and defines all subsequent evolutionary events within the tree (Brady *et al.*, 2011; Williams, 2014). In addition to providing evolutionary information of the ingroup, the outgroup has other additional functions. It allows the identification of distinct features within ingroup sequences (Wheeler, 1990). An important aspect of outgroup method is the need for a priori knowledge on the appropriate outgroup to use for the set of sequences (Wheeler, 1990; Hess and De Moraes Russo, 2007). However, this is also the main bottleneck for this method, especially within higher taxonomic groups such as angiosperms, birds, and mammals where a consensus outgroup is lacking (Qiu *et al.*, 2001). As a consequence, many authors are forced to choose between different sorts of outgroups that are either phylogenetically close or phylogenetically distant (Rota-Stabelli and Telford, 2008).

Lack of an appropriate outgroup results in drawbacks such as the long branch attraction (LBA). LBA occurs mainly when the outgroup taxa are distantly related to the ingroup due to either large divergence time and/or increased rate of evolution (Tarrío *et al.*, 2000). This results in homoplastic changes occurring at rapidly evolving sites thus resulting in artifactual rooting (random rooting) (Wheeler, 1990; Hendy and Penny, 2011; Maddison *et al.*, 1984). Several criteria have been proposed to prevent LBA within phylogenetic trees, through a multistep process as proposed by Rota-Stabelli and Telford (2008) to assist in outgroup selection especially in the case of arthropod classes. They include: (1) low substitution rate; (2) ingroup like G + C composition; (3) new strand bias estimators 'skew index'; (4) the tendency of the outgroup to avoid 'random branding effect'; and (5) phylogenetic proximity to the arthropod.

An alternative approach to assess the importance of an outgroup in rooting the tree is explored by Graham *et al.* (2002); this is by establishing whether the outgroup provides sufficient signal in response to root location, indicative of historic linkage or due to LBA. Using Pontederiaceae, an aquatic monocot, as the case study they assessed how the nearest outgroup provides for rooting Pontederiaceae compared to those less closely related relatives and further investigate the role of LBA when determining the optimal rooting of Pontederiaceae. However, they concluded that LBA may influence rooting, and may be supporting the wrong outgroup. To further reduce LBA and to ensure robustness of the outgroup rooting method they recommend multiple sampling of outgroups within the sister group rather than sampling within less closely related taxa.



(c)

Figure 1 Continued.

Midpoint Rooting

Midpoint rooting calculates tip to tip distances and then places the root halfway between the two longest tips (Swofford *et al.*, 1996). The ancestral point will be identified if the tree has constant rates of evolution. The method is exclusively dependent on the branch length of the phylogenetic tree and the assumption of the molecular clock (Holland *et al.*, 2003). Homogeneity is assumed across the branch and that the two most divergent taxa evolved at equal rates (Holland *et al.*, 2003; Swofford *et al.*, 1996). If the tree is balanced, midpoint rooting works well. However, a major limitation of midpoint rooting is the dependence on having clocklike data and a balanced topology.

The midpoint rooting method is often applied to viral genetic datasets because in many cases outgroups are unknown. For example, Stavrinides and Guttman (2004) utilize midpoint rooting to establish the evolutionary relationship of severe acute respiratory syndrome (SARS) coronaviruses. They carried out phylogenetic analysis of the viral genes encoding for viral structural proteins specifically, envelope matrix (M) and nucleocapsid (N) proteins. The midpoint rooting of the

Table 1 Four methods for rooting phylogentic trees

Rooting method	*Pros*	*Cons*	*Software*
Outgroup	Accurate	Must have an outgroup Long branch attraction	PAUP Figtree R package: ape (http://www.inside-r.org/packages/cran/ape/docs/unroot)
Midpoint	Fast No outgroup needed	Dependent on clocklike data Not good with unbalanced trees	PAUP R package: phangorn (https://cran.r-project.org/web/packages/phangorn/phangorn.pdf)
Molecular clock	No outgroup needed Robust to violations of the clock	Computationally intense	PAUP PAML
Bayesian molecular clock	Alternative rootings uncovered No outgroup needed	Must customize the prior	Post Root (http://www.stat.osu.edu/~lkubatko/software/phyl_util.html) Root annotator (http://sourceforge.net/projects/rootannotator/)

trees generated using the M protein data shows two groups, one that consisted of porcine, feline, and canine and one that contained bovine and murine coronaviruses (*Coronaviridae*). On the other hand, the N protein tree is midpoint rooted on the branch leading to the group 1 coronaviruses. Moreover, the appropriateness of midpoint rooting is supported by the results of Tajima's relative rate test (Tajima, 1993), indicating no rate heterogeneity among the coronavirus groups (Stavrinides and Guttman, 2004).

This method should be used as an alternative to the outgroup rooting method and could be adopted as the default method when the outgroup method is difficult to apply either due to problems with available outgroups, such as LBA or lack of a priori knowledge of the outgroup (Hess and De Moraes Russo, 2007).

Molecular Clock Rooting

The molecular clock rooting method has one assumption: the rate of evolution is constant for the sequences of interest (Yang and Rannala, 2012). The rate is typically expressed in substitutions per site per year or substitutions per site per million years (Brown and Yang, 2011). The strict clock is often used in analyses of sequences sampled at the intraspecific level, for which usually there is an exceptionally low rate of variation (Brown and Yang, 2011; Ho and Duchêne, 2014). The molecular clock assumption becomes problematic for distantly related species because there is a linear relationship between the genetic distances and approximate divergence. The slope of the line directly corresponds to the evolutionary rate variation among species especially among divergent taxa (Welch and Bromham, 2005). Before utilizing the molecular clock method for rooting a phylogenetic tree users should test if a molecular clock is appropriate to describe the data. Testing for the molecular clock entails generating two maximum likelihood trees, one computed with the molecular clock enforced and one without the molecular clock enforced and then utilizing the likelihood ratio test (Felsenstein, 1983; Holder and Lewis, 2003).

Bayesian Molecular Clock Rooting

Huelsenbeck *et al.* (2002) proposed the use of Bayesian inference under the molecular clock assumption to infer the root of a phylogenetic tree. After obtaining the posterior distribution of trees under Bayesian inference, the root of the tree is inferred to be the root position with the highest posterior probability. This method also provides the posterior probability that the root lies on any branch of the ingroup topology. Another advantage of the Bayesian method is that it allows the user to evaluate alternative rootings. Other rooting methods only return one rooting for a particular dataset, without any numerical assessment of confidence in that rooting. A Bayesian molecular clock analysis successfully identified the root of Orcuttieae (Poaceae) (Boykin *et al.*, 2010) when all other methods failed. Post_root was developed to analyze the output from MrBayes (Ronquist *et al.*, 2012) or ExaBayes (Aberer *et al.*, 2014) runs. The output from Post_Root will give the number of unique roots and also the most probable root position.

Most recently, Calvignac-Spencer *et al.* (2014) have further developed Post_Root to a web-based interface in their quest to identify the branch root posterior probability (RPP) of the most recent Ebola outbreak in West Africa. They were forced to rely on Bayesian molecular clock rooting because there is no known outgroup for Ebola. It is often the case when analyzing viral sequences that no outgroup is known; therefore the Bayesian molecular clock rooting is a very useful alternative, especially when rooting is crucial for viral outbreaks.

Acknowledgment

Research in the Boykin lab is provided by the Bill & Melinda Gates Foundation via a subcontract (B04265) from the National Research Institute, University of Greenwich, UK. An Australian PhD Award to J Wainaina. The Plant Biosecurity Cooperative Research Centre, Australia (61056) supported S Maina.

See also: Bayesian Phylogenetic Methods. Phylogenetic Tree

References

Aberer, A.J., Kobert, K., Stamatakis, A., 2014. ExaBayes: Massively parallel bayesian tree inference for the whole-genome era. Molecular Biology and Evolution 31 (10), 1–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25135941 (accessed 13.10.15).

Baldauf, S.L., 2003. The deep roots of eukaryotes. Science 300 (5626), 1703–1706.

Boykin, L.M., Bell, C.D., Evans, G., Small, I., DeBarro, P., 2013. Is agriculture driving the diversification of the *Bemisia tabaci* species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)?: Dating, diversification and biogeographic evidence revealed. BMC Evolutionary Biology 13, 228. doi:10.1186/1471-2148-13-228.

Boykin, L.M., Kubatko, L.S., Lowrey, T.K., 2010. Comparison of methods for rooting phylogenetic trees: A case study using *Orcuttieae* (Poaceae: Chloridoideae). Molecular Phylogenetics and Evolution 54 (3), 687–700. Available at: http://dx.doi.org/10.1016/j.ympev.2009.11.016 (accessed 13.10.15).

Brady, S.G., Litman, J.R., Danforth, B.N., 2011. Rooting phylogenies using gene duplications: An empirical example from the bees (Apoidea). Molecular Phylogenetics and Evolution 60 (3), 295–304.

Brown, R.P., Yang, Z., 2011. Rate variation and estimation of divergence times using strict and relaxed clocks. BMC Evolutionary Biology 11 (1), 271.

Calvignac-Spencer, S., Schulze, J.M., Zickmann, F., Renard, B.Y., 2014. Clock rooting further demonstrates that Guinea 2014 EBOV is a member of the Zaïre lineage. PLoS Currents. doi:10.1371/currents.outbreaks.c0e035c86d721668a6ad7353f7f6fe86.

Degnan, J.H., 2013. Anomalous unrooted gene trees. Systematic Biology 62 (4), 574–590.

Felsenstein, J., 1983. Statistical inference of phylogenies. Journal of the Royal Statistical Society A 146, 246–272.

Graham, S.W., Olmstead, R.G., Barrett, S.C.H., 2002. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. Molecular Biology and Evolution 19 (10), 1769–1781.

Hendy, M.D., Penny, D., 2011. A framework for the quantitative study of evolutionary trees. Systematic Zoology 38 (4), 297–309. Society of systematic biologists a framework for the quantitative study of evolutionary trees.

Hess, P.N., De Moraes Russo, C.A., 2007. An empirical test of the midpoint rooting method. Biological Journal of the Linnean Society 92 (4), 669–674.

Ho, S.Y.W., Duchêne, S., 2014. Molecular-clock methods for estimating evolutionary rates and timescales. Molecular Ecology 23 (24), 5947–5965.

Holder, M., Lewis, P.O., 2003. Phylogeny estimation: Traditional and Bayesian approaches. Nature Reviews. Genetics 4 (4), 275–284.

Holland, B.R., Penny, D., Hendy, M.D., 2003. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock – a simulation study. Systematic Biology 52 (2), 229–238.

Huelsenbeck, J.P., Larget, B., Miller, R.E., Ronquist, F., 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. Systematic Biology 51 (5), 673–688.

Liu, L., Yu, L., 2011. Estimating species trees from unrooted gene trees. Systematic Biology 60 (5), 661–667.

Maddison, W.P., Donoghue, M.J., Maddison, D.R., 1984. Outgroup analysis and parsimony. Systematic Zoology 33 (1), 83–103. Society of systematic biologists outgroup analysis and parsimony.

Pearson, T., Hornstra, H.M., Sahl, J.W., *et al*., 2013. When outgroups fail; Phylogenomics of rooting the emerging pathogen, *Coxiella burnetii*. Systematic Biology 62 (5), 752–762.

Qiu, Y.L., Lee, J., Whitlock, B.A., Bernasconi-Quadroni, F., Dombrovska, O., 2001. Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? Amborella, Nymphaeales, Illiciales, Trimeniaceae, and Austrobaileya. Molecular Biology and Evolution 18 (9), 1745–1753.

Ronquist, F., Teslenko, M., van der Mark, P., *et al*., 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61 (3), 539–542.

Rota-Stabelli, O., Telford, M.J., 2008. A multi criterion approach for the selection of optimal outgroups in phylogeny: Recovering some support for Mandibulata over Myriochelata using mitogenomics. Molecular Phylogenetics and Evolution 48 (1), 103–111.

Stavrinides, J., Guttman, D.S., 2004. Mosaic evolution of the severe acute respiratory syndrome coronavirus mosaic evolution of the severe acute respiratory syndrome coronavirus. Journal of Virology 78 (1), 76–82. Available at: http://jvi.asm.org/content/78/1/76.full.pdf + html (accessed 13.10.15).

Steel, M., 2012. Root location in random trees: A polarity property of all sampling consistent phylogenetic models except one. Molecular Phylogenetics and Evolution 65 (1), 345–348. Available at: http://dx.doi.org/10.1016/j.ympev.2012.06.022 (accessed 13.10.15).

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hiillis, D.M., Moritz, D., Mable, B.K. (Eds.), Molecular Systematics. Sunderland, MA: Sinauer Associates, pp. 407–514.

Tajima, F., 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135 (2), 599–607.

Tarrío, R., Rodríguez-Trelles, F., Ayala, F.J., 2000. Tree rooting with outgroups when they differ in their nucleotide composition from the ingroup: The *Drosophila saltans* and *willistoni* groups, a case study. Molecular Phylogenetics and Evolution 16 (3), 344–349.

Welch, J.J., Bromham, L., 2005. Molecular dating when rates vary. Trends in Ecology and Evolution 20 (6), 320–327.

Wheeler, W.C., 1990. Nucleic acid sequence phylogeny and random outgroups. Cladistics 6 (4), 363–367. Available at: http://doi.wiley.com/10.1111/j.1096–0031.19900.tb00550.x (accessed 13.10.15).

Williams, T.A., 2014. Evolution: Rooting the eukaryotic tree of life. Current Biology 24 (4), R151–R152. Available at: http://dx.doi.org/10.1016/j.cub.2014.01.026 (accessed 13.10.15).

Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. Nature Reviews Genetics 13 (5), 303–314. Available at: http://dx.doi.org/10.1038/nrg3186 (accessed 13.10.15).

Chapter 3: Recombination Detected in the Heat Shock Protein 90 (HSP90) of the *Bemisia tabaci* Species Complex

This article was submitted to BMC Evolutionary Biology on 5th Feb 2019.

Tonny Kinene*[1], Bruno Rossito De Marchi[2], Titus Alicai[3], Livingstone S. Luboobi[4], Christopher Abu Omongo[3], Anders Savill[1, 5] and Laura M. Boykin[1] (2019). Recombination detected in the Heat Shock Protein 90 (HSP90) of the *Bemisia tabaci* species complex

Title: Recombination detected in the Heat Shock Protein 90 (HSP90) of the *Bemisia tabaci* species complex

Authors: Tonny Kinene*[1], Bruno Rossito De Marchi[2], Titus Alicai[3], Livingstone S. Luboobi[4], Christopher Abu Omongo[3], Anders Savill[1, 5] and Laura M. Boykin[1]

Addresses

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

[2]São Paulo State University (UNESP), School of Agriculture, Dept. of Plant Protection, CEP 18610-307, Botucatu (SP), Brazil

[3]National Crops Resources Research Institute, P.O. Box 7084, Kampala, Uganda

[4]Department of Mathematics, Makerere University, P.O. Box 7062, Kampala, Uganda

[5]Pawsey Supercomputing Centre Technology Park, 26 Dick Perry Ave Kensington 6151, Australia

*Author for correspondence: Laura M. Boykin: laura.boykin@uwa.edu.au

Authors' email addresses:

tonny.kinene@research.uwa.edu.au

Laura M. Boykin: laura.boykin@uwa.edu.au

Livingstone Serwadda Luboobi: luboobi@gmail.com

Bruno Rossito De Marchi: bruno_dmarchi@hotmail.com

Christopher Abu Omongo: chrisomongo@yahoo.com

Anders Savill: goderiich@gmail.com

Titus Alicai: talicai@hotmail.com

3.1 Abstract

Background

Bemisia tabaci (whiteflies) are a global insect pest causing billions of dollars in damage each year by leaving farmers with low yields. In East Africa, whiteflies are superabundant and present on cassava plants throughout the year. Whiteflies do not decrease in number in the hot dry seasons in East Africa; therefore, it has been suggested that the synthesis of Heat Shock Protein (HSP) may protect the whitefly from heat stress and other biotic factors. In this study we used sequence data generated from individual whiteflies to assess the variability and recombination of the HSP90 gene in members of the *B. tabaci* species complex.

Results

A total of 21 samples were sequenced on Illumina Hiseq 2500 and Hiseq 4000. These included eight genetic groups of *B. tabaci*: 8 SSA1, 5 SSA2, 2 Australia I (AUSI), 2 New World Africa (NWAfrica), *B. afer*, Uganda, Mediterranean (MED), and Middle East Asia Minor 1 (MEAM1). An alignment of 21 HSP90 sequences was generated after mapping and *de novo* assembly. Recombination analysis was performed on an alignment of 27 HSP90 sequences (we added an additional 6 sequences from GenBank). There were 18 recombination events detected in the HSP90 gene of the *B. tabaci* species complex, 7 of which were regarded as events that could be caused by evolutionary mechanisms other than recombination, such as gene duplication. The phylogenetic analysis carried out on a dataset without recombination events revealed a tree pattern with short terminal branches.

Conclusion

Recombination events were detected for members of the *B. tabaci* species complex in the HSP90 gene. This could explain the variability in the HSP90 gene of the *B. tabaci* species complex and highlight the phenomenon of the increased chance of survival and reproductive abundance of whiteflies in hot conditions in East Africa, since recombination is a major driving force of evolution.

Keywords: Recombination, HSP90 gene, Climate change, *Bemisia tabaci.* Food security

3.2 Background

The *Bemisia tabaci* species complex (whiteflies) is a group of small phloem sap-feeding insects capable of causing extensive crop damage globally (Brown et al., 1995a; Hoddle, 1999). The species complex is composed of at least 34 species that are morphologically indistinguishable (Boykin et al., 2012; Boykin et al., 2013). These species cause extensive damage to plants through direct feeding by both juvenile and adult stages of the whitefly (Boykin, 2014; Boykin and De Barro, 2014; Liu et al., 2012). They also cause indirect damage through the excretion of honeydew, which covers the leaf and fruit surfaces promoting the growth of black sooty mold. This mold interferes with plant photosynthesis, affects plant growth, and leads to poor quality fruits (Omongo et al., 2012). Whiteflies also transmit plant viruses in the process of feeding from one plant to another (Brown and Czosnek, 2002). Yield losses due to direct feeding and black sooty mold are estimated at 40% (Omongo et al., 2012). In Africa, whiteflies are a major threat to food security (Patil et al., 2015).

Many smallholder farmers in East Africa rely heavily on cassava as a food security and cash crop because of its low labour intensity, long harvesting window, and tolerance to

low fertility soils and low rainfall (Alicai et al., 2007; Colvin et al., 2004; Legg et al., 2014b). Cassava roots are eaten and are rich in carbohydrates, and the leaves are rich in protein and are eaten as vegetables. Cassava can be processed into other commercial products such as flour, cakes and alcohol. Unfortunately, cassava is highly vulnerable to both whiteflies and the viruses they transmit. Attempts to protect cassava from whiteflies with chemical pesticides have been ineffective and costly (Cahill et al., 1995; Howeler et al., 2013; Kinene et al., 2015). The honeydew excreted by superabundant populations of whiteflies in East Africa interferes with photosynthesis and reduces the effectiveness of insecticides (Boykin, 2014; Brown et al., 1995a; Liu et al., 2012; Omongo et al., 2012). In Africa, whiteflies transmit two cassava viral diseases: cassava brown streak disease (CBSD) and cassava mosaic disease (CMD) (Alicai et al., 2016; Colvin et al., 2004; Maruthi et al., 2005; Sseruwagi et al., 2004; Ateka et al., 2017).

Agriculture development in Africa has been constrained by rising temperatures due to global warming (Pereira, 2017). Impacts are economically significant in countries in East Africa that are heavily reliant on natural resources, such as rainwater, for agriculture. Increasing temperature has resulted in prolonged drought seasons, which have left most plants dry. However, the drought-tolerant plants like cassava, which save farmers from famine, are now heavily infested with whiteflies. This has caused loss of food and income for farmers due to the viral diseases associated with whiteflies that infect their crops. The high numbers of whiteflies seen in cassava fields in East Africa, termed as superabundant populations, have increased with increasing temperatures over time, which may be due to the fact that temperature determines the geographical distribution and reproductive abundance of species (Cossins and Bowler, 1987; Hoffmann et al., 2003; Kellermann et al., 2012).

The capacity of Sub-Saharan Africa (SSA) whitefly species to breed in large numbers during periods of high temperatures in East Africa indicates that SSA species might have developed a molecular response to tolerate heat stress. It is known that high temperatures in the environment can cause organisms to respond with physiological, biochemical or behavioural traits that enhance their chances of survival (Fasolo and Krebs, 2004; Hoffmann et al., 2003). During stressful conditions the cells of an organism produce a set of conserved proteins called heat shock proteins (HSPs). HSPs are molecular chaperones that bind to and stabilise unfolded proteins and are found in nearly all living organisms (Parsell and Lindquist, 1993; Somero, 1995). HSPs are named according to their molecular weight (for example, HSP100, 90, 70, 60, and 40) and not all are associated with the heat shock response (Parsell and Lindquist, 1993). HSP70 and HSP90 are the most widely studied families and are synthesized when the environmental temperature exceeds the optimal temperature of an organism (Gorovits and Czosnek, 2017; Parsell and Lindquist, 1993).

Gorovits & Czosnek (2017) revealed that HSP70 and HSP90 in plants and vectors are very important for efficient virus infection. HSP90 is also considered to promote virus replication by interacting with virus replicase in *Bamboo mosaic* virus (Huang et al., 2012). Differential expression studies on HSP on Middle East Asia Minor 1 (MEAM1), one of the whitefly species in the complex, indicated that heat shock proteins (HSP20, 70 & 90) were up-regulated when whiteflies were subjected to a high temperature of 40^0 C, and this increased the fitness of whiteflies following heat stress (Diaz et al., 2015). Other studies indicate that up-regulation of HSP could be key in the determination of the natural geographical distribution of natural populations by seasonal dynamics (Yu and Wan, 2009). HSP response was examined on silverleaf whitefly and it was found that HSP70 and HSP90 were the major polypeptides synthesized by whiteflies in

response to heat stress (Salvucci et al., 2000). There is more evidence that HSPs are up-regulated by stress conditions in *B. tabaci* (Cui et al., 2008; Lü and Wan, 2011, Mahadav et al., 2009). However, none of these studies have been carried out on SSA whitefly species.

In this study we investigated the variability of the HSP90 gene in the *B. tabaci* species complex and found evidence of recombination in the coding region of HSP90 gene in the *B. tabaci* species complex.

3.3 Results

A total of 21 samples were sequenced, which included eight genetic groups of *B. tabaci*: 8 SSA1, 5 SSA2, 2 Australia I (AUSI), 2 New World Africa (NWAfrica), *B. afer*, Uganda, Mediterranean (MED), and Middle East Asia Minor 1 (MEAM1). Fifteen of these samples were sequenced on an Illumina Hiseq 2500 on a rapid run mode, resulting in 8,000,000 raw reads, and 6 samples were sequenced on an Illumina Hiseq 4000 on a high throughput mode, which yielded raw reads ranging from 43,194,264 to 37,306,246. After trimming for quality, the reads reduced to a range of 7,876,283 to 7,816,022 for the first 15 samples and 42,953,727 to 37,071,584 for the 6 samples (additional file 2). *De novo* assembly produced contigs ranging from 2530–30716, which were then mapped to the HSP 90 EU934241 reference sequence. The final coding sequence of a length of 2160 base pairs was obtained and used in this study as the HSP90 sequence. The final HSP90 sequence consisted of a consensus between the *de novo* contig of interest and the mapped consensus.

Eighteen recombinant events were identified by RDP4 (Martin et al., 2015) and 7 of these were flagged as events caused by evolutionary processes other than

recombination, such as gene duplication (Table 1). These included: SSA2_191_HSP90, SSA2_193_HSP90, AUSI_347_1_HSP90, AUSI_347_2_HSP90, NWI_156_2, NWI_156_3_HSP90 and are shown in Figure 1.

Figure 1 shows the results of a recombination analysis of an alignment of 27 sequences of HSP90 gene conducted in RDP4 (Martin et al., 2015) software. The sequence segments with recombination events are indicated in white and the ones without recombination in black. We observe that most of the SSA species did not have recombination events apart from 2 sequences from Uganda (SSA1_70_HSP90 and SSA1_65_HSP90), 2 sequences from Kenya (SSA1_193_HSP90 and SSA1_191_HSP90) and *B. afer* from Malawi. Recombination events were also detected in NewWorld1 (NW1), MED, and AUS1 species. We observed that all the MEAM1 samples previously collected did not have recombinant sequences apart from the new sample collected from Brazil (MEAM_153_HSP90) and the MED_320_3_HSP90.

Figure 2 Shows the topology of the Bayesian estimate of the phylogenetic tree for the HSP90 gene created from an alignment that is free from recombination events. We observe a long branch leading to clade C that is full of species that had recombination events; this shows that recombination has an effect on the topology of the phylogenetic tree and hence influences evolution. The majority of the species that never had recombination events have short branches as seen in clades A and B. We see sequences from Australia having a relatively long branch in clade B because we detected small segments of recombination in these sequences.

3.4 Discussion

Recombination events were detected for members of the *B. tabaci* species complex in the HSP90 gene. This could explain the phenomenon of the increased chance of survival and reproductive abundance of whiteflies in hot conditions in East Africa since recombination is a major driving force of evolution (Cangi et al., 2016). Experimental studies on the members of the *B. tabaci* species complex have demonstrated that HSP plays an important role in thermal adaption, and there is convincing evidence based on the HSP90 gene indicating that MED species are adapted to a wider range of temperature than MEAM1 (Jiang et al., 2017; Xiao et al., 2016; Díaz et al., 2014; Ma et al., 2014). We identified recombinant events in the MED recently collected from the tropical region of Brazil and additional studies show that the MEAM1 cryptic species are being replaced by MED cryptic species in China (Xiao et al., 2016, Chu et al., 2010). Based on our results, we hypothesise that recombination in the HSP90 gene may not have caused significant evolution of MEAM1 species to adapt to climate changes. However, thermal adaptation of whiteflies may depend on many other factors, such as insecticide resistance, host plants, endosymbiont composition and viruses (Sun et al., 2013; Aregbesola et al., 2018). The invasive MED cryptic species is thought to have originated from North Africa and they could have a connection with the SSA species (Boykin et al., 2007).

We observed that the some SSA species in Figure 1 had recombination events while others did not have any recombination events. This could be explained by a number of factors; i.e., whitefly host plants, the viruses carried by whiteflies in East Africa (which are different from those in other geographic regions) and, lastly, the endosymbiont composition of the whiteflies (Tsueda and Tsuchida, 2011; Brumin et al., 2011; Pusag

et al., 2012). Further research is needed to evaluate these interesting patterns on the *B. tabaci* SSA species.

Figure 2 represents a Bayesian estimate of a phylogenetic tree derived from an alignment of sequences without recombination events. We did not estimate the Bayesian phylogenetic tree for the dataset with recombination events because different regions of the alignment might have different underlying evolutionary histories due to recombination (Posada and Crandall, 2002; Schierup and Hein, 2000). Therefore, ignoring recombination might lead to overestimation of the substitution rate heterogeneity and loss of molecular clock (Schierup and Hein, 2000). Usually phylogenetic trees estimated without removing recombination events have larger terminal branches similar to those of exponentially growing populations (Schierup and Hein, 2000). This is different for our situation in Figure 2 since we have short terminal branches. Hence, all portions of recombination events were removed prior to phylogenetic tree estimation. Therefore, recombination drives evolution since it has an effect on the shape of the phylogenetic tree.

Despite the increase in temperature, whitefly numbers have not decreased (Xiao et al., 2016). We therefore hypothesise that recombination in the HSP90 gene of whiteflies could have played a role in whitefly thermal adaptation. The optimal temperature for whitefly development ranges from 28.0^0 C to 32.0^0 C, which is the current temperature range in the East African region (Bonato et al., 2007; Curnutte et al., 2014). The chaperone function of HSP90 is also a factor in the survival of whiteflies, as it is triggered by temperature changes (Hoffmann et al., 2003; Diaz et al., 2015). While this is not a conclusive study on climate change in the region, the findings help us understand why there are high numbers of whiteflies in the cassava fields in East Africa,

thus forming a basis for future studies. Therefore, we recommend that gene expression studies be carried out on SSA species in East Africa.

3.5 Conclusion

Temperature is a key determinant is the geographical distribution and survival of species. Theoretical and experimental results from other studies confirm that HSP90 plays an important role in thermal adaptation but none of these has been conducted on SSA whitefly species. We conclude that recombination events are present in the HSP90 gene of *B. tabaci* and this could be the major driver for superabundance or adaptive evolution of whiteflies in regions with varying environmental temperatures.

3.6 Methods

3.6.1 Whitefly sampling

Seven whitefly samples were collected from whitefly 'hot spots' in Uganda in a country-wide survey conducted in 2013 from the Nakaseke, Luwero and Nakasongola districts of Uganda. Adult whiteflies from symptomatic cassava were collected from the top five leaves using an aspirator, transferred immediately into 70% ethanol in an Eppendorf tube, and then later exported to the laboratory at the University of Western Australia (UWA) where they were kept at -20^0 C before analysis. Five whiteflies from Kenya were collected from symptomatic common beans with an aspirator. Collaborators from Brazil, Australia and Malawi also sent whitefly samples already preserved in 70% ethanol and they were stored at -20^0 C upon arrival at UWA. See additional file 1.

3.6.2 Total RNA extraction from individual whiteflies

RNA extractions were performed in the UWA genomics laboratory using an Arcturus PicoPure RNA Isolation Kit (Arcturus, CA, USA) following extraction procedures from Sseruwagi et al. (2017). To remove contaminating DNA, DNase and divalent cations from the extracted RNA, we used the TURBO DNA free kit as described by the manufacturer (Ambion Life Technologies, CA, USA). To increase the concentration of the RNA, we used a vacuum centrifuge (Eppendorf, Germany) set at room temperature for one hour, then we resuspended the pellet in 18 µl of RNase free water and stored the prepared RNA immediately at -80^0 C awaiting further analysis. RNA integrity was quantified by 2100 Bio-analyser (Agilent Technologies).

3.6.3 cDNA and Illumina library preparation

An Illumina TruSeq Stranded Total RNA preparation kit was used to make cDNA libraries from the RNA of each individual whitefly, as described by the manufacturer (Illumina, San Diego, CA, USA). Libraries were sent to Macrogen, Korea (www.macrogen.com), where 15 samples were sequenced on a HiSeq 2500 on a rapid run mode and 6 samples on HiSeq 4000 on a high-throughput mode (see additional file 2). Sequencing control software HCS v2.2 and HCS v3.3 were used for base calling and quality assessment respectively.

3.6.4 NGS data analysis (*de novo* sequence assembly and mapping)

All samples were trimmed using CLC Genomic workbench 8.5.1 (CLCGW) with the quality score limit set to 0.01, and the ambiguous limit set to 2. Trimmed reads were assembled into contigs using the *de novo* assembly function of CLCGW with automatic word size, automatic bubble size, minimum contig length 500, mismatch cost 2,

insertion cost 3, deletion cost 3, length fraction 0.5 and similarity fraction 0.9 (additional file 2). Contigs were imported in Geneious 9.1.8 (Kearse et al., 2012) on Mac OS 10.6, and contigs were then mapped onto the reference sequence for HSP90 MEAM1 EU934241 from GenBank. Mapping was performed with the following settings in Geneious: minimum overlap 10%, minimum overlap identity 80%, allow gaps 10%, fine tuning set to iterate up to 10 times at custom sensitivity. The consensus contig from mapping was aligned to the *de novo* contig using MAFFT (Katoh et al., 2002). The resulting alignment consensus was manually inspected for ambiguities and gave rise to the new HSP90 sequences covering the full coding region of a length of 2160 base pairs. Open reading frames (ORFs) were predicted in Geneious and other HSP90 sequences used in this study were downloaded from Genbank (i.e., HM013713, HM367080, HM01370, DQ093380 and DQ093381). Alignment of the new HSP90 sequences with Genbank sequences was performed in Geneious using the MAFFT plugin (Katoh et al., 2002). A total of 27 sequences of 2160 base pairs were aligned.

3.6.5 Recombination analysis

An alignment of 27 sequences of the coding region of the HSP90 gene was screened for recombination using the recombination detection program RDP4 (Martin et al., 2015). An extensive array of methods implemented in RDP4 was used for detecting and visualising recombination events; for example, RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), Bootscan (Martin et al., 2005), MaxChi (Smith, 1992), Chimaera (Posada and Crandall, 2001) and SiScan (Gibbs et al., 2000). Default parameter settings were used with a Bonferroni corrected p-value cut-off of 0.05. Recombination events detected by four or more methods with significant phylogenetic support were considered reliable evidence for recombination (Table 1). A recombinant-free dataset was generated following the recombination analysis by removal of all parts

of the alignment that were detected as recombinant events and replaced with gap characters in RDP4 software. The recombination-free dataset was later used for phylogenetic analysis.

3.6.6 Phylogenetic analysis

The recombination-free dataset generated from RDP4 software was used in phylogenetic analysis. We ran jModelTest 2 (Darriba et al., 2012) on the dataset to statistically select the best-fit model of nucleotide substitution. The best model was GTR+I+G and this model was then used in a Bayesian phylogenetic analysis. We ran MrBayes 3.2.2 (Ronquist et al., 2012) in parallel on Magnus (Pawsey supercomputing Centre, Perth, Western Australia) and a phylogenetic tree was constructed. Four Markov chains were run for 50 million generations: trees were sampled every 1000 generations, and 12,500 sub-optimal trees were discarded as burn-in at the beginning of the MCMC run. No runs indicated a lack of convergence and the potential scale reduction factor for all parameters approached one. The effective sample size for all parameters was above 200 for each run.

3.7 List of abbreviations

ASIAII: Asia II

AUSI: Australia I

CBSD: Cassava brown streak disease

CMD: Cassava mosaic disease

MCMC: Markov chain Monte Carlo

MEAMI: Middle East Asia Minor I

MED: Mediterranean

NWI: New World I

NWAfrica: New World Africa

SSA: Sub-Saharan Africa

UGA: Uganda

UWA: University of Western Australia

3.8 Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon request and sequence data were submitted to Genbank with accession numbers (MH383308–MH383328).

Competing interests

Not applicable.

Funding

The Natural Resources Institute, University of Greenwich, supported sequencing of the whitefly samples collected in this study, fieldwork and the first 2 years of TK's PhD

from the grant provided by the Bill and Melinda Gates Foundation (Investment ID OPP1058938).

Author's contributions

TK carried out the recombination analysis with inputs from LMB, LSL and AS. TK, BRD, JMW participated in wet laboratory work, TK, JMW, BRD, CAO and TA provided samples. TK drafted the manuscript and all authors read and approved the manuscript.

3.9 Acknowledgements

TK is supported by a University of Western Australia Scholarship for International Research Fees (SIRF), University Postgraduate Awards (UPA) and a University of Western Australia Safety Net top-up scholarship. This work forms part of his PhD. The Natural Resources Institute, University of Greenwich, supported the sequencing, fieldwork and the first 2 years of TK's PhD from the grant provided by the Bill and Melinda Gates Foundation (Investment ID OPP1058938). The National Crop Resources Research Institute (NaCRRI) in Uganda provided samples and facilitated sample collection. We thank Paul De Barro, JMW, BRD, and Donald Kachigamba for providing samples. Resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia supported computation analysis. We thank Joanne Edmondston from the Graduate Research School at UWA, Joe Bielawski and Laura Kubatko for reviewing the earlier versions of this manuscript. Finally, we thank the farmers in East Africa for allowing us to collect samples on their farms.

3.10 References

1. Brown J, Coats S, Bedford I, Markham P, Bird J, Frohlich D: **Characterization and distribution of esterase electromorphs in the whitefly, Bemisia tabaci (Genn.)(Homoptera: Aleyrodidae)**. *Biochemical genetics* 1995, **33**(7-8):205-214.

2. Hoddle M: **The biology and management of silverleaf whitefly, Bemisia argentifolii Bellows and Perring (Homoptera: Aleyrodidae) on greenhouse grown ornamentals**. *Department of Entomology, University of California, Riverside USA Applied Biological Control Research* 1999.

3. Boykin LM, Armstrong KF, Kubatko L, De Barro P: **Species delimitation and global biosecurity**. *Evolutionary bioinformatics online* 2012, **8**:1.

4. Boykin LM, Bell CD, Evans G, Small I, De Barro PJ: **Is agriculture driving the diversification of the Bemisia tabaci species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)?: Dating, diversification and biogeographic evidence revealed**. *BMC Evol Biol* 2013, **13**:228.

5. Boykin LM: **Bemisia tabaci nomenclature: lessons learned**. *Pest management science* 2014, **70**(10):1454-1459.

6. Boykin LM, De Barro PJ: **A practical guide to identifying members of the Bemisia tabaci species complex: and other morphologically identical species**. *Frontiers in Ecology and Evolution* 2014, **2**:45.

7. Liu S-s, Colvin J, De Barro PJ: **Species concepts as applied to the whitefly Bemisia tabaci systematics: how many species are there?** *Journal of Integrative Agriculture* 2012, **11**(2):176-186.

8. Omongo CA, Kawuki R, Bellotti AC, Alicai T, Baguma Y, Maruthi M, Bua A, Colvin J: **African cassava whitefly, Bemisia tabaci, resistance in African and South American cassava genotypes**. *Journal of integrative agriculture* 2012, **11**(2):327-336.

9. Brown JK, Czosnek H: **Whitefly transmission of plant viruses**. 2002.

10. Patil BL, Legg JP, Kanju E, Fauquet CM: **Cassava brown streak disease: a threat to food security in Africa**. *Journal of General Virology* 2015, **96**(5):956-968.

11. Alicai T, Omongo C, Maruthi M, Hillocks R, Baguma Y, Kawuki R, Bua A, Otim-Nape G, Colvin J: **Re-emergence of cassava brown streak disease in Uganda**. *Plant Disease* 2007, **91**(1):24-29.

12. Colvin J, Omongo C, Maruthi M, Otim‐Nape G, Thresh J: **Dual begomovirus infections and high Bemisia tabaci populations: two factors driving the spread of a cassava mosaic disease pandemic**. *Plant Pathology* 2004, **53**(5):577-584.

13. Legg JP, Sseruwagi P, Boniface S, Okao-Okuja G, Shirima R, Bigirimana S, Gashaka G, Herrmann H-W, Jeremiah S, Obiero H: **Spatio-temporal patterns of genetic change amongst populations of cassava Bemisia tabaci whiteflies driving virus pandemics in East and Central Africa**. *Virus research* 2014, **186**:61-75.

14. Cahill M, Byrne FJ, Gorman K, Denholm I, Devonshire AL: **Pyrethroid and organophosphate resistance in the tobacco whitefly Bemisia tabaci**

(Homoptera: Aleyrodidae). *Bulletin of Entomological Research* 1995, **85**(02):181-187.

15. Howeler RL, Thomas N, Arrieta Bolaños S, Ochoa Henríquez JM, Odio Echeverría F, Moreno V, Simonetti LLA, Dardón O, RamiroMejía FSC, Preissing X: **Save and grow: cassava. A guide to sustainable production intensificationProduire plus avec moins Ahorrar para crecer**: FAO, Roma (Italia); 2013.

16. Kinene T, Luboobi LS, Nannyonga B, Mwanga GG: **A mathematical model for the dynamics and cost effectiveness of the current controls of cassava brown streak disease in Uganda**. *Journal of Mathematical and Computational Science* 2015, **5**(4):567.

17. Alicai T, Ndunguru J, Sseruwagi P, Tairo F, Okao-Okuja G, Nanvubya R, Kiiza L, Kubatko L, Kehoe MA, Boykin LM: **Cassava brown streak virus has a rapidly evolving genome: implications for virus speciation, variability, diagnosis and host resistance**. *Scientific reports* 2016, **6**.

18. Maruthi M, Hillocks R, Mtunda K, Raya M, Muhanna M, Kiozia H, Rekha A, Colvin J, Thresh J: **Transmission of Cassava brown streak virus by Bemisia tabaci (Gennadius)**. *Journal of Phytopathology* 2005, **153**(5):307-312.

19. Sseruwagi P, Sserubombwe W, Legg J, Ndunguru J, Thresh J: **Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: a review**. *Virus research* 2004, **100**(1):129-142.

20. Ateka E, Alicai T, Ndunguru J, Tairo F, Sseruwagi P, Kiarie S, Makori T, Kehoe MA, Boykin LM: **Unusual occurrence of a DAG motif in the**

Ipomovirus Cassava brown streak virus and implications for its vector transmission. *PloS one* 2017, **12**(11):e0187883.

21. Pereira L: **Climate change impacts on agriculture across Africa**. *Oxford Research Encyclopedia of Environmental Science* 2017.

22. Cossins A, Bowler K: **Rate compensations and capacity adaptations**. In: *Temperature biology of animals*. Springer; 1987: 155-203.

23. Hoffmann AA, Sørensen JG, Loeschcke V: **Adaptation of Drosophila to temperature extremes: bringing together quantitative and molecular approaches**. *Journal of Thermal Biology* 2003, **28**(3):175-216.

24. Kellermann V, Overgaard J, Hoffmann AA, Fløjgaard C, Svenning J-C, Loeschcke V: **Upper thermal limits of Drosophila are linked to species distributions and strongly constrained phylogenetically**. *Proceedings of the National Academy of Sciences* 2012, **109**(40):16228-16233.

25. Fasolo AG, Krebs RA: **A comparison of behavioural change in Drosophila during exposure to thermal stress**. *Biological Journal of the Linnean Society* 2004, **83**(2):197-205.

26. Parsell D, Lindquist S: **The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins**. *Annual review of genetics* 1993, **27**(1):437-496.

27. Somero GN: **Proteins and temperature**. *Annual review of physiology* 1995, **57**(1):43-68.

28. Gorovits R, Czosnek H: **The involvement of heat shock proteins in the establishment of Tomato yellow leaf curl virus infection**. *Frontiers in Plant Science* 2017, **8**.

29. Huang YW, Hu CC, Liou MR, Chang BY, Tsai CH, Meng M, Lin NS, Hsu YH: **Hsp90 interacts specifically with viral RNA and differentially regulates replication initiation of Bamboo mosaic virus and associated satellite RNA**. *PLoS pathogens* 2012, **8**(5):e1002726.

30. Diaz F, Orobio RF, Chavarriaga P, Toro-Perea N: **Differential expression patterns among heat-shock protein genes and thermal responses in the whitefly Bemisia tabaci (MEAM 1)**. *Journal of thermal biology* 2015, **52**:199-207.

31. Yu H, Wan FH: **Cloning and expression of heat shock protein genes in two whitefly species in response to thermal stress**. *Journal of Applied Entomology* 2009, **133**(8):602-614.

32. Salvucci ME, Stecher DS, Henneberry TJ: **Heat shock proteins in whiteflies, an insect that accumulates sorbitol in response to heat stress**. *Journal of Thermal Biology* 2000, **25**(5):363-371.

33. Cui X, Wan F, Xie M, Liu T: **Effects of heat shock on survival and reproduction of two whitefly species, Trialeurodes vaporariorum and Bemisia tabaci biotype B**. *Journal of Insect Science* 2008, **8**(24):1-10.

34. Lü Z-C, Wan F-H: **Using double-stranded RNA to explore the role of heat shock protein genes in heat tolerance in Bemisia tabaci (Gennadius)**. *Journal of Experimental Biology* 2011, **214**(5):764-769.

35. Mahadav A, Kontsedalov S, Czosnek H, Ghanim M: **Thermotolerance and gene expression following heat stress in the whitefly Bemisia tabaci B and Q biotypes**. *Insect biochemistry and molecular biology* 2009, **39**(10):668-676.

36. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B: **RDP4: Detection and analysis of recombination patterns in virus genomes**. *Virus evolution* 2015, **1**(1).

37. Cangi N, Gordon JL, Bournez L, Pinarello V, Aprelon R, Huber K, Lefrançois T, Neves L, Meyer DF, Vachiéry N: **Recombination is a major driving force of genetic diversity in the anaplasmataceae Ehrlichia ruminantium**. *Frontiers in cellular and infection microbiology* 2016, **6**:111.

38. Jiang R, Qi L-D, Du Y-Z, Li Y-X: **Thermotolerance and Heat-Shock Protein Gene Expression Patterns in Bemisia tabaci (Hemiptera: Aleyrodidae) Mediterranean in Relation to Developmental Stage**. *Journal of economic entomology* 2017, **110**(5):2190-2198.

39. Xiao N, Pan L-L, Zhang C-R, Shan H-W, Liu S-S: **Differential tolerance capacity to unfavourable low and high temperatures between two invasive whiteflies**. *Scientific reports* 2016, **6**:24306.

40. Díaz F, Muñoz‐Valencia V, Juvinao‐Quintero D, Manzano‐Martínez M, Toro‐Perea N, Cárdenas‐Henao H, Hoffmann A: **Evidence for adaptive divergence of thermal responses among B emisia tabaci populations from tropical C olombia following a recent invasion**. *Journal of evolutionary biology* 2014, **27**(6):1160-1171.

41. Ma F-Z, Lü Z-C, Wang R, Wan F-H: **Heritability and evolutionary potential in thermal tolerance traits in the invasive Mediterranean cryptic species of Bemisia tabaci (Hemiptera: Aleyrodidae)**. *PloS one* 2014, **9**(7):e103279.

42. Chu D, Wan FH, Zhang YJ, Brown JK: **Change in the biotype composition of Bemisia tabaci in Shandong Province of China from 2005 to 2008**. *Environmental entomology* 2010, **39**(3):1028-1036.

43. Sun D-B, Liu Y-Q, Qin L, Xu J, Li F-F, Liu S-S: **Competitive displacement between two invasive whiteflies: insecticide application and host plant effects**. *Bulletin of entomological research* 2013, **103**(3):344-353.

44. Aregbesola OZ, Legg JP, Sigsgaard L, Lund OS, Rapisarda C: **Potential impact of climate change on whiteflies and implications for the spread of vectored viruses**. *Journal of Pest Science* 2018:1-12.

45. Boykin LM, Shatters RG, Jr., Rosell RC, McKenzie CL, Bagnall RA, De Barro P, Frohlich DR: **Global relationships of Bemisia tabaci (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences**. *Mol Phylogenet Evol* 2007, **44**(3):1306-1319.

46. Tsueda H, Tsuchida K: **Reproductive differences between Q and B whiteflies, Bemisia tabaci, on three host plants and negative interactions in mixed cohorts**. *Entomologia Experimentalis et Applicata* 2011, **141**(3):197-207.

47. Brumin M, Kontsedalov S, Ghanim M: **Rickettsia influences thermotolerance in the whitefly Bemisia tabaci B biotype**. *Insect Science* 2011, **18**(1):57-66.

48. Pusag JCA, Jahan SH, Lee K-S, Lee S, Lee K-Y: **Upregulation of temperature susceptibility in Bemisia tabaci upon acquisition of Tomato yellow leaf curl virus (TYLCV)**. *Journal of insect physiology* 2012, **58**(10):1343-1348.

49. Posada D, Crandall KA: **The effect of recombination on the accuracy of phylogeny estimation**. *Journal of molecular evolution* 2002, **54**(3):396-402.

50. Schierup MH, Hein J: **Consequences of recombination on traditional phylogenetic analysis**. *Genetics* 2000, **156**(2):879-891.

51. Bonato O, Lurette A, Vidal C, Fargues J: **Modelling temperature‐dependent bionomics of Bemisia tabaci (Q‐biotype)**. *Physiological Entomology* 2007, **32**(1):50-55.

52. Curnutte LB, Simmons AM, Abd-Rabou S: **Climate change and Bemisia tabaci (Hemiptera: Aleyrodidae): impacts of temperature and carbon dioxide on life history**. *Annals of the Entomological Society of America* 2014, **107**(5):933-943.

53. Sseruwagi P, Wainaina J, Ndunguru J, Tumuhimbise R, Tairo F, Guo J-Y, Vrielink A, Blythe A, Kinene T, De Marchi B: **The first transcriptomes from field-collected individual whiteflies (Bemisia tabaci, Hemiptera: Aleyrodidae)**. *Gates Open Research* 2017, **1**.

54. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data**. *Bioinformatics* 2012, **28**(12):1647-1649.

55. Katoh K, Misawa K, Kuma Ki, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform**. *Nucleic acids research* 2002, **30**(14):3059-3066.

56. Martin D, Rybicki E: **RDP: detection of recombination amongst aligned sequences**. *Bioinformatics* 2000, **16**(6):562-563.

57. Padidam M, Sawyer S, Fauquet CM: **Possible emergence of new geminiviruses by frequent recombination**. *Virology* 1999, **265**(2):218-225.

58. Martin D, Posada D, Crandall K, Williamson C: **A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints**. *AIDS Research & Human Retroviruses* 2005, **21**(1):98-102.

59. Smith JM: **Analyzing the mosaic structure of genes**. *Journal of molecular evolution* 1992, **34**(2):126-129.

60. Posada D, Crandall KA: **Evaluation of methods for detecting recombination from DNA sequences: computer simulations**. *Proceedings of the National Academy of Sciences* 2001, **98**(24):13757-13762.

61. Gibbs MJ, Armstrong JS, Gibbs AJ: **Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences**. *Bioinformatics* 2000, **16**(7):573-582.

62. Darriba D, Taboada GL, Doallo R, Posada D: **jModelTest 2: more models, new heuristics and parallel computing**. *Nature methods* 2012, **9**(8):772-772.

63. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP: **MrBayes 3.2: efficient**

Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* 2012, **61**(3):539-542.

3.11 Figure legends

Figure 1. Shows results of the recombination analysis performed on an HSP90 sequence alignment of a length of 2160 base pairs. Recombination events are shown in white, and black represents the segment without recombination events.

Figure 2. Bayesian phylogenetic tree for the HSP90 gene created from an alignment where positions identified as recombinant were removed. A long branch with a mixture of sequences from groups A, C and D was formed. The branch length is measured as the expected number of nucleotide substitutions per site.

3.12 Table legends

Table 1. A summary of the results of recombination analysis from various recombination-detection tests, using RDP4 software

Additional file 1. Table showing species ID, Genbank accession numbers, sample locations and host plants where each species was found

Additional file 2. A summary of the next generation sequencing (NGS) data analysis results in this study

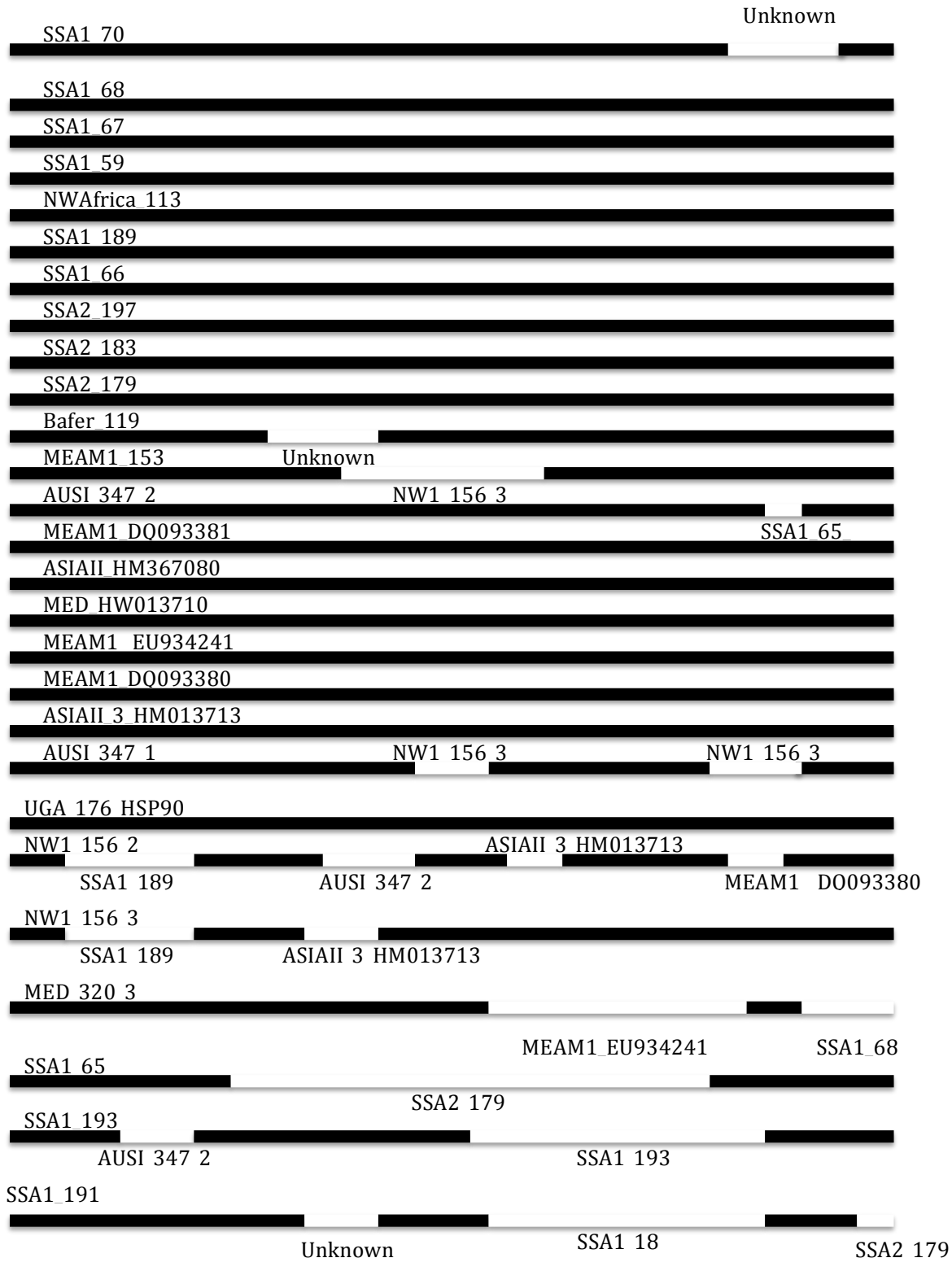


Figure 1.

Shows results of the recombination analysis performed on an HSP90 sequence alignment of a length of 2160 base pairs. Recombination events are shown in white, and black represents the segment without recombination events.

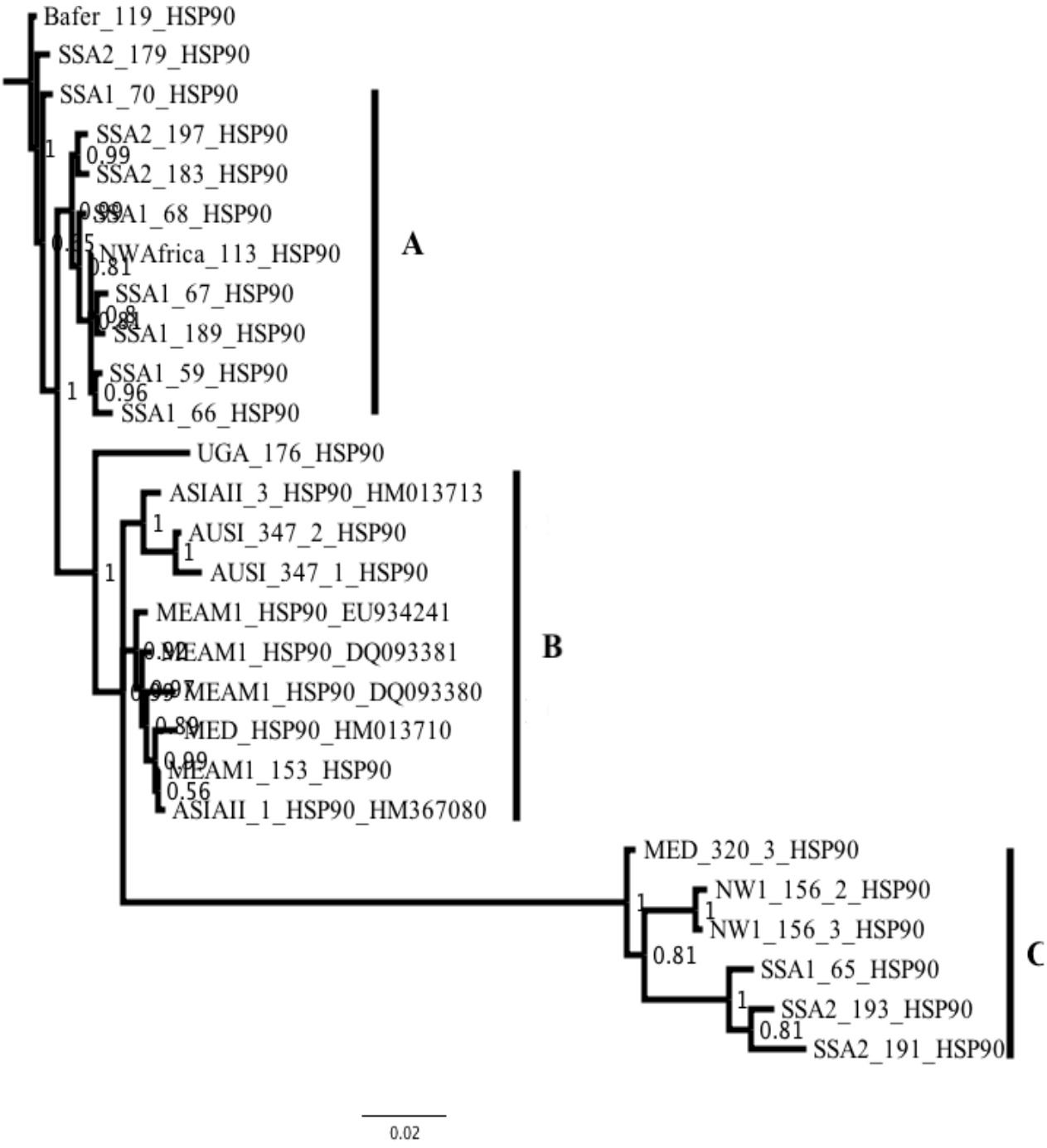


Figure 2.

Bayesian phylogenetic tree for the HSP90 gene created from an alignment where positions identified as recombinant were removed. A long branch with a mixture of sequences from groups A, C and D was formed. The branch length is measured as the expected number of nucleotide substitutions per site.

Table 1.

A summary of the results of recombination analysis from various recombination-detection tests, using RDP4 software

Event	Recombinant Sequence(s)	Major parent	Minor parent	Detected method in RDP4	Detected break-point positions	Highest P-value
1	SSA1_65_HSP90	NW1_156_3_HSP90	SSA2_179_HSP90	R G B M C ST	569 - 1682	1.465x10-25
2	^SSA2_191_HSP90	Unknown	SSA1_189_HSP90	R G _ M C ST	1101 - 1762	8.489x10-20
3	^NW1_156_3_HSP90	Unknown	SSA1_189_HSP90	R G B M CS T	132* - 542	4.575x10-14
4	SSA1_70_HSP90	SSA1_189_HSP90	Unknown	R G B M CS T	1927 - 2077*	1.134x10-13
5	MED_320_3_HSP90	SSA2_193_HSP90	SSA1_68_HSP90	R G _ M C ST	1940 - 2160*	4.243x10-13
6	^SSA2_191_HSP90	NW1_156_3_HSP90	Unknown (ASIAII_3_HSP90_HM013713)	RG_BMCST	674 - 856	2.60x10 -11
7	AUS1_347_1_HSP90	MEAM1_153_HSP90	NW1_156_3_HSP90	R G B M CS T	1884 - 2074*	4.730x10-11
8	MEAM1_153_HSP90	MED_HSP90_HM013710	MW1_156_3_HSP90	R G B M C ST	574* -1030	5.032x10-10
9	AUSI_347_1_HSP90	AUSI_347_2_HSP90	NW1_156_3_HSP90	_G B M C S T	896 - 1036	9.11x 10 -10
10	MED_320_3_HSP90	NW1_156_3_HSP90	MEAM1_HSP90_EU934241	_ G B M C ST	1091 - 1752	9.735x10-10
11	SSA2_191_HSP90	SSA2_193_HSP90	SSA2_179_HSP90	R G _ M C _ T	2057 – 2160*	2.50x 10 -8
12	^SSA2_193_HSP90	MED_320_3_HSP90	AUSI_347_2_HSP90	R G B M C _ T	318*- 504	2.95x 10-8
13	^NW1_156_2_HSP90	NM1_156_3_HSP90	ASIAII_3_HSP90_HM013713	R G _MC_ _ T	1076 - 1257	4.43 x 10 - 8
14	NW1_156_2_HSP90	NW1_156_3_HSP90	MEAM1_HSP90_DQ093380	_G B M C S T	1730-1888	2.93 x 10 -8
15	Bafer_119_HSP90	SSA1_66_HSP90	Unknown	R G _ M CS T	405* - 616	4.376x10-06
16	AUSI_347_2_HSP90	UGA_176_HSP90	SSA1_65_HSP90	R G _ _ M _ _ T	1970 - 2026	1.90 x 10 -5
17	^NW1_156_2_HSP90	MED_320_3_HSP90	AUS1_347_2_HSP90	_G _ _ _ T	738 - 886	1.292x10-03
18	NW1_156_3_HSP90	Unknown (MEAM1_HSP90_DQ093380)	ASIAII_3_HSP90_HM013713	_ G _ _ _ _ _	694 - 852	1.352 x 10 -3

The table represents recombination events, the abbreviations; G- GENECONV, B-Bootscan, M- Maxchi, C- Chimaera and S- Siscan represent recombination detection packages implemented in RDP4.

* The actual breakpoint position is undetermined (It was most likely over-printed by a subsequent recombination event).

^ The recombinant sequence may have been misidentified (one of the identified parents might be the recombinant).

Unknown The sequence listed as ‘unknown’ was used to infer the existence of a missing parent sequence.

The red colour represents the possibility that an evolutionary process other than recombination could have caused this apparent recombination signal.

Additional file 1

Table showing species ID, Genbank accession numbers, sample locations and host plants where each species was found

Sample ID	Species ID	Genbank Accession #	Country	Latitude	Longitude	Altitude	Host plant
59	SSA1	MH383311	Uganda	0.85070	32.41920	1078	Cassava
65	SSA1	MH383326	Uganda	0.87345	32.64796	1085	Cassava
66	SSA1	MH383314	Uganda	0.86932	32.51936	1113	Cassava
67	SSA1	MH383310	Uganda	1.41966	32.43936	1064	Cassava
68	SSA1	MH383309	Uganda	1.19876	32.73804	1069	Cassava
70	SSA1	MH383308	Uganda	1.31247	32.58484	1042	Cassava
113	NWAfrica	MH383312	Uganda	0.54896	31.44420	1283	Mujaja
119	*B.afer*	MH383318	Malawi	-12.792975	34.200964	500	Cassava
153	MEAM1	MH383320	Brazil	-22.844553	-48.435439	800	Cabbage
156_2	NWI	MH383323	Brazil	-21.980278	-50.788312	397	Euphorbia heterophylla
156_3	NWI	MH383324	Brazil	-21.980278	-50.788312	397	Euphorbia heterophylla
176	UGA	MH383319	Kenya	00.00622	034.73450	1510	Common bean
179	SSA2	MH383317	Kenya	00.02544	034.62852	1561	Common bean
189	SSA1	MH383313	Kenya	00.33262	034.25285	1322	Common bean
191	SSA2	MH383328	Kenya	00.56222	034.17064	1165	Common bean
193	SSA2	MH383327	Kenya	00.61071	037.37062	1167	Common bean
197	SSA2	MH383315	Kenya	00.59737	034.43888	1331	Common bean
320_3	MED	MH383325	Brazil	-23.01694	-048.06861		Cassava
347_1	AUSI	MH383322	Australia				
347_2	AUSI	MH383321	Australia				
183	SSA2	MH383316	Kenya	00.00622	034.55100	1561	Common bean

Additional file 2

A summary of the next generation sequencing (NGS) data analysis results in this study

Sample ID	Number of reads obtained	Reads after trimming	Number of Contigs	Average Contig length	N50
59	8,000,000	7,848,046	16,473	1,102	1,195
65	8,000,000	7,857,579	17,146	1,024	1,071
66	8,000,000	7,857,564	16,489	1,045	1,095
67	8,000,000	7,876,283	18,271	1,165	1,289
68	8,000,000	7,861,938	19,852	1,201	1,368
70	8,000,000	7,861,502	20,182	1,113	1,205
113	8,000,000	7,850,359	13,156	1,023	1,050
119	8,000,000	7,845,846	20,368	1,238	1,400
153	36,449,340	36,232,156	25,435	1,235	1,432
156_2	40,826,418	40,604,557	27,466	1,356	1,689
156_3	43,194,264	42,953,727	30,716	1,346	1,655
176	8,000,000	7,835,023	19,865	1,366	1,616
179	8,000,000	7,842,509	15,570	1,194	1,295
189	8,000,000	7,823,762	2,530	927	910
191	8,000,000	7,833,877	19,770	1,124	1,222
193	8,000,000	7,852,051	20,061	1,122	1,218
197	8,000,000	7,871,950	19,224	1,230	1,408
320_3	39,894,296	39,686,374	27,867	1,144	1,250
347_1	39,142,358	38,894,512	30,496	1,372	1,723
347_2	37,306,246	37,071,584	27,343	1,314	1,614
183	8,000,000	7,816,022	20,066	1,298	1,512

Chapter 4: Patterns of Conflict between Nuclear and Mitochondrial Cytochrome Oxidase I (mtCOI) Phylogenies in *Bemisia tabaci* Species Complex

This article will be submitted for publication in April 2019 in PlosOne Journal.

Tonny Kinene*[1], Laura Kubatko[2], Titus Alicai[3], Christopher Abu Omongo[3], Livingstone S. Luboobi[4], and Laura M. Boykin[1] (2019) Patterns of conflict between nuclear and mitochondrial cytochrome oxidase I (mtCOI) phylogenies in *Bemisia tabaci* species complex

Title: Patterns of conflict between nuclear and mitochondrial cytochrome oxidase I (mtCOI) phylogenies in *Bemisia tabaci* species complex

Authors: Tonny Kinene*[1], Laura Kubatko[2], Titus Alicai[3], Christopher Abu Omongo[3], Livingstone S. Luboobi[4], and Laura M. Boykin[1] (2019). Patterns of conflict between nuclear and mitochondrial cytochrome oxidase I (mtCOI) phylogenies in *Bemisia tabaci* species complex

Addresses

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

[2]The Ohio State University, 154W 12th Avenue, Columbus, Ohio 43210, USA

[3]National Crops Resources Research Institute, P.O. Box 7084, Kampala, Uganda

[4]Department of Mathematics, Makerere University, P.O. Box 7062, Kampala, Uganda

*Author for correspondence: Tonny Kinene, Email: kinenetonny@gmail.com; tonny.kinene@research.uwa.edu.au

Authors' email addresses:

Laura Kubatko: lkubatko@stat.osu.edu

Titus Alicai: talicai@hotmail.com

Christopher Abu Omongo: chrisomongo@yahoo.com

Livingstone Serwadda Luboobi: luboobi@gmail.com

Laura M. Boykin: laura.boykin@uwa.edu.au

4.1 Abstract

The genetic diversity of the *Bemisia tabaci* species complex has been assessed based on mtCOI gene, restriction-site-associated DNA sequencing (RAD) and mating experiments. The advent of new sequencing technology has made it possible to collect an immense number of unlinked nuclear loci. It has been suggested that the mtCOI locus might not be an effective molecular marker for delimiting *B. tabaci* species. Therefore, a combination of the mtCOI and multiple unlinked nuclear loci might be important in defining species boundaries. In this study we generated 137 mtCOI gene sequences, 69 shaker cognate w sequences and 44 RNA polymerase II gene sequences from East Africa *B. tabaci* species using Sanger sequencing.

We investigated whether nuclear genes would be as useful as the mtCOI gene to delimit members of the *B. tabaci* species complex. We found that the Bayesian estimate for the phylogenies for the nuclear genes were not concordant with the mtCOI gene tree topology. This could be due to incomplete lineage sorting, horizontal gene transfer or gene duplication. We measured the difference between nuclear gene tree topologies and mtCOI topologies using taxonomic distinctiveness methods and coalescent-based methods. From our results we found that combining these nuclear loci would resolve some conflicts. Thus, a single nuclear molecular marker cannot delimit members of the *B. tabaci* species complex.

From our analysis, the mtCOI molecular marker is still superior to other markers when used as a stand-alone marker. We recommend the use of more genes to generate a robust species tree and this should be done using methods that incorporate the coalescent process.

Keywords: gene tree, mtCOI, RNA Polymerase II, shaker cognate w

4.2 Introduction

Bemisia tabaci (whitefly) is a species complex (Boykin et al., 2007) whose members are known as global pests that are morphologically indistinguishable (Boykin et al., 2007; De Barro et al., 2011b; Liu et al., 2012; Boykin et al., 2012). Whiteflies are polyphagous in nature; they transmit several plant viruses (Polston and Capobianco, 2013; Brown and Czosnek, 2002; Maruthi et al., 2005) and cause physiological disorders on plants. In East Africa the members of the *B. tabaci* species complex commonly transmit cassava mosaic *Geminiviruses* genus *Begomovirus* and cassava brown streak virus genus *Ipomovirus* of the *Potyviridae* family (Maruthi et al., 2005; Legg et al., 2006; Ndunguru et al., 2015; Alicai et al., 2016; Alicai et al., 2007). Yield losses due to these viruses are estimated at 47% (Omongo et al., 2012; Legg et al., 2006). This has resulted in recurrent famine and hunger in the East African region.

The genetic diversity of members of the *B. tabaci* species complex has been assessed based on NextRAD sequencing (Wosula et al., 2017), mating experiments (Liu et al., 2012; Xu et al., 2010; Wang et al., 2010; Wang et al., 2011; Elbaz et al., 2010) and sequence data of the partial mitochondrial cytochrome oxidase I (mtCOI) gene (Boykin et al., 2007; Boykin and De Barro, 2014; De Barro, 2012). These analyses have yielded insights into several relationships between *B. tabaci* genetic groups; however, it has been debated that the mtCOI gene alone might not be sufficient for species delimitation and might produce misleading signals (Wiens et al., 2010; Shaw, 2002; Wosula et al., 2017). Despite the pros and cons of using a single locus (mtCOI), many phylogenetic papers published to date on *B. tabaci* species continue using the partial mtCOI gene for

inferring species relationships (Mugerwa et al., 2018; Wosula et al., 2017; Tocko-Marabena et al., 2017; Boykin et al., 2007).

The advent of next generation sequencing technology has made it easy to collect an immense number of nuclear loci (Hsieh et al., 2014) and it has been suggested that data from unlinked loci – for example nuclear genes – may be beneficial in overcoming misleading signals from individual genes (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007; Kubatko et al., 2011; Liu et al., 2009). Discordance between gene trees and species trees may be due to evolutionary processes such as incomplete lineage sorting, horizontal gene transfer, gene duplication and extinction (Maddison, 1997), and thus any single gene tree may not give a good indication of the species-level phylogenetic relationships.

In this study, we investigated whether nuclear genes would be as useful as the mtCOI gene to delimit members of *B. tabaci* species complex. We did this by generating Bayesian estimates of phylogenies of two nuclear genes (shaker cognate gene w and RNA polymerase II) and one mtCOI gene. We found that nuclear gene tree topologies were not concordant with the mtCOI gene tree topology. We measured how much nuclear gene tree topologies differ from mtCOI gene tree topology using monopoly-based tests for taxonomic distinctiveness (Rosenberg, 2007) and coalescent-based methods (Edwards et al., 2007; Yang, 2015). From our results we found that combining the two nuclear loci would resolve some of the conflicts. Thus, a single nuclear molecular marker cannot delimit members of the *B. tabaci* species complex, but a combination of several nuclear markers would be advantageous in generating a robust species tree. From our analysis, the mtCOI molecular marker is still superior to other markers when used as a stand-alone marker. We recommend the use of more nuclear

genes to generate a robust species tree estimate using methods that incorporate the coalescent process.

4.3 Materials and Methods

Sample collection was carried out on farms in Uganda, Tanzania and Malawi between 2015 and 2018. Adult whiteflies were collected from the uppermost five leaves of the cassava plant using an aspirator and immediately stored in 70% ethanol in a 1.5 ml eppendorf tube. At each site, the name of the village was recorded together with the corresponding geographical co-ordinates using the Global positioning system (GPS) (Figure 5). Samples were then shipped to the laboratory at the University of Western Australia (UWA) where they were kept at -20^0 C awaiting molecular analysis.

4.3.1 Genomic DNA Isolation

Genomic DNA was extracted from individual whiteflies using Zymo Research Tissue and Insect DNA extraction kit (Zymo, USA) as described by the manufacturer. To obtain maximal yields of the genomic DNA, the last step of the extraction protocol was modified and thus genomic DNA was eluted using 30 μl of nuclease-free water followed by a centrifugation step of 10,000 $x\ g$ for 30s. The eluted DNA was then stored at -20^0 C awaiting further analysis.

4.3.2 Polymerase chain reaction PCR

Three genes were targeted for amplification: mitochondrial cytochrome oxidase subunit I (mtCOI) and the two nuclear genes – shaker cognate w (shaw) and RNA polymerase II (RNApyII). Amplification of the partial fragment (~820bp) of the mtCOI was performed with 4 primer sets (Table 1): C1-J-2195 and TL_2 –N-3014 (Simon et al.,

1994), Bt-Forward and Bt-reverse (Mugerwa et al., 2018), Btab-Uni primers (Shatters Jr et al., 2009) and 775F and 1489R primers designed at UWA specifically for amplification of the mtCOI region of the *Bemisia afer* whitefly species. The PCR reaction mixture ($20\ \mu l$) contained $5\ \mu l$ of whitefly DNA template, $10\ \mu l$ Bulk AccuPower® PCR PreMix ($1\ U$ of Top DNA polymerase, $250\ \mu M$ of dNTP, $10\ mM$ Tris-HCl (pH 9.0), $30\ mM$ KCl and $1.5\ mM$ $MgCl_2$), $1\ \mu l$ of $5\ \mu M$ forward and $5\ \mu M$ reverse primers. The PCR reaction was run in an eppendorf thermocycler (Eppendorf, Hamburg, Germany) under conditions in Table 1. PCR amplicons were separated in a 2% agarose gel stained with gel red (Biotium, CA, USA) in a 0.5 x Tris-Boric EDTA (TBE) electrophoresis buffer at 100 volts for 30 minutes. PCR products on the gel were visualised under UV light (Biorad, CA, USA) and those with expected size (~820bp) were purified using Qiaqucik PCR purification kit (Qiagen Inc) as described by the manufacturer. The purified PCR products were then sent to Macrogen Korea for bi-direction Sanger sequencing.

The nuclear genes were amplified using two nuclear markers; that is to say, the partial shaker cognate w (Shaw) gene (~510bp) was successfully amplified by PCR with primer set ShawF and ShawR as seen in Table 1, and the partial RNA polymerase II (RNApyII) gene (~ 961bp) was amplified by PCR with primer set RNApyIIF and RNApyIIR. The PCR reaction mixtures were the same as the ones described for mtCOI above and the amplification conditions are shown in Table 1.

4.3.3 Sequence analysis and phylogenetic reconstruction

A total of 137 mtCOI sequences obtained were inspected for ambiguities by looking at their chromatograms in Geneious v. 9.1 (Kearse et al., 2012). Consensus sequences were generated from the forward and reverse sequences where necessary; for some

sequences only the forward direction was considered. A well-curated mtCOI dataset from Boykin et al. (2017) was aligned with the 137 sequences in Geneious using the MAFFT plugin (Katoh et al., 2002). The final combined alignment was translated to ensure sequence alignments were in the correct reading frame and trimming of overhangs was carried out.

For the nuclear genes, a total of 69 Shaw sequences were obtained and aligned with the Shaw dataset from Hsieh et al. (2014) using the MAFFT plugin in Geneious. Similarly, a total of 44 RNApyII sequences were aligned to the RNApyII dataset from Hsieh et al. (2014). Nucleotide substitution models for phylogenetic reconstruction were specified by Akaike information Criterion (AIC) (Posada & Buckley, 2004) using jModeltest 2 (Darriba et al., 2012) with 11 substitution schemes. The optimal model selected for mtCOI was GTR+I+G, TVM+I for Shaw and TIM3+I+G for RNApyII.

Phylogenetic analysis of the three partial genes was carried out in a Bayesian framework. MrBayes 3.2.7 (Ronquist et al., 2012; Ronquist and Huelsenbeck, 2003) was run for each gene for 50 million generations on the Magnus supercomputer (Pawsey Supercomputer Centre at the University of Western Australia). Trees were sampled every 1000 generations and the first 25% of the trees were discarded as burn in. Convergence of the runs was evaluated in Tracer v1.6 (Rambaut et al., 2015) and all runs reached a plateau in likelihood. The effective sample size (ESS) values were above 200 and the potential scale reduction factor (PSRF) was close to one, all of which indicate convergence of runs. A 50% majority rule consensus tree was generated for each gene (Figures 1, 2 and 3) and the length and variability of each gene is shown in Table 2.

4.3.4 Species delimitation

We used the species delimitation plugin in Geneious (Masters et al., 2011) for exploration of species boundaries in the gene trees (Table 3). The plugin was used to compute statistics for Rosenberg's reciprocal monophyly P (AB) for observation of monophyly (Rosenberg, 2007), Kimura two-parameter (K2P) distance and the probability P (Randomly distinct) that a clade has the observed degree of distinctiveness due to random coalescent process (Rodrigo et al., 2008). Bayesian posterior probability for clade support (PP) was used to determine putative species boundaries (Table 3).

*BEAST was used to generate a species tree from the two nuclear genes through a MCMC procedure. Substitution model GTR as determined by JModelTest2 was used and a strict molecular clock model was considered. *BEAST was run for 500 million iterations, assuming a Yule process of species tree prior. The first 50 million trees (10%) were discarded as burn-in and every 10,000 trees were stored. To assess the convergence of the algorithm, we used tracer v1.6 (Rambaut et al., 2015) to visualise trace plots, histograms and marginal posterior density. We observed that the ESS value for most of the parameters was above 200 apart from that of the prior. Tree-Annotator was used to summarise the sample trees produced by *BEAST and a maximum clade credibility tree annotated in Figtree was constructed (Figure 4). The species tree (Figure 4) from the nuclear genes was compared with the mtCOI gene tree (Figure 1) for concordance.

Bayesian species delimitation was conducted using the Bayesian Phylogenetic and Phylogeography (BPP 3.3a) program (Yang, 2015). The two nuclear loci were passed to BPP as the source data with a Bayesian mtCOI phylogenetic tree as the guide tree inferred by MrBayes 3.2.7. The mtCOI guide tree was partitioned into 3 clades (Figure

1) since BPP doesn't take on more than ten taxa. Therefore, each clade in Figure 1 (i.e., the subtrees descending from the 3 numbered clades in Figure 1) was used as a guide tree in the analysis. BPP uses the multispecies coalescent model to sample different species delimitation models that are compatible with the fixed guide tree through the reversible-jump Markov Chain Monte Carlo (rjMCMC) algorithm (Rannala and Yang, 2013). The rjMCMC was run for 500,000 generations, sampled every generation with a burn-in of 50,000. Algorithm 1 was used with $\propto= 2$ and $m = 1,$, and the proposal parameters were set to automatic fine-tuning.

To evaluate the effect of the priors on the ancestral population size (θ) and the root age (τ) on the posterior distribution of the tested models (Yang and Rannala, 2010), we assumed four combinations of diffusion priors (Hsieh et al., 2014). That is to say: relatively small ancestral population size and shallow divergence $\theta \sim G(2, 2000)$ and $\tau \sim G(2, 2000)$; relatively large ancestral population size with deep divergence $\theta \sim G(1, 10)$ and $\tau \sim G(1, 10)$; large ancestral population size with shallow divergence $\theta \sim G(1, 10)$ and $\tau \sim G(2, 2000)$; and finally small ancestral size with deep divergence $\theta \sim G(2, 2000)$ and $\tau \sim G(1, 10)$, where $G(\propto, \beta)$ denotes the gamma distribution with parameters $\propto$ and β. The results are shown in Table 5.

4.4 Results

Figures 1, 2 and 3 represent phylogenetic gene trees in this study based on Bayesian inference for mtCOI, RNApyII, and Shaw genes respectively. The mtCOI gene tree has strong clade support with posterior probability (PP> 90%) for the different genetic groups of the *B. tabaci* species complex. We didn't include *B. formosana, B. emiliae, B.*

breyniae or *B. vernoniae* in the mtCOI analyses because we did not have the specimens, and the sequences in Genbank were from a different region of the mtCOI gene.

All clades in the mtCOI gene tree are well resolved to monophyletic groups as seen in Figure 1. We partition the mtCOI gene tree into three clades as in Hsieh et al. (2014). Clade 1 consists of SSA1, SSA2, SSA3, SSA4 and NWAfrica, Clade 2 includes New World I, MEAM1, Indian Ocean and MED, and Clade 3 has all Asian subgroups, all China subgroups, Italy and Australia (Figure 1). The three clades of the mtCOI gene are compared to each of the nuclear gene tree topologies and they were not resolved in RNApyII and Shaw gene trees (Figure 2, 3). Only 56% of the clades in the RNApyII gene tree and 35% of the clades in Shaw gene tree formed monophyletic groups.

Taxonomic distinctiveness examined by methods implemented in the species delimitation plugin in Geneious (Masters et al., 2011) reveal that the RNApyII gene tree topology has 7 clades that are confirmed distinct across all the four measures of distinctiveness: P (RD), P (AB), PP and K2P distances. For the Shaw gene tree, however, only one clade (MEAM1) is confirmed as distinct across all four measures of taxonomic distinctiveness implemented in the species delimitation plugin (Table 3).

Significant values in Table 3 across the four measures are indicated with bold values. A species defined as not significant in one of the methods doesn't necessary mean discordance; however, if all four measures are not significant then such species are poorly resolved in a particular gene and this might be due to incomplete lineage sorting or horizontal gene transfer (Maddison, 1997). Therefore, the shaw gene tree topology was congruent to RNApyII gene tree topology in only four clades: *B. porteri*, MEAM1, *B. emiliae* and Australia. Table 4 shows a summary of the congruence and discordance

of nuclear genes with the mtCOI gene tree in each clade. Thus, the mtCOI gene was discordant with the nuclear gene tree topologies.

The species tree topology from the two nuclear genes generated from *BEAST analysis (Figure 4) is compared with the mtCOI gene tree topology (Figure 1); clade 1 in both trees is monophyletically resolved with high posterior probability for clade support. Clade 2 was not monophyletic in the nuclear species tree topology and had a high clade support of 97%. Clade 3 was resolved with similar entities as the mtCOI gene but with very low clade support (31%). Therefore, since the mtCOI locus has a higher mutation rate compared to the nuclear loci, it would take more nuclear genes to really compare the nuclear phylogenetic signal to the mtCOI phylogenetic signal.

For the Bayesian species delimitation with BPP, different prior distributions for (θ) and (τ) affected the outcome in clade 1 of the mtCOI gene tree (Figure 1). Thus, the small ancestral population with shallow divergence – $\theta \sim G\ (2, 2000)$ and $\tau \sim G(2, 2000)$ – and the small ancestral population with deep divergence – $\theta \sim G\ (2, 2000)$ and $\tau \sim G(1, 10)$ – had low posterior probability of 0.487 and 0.517 respectively. However, the relatively large ancestral population size with deep divergence – $\theta \sim G\ (1, 10)$ and $\tau \sim G(1, 10)$ – and large ancestral population size with shallow divergence – $\theta \sim G\ (1, 10)$ and $\tau \sim G(2, 2000)$ – had consistent results across all the three clades with posterior probability greater than 90% (Table 5).

4.5 Discussion

B. tabaci species complex boundaries have been based on the mtCOI gene (Boykin et al., 2012; Boykin et al., 2007; De Barro, 2012). However, the use of a single locus to

define species boundaries has been a question of debate for decades. Therefore, it has been suggested that the use of multiple unlinked nuclear loci may be advantageous in resolving the tips of the phylogenetic tree (Hsieh et al., 2014; Heled and Drummond, 2009). In this study we addressed answers to the question: Are nuclear genes as useful as the mtCOI gene to delimit members of the *B. tabaci* species complex?

We found clear conflicts between the nuclear gene tree topologies and the mtCOI gene tree topologies. These were observed in clades 1 and 2 of the mtCOI gene tree that were poorly resolved in the RNAPyII and Shaw gene trees (Figure 2, 3). Phylogenies from various unlinked genes may disagree because of tree construction errors (e.g., few characters sampled or long branch attraction), incomplete lineage sorting, horizontal gene transfer, gene duplication and extinction (Wiens et al., 2010; Scornavacca and Galtier, 2017; Maddison, 1997). Other variations in gene tree phylogenies may come from a number of variable sites (Table 2) and the rate of mutation in a particular gene (Rokas et al., 2003). In our case Table 2 shows that the mtCOI gene had the highest number of variable characters and was well-resolved (Figure 1), followed by RNApyII and Shaw (Figure 2, 3).

We also noticed patterns of conflict between the two nuclear gene tree topologies (Figure 2, 3). Congruence between the nuclear gene tree topologies was only present in four clades: *B. porteri*, MEAM1, *B. emiliae* and Australia. The rest of the species had conflicting patterns between the two nuclear genes. Therefore, we decided to see the effect of using a coalescent-based method (*BEAST) on the two nuclear genes to generate a species tree (Figure 4).

We observed an improvement in the resolution of species tree topology compared to the individual gene trees. For example, clade 1 entities were well resolved in the species tree (Figure 4) with high clade support. Clade 2 entities were not well resolved compared to the entities of the mtCOI gene tree topology (Figure 1); thus, we did not observe the monophyletic origin of this clade. Clade 3 was topologically well resolved compared to the scenarios observed in the individual nuclear gene tree topologies, although the clade support for this clade was very low. Generally there is a bright future for using more unlinked multiple loci to generate robust species tree estimates.

The guide tree in Bayesian species delimitation using BPP plays an important role in the output of the species delimitation model (Leaché and Fujita, 2010). Therefore, we used the mtCOI gene tree topology as the guide tree. We used a combination of 4 diffuse priors as in Hsieh et al. (2014) in our analysis (Table 5) and two of the priors sets had an effect on the models tested (Yang and Rannala, 2010). Our results show that large θ priors gave consistent results across all the clades. This may be because we portioned the mtCOI guide tree to handle a few species at a time, since analysis with large θ favour fewer species compared to analysis with small θ (McKay et al., 2013). In general, BPP results support results from the taxonomic distinctiveness methods used above. Hence, the mtCOI gene remains robust in *B. tabaci* species delimitation.

4.6 Conclusion

There are clear conflicting patterns between the mtCOI gene tree topology and nuclear gene nuclear gene topologies; the conflicts are also observed between the individual nuclear gene tree topologies, i.e. RNApyII and Shaw. Therefore, a single nuclear gene alone cannot be used to define species boundaries. However, we have observed that the

use of multiple nuclear unlinked loci might be key to generation of a robust species tree under coalescent-based methods. The mtCOI gene tree remains useful for *B. tabaci* species delimitation.

4.7 Declarations

Author's contributions

TK carried out the phylogenetic analysis with inputs from LMB and LSL. TK carried out the wet laboratory work; TK, CAO and TA provided samples. TK drafted the manuscript and all authors read and approved the manuscript.

Conflict of interest

Not applicable

4.8 Acknowledgements

TK is supported by a University of Western Australia Scholarship for International Research Fees (SIRF), University Postgraduate Awards (UPA) and the University of Western Australia Safety Net top-up scholarship. This work forms part of his PhD. The Natural Resources Institute, University of Greenwich, supported sequencing, fieldwork and the first 2 years of TK's PhD from the grant provided by the Bill and Melinda Gates Foundation (Investment ID OPP1058938). The National Crop Resources Research Institute (NaCRRI) in Uganda provided samples and facilitated sample collection. We thank Donald Kachigamba for providing samples. Resources provided by the Pawsey Supercomputing Centre with Funding from the Australian Government and the Government of Western Australia supported computation analysis. Finally, we thank the farmers in East Africa for allowing us to collect samples on their farms.

4.9 References

ALICAI, T., NDUNGURU, J., SSERUWAGI, P., TAIRO, F., OKAO-OKUJA, G., NANVUBYA, R., KIIZA, L., KUBATKO, L., KEHOE, M. A. & BOYKIN, L. M. 2016. Cassava brown streak virus has a rapidly evolving genome: implications for virus speciation, variability, diagnosis and host resistance. *Scientific reports,* 6.

ALICAI, T., OMONGO, C., MARUTHI, M., HILLOCKS, R., BAGUMA, Y., KAWUKI, R., BUA, A., OTIM-NAPE, G. & COLVIN, J. 2007. Re-emergence of cassava brown streak disease in Uganda. *Plant Disease,* 91**,** 24-29.

BOYKIN, L. M., ARMSTRONG, K. F., KUBATKO, L. & DE BARRO, P. 2012. Species delimitation and global biosecurity. *Evolutionary bioinformatics online,* 8**,** 1.

BOYKIN, L. M. & DE BARRO, P. J. 2014. A practical guide to identifying members of the Bemisia tabaci species complex: and other morphologically identical species. *Frontiers in Ecology and Evolution,* 2**,** 45.

BOYKIN, L. M., SAVILL, A. & DE BARRO, P. 2017. Updated mtCOI reference dataset for the Bemisia tabaci species complex. *F1000Research,* 6.

BOYKIN, L. M., SHATTERS, R. G., JR., ROSELL, R. C., MCKENZIE, C. L., BAGNALL, R. A., DE BARRO, P. & FROHLICH, D. R. 2007. Global relationships of Bemisia tabaci (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol Phylogenet Evol,* 44**,** 1306-19.

BROWN, J. K. & CZOSNEK, H. 2002. Whitefly transmission of plant viruses.

DARRIBA, D., TABOADA, G. L., DOALLO, R. & POSADA, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods,* 9**,** 772-772.

DE BARRO, P. J. 2012. The Bemisia tabaci species complex: questions to guide future research. *Journal of Integrative Agriculture,* 11**,** 187-196.

DE BARRO, P. J., LIU, S. S., BOYKIN, L. M. & DINSDALE, A. B. 2011. Bemisia tabaci: a statement of species status. *Annu Rev Entomol,* 56**,** 1-19.

DEGNAN, J. H. & ROSENBERG, N. A. 2006. Discordance of species trees with their most likely gene trees. *PLoS genetics,* 2**,** e68.

EDWARDS, S. V., LIU, L. & PEARL, D. K. 2007. High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences,* 104**,** 5936-5941.

ELBAZ, M., LAHAV, N. & MORIN, S. 2010. Evidence for pre-zygotic reproductive barrier between the B and Q biotypes of Bemisia tabaci (Hemiptera: Aleyrodidae). *Bulletin of entomological research,* 100**,** 581-590.

HELED, J. & DRUMMOND, A. J. 2009. Bayesian inference of species trees from multilocus data. *Molecular biology and evolution,* 27**,** 570-580.

HSIEH, C.-H., KO, C.-C., CHUNG, C.-H. & WANG, H.-Y. 2014. Multilocus approach to clarify species status and the divergence history of the Bemisia tabaci (Hemiptera: Aleyrodidae) species complex. *Molecular phylogenetics and evolution,* 76**,** 172-180.

KATOH, K., MISAWA, K., KUMA, K. I. & MIYATA, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research,* 30**,** 3059-3066.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S. & DURAN, C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-1649.

KUBATKO, L. S. & DEGNAN, J. H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology,* 56**,** 17-24.

KUBATKO, L. S., GIBBS, H. L. & BLOOMQUIST, E. W. 2011. Inferring species-level phylogenies and taxonomic distinctiveness using multilocus data in Sistrurus rattlesnakes. *Systematic Biology,* 60**,** 393-409.

LEACHÉ, A. D. & FUJITA, M. K. 2010. Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings of the Royal Society of London B: Biological Sciences***,** rspb20100662.

LEGG, J., OWOR, B., SSERUWAGI, P. & NDUNGURU, J. 2006. Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Advances in virus research,* 67**,** 355-418.

LIU, L., YU, L., KUBATKO, L., PEARL, D. K. & EDWARDS, S. V. 2009. Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution,* 53**,** 320-328.

LIU, S.-S., COLVIN, J. & DE BARRO, P. J. 2012. Species concepts as applied to the whitefly Bemisia tabaci systematics: how many species are there? *Journal of Integrative Agriculture,* 11**,** 176-186.

MADDISON, W. P. 1997. Gene trees in species trees. *Systematic biology,* 46**,** 523-536.

MARUTHI, M., HILLOCKS, R., MTUNDA, K., RAYA, M., MUHANNA, M., KIOZIA, H., REKHA, A., COLVIN, J. & THRESH, J. 2005. Transmission of Cassava brown streak virus by Bemisia tabaci (Gennadius). *Journal of Phytopathology,* 153**,** 307-312.

MASTERS, B. C., FAN, V. & ROSS, H. A. 2011. Species delimitation–a geneious plugin for the exploration of species boundaries. *Molecular Ecology Resources,* 11**,** 154-157.

MCKAY, B. D., MAYS JR, H. L., WU, Y., LI, H., YAO, C. T., NISHIUMI, I. & ZOU, F. 2013. An empirical comparison of character‐based and coalescent‐based approaches to species delimitation in a young avian complex. *Molecular Ecology,* 22**,** 4943-4957.

MUGERWA, H., SEAL, S., WANG, H.-L., PATEL, M. V., KABAALU, R., OMONGO, C. A., ALICAI, T., TAIRO, F., NDUNGURU, J. & SSERUWAGI, P. 2018. African ancestry of New World, Bemisia tabaci-whitefly species. *Scientific reports,* 8**,** 2734.

NDUNGURU, J., SSERUWAGI, P., TAIRO, F., STOMEO, F., MAINA, S., DJINKENG, A., KEHOE, M. & BOYKIN, L. M. 2015. 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,* 10**,** e0139321.

OMONGO, C. A., KAWUKI, R., BELLOTTI, A. C., ALICAI, T., BAGUMA, Y., MARUTHI, M., BUA, A. & COLVIN, J. 2012. African cassava whitefly, Bemisia tabaci, resistance in African and South American cassava genotypes. *Journal of integrative agriculture,* 11**,** 327-336.

POLSTON, J. E. & CAPOBIANCO, H. 2013. Transmitting plant viruses using whiteflies. *Journal of visualized experiments: JoVE*.

RAMBAUT, A., SUCHARD, M. A., XIE, D. & DRUMMOND, A. J. 2015. Tracer v1. 6. 2014.

RANNALA, B. & YANG, Z. 2013. Improved reversible jump algorithms for Bayesian species delimitation. *Genetics***,** genetics. 112.149039.

RODRIGO, A., BERTELS, F., HELED, J., NODER, R., SHEARMAN, H. & TSAI, P. 2008. The perils of plenty: what are we going to do with all these genes?

Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363**,** 3893-3902.

ROKAS, A., KING, N., FINNERTY, J. & CARROLL, S. B. 2003. Conflicting phylogenetic signals at the base of the metazoan tree. *Evolution & development,* 5**,** 346-359.

RONQUIST, F. & HUELSENBECK, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19**,** 1572-1574.

RONQUIST, F., TESLENKO, M., VAN DER MARK, P., AYRES, D. L., DARLING, A., HÖHNA, S., LARGET, B., LIU, L., SUCHARD, M. A. & HUELSENBECK, J. P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology,* 61**,** 539-542.

ROSENBERG, N. A. 2007. Statistical tests for taxonomic distinctiveness from observations of monophyly. *Evolution,* 61**,** 317-323.

SCORNAVACCA, C. & GALTIER, N. 2017. Incomplete lineage sorting in mammalian phylogenomics. *Systematic biology,* 66**,** 112-120.

SHATTERS JR, R. G., POWELL, C. A., BOYKIN, L. M., LIANSHENG, H. & MCKENZIE, C. 2009. Improved DNA barcoding method for Bemisia tabaci and related Aleyrodidae: development of universal and Bemisia tabaci biotype-specific mitochondrial cytochrome c oxidase I polymerase chain reaction primers. *Journal of Economic Entomology,* 102**,** 750-758.

SHAW, K. L. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences,* 99**,** 16122-16127.

SIMON, C., FRATI, F., BECKENBACH, A., CRESPI, B., LIU, H. & FLOOK, P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the entomological Society of America,* 87**,** 651-701.

TOCKO-MARABENA, B. K., SILLA, S., SIMIAND, C., ZINGA, I., LEGG, J., REYNAUD, B. & DELATTE, H. 2017. Genetic diversity of Bemisia tabaci species colonizing cassava in Central African Republic characterized by analysis of cytochrome c oxidase subunit I. *PloS one,* 12**,** e0182749.

WANG, P., RUAN, Y. M. & LIU, S. S. 2010. Crossing experiments and behavioral observations reveal reproductive incompatibility among three putative species of the whitefly Bemisia tabaci. *Insect Science,* 17**,** 508-516.

WANG, P., SUN, D. B., QIU, B. L. & LIU, S. S. 2011. The presence of six cryptic species of the whitefly Bemisia tabaci complex in China as revealed by crossing experiments. *Insect Science,* 18**,** 67-77.

WIENS, J. J., KUCZYNSKI, C. A. & STEPHENS, P. R. 2010. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. *Biological Journal of the Linnean Society,* 99**,** 445-461.

WOSULA, E. N., CHEN, W., FEI, Z. & LEGG, J. P. 2017. Unravelling the genetic diversity among cassava Bemisia tabaci whiteflies using NextRAD sequencing. *Genome biology and evolution,* 9**,** 2958-2973.

XU, J., DE BARRO, P. & LIU, S. 2010. Reproductive incompatibility among genetic groups of Bemisia tabaci supports the proposition that the whitefly is a cryptic species complex. *Bulletin of entomological research,* 100**,** 359-366.

YANG, Z. 2015. The BPP program for species tree estimation and species delimitation. *Current Zoology,* 61**,** 854-865.

YANG, Z. & RANNALA, B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, 200913022.

4.10 Figure Legends

Figure 1. The mtCOI consensus gene tree obtained from the posterior distribution estimated by MrBayes 3.2.2 run on the Pawsey supercomputer. The numbers on the branches represent posterior probability for clade support and the numbers on a black background represent clades of interest. For convenience, the cartoon option was used to reduce the crowding of taxa. Each cartoon in the tree represents a monophyletic group and thus a particular species. The tree was rooted with *B. afer*

Figure 2. The RNApyII consensus gene tree topology obtained from posterior distribution estimated from MrBayes 3.2.2 run on the Pawsey supercomputer. The numbers on the branches represent posterior probability for clade support. Species with monophyletic groups have bold names and those that were not resolved to monephyletic groups are not bold. The tree was rooted with *B. porteri* as in (Hsieh et al., 2014).

Figure 3. The Shaw consensus gene tree topology obtained from the posterior distribution estimated from MrBayes 3.2.2 run on the Pawsey supercomputer. The numbers on the branches represent posterior probability for clade support. Species with monophyletic groups have bold names and those that were not resolved to monephyletic groups are not bold. The tree was rooted with *B. porteri* as in (Hsieh et al., 2014)**.**

Figure 4. Maximum clade credibility species tree generated from two nuclear genes (Shaw and RNApyII) in *BEAST under the Yule model. The posterior probability at the clades represents support from *BEAST analysis and the numbers on a black background represent clades of interest. The species tree was rooted with *B. porteri*

Figure 5. A map of East Africa showing sample collection sites

4.10 Table Legends

Table 1. PCR primer sequences and amplification conditions used in this study

Table 2. Summary of genes used in phylogenetic analysis, showing length and number of phylogenetically informative characters

Table 3. *B. tabaci* species distinctiveness from two nuclear genes (RNApyII and Shaw)

Table 4. Congruence and discordance of three gene trees topologies (Figure 1, Figure 2 and Figure 3), * represents strong clade support of PP > 90 %

Table 5. Results of Bayesian species delimitation using BPP under a guide tree

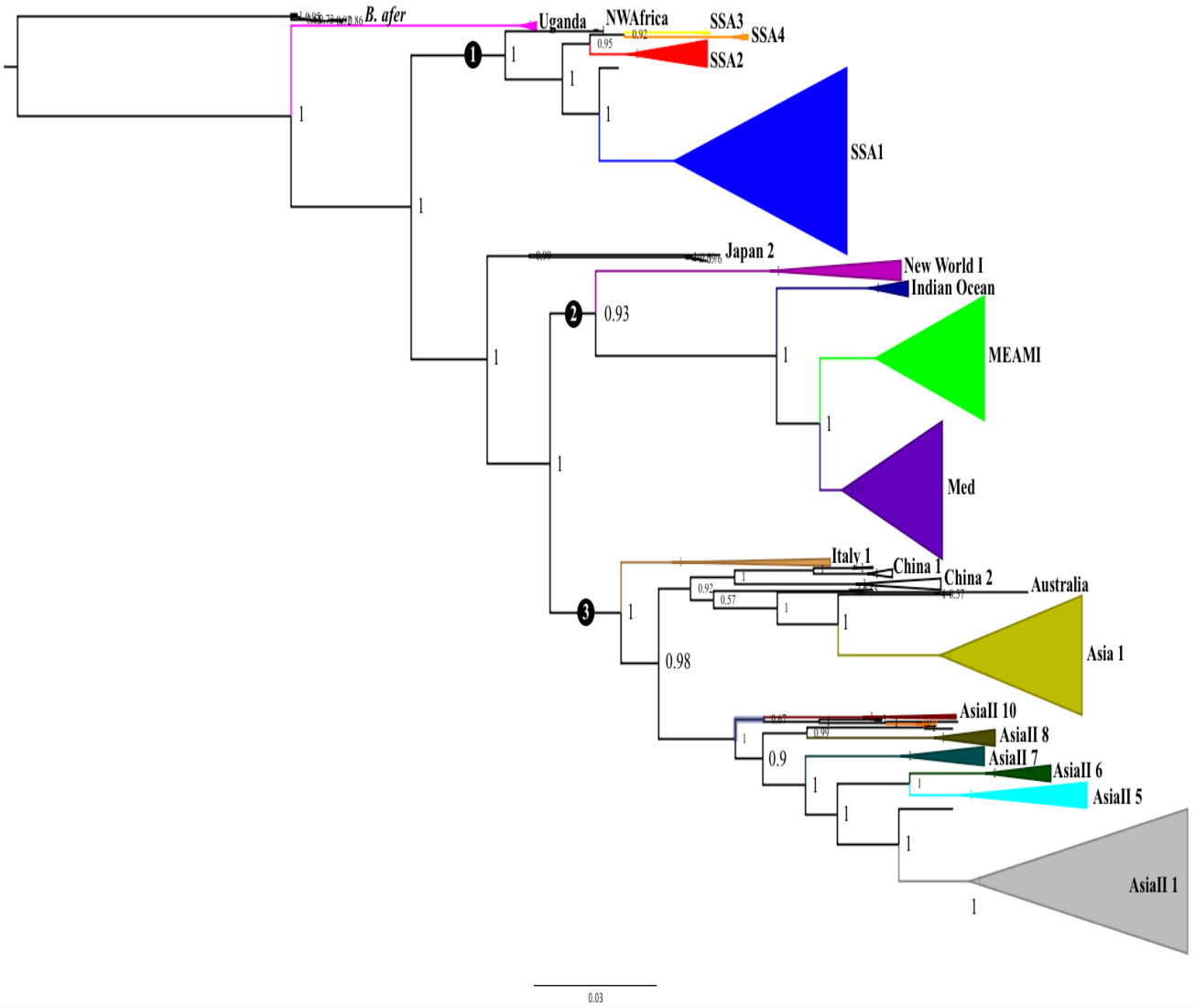


Figure 1

mtCOI consensus gene tree obtained from the posterior distribution estimated by MrBayes 3.2.2 run on the Pawsey supercomputer. The numbers on the branches represent posterior probability for clade support and the numbers on a black background represent clades of interest. For convenience, the cartoon option was used to reduce the crowding of taxa. Each cartoon in the tree represents a monophyletic group and thus a particular species. The tree was rooted with *B. afer*

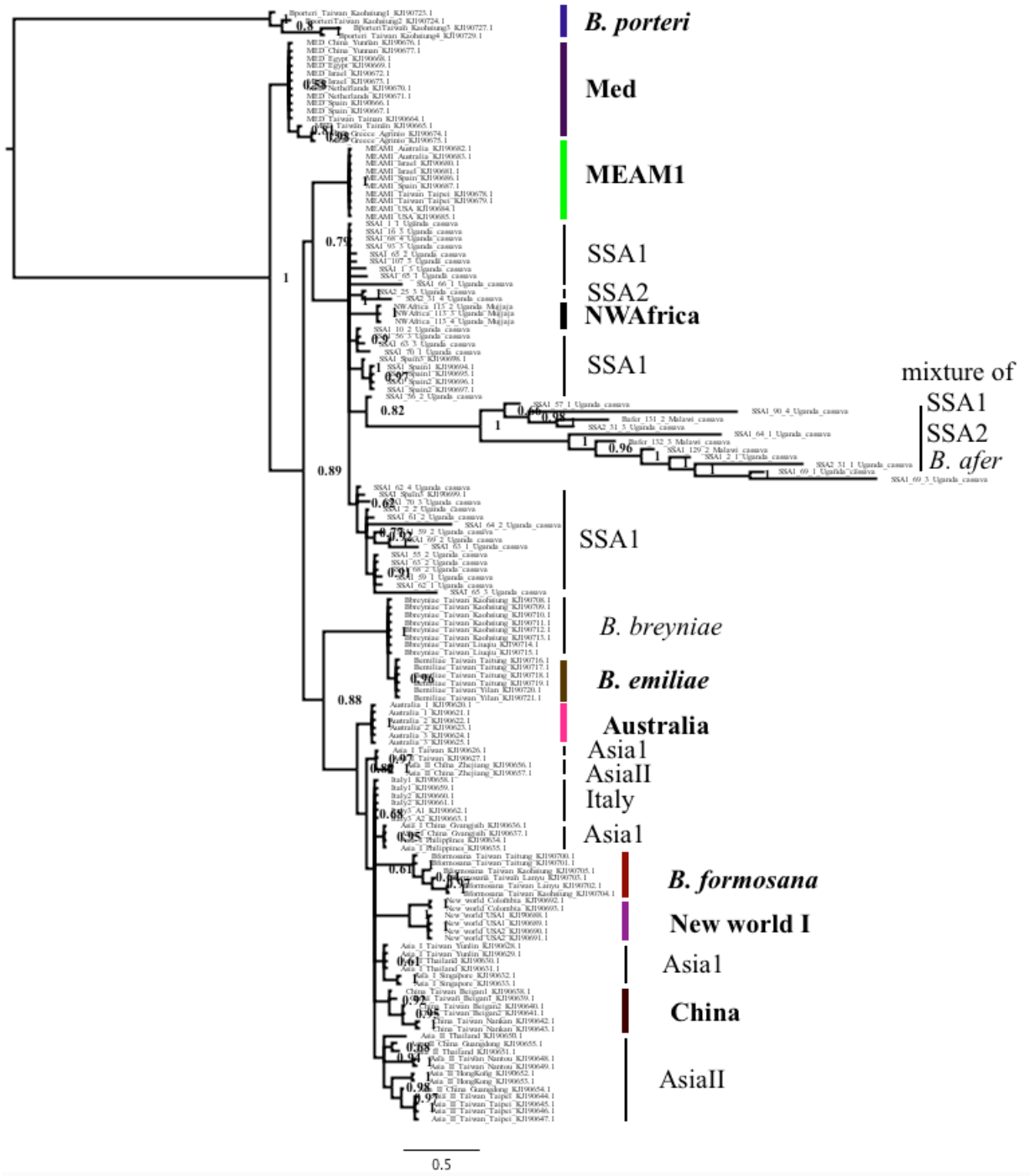


Figure 2

RNApyII consensus gene tree topology obtained from posterior distribution estimated from MrBayes 3.2.2 run on the Pawsey supercomputer. The numbers on the branches represent posterior probability for clade support. Species with monophyletic groups have bold names and those that were not resolved to monophyletic groups are not bold. The tree was rooted with *B. porteri* as in Hsieh et al. (2014).

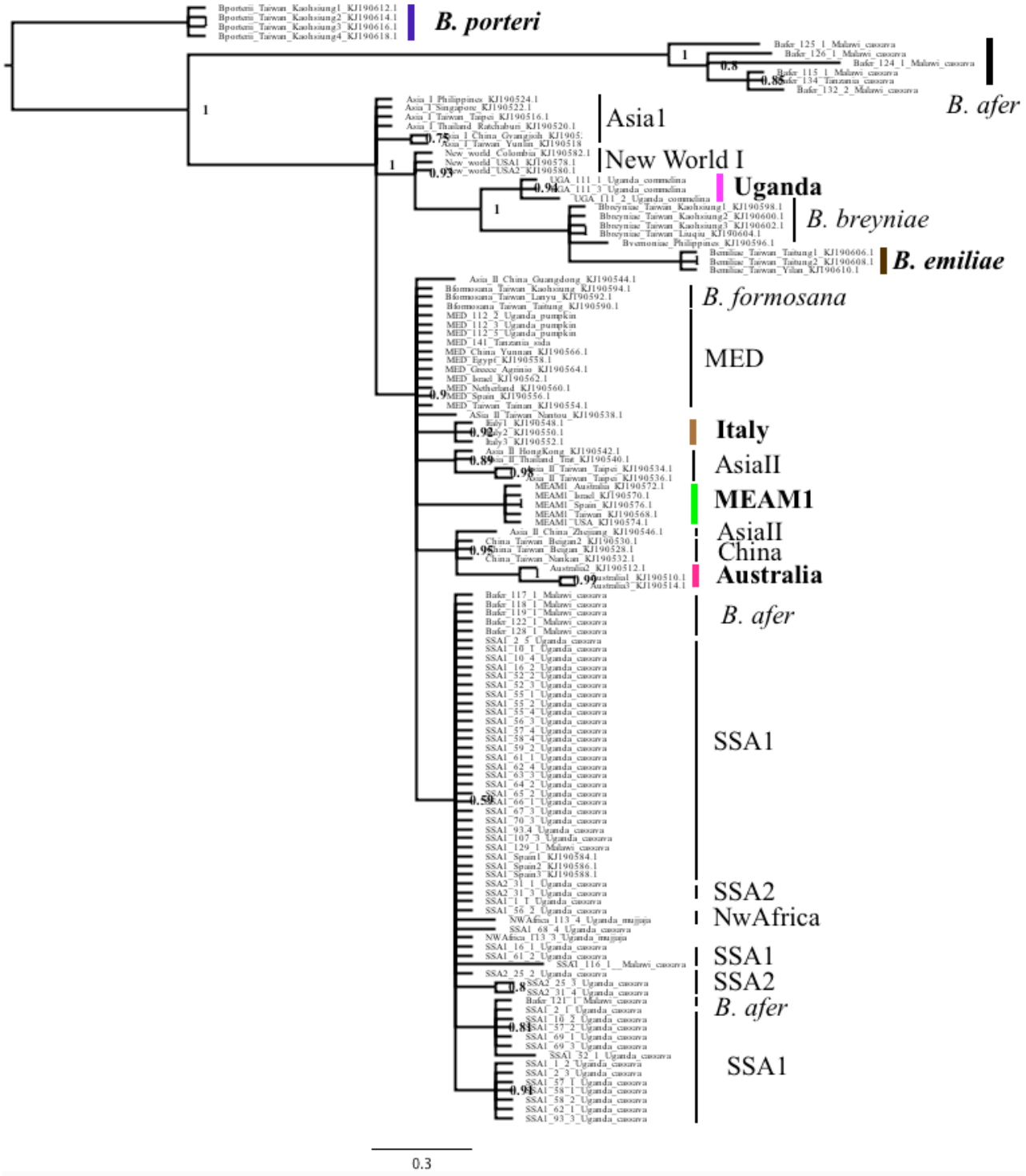


Figure 3

The Shaw consensus gene tree topology obtained from the posterior distribution estimated from MrBayes 3.2.2 run on the Pawsey supercomputer. The numbers on the branches represent posterior probability for clade support. Species with monophyletic groups have bold names and those that were not resolved to monophyletic groups are not bold. The tree was rooted with *B. porteri* as in Hsieh et al. (2014).

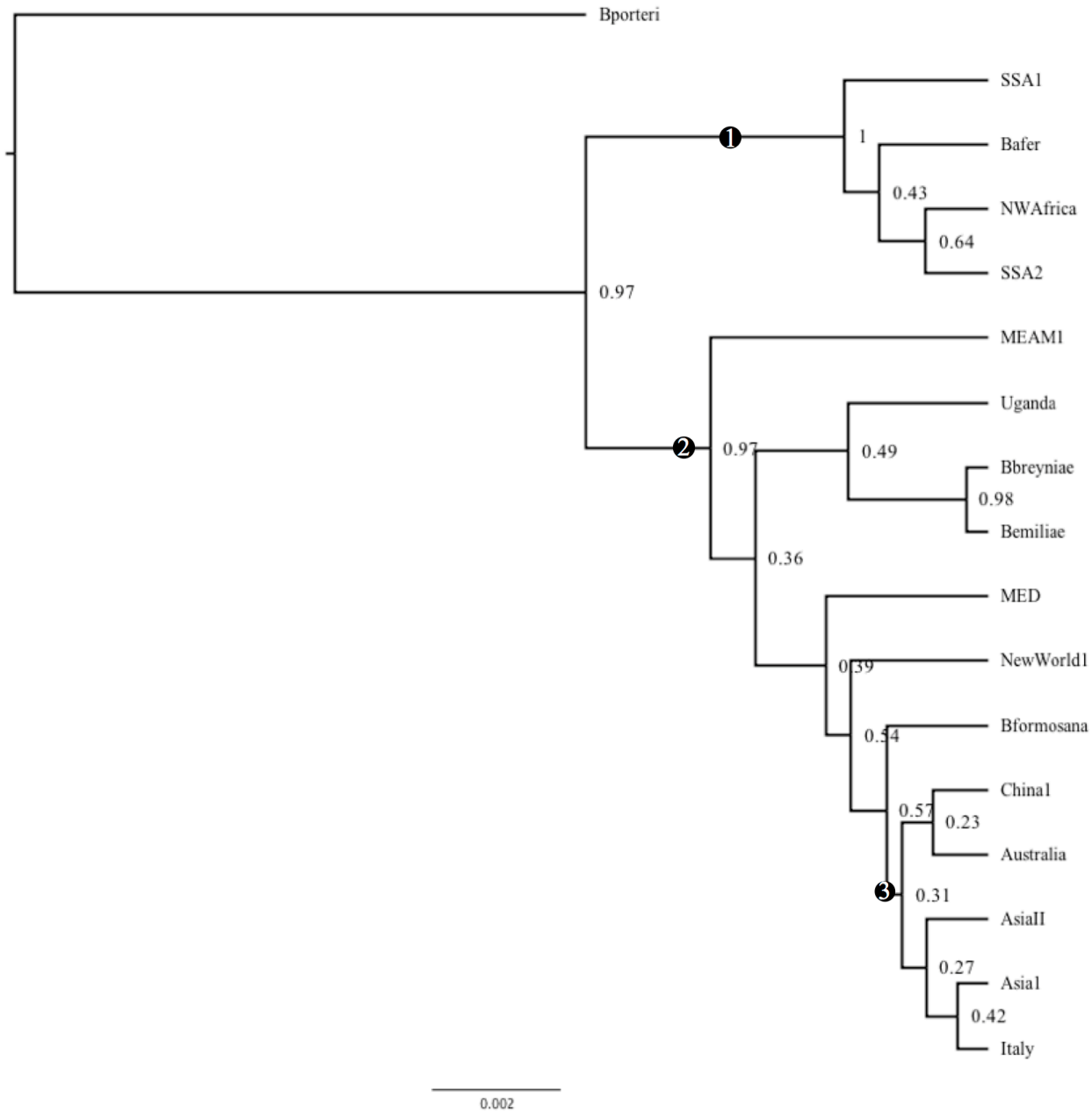


Figure 4

Maximum clade credibility species tree generated from two nuclear genes (Shaw and RNApyII) in *BEAST under the Yule model. The posterior probability at the clades represents support from *BEAST analysis and the numbers on a black background represent clades of interest. The species tree was rooted with *B. porteri*

Table 1. PCR primer sequences and amplification conditions used in this study

Gene	Reference	Direction	Primer sequence (5' to 3')	PCR amplification conditions
RNApyII	(Hsieh et al., 2014)	RNApyIIF RNApyIIR	CACAAAATGAGTATGATGGGTC CTTCACGACCTCCCATAGC	94^0C for 5 min, 94^0C for 20 sec, 60^0C for 1 min, 72^0C for 2 min (35 cycles) 72^0C for 10 min 20^0C hold.
Shaw	(Hsieh et al., 2014)	ShawF ShawR	ATCTGCATCTCGATCCTCTC AGTTGAAGTCGTTGTGCGG	94^0C for 5 min, 94^0C for 20 sec, 60^0C for 1 min, 72^0C for 1 min (35 cycles) 72^0C for 10 min 20^0C hold
mtCOI	(Simon et al., 1994)	C1-J-2195 TL2-N-3014	TTGATTTTTTGGTCATCCAGAAGT TCCAATGCACTAATCTGCCATATTA	94^0C for 5 min, 94^0C for 20 sec, 52^0C for 30 sec, 72^0C for 1 min (35 cycles) 72^0C for 10 min 20^0C hold.
mtCOI	(Mugerwa et al., 2018)	Bt Forward Bt Reverse	TGRTTTTTTGGTCATCCRGAAGT TTTACTGCACTTTCTGCC	94^0C for 5 min, 94^0C for 20 sec, 53.1^0C for 30 sec, 72^0C for 1 min (35 cycles) 72^0C for 10 min 20^0C hold.
mtCOI	(Shatters Jr et al., 2009)	Btab-UniR Btab-UniL	CTAATRTGGCAGAAAGTGCAGTAAATTTAAG GAGGCTGRAAAATTARAAGTATTTGG	94 ^{0}C 2 min, 35 cycles of 94 C 30 s, 53^0C for 1 min, 72^0C for 1 min, 72 for 10 min
mtCOI	Bafer (UWA) unpublished	775F 1489R	AGGCCGGTAAACTCGAAGTG ACGCTGGGCTTGTTGATCTT	94^0C for 5 min, 94^0C for 20 sec, 60^0C for 30 sec, 72^0C for 1 min (35 cycles) 72^0C for 10 min 20^0C hold.

Table 2. Summary of genes used in phylogenetic analysis, showing length and number of phylogenetically informative characters

Gene	Length	Phylogenetically informative characters	Taxa
mtCOI	559	405	1211
RNApyII	865	215	148
Shaw	455	60	124

Table 3. *B. tabaci* species distinctiveness from two nuclear genes (RNApyII and Shaw)

	RNApyII					Shaw				
Species	**Monophyletic?**	**K2P**	**P (RD)**	**PP**	**P (AB)**	**Monophyletic?**	**K2P**	**P (RD)**	**PP**	**P (AB)**
B.porteri	Yes	**3.824**	0.13	**1**	**7.0E-10**	Yes	**1.779**	0.13	**1**	**1.7E-9**
MED	Yes	**0.723**	**0.05**	0.58	**1.4E-21**	No	**0.100**	NA	NA	NA
MEAM1	Yes	**0.713**	**0.05**	**1**	**4.4E-13**	Yes	**0.370**	**0.05**	**1**	**3.9E-10**
SSA1	No	**0.776**	NA	NA	NA	No	**0.177**	NA	NA	NA
SSA2	No	**1.446**	NA	NA	NA	No	**0.186**	NA	NA	NA
NWAfrica	Yes	**0.713**	**0.05**	**1**	**2.0E-6**	No	**0.177**	NA	NA	NA
B.breyniae	No	**0.105**	NA	NA	NA	No	**0.451**	NA	NA	NA
B.emiliae	Yes	**0.105**	0.91	**0.96**	**5.1E-5**	Yes	**0.451**	**0.05**	**1**	0.01
Australia	Yes	**0.269**	**0.05**	**1**	**1.1E-9**	Yes	**0.375**	0.20	**1**	0.01
Asia_I	No	**0.131**	NA	NA	NA	No	**0.257**	NA	NA	NA
Asia_II	No	**0.270**	NA	NA	NA	No	**0.252**	NA	NA	NA
Italy	No	**0.131**	NA	NA	NA	Yes	**0.222**	**0.05**	**0.92**	**1.6E-7**
B.formosana	Yes	**0.413**	**0.05**	**1**	**4.7E-9**	No	**0.100**	NA	NA	NA
New_world	Yes	**0.404**	**0.05**	**1**	**4.7E-9**	No	**0.257**	NA	NA	NA
China	Yes	**0.236**	0.93	**0.92**	**4.7E-9**	No	**0.222**	NA	NA	NA
Bafer	No	**1.785**	NA	NA	NA	No	**1.390**	NA	NA	NA
UGA						Yes	**0.443**	0.14	**0.94**	1.21E-03

Notes: K2P distance, species were considered significant for K2P > 1%, P(RD) is the probability that a clade has some degree of distinctiveness (Rodrigo et al., 2008) and species were considered significant with P <0.05. P(AB) is the probability that monophyly happens by chance (Rosenberg, 2007) and here species were considered significant with $P < 10^{-5}$. Finally PP is the Bayesian posterior probability for clade support and we considered species with probability greater than or equal to 90% significant.

Table 4. Congruence and discordance of three gene trees topologies (Figure 1, Figure 2 and Figure 3), * represents strong clade support of PP > 90 %

Clade / species	Gene supporting	Gene rejecting	% Supporting
SSA1	*mtCOI	*RNApyII, Shaw	33
SSA2	*mtCOI	*RNApyII, Shaw	33
NWAfrica	*mtCOI, *RNApyII	Shaw	67
B. afer	*mtCOI	*RNApyII, *Shaw	33
MEAM1	*mtCOI, *RNApyII, *Shaw		100
Uganda	*mtCOI, *Shaw		100
B. breyniae		*RNApyII, *Shaw	0
B. emiliae	*RNApyII, *Shaw		100
MED	*mtCOI, RNApyII,	*Shaw	67
Newworld1	*mtCOI, *RNApyII	*Shaw	67
B. formosana	RNApyII	*Shaw	50
China1	*mtCOI, *RNApyII	*Shaw	67
Australia	*mtCOI, *RNApyII, *Shaw		100
Asia II	*mtCOI,	*RNApyII, *Shaw	33
Asia I	*mtCOI	*RNApyII, *Shaw	33
Italy	*mtCOI, Shaw	RNApyII	67
B. porteri	*RNApyII, *Shaw		100

Table 5. Results of Bayesian species delimitation using BPP under a guide tree

Delimited species	**Run**	**Posterior**	θ	θ **Priors**	τ **Prior**
5 species/clade 1					
	1	0.487	0.0316	$\theta \sim G(2, 2000)$	$\tau \sim G(2, 2000)$
	2	**0.949**	0.0819	$\boldsymbol{\theta \sim G(\ ,\ )}$	$\boldsymbol{\tau \sim G(\ ,\ )}$
	3	**0.920**	0.0760	$\boldsymbol{\theta \sim G(\ ,\ )}$	$\boldsymbol{\tau \sim G(\ ,\ \ )}$
	4	0.517	0.0302	$\theta \sim G(2, 2000)$	$\tau \sim G(1, 10)$
3 species/clade 2					
	1	1.000	0.0020	$\theta \sim G(2, 2000)$	$\tau \sim G(2, 2000)$
	2	**1.000**	0.0136	$\boldsymbol{\theta \sim G(\ ,\ )}$	$\boldsymbol{\tau \sim G(\ ,\ )}$
	3	**1.000**	0.0219	$\boldsymbol{\theta \sim G(\ ,\ )}$	$\boldsymbol{\tau \sim G(\ ,\ \ )}$
	4	1.000	0.0010	$\theta \sim G(2, 2000)$	$\tau \sim G(1, 10)$
5 species/clade 3					
	1	1.000	0.0011	$\theta \sim G(2, 2000)$	$\tau \sim G(2, 2000)$
	2	**0.999**	0.0029	$\boldsymbol{\theta \sim G(\ ,\ )}$	$\boldsymbol{\tau \sim G(\ ,\ )}$
	3	**0.999**	0.0033	$\boldsymbol{\theta \sim G(\ ,\ )}$	$\boldsymbol{\tau \sim G(\ ,\ \ )}$
	4	1.000	0.0010	$\theta \sim G(2, 2000)$	$\tau \sim G(1, 10)$

Posterior represents the posterior probability for the best species tree model. θ Represents the posterior mean.

Chapter 5: Species Tree Estimation for the Global *Bemisia tabaci* Species Complex Using 3000 Nuclear Genes

5.1 Species tree estimation for the global *Bemisia tabaci* species complex using 3000 nuclear genes

This article will be submitted for publication in Molecular Phylogenetics and Evolution Journal in April 2019.

Tonny Kinene*[1], Laura Kubatko[2], Titus Alicai[3], Bruno Rossito De Marchi[4], James Mbora Wainaina[1], Livingstone S. Luboobi[5], Peter Sseruwagi[6], Joseph Ndunguru[6] and Laura M. Boykin[1]. (2019). Species tree Estimation for the Global *Bemisia tabaci* species complex using 3000 nuclear genes**.**

Title: Species tree estimation for the global *Bemisia tabaci* species complex using 3000 nuclear genes.

Authors: Tonny Kinene*[1], Laura Kubatko[2], Titus Alicai[3], Bruno Rossito De Marchi[4], James Mbora Wainaina[1], Livingstone S. Luboobi[5], Peter Sseruwagi[6], Joseph Ndunguru[6] and Laura M. Boykin[1]

Addresses

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

[2]The Ohio State University, 154W 12th Avenue, Columbus, Ohio 43210, USA

[3]National Crops Resources Research Institute, P.O. Box 7084, Kampala, Uganda

[4]São Paulo State University (UNESP), School of Agriculture, Dept. of Plant Protection, CEP 18610-307, Botucatu (SP), Brazil.

[5]Department of Mathematics, Makerere University, P.O. Box 7062, Kampala, Uganda

[6]Mikocheni Agricultural Research Institute (MARI), P.O. Box 6226, Dar es Salaam, Tanzania.

*Author for correspondence: Tonny Kinene, Email: kinenetonny@gmail.com; tonny.kinene@research.uwa.edu.au

Authors' email addresses:

Laura Kubatko: lkubatko@stat.osu.edu

Titus Alicai: talicai@hotmail.com

Bruno Rossito De Marchi: bruno_dmarchi@hotmail.com

Christopher Abu Omongo: chrisomongo@yahoo.com

Livingstone Serwadda Luboobi: luboobi@gmail.com

Laura M. Boykin: laura.boykin@uwa.edu.au

James Mbora Wainaina: james.wainaina@research.uwa.edu.au

Abstract

The whitefly *Bemisia tabaci* (Gennadius) (Hemiptera; Aleyrodidae), is a species complex and a global pest that threatens global food security. It infests more than 1000 plant species, including several major crops in Africa such as cassava, sweet potatoes and tomatoes. The whitefly transmits more than 200 plant viruses including cassava mosaic virus and cassava brown streak virus, which have caused substantial yield losses for farmers in East Africa. The genetic diversity of *B. tabaci* species has previously been assessed based on the partial mitochondrial cytochrome oxidase I (mtCOI) gene alone. In this study, we used transcriptome data for 49 individual whiteflies from 12 genetic groups of the *B. tabaci* species complex to generate 3000 orthologous genes that we used to estimate the species tree of the *B. tabaci* species complex with ASTRAL and SVDQuartets. Our results show congruence of the species tree with the mtCOI gene tree in 10 genetic groups. However, the placement of *B. tabaci* species from Uganda was not congruent with its usual placement in the conventional mtCOI gene tree topology. This highlights the importance of using multi-loci datasets to delimit species, since a single locus might not give the true history of the species relationships. It is becoming easy to sequence transcriptomes of individual *B. tabaci* species; thus, a large number of genes are being generated and more will be available in the future of *B. tabaci* systematics. Combining the 3000 genes has provided a well-supported estimate of the species relationships for the complex, which can be used to develop sustainable management strategies.

Keywords: *Bemisia tabaci,* species tree inference, transcriptome, multi-loci, ASTRAL, SVDQuartets.

5.2 Introduction

The whitefly *Bemisia tabaci* is a species complex consisting of more than 34 morphologically indistinguishable species (Boykin et al., 2012; Boykin and De Barro, 2014). The whitefly is considered polyphagous, and is a global pest (De Barro et al., 2011a) that threatens food security because it vectors more than 200 plant viruses (Polston et al., 2014). In particular, begomoviruses (family *Geminiviridae*) are the most common plant viruses transmitted by a whitefly; however, the whiteflies also transmit other viruses, such as ipomoviruses, carlaviruses and criniviruses (Zaidi et al., 2017; Jones, 2003). These viruses have caused tremendous yield losses to economically important crops such as cassava, cotton, beans, tomato and cucurbit (Mansoor et al., 2006; Legg et al., 2006; Ndunguru et al., 2015).

Assessment of the genetic diversity of the *B. tabaci* species complex has been based primarily on analysis of the mitochondrial cytochrome oxidase I (mtCOI) gene (Boykin et al., 2012; Boykin and De Barro, 2014; Boykin et al., 2007; Dinsdale et al., 2010), data derived from Restriction-site Associated DNA sequencing (RAD-seq) (Wosula et al., 2017; Elfekih et al., 2018), and mating experiments (Liu et al., 2012; Xu et al., 2010; Wang et al., 2010). According to the mtCOI gene, Boykin et al. (2007) identified 12 well-resolved genetic groups of *B. tabaci* species complex. These included: Mediterranean (MED), Middle East Asia Minor 1 (MEAM1), Indian Ocean, Sub-Saharan Africa 1 (SSA1), Sub-Saharan Africa 2 (SSA2), Asia I, Asia II, Australia (AUS), China, Italy, New World 1 (NW1) and Uganda. These were later confirmed by Dinsdale et al. (2010) at 3.5% mtCOI sequence pairwise genetic distance divergence limit.

Among the 12 genetic groups, MEAM1 and MED are identified as invasive and cryptic species (Vyskočilová et al., 2018). It has been suggested that both species have signatures of introgression with Indian Ocean species (Elfekih et al., 2018). The MED species have a broader home range than MEAM1 species, which stretches up to Western African and the countries bordering the Mediterranean basin (Elfekih et al., 2018), whereas the Indian Ocean species dominate the Indian Ocean islands and East Africa (Boykin et al., 2013). The SSA1 species appear throughout Sub-Saharan Africa, while SSA2 is mostly found in East Africa and West Africa. The SSA3 and SSA4 species dominate both Central and Western Africa (Legg et al., 2014b). All the SSA species (SSA 1–4) are putative species, meaning they are not expected to interbreed; however, some mating studies have demonstrated successful interbreeding between SSA2 and SSA1 individual whiteflies collected from Uganda (Maruthi et al., 2004).

New studies using Next RAD-seq suggest that the mtCOI gene may not be an efficient taxonomic marker because of its failure to accurately provide a full representation of the phylogenetic history (Wosula et al., 2017). Other researchers have argued that the mtCOI gene is always under strong selection pressure and it is hard to differentiate nuclear mitochondrial DNA segments from genes in mitochondrial DNA (Tay et al., 2017). To date, no species tree for the *B. tabaci* complex has been published.

The advancement in next generation sequencing (NGS) technology has led to a reduction in the cost of sequencing, resulting in massive generation of genomic data and a subsequent improvement in the methods for inferring species-level phylogenies (Chifman and Kubatko, 2014; Washburn et al., 2017; Mirarab et al., 2014; Dunn et al., 2013). Previously, phylogenetic inference of the *B. tabaci* species complex was based

on one gene or a few genes (Boykin et al., 2012; Hsieh et al., 2014; Mugerwa et al., 2018; Dinsdale et al., 2010; Boykin et al., 2007); recently, phylogenetic inferences for the species complex have been made on genomic-level variation inferred from SNP data (Elfekih et al., 2018; Washburn et al., 2017).

The ability to generate transcriptome data has revolutionised biology. This has resulted in the production of a large number of genes that can be used to provide a well-supported estimate of a species tree (Maddison and Knowles, 2006). New software packages like Agalma (Dunn et al., 2013) provide an automated pipeline for generating thousands of orthologous genes from transcriptome data. These genes can be used to estimate a reliable species tree.

Phylogenies estimated from single genes are often incongruent with the species phylogenies (Liu et al., 2009). To account for gene tree incongruence due to incomplete lineage sorting, coalescent-based methods – such as SVDQuartet (Chifman and Kubatko, 2014) and ASTRAL (Mirarab et al., 2014; Zhang et al., 2018) – are recommended, among others.

In this study we use some of the computationally efficient methods to estimate the *B. tabaci* species tree using thousands of genes from transcriptome data. We used the Agalma pipeline (Dunn et al., 2013) to generate an orthologous genes dataset, these genes were then used as source data for RAxML (Stamatakis, 2014) to estimate maximum likelihood gene trees. The gene trees were then used in ASTRAL (Mirarab et al., 2014; Zhang et al., 2018) as source data to estimate the *B. tabaci* species tree. The orthologous genes from Agalma were also used as source data in SVDQuartets

(Chifman and Kubatko, 2014) to estimate another *B. tabaci* species tree. We compared our results to the previous estimates of the phylogeny based on the mtCOI gene.

5.3 Results

The 17 cDNA libraries sequenced on a Hi-seq 2500 produced 8,000,000 raw reads before trimming and between 7,816,022 and 7,876,283 raw reads after trimming in CLC genomic workbench. The 32 cDNA libraries sequenced on a Hi-seq 4000 produced raw reads ranging from 29,665,109 to 64,080,188 before trimming and 29,106,192 to 57,947,512 after trimming (Table 1).

The analysis from Agalma revealed a supermatrix of 7362 orthologous gene sequences with 4,945,709 sites. The actual gene occupancy across the taxa was 49.50%, and the number and percentage of genes present per taxon in the supermatrix are shown in Table 2. Three thousand genes, representing 80% of the gene occupancy (Figure 1), were passed to RAxML as source data from which 3000 maximum likelihood gene trees were inferred. These were then used as input for ASTRAL. The species relationships from ASTRAL had high support (local posterior probability) as indicated in Figure 2. All nodes had a high ASTRAL support value (local posterior probability) of 100% apart from the node leading to the sequences from Italy (89%) and the node between the NWAfrica and SSA2 sequences (74%).

SVDQuartets was applied to a dataset consisting of all 3000 genes. In the species tree estimated by SVDQuartets, the Italy and Newworld1 sequences share a recent common ancestor with a high bootstrap support value of 100% (Figure 3). This clade is congruent with the species tree inferred by ASTRAL in which the Newworld1 sequence is at the base of this group. We also observed low bootstrap support on the node leading

to the sequences from Uganda (55%), and this support value was lower than that of ASTRAL (100%). Lastly, there was low bootstrap support for the clade that included the sequences from SSA2 and NWAfrica (59%), which was consistent with the ASTRAL species tree topology (Figure 2), which had a slightly higher support value of 74%.

Clade 1 in the mtCOI gene tree (Figure 4) shows that SSA2 is more closely related to SSA1 than NWAfrica; this is not consistent with clade 1 in both species trees where SSA2 and NWAfrica are seen to share a recent common ancestor, but this relationship had low support values in both species trees (Figures 2 and 3). Clade 2 is consistent across both species trees and the mtCOI gene tree, with the MED and MEAM1 species sharing a common ancestor and closely related to the Indian Ocean species. Clade 3 in the species trees (Figures 2 and 3) shows that the ASIA II_1 species are more closely related to AUSI and AUSII species than to ITALY and New World 1 species. This is consistent across all topologies.

Generally, both tree topologies (Figures 2 and 3) are consistent with the species' placements. We observed that both topologies place Uganda sequences as sister to the invasive group (MED and MEAM1) and Indian Ocean. However, this finding is not congruent with the conventional mtCOI tree topology (Figure 4) where Uganda sequences are seen at the base of the tree. This might be one of the failures of mtCOI marker to accurately represent the Ugandan species.

5.4 Discussion

In this study, we used transcriptome data from 49 individual libraries of individual whiteflies to generate 7362 orthologous genes, of which 3000 genes were used to

estimate the *B. tabaci* species tree. Two coalescent-based software packages for species tree inference, ASTRAL (Mirarab et al., 2014) and SVDQuartets (Chifman and Kubatko, 2014), were used. The species trees estimated by the two methods were congruent when compared and incongruent to the tree inferred from the mtCOI locus. In particular, the placement of the Uganda sequences in the estimated *B. tabaci* species trees (Figures 2 and 3) differed from the conventional mtCOI tree groupings. The placement of the other 11 genetic groups were consistent with the mtCOI placement; i.e., ASIAII1, Australia I (AUS I), Australia II (AUSII), ITALY, New World 1, Indian Ocean, Middle East Asia Minor 1 (MEAM1), Mediterranean (MED), New World Africa (NWAfrica), Sub-Saharan Africa 1 (SSA1), SSA2 and *B. afer*.

We firstly compared the output from the two species trees estimated from ASTRAL and SVDQuartets, and Figures 2 and 3 both show that the Ugandan *B. tabaci* sequences separate the SSA species from the rest of the species. We note that ASTRAL has a high support value of 100% (local posterior probability) for the placement of the Ugandan species, but SVDQuartets bootstrap support value is only 55%. To examine this finding more carefully, we re-ran the SVDQuartets analysis with more genes (5836) and 1000 bootstrap replicates. In the tree estimated for this larger dataset, the bootstrap value for the Ugandan *B. tabaci* sequences increased to 71%. These findings are not consistent with the conventional mtCOI tree topology (see Figure 4) in which the mtCOI gene tree topology shows that the Uganda sequences are at the base of the tree. It is well known that phylogenies estimated from single genes are often incongruent with the species phylogeny (see, e.g., Liu et al., 2009; Maddison 2001). We thus expect that the species tree methods used here provide a more robust estimate of the species-level phylogenetic relationships than those given by the mtCOI gene tree.

The Ugandan *B. tabaci* species were first recognised by Boykin et al. (2007) as a special group that was genetically different from all other groups. The basal placement of these species and the other SSA species in that phylogeny led to the hypothesis that Africa was the origin of the members of the *B. tabaci* species complex. Dinsdale et al. (2010) also identified the Uganda *B. tabaci* species and called them an unusual group; however, authors (Dinsdale et al., 2010) did not incorporate the Uganda *B. tabaci* sequences in their phylogeny because its placement would cause some conflict. Instead they indicated what would have been its position in the phylogenetic tree with dashed lines. Unfortunately the annual review (De Barro et al., 2011a) carried on the phylogeny from Dinsdale et al. (2010) with a contested location of *B. tabaci* Uganda and used it in their paper. Therefore, we are not the first ones to have seen this pattern; previous researchers saw the pattern and ignored it. The Ugandan *B. tabaci* is placed as sister to the invasive species (MED, MEAM1) and Indian Ocean species of *B. tabaci* (Figures 2 and 3). In other research publications for mtCOI phylogeny, *B. tabaci* Uganda has been seen at the base of the tree (Mugerwa et al., 2012; Sseruwagi et al., 2005; Boykin et al., 2007).

The placement of the *B. tabaci* Uganda sequences in the species trees in Figures 2 and 3 were supported by Campbell et al. (1996), who suggested that *B. tabaci* species originated from tropical Africa and spread to Northern Africa, the Mediterranean basin, Asia minor, Asia, Australia and others. Therefore, according to our observation, *B. tabaci* Uganda is at the base of the invasive species group (Indian Ocean, Med and MEAM1). Based on this new placement in the global phylogeny, Uganda *B. tabaci* appears to be the source of the global invasion of *B. tabaci*. This claim has been supported by genome-wide analysis using SNPs where Elfekih et al. (2018) reported that the invasive species (MED, MEAM1), and Indian Ocean evolved in Sub–Saharan

Africa before spreading to the Mediterranean basin. This has also been supported by other studies using the mtCOI gene (Delatte et al., 2011; Boykin et al., 2007).

Accurate identification of the members of the *B. tabaci* species complex is key for sustainable control strategies. In order for whitefly species outbreaks to be managed we must think of using other molecular markers in addition to mtCOI to identify species-level boundaries. One of the drawbacks of the mtCOI locus is that it is maternally inherited and that it is non-recombining (Macfadyen et al., 2018; Wosula et al., 2017). Therefore, the process of hybridisation cannot be detected with only the mtCOI locus (Macfadyen et al., 2018). This may be an important oversight, as several studies have pointed out hybridisation in the members of the *B. tabaci* species complex (Elfekih et al., 2018; Wosula et al., 2017; Maruthi et al., 2004).

5.5 Conclusion

We used 3000 orthologous genes to estimate the *B. tabaci* species tree; combining the information in all the genes generated provides a well-supported estimate of the species tree. As more species are added to this framework we will get a better understanding of the *B. tabaci* species relationships. Our work serves as a platform for future understanding of *B. tabaci* species relationships, enabling the development of sustainable management strategies.

5.6 Materials and Methods

Seventeen whitefly samples were collected from Uganda, Tanzania and Malawi between 2015 and 2018. Adult whiteflies were collected from the top five leaves of symptomatic cassava plants using an aspirator and immediately preserved in RNAlater.

The name of the village was recorded together with the corresponding geographical co-ordinates at each site using the Global positioning system (GPS). Collaborators sent 32 samples preserved in 70% ethanol from Brazil, Kenya, Australia, Peru, Zambia, Thailand and Vietnam. Samples were kept at -20^0 C at the University of Western Australia (UWA) laboratory awaiting molecular analysis.

5.6.1 Total RNA extraction from individual whiteflies

The Arcturus Picopure RNA isolation kit was used to extract total RNA from the 49 individual whiteflies following extraction procedures from Sseruwagi et al. (2017). The TURBO DNA free kit (Ambion Life Technologies, CA, USA) was used to digest away trace amounts of unwanted DNA and divalent cations as described by the manufacturer. A vacuum centrifuge (Eppendof, Germany) was used as a measure to increase RNA concentration and this was set at room temperature for one hour. The resulting pellet was resuspended in 18 µl of RNase free water and stored immediately at -80^{0}C awaiting further analysis. The integrity of RNA was quantified by 2100 Bio analyser (Agilent Technologies).

5.6.2 cDNA and Illumina library preparation

Libraries were made from RNA of each individual whitefly using the Illumina TruSeq Stranded Total RNA preparation Kit, as described by the manufacturer (Illumina, San Diego, CA, USA). Libraries were then sent off to Macrogen Korea (www.macrogen.com) for sequencing. Seventeen samples were sequenced on a HiSeq 2500 on a rapid run mode, while 32 samples were sequenced on HiSeq 4000 on a high throughput mode (Table 1). Quality assessment and base calling were done using sequencing control software HCS V2.2 and HCS v3.3 respectively. The raw reads generated were assessed for quality using FastQC tool (Andrews, 2010).

5.6.3 Next Generation Sequencing (NGS) data analysis (*de novo* sequence assembly)

The CLC Genomic workbench (version 8.5.1) software was used for visualising and trimming of the raw reads. The quality score limit was set to 0.01 and the ambiguous limit set to 2. The trimmed reads were assembled into contigs using the *de novo* assembly function of the CLC genomic workbench. The following settings were used: automatic word size, automatic bubble size, minimum contig length 500, mismatch cost 2, insertion cost 3, deletion cost 3, length fraction 0.5 and similarity fraction 0.9. The assembly results are shown in Table 1.

5.6.4 Transcriptome analysis and phylogenetics

The assembled transcripts from CLC genomic workbench were catalogued into the Agalma pipeline (Dunn et al., 2013). Agalma automatically loads the gene predictions from other sources into the database, identifies homologous sequences across multiple species based on similarity search, builds gene trees for each set of aligned homologous sequence with RAxML (Stamatakis, 2014), identifies gene trees and gene subtrees with orthologs, creates a nucleotide and protein alignment for each and constructs a supermatrix of orthologous genes. Agalma generates a preliminary species tree in RAxML that we ignored in this study. A supermatrix of 7362 genes was generated.

Of the 7362 genes identified in Agalma, 3000 were selected for inclusion in the subsequent analysis because they had sequence information for at least 80% of the taxa. The 3000 genes were then passed to RAxML version 8.2.11 (Stamatakis, 2014) and SVDquatets (Chifman and Kubatko, 2014). For each gene, RAxML was run with

command raxmlHPC on 4 compute nodes, with a maximum of 28 tasks assigned per node, on Zeus (Pawsey Supercomputing Centre, Perth, Western Australia). The raxmlHPC command had the following input set: the rapid bootstrap analysis and search for the best scoring Maximum Likelihood (ML) tree set in one program run, GTRGAMMA model was automatically assigned based on the ML scores, number of bootstrap replicates set to 100, 12345 set as the random seed number for the parsimony inference and the random seed number for the rapid bootstrap. The estimated maximum likelihood trees (bestML) were used as the input data for ASTRAL (Mirarab et al., 2014), which was then run with default settings to generate a species tree from the 3000 genes. The species tree was compared with the bootstrap consensus tree generated from SVDQuartets as implemented in the PAUP* software (Swofford, 2001). In both species trees, individual sequences were mapped to their respective species identity; for example, *B. afer_119, B.afer_121* (Table 2) were mapped to *B. afer* sequences as indicated in Figures 2 and 3. Therefore, *B. afer* is representing two individuals, and the same was applied to the rest of the species in Table 2.

5.7 Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon request and sequence data will be submitted to a gene bank prior to

publication of this study.

Competing interests

Not applicable.

Funding

The Natural Resources Institute, University of Greenwich, supported sequencing, fieldwork and the first 2 years of TK's PhD from the grant provided by the Bill and Melinda Gates Foundation (Investment ID OPP1058938). A University of Western Australia Scholarship for International Research Fees (SIRF) F51416, University Postgraduate Award (UPA) F51204 and UWA Safety Net Top-Up Scholarship support TK's PhD.

Author's contributions

TK carried out the analysis with inputs from LK, LMB and LSL. TK, BRD participated in wet laboratory work, TK, BRD, PS, JN and TA provided samples. TK drafted the manuscript and all authors read and approved the manuscript.

5.8 Acknowledgements

TK is supported by a University of Western Australia Scholarship for International Research Fees (SIRF), University Postgraduate Award (UPA) and University of Western Australia Safety Net Top-Up Scholarship. This work forms part of his PhD. The Natural Resources Institute, University of Greenwich, supported sequencing, fieldwork and the first 2 years of TK's PhD from the grant provided by the Bill and Melinda Gates Foundation (Investment ID OPP1058938). National Crop Resources

Research Institute (NaCRRI) in Uganda provided samples and facilitated sample collection. We thank Paul De Barro, James Wainaina, Christopher Abu Omongo, Sunnada Chaovalit, Donald Kachigamba, and the University of Greenwich for providing samples. Resources provided by the Pawsey Supercomputing Centre with funding from the Australia Government and the Government of Western Australia supported computation analysis. Finally, we thank the farmers in East Africa for allowing us to collect samples on their farms.

5.9 References

ANDREWS, S. 2010. FastQC: a quality control tool for high throughput sequence data.

BOYKIN, L. M., ARMSTRONG, K. F., KUBATKO, L. & DE BARRO, P. 2012. Species delimitation and global biosecurity. *Evolutionary bioinformatics online,* 8**,** 1.

BOYKIN, L. M., BELL, C. D., EVANS, G., SMALL, I. & DE BARRO, P. J. 2013. Is agriculture driving the diversification of the Bemisia tabaci species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)?: Dating, diversification and biogeographic evidence revealed. *BMC Evol Biol,* 13**,** 228.

BOYKIN, L. M. & DE BARRO, P. J. 2014. A practical guide to identifying members of the Bemisia tabaci species complex: and other morphologically identical species. *Frontiers in Ecology and Evolution,* 2**,** 45.

BOYKIN, L. M., SHATTERS, R. G., JR., ROSELL, R. C., MCKENZIE, C. L., BAGNALL, R. A., DE BARRO, P. & FROHLICH, D. R. 2007. Global relationships of Bemisia tabaci (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol Phylogenet Evol,* 44**,** 1306-19.

CAMPBELL, B., STEFFEN-CAMPBELL, J. & GILL, R. 1996. Origin and radiation of whiteflies: an initial molecular phylogenetic assessment. *Bemisia: 1995, taxonomy, biology, damage, control and management*.

CHIFMAN, J. & KUBATKO, L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics***,** btu530.

DE BARRO, P. J., LIU, S.-S., BOYKIN, L. M. & DINSDALE, A. B. 2011. Bemisia tabaci: a statement of species status. *Annual review of entomology,* 56**,** 1-19.

DELATTE, H., HOLOTA, H., WARREN, B. H., BECKER, N., THIERRY, M. & REYNAUD, B. 2011. Genetic diversity, geographical range and origin of Bemisia tabaci (Hemiptera: Aleyrodidae) Indian Ocean Ms. *Bulletin of entomological research,* 101**,** 487-497.

DINSDALE, A., COOK, L., RIGINOS, C., BUCKLEY, Y. & DE BARRO, P. 2010. Refined global analysis of Bemisia tabaci (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Annals of the Entomological Society of America,* 103**,** 196-208.

DUNN, C. W., HOWISON, M. & ZAPATA, F. 2013. Agalma: an automated phylogenomics workflow. *BMC bioinformatics,* 14**,** 330.

ELFEKIH, S., ETTER, P., TAY, W., FUMAGALLI, M., GORDON, K., JOHNSON, E. & DE BARRO, P. 2018. Genome-wide analyses of the Bemisia tabaci species complex reveal contrasting patterns of admixture and complex demographic histories. *PloS one,* 13**,** e0190555.

HSIEH, C.-H., KO, C.-C., CHUNG, C.-H. & WANG, H.-Y. 2014. Multilocus approach to clarify species status and the divergence history of the Bemisia tabaci (Hemiptera: Aleyrodidae) species complex. *Molecular phylogenetics and evolution,* 76**,** 172-180.

JONES, D. R. 2003. Plant viruses transmitted by whiteflies. *European Journal of Plant Pathology,* 109**,** 195-219.

LEGG, J., OWOR, B., SSERUWAGI, P. & NDUNGURU, J. 2006. Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Advances in virus research,* 67**,** 355-418.

LEGG, J. P., SSERUWAGI, P., BONIFACE, S., OKAO-OKUJA, G., SHIRIMA, R., BIGIRIMANA, S., GASHAKA, G., HERRMANN, H.-W., JEREMIAH, S. &

OBIERO, H. 2014. Spatio-temporal patterns of genetic change amongst populations of cassava Bemisia tabaci whiteflies driving virus pandemics in East and Central Africa. *Virus research,* 186**,** 61-75.

LIU, L., YU, L., KUBATKO, L., PEARL, D. K. & EDWARDS, S. V. 2009. Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution,* 53**,** 320-328.

LIU, S.-S., COLVIN, J. & DE BARRO, P. J. 2012. Species concepts as applied to the whitefly Bemisia tabaci systematics: how many species are there? *Journal of Integrative Agriculture,* 11**,** 176-186.

MACFADYEN, S., PAULL, C., BOYKIN, L., DE BARRO, P., MARUTHI, M., OTIM, M., KALYEBI, A., VASSÃO, D., SSERUWAGI, P. & TAY, W. T. 2018. Cassava whitefly, Bemisia tabaci (Gennadius)(Hemiptera: Aleyrodidae) in East African farming landscapes: a review of the factors determining abundance. *Bulletin of entomological research***,** 1-18.

MADDISON, W. P. & KNOWLES, L. L. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic biology,* 55**,** 21-30.

MANSOOR, S., ZAFAR, Y. & BRIDDON, R. W. 2006. Geminivirus disease complexes: the threat is spreading. *Trends in plant science,* 11**,** 209-212.

MARUTHI, M., COLVIN, J., THWAITES, R. M., BANKS, G. K., GIBSON, G. & SEAL, S. E. 2004. Reproductive incompatibility and cytochrome oxidase I gene sequence variability amongst host‐adapted and geographically separate Bemisia tabaci populations (Hemiptera: Aleyrodidae). *Systematic Entomology,* 29**,** 560-568.

MIRARAB, S., REAZ, R., BAYZID, M. S., ZIMMERMANN, T., SWENSON, M. S. & WARNOW, T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics,* 30**,** i541-i548.

MUGERWA, H., REY, M. E., ALICAI, T., ATEKA, E., ATUNCHA, H., NDUNGURU, J. & SSERUWAGI, P. 2012. Genetic diversity and geographic distribution of Bemisia tabaci (Gennadius)(Hemiptera: Aleyrodidae) genotypes associated with cassava in East Africa. *Ecology and evolution,* 2**,** 2749-2762.

MUGERWA, H., SEAL, S., WANG, H.-L., PATEL, M. V., KABAALU, R., OMONGO, C. A., ALICAI, T., TAIRO, F., NDUNGURU, J. & SSERUWAGI, P. 2018. African ancestry of New World, Bemisia tabaci-whitefly species. *Scientific reports,* 8**,** 2734.

NDUNGURU, J., SSERUWAGI, P., TAIRO, F., STOMEO, F., MAINA, S., DJINKENG, A., KEHOE, M. & BOYKIN, L. M. 2015. 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,* 10**,** e0139321.

POLSTON, J. E., DE BARRO, P. & BOYKIN, L. M. 2014. Transmission specificities of plant viruses with the newly identified species of the Bemisia tabaci species complex. *Pest Manag Sci,* 70**,** 1547-52.

SSERUWAGI, P., LEGG, J., MARUTHI, M., COLVIN, J., REY, M. & BROWN, J. K. 2005. Genetic diversity of Bemisia tabaci (Gennadius)(Hemiptera: Aleyrodidae) populations and presence of the B biotype and a non‐B biotype that can induce silverleaf symptoms in squash, in Uganda. *Annals of Applied Biology,* 147**,** 253-265.

SSERUWAGI, P., WAINAINA, J., NDUNGURU, J., TUMUHIMBISE, R., TAIRO, F., GUO, J.-Y., VRIELINK, A., BLYTHE, A., KINENE, T. & DE MARCHI, B. 2017. The first transcriptomes from field-collected individual whiteflies (Bemisia tabaci, Hemiptera: Aleyrodidae). *Gates Open Research,* 1.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-1313.

SWOFFORD, D. L. 2001. Paup*: Phylogenetic analysis using parsimony (and other methods) 4.0. B5.

TAY, W. T., ELFEKIH, S., COURT, L. N., GORDON, K. H., DELATTE, H. & DE BARRO, P. J. 2017. 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,* 9**,** 2732-2738.

VYSKOČILOVÁ, S., TAY, W. T., VAN BRUNSCHOT, S., SEAL, S. & COLVIN, J. 2018. An integrative approach to discovering cryptic species within the Bemisia tabaci whitefly species complex. *Scientific reports,* 8**,** 10886.

WANG, P., RUAN, Y. M. & LIU, S. S. 2010. Crossing experiments and behavioral observations reveal reproductive incompatibility among three putative species of the whitefly Bemisia tabaci. *Insect Science,* 17**,** 508-516.

WASHBURN, J. D., SCHNABLE, J. C., CONANT, G. C., BRUTNELL, T. P., SHAO, Y., ZHANG, Y., LUDWIG, M., DAVIDSE, G. & PIRES, J. C. 2017. Genome-guided phylo-transcriptomic methods and the nuclear phylogentic tree of the Paniceae grasses. *Scientific Reports,* 7**,** 13528.

WOSULA, E. N., CHEN, W., FEI, Z. & LEGG, J. P. 2017. Unravelling the genetic diversity among cassava Bemisia tabaci whiteflies using NextRAD sequencing. *Genome biology and evolution,* 9**,** 2958-2973.

XU, J., DE BARRO, P. & LIU, S. 2010. Reproductive incompatibility among genetic groups of Bemisia tabaci supports the proposition that the whitefly is a cryptic species complex. *Bulletin of entomological research,* 100**,** 359-366.

ZAIDI, S. S.-E.-A., BRIDDON, R. W. & MANSOOR, S. 2017. Engineering dual begomovirus-Bemisia tabaci resistance in plants. *Trends in plant science,* 22**,** 6-8.

ZHANG, C., RABIEE, M., SAYYARI, E. & MIRARAB, S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics,* 19**,** 153.

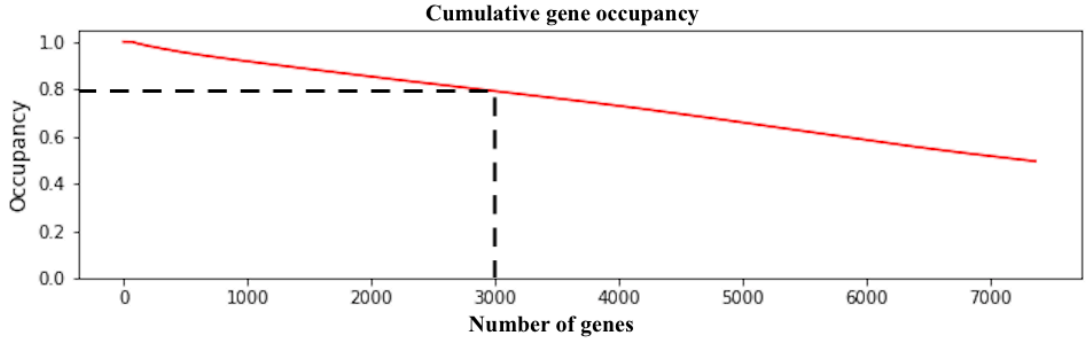


Figure 1
A plot of gene occupancies for genes in the supermartix created in Agalma ordered by the most complete genes. Of the 7362 genes identified, we used 3000 genes for downstream analysis, representing 80% occupancy as indicated with the dashed line.

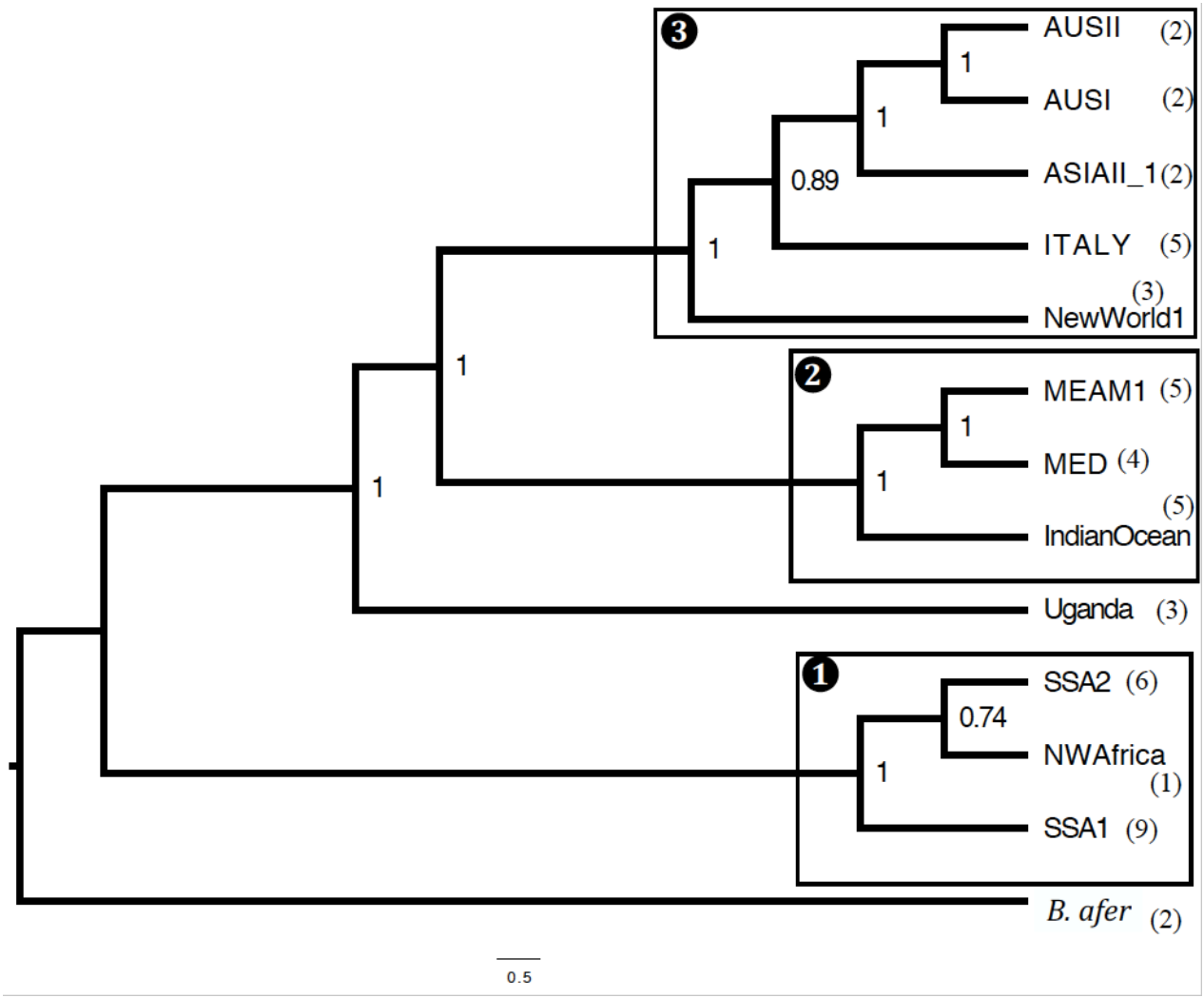


Figure 2

The *B. tabaci* species tree estimated from ASTRAL. Three thousand genes of 49 individual whiteflies generated from transcriptome data were used as source data in RAxML to generate 3000 maximum likelihood gene trees, which were used as input data for ASTRAL. The node support values are local posterior probability and the tree is rooted with *B. afer*. The numbers alongside each taxon indicate the number of individual whiteflies used for a particular species.

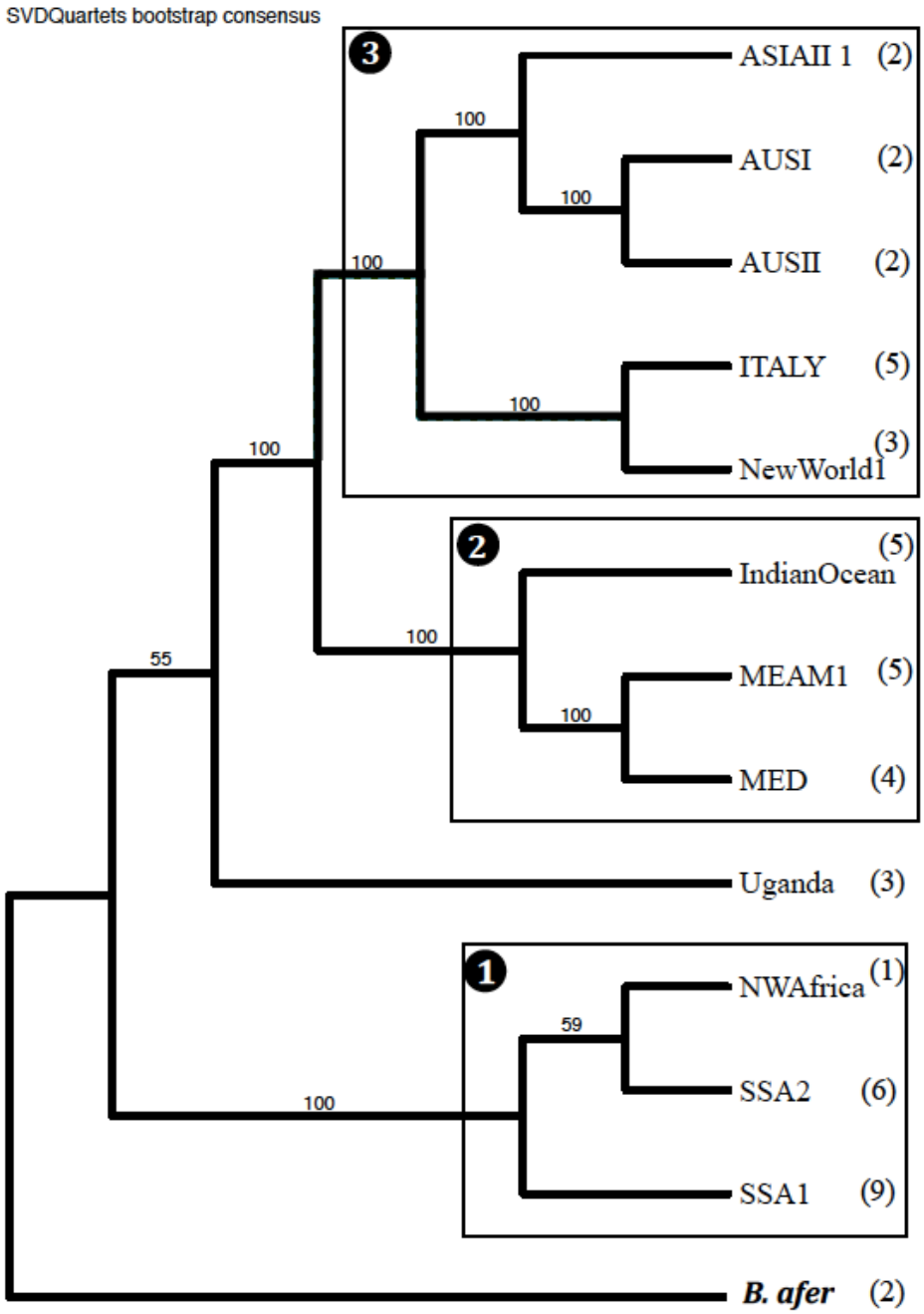


Figure 3

The *B. tabaci* SVDQuartets bootstrap consensus tree generated from 3000 genes of 49 individual whiteflies. The genes were extracted from transcriptome data of individual whiteflies. The alignment of the genes from Agalma supermatrix was composed of 2,360,189 base pairs. The numbers alongside each taxon indicate the number of individual whiteflies used for that species. One hundred bootstrap replicates were set and a multispecies tree model was used. The tree was rooted with *B. afer*

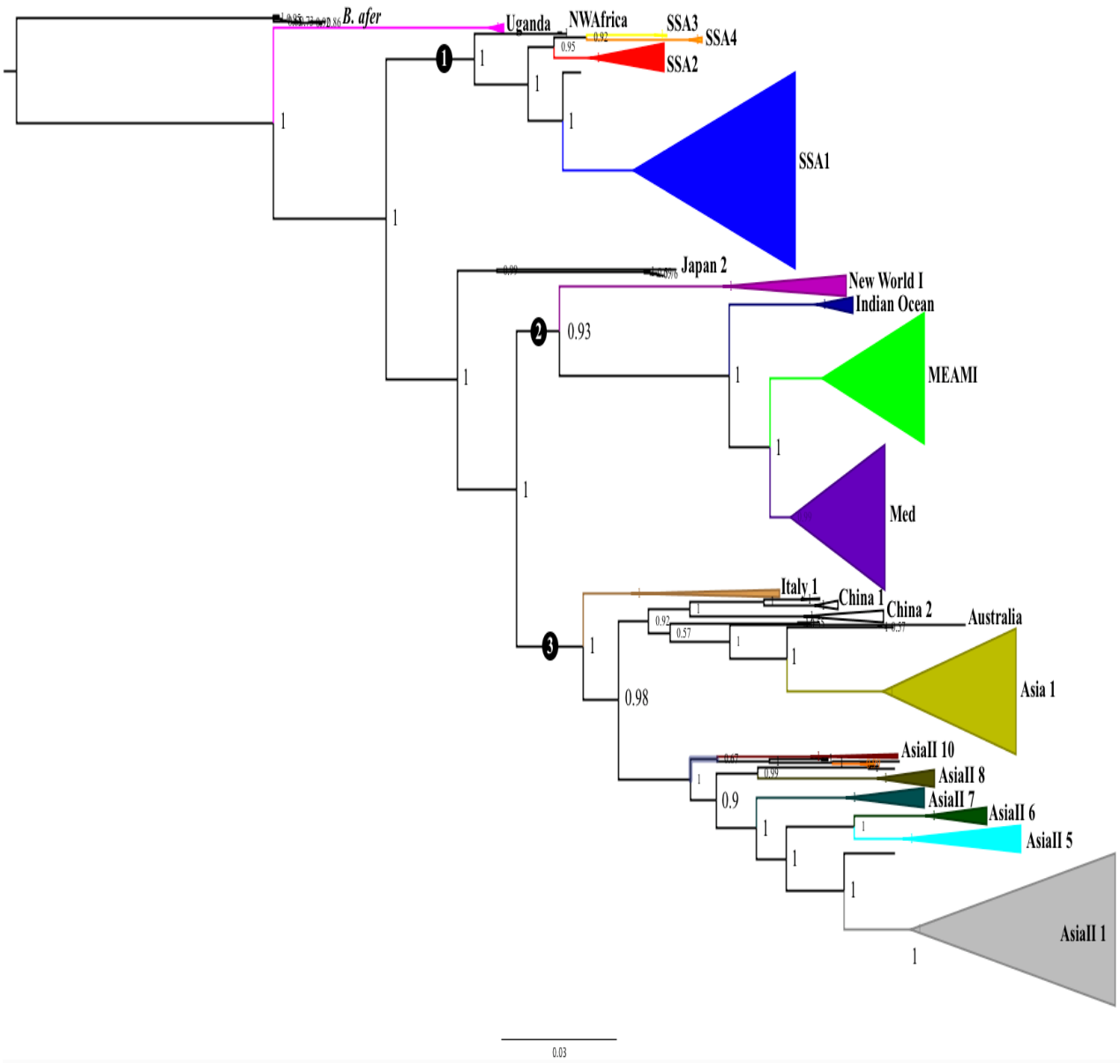


Figure 4

The *B. tabaci* mtCOI consensus gene tree obtained from the posterior distribution estimated by MrBayes 3.2.2. An alignment of 599 base pairs with 1211 taxa was used to generate the tree. The tree was rooted with *B. afer*. (Unpublished data by Kinene from chapter 3).

Table 1

Sample information, including summaries of results from *de novo* assembly and geographical positioning system coordinates for sample collection sites.

Sample ID	Number of reads obtained	Reads after trimming	Number of Contigs	Average Contig length	N50	Country
SSA1_59	8,000,000	7,848,046	16,473	1,102	1,195	Uganda
SSA1_65	8,000,000	7,857,579	17,146	1,024	1,071	Uganda
SSA1_66	8,000,000	7,857,564	16,489	1,045	1,095	Uganda
SSA1_67	8,000,000	7,876,283	18,271	1,165	1,289	Uganda
SSA1_68	8,000,000	7,861,938	19,852	1,201	1,368	Uganda
SSA1_70	8,000,000	7,861,502	20,182	1,113	1,205	Uganda
NWAfrica_113	8,000,000	7,850,359	13,156	1,023	1,050	Uganda
Bafer_119	8,000,000	7,845,846	20,368	1,238	1,400	Malawi
Bafer_121	8,000,000	7,837,796	20,006	1,199	1,340	Malawi
NewWorld1_156_2	40,826,418	40,604,557	27,466	1,356	1,689	Brazil
NewWorld1_156_3	43,194,264	42,953,727	30,716	1,346	1,655	Brazil
UGA_176	8,000,000	7,835,023	19,865	1,366	1,616	Kenya
SSA2_179	8,000,000	7,842,509	15,570	1,194	1,295	Kenya
SSA2_191	8,000,000	7,833,877	19,770	1,124	1,222	Kenya
SSA2_193	8,000,000	7,852,051	20,061	1,122	1,218	Kenya
SSA2_197	8,000,000	7,871,950	19,224	1,230	1,408	Kenya
MED_320_3	39,894,296	39,686,374	27,867	1,144	1,250	Brazil
AUSI_347_1	39,142,358	38,894,512	30,496	1,372	1,723	Australia
AUSI_347_2	37,306,246	37,071,584	27,343	1,314	1,614	Australia
SSA2_183	8,000,000	7,816,022	20,066	1,298	1,512	Kenya
UGA_184	8,000,000	7,825,633	16,265	1,028	1,059	Kenya
SSA2_194	8,000,000	7,856,677	17,838	1,094	1,179	Kenya
UGA_113_2	29,106,192	28,941,853	21,061	1,127	1,230	Uganda
MEAM1_153	36,449,340	36,232,156	25,435	1,235	1,432	Brazil
MED_154_1	37,206,396	37,006,799	27,841	1,149	1,261	Brazil
MED_154_2	29,840,288	29,665,109	20,831	1,158	1,288	Brazil
MEAM1_314_1	46,956,932	46,678,144	17,814	921	1,375	Peru
MEAM1_314_2	42,735,738	42,502,178	27,768	1,232	1,424	Peru
MEAM1_314_3	43,236,834	42,891,689	39,423	1,140	1,242	Peru
SSA1_334_1	38,748,628	38,556,885	20,528	986	1,011	Zambia
SSA1_139	43,488,848	40,262,056	33,784	1,234	1,449	Tanzania
IndianOcean_145_1	57,297,138	52,925,636	37,900	1,144	1,270	Tanzania
IndianOcean_145_2	51,983,978	48,379,504	32,281	1,127	1,247	Tanzania
IndianOcean_145_4	47,897,980	44,804,753	34,298	1,113	1,226	Tanzania
IndianOcean_145_5	42,841,660	39,076,561	33,093	1,058	1,138	Tanzania
IndianOcean_145_6	43,379,908	39,699,335	30,016	1,108	1,209	Tanzania
SSA1_151	47,750,214	41,384,692	29,451	1,052	1,119	Uganda
MEAM1_153_2	44,093,422	41,041,633	31,072	1,228	1,429	Brazil
NewWorld1_156_4	48,800,318	45,102,719	34,090	1,298	1,608	Brazil
MED_320	64,080,188	57,947,512	41,165	1,240	1,447	Brazil
ASIAII_1 329_1	45,838,350	41,816,776	50,980	1,030	1,078	Thailand
ASIAII_1 332_2	45,492,590	40,960,459	35,098	1,352	1,703	Vietnam
AUSII_348_1	47,454,764	43,338,895	16,719	1,231	1,452	Australia
AUSII_348_3	44,354,710	31,106,521	9,244	982	1,019	Australia
ITALY_382_1	47,359,732	43,297,176	30,653	1,276	1,508	
ITALY_382_2	54,547,202	49,921,245	33,704	1,362	1,719	
ITALY_382_3	57,644,056	53,065,972	33,428	1,351	1,724	
ITALY_382_4	58,336,180	53,803,301	29,305	1,232	1,459	
ITALY_382_5	56,140,048	51,120,827	30,131	1,246	1,497	

Table 2

Number and per cent of genes in the supermatrix from Agalma for each taxon.

Species ID	Number of genes	Percentage (%)
SSA1_139	5214	70.80
ASIAII_1_332_2	5213	70.80
ITALY_382_2	4736	64.30
SSA2_183	4733	64.30
IndianOcean_145_1	4661	63.30
Bafer_119	4595	62.40
SSA2_197	4511	61.30
Bafer_121	4501	61.10
AUSI_347_1	4483	60.90
IndianOcean_145_4	4465	60.60
MED_320	4432	60.20
UGA_176	4424	60.10
ITALY_382_3	4384	59.50
NewWorld1_156_3	4351	59.10
AUSI_347_2	4295	58.30
SSA1_68	4284	58.20
ITALY_382_1	4251	57.70
IndianOcean_145_2	4199	57.00
SSA1_67	4188	56.90
NewWorld1_156_2	4094	55.60
IndianOcean_145_6	4012	54.50
ITALY_382_5	3886	52.80
MEAM1_153_2	3864	52.50
AUSII_348_1	3863	52.50
NewWorld1_156_4	3859	52.40
SSA1_59	3804	51.70
SSA2_193	3524	47.90
SSA1_70	3460	47.00
SSA1_151	3421	46.50
ASIAII_1_329_1	3400	46.20
IndianOcean_145_5	3389	46.00
MEAM1_153	3370	45.80
SSA2_179	3292	44.70
MEAM1_314_2	3172	43.10
NWAfrica_113	3075	41.80
MED_154_1	3047	41.40
SSA1_65	3040	41.30
MED_320_3	2910	39.50
MEAM1_314_3	2908	39.50
ITALY_382_4	2864	38.90
SSA2_191	2805	38.10
MED_154_2	2795	38.00
SSA1_66	2729	37.10
UGA_113_2	2727	37.00
AUSII_348_3	2572	34.90
SSA2_194	2193	29.80
UGA_184	1632	22.20
MEAM1_314_1	1565	21.30
SSA1_334_1	1434	19.50

Chapter 6: General discussion

6.1 Discussion

This study presents the role of the heat shock protein 90 (HSP90) in *B. tabaci's* ability to adapt to varying climate and the relation of HSP90 to the superabundance phenomenon of whiteflies. The study further discusses the phylogenetic relationship of the various *B. tabaci* species using common nuclear genes (RNA polymerase II and Shaker cognate W gene). Finally, the study presents the framework for *B. tabaci* systematics; i.e., we go beyond using a single locus (mtCOI) for explaining *B. tabaci* systematics to using 3000 orthologous nuclear genes. Therefore, we present the first estimate for a *B. tabaci* species tree.

We first considered the two common nuclear single copy genes (RNA polymerase II and Shaker cognate W gene) and compared them with the mtCOI gene tree topology to define species-level boundaries. We found the nuclear gene tree topologies well resolved in some species and not well resolved in others, especially the SSA species; thus, the conflicting pattern with the mtCOI gene tree topology (Figure 2 and Figure 3, chapter 4). This pattern of conflict can be caused by many factors, for example sampling error or evolutionary processes such as gene duplication, horizontal gene transfer and incomplete lineage sorting (Maddison, 1997). Some species were well resolved in the nuclear genes and were consistent with the mtCOI gene tree. These included MEAM1, Australia, and *B. porteri* (Table 4, chapter 4). Other species were well resolved in one nuclear gene and poorly resolved in another, but when a species tree was estimated in *BEAST for the two nuclear genes, some of the conflicts were resolved (Figure 4, chapter 4). This indicated the need for more nuclear loci to improve

the phylogenetic signal for comparison with the mtCOI locus. Hsieh et al. (2014) used three nuclear loci and had a better phylogenetic signal for comparison with mtCOI locus. Therefore, we attempted to use more nuclear loci in Chapter 5 where we estimated the *B. tabaci* species tree using 3000 orthologous genes.

We estimated the first *B. tabaci* species tree using 3000 orthologous genes generated from 49 individual whitefly transcriptomes representing 12 genetic groups, namely: AUSI, AUSII, ASIAII_1, ITALY, New world 1, MEAM1, MED, Indian Ocean, Uganda, SSA1, SSA2 and *B. afer* (Figure 2 and Figure 3, chapter 5). The *B. tabaci* Uganda placement in the species tree was not consistent with the conventional placement in the mtCOI tree topology. However, the rest of the species placements were consistent with the mtCOI tree topology. We think this might be one of the limitations of the mtCOI locus to accurately represent all species relationships (Wosula et al., 2017). In other studies based on mtCOI locus alone, *B. tabaci* Uganda is found at the base of the tree with *B. afer* (Mugerwa et al., 2012; Sseruwagi et al., 2005). It is therefore important to properly understand species relationships because misidentification of species might lead to scientific miscommunication and may have tremendous implications on proposed management strategies.

The superabundance phenomenon of the whiteflies is explained in relation to the roles and importance of the HSP90 gene. Originally, our aim was to measure the rates of molecular evolution of HSP90 gene using codon models in order to establish the whitefly species that were under selection and the region they represented (Appendix A1). However, on further analysis of the dataset, we detected recombination events in the HSP90 gene of the *B. tabaci* species. The presence of recombination events in my dataset meant that there were different underlying phylogenetic histories in the

recombinant regions (Posada and Crandall, 2002; Schierup and Hein, 2000); yet, the codon models used in the selection pressure analysis assume a single phylogenetic tree for all sites. Violating this assumption in codon models would elevate the false positive rate. Therefore, the focus of the paper indicated in appendix A1 was changed to recombination analysis, discussed in chapter 3.

Based on the findings in chapter 3, recombination provides evidence for adaptive evolution of whiteflies in hot weather conditions because recombination events were detected in some SSA species. In contrast, in the other SSA species there were no traces of recombination detected (Figure 1, chapter 3) and this could be attributed to viruses present or the endosymbiont composition, and not temperature (Aregbesola et al., 2018; Sun et al., 2013). Apart from recombination in the HSP90 gene, other factors, such as resistance to insecticide, might also be contributing to the increasing populations of whiteflies in the region, and somehow the HSP90 gene is responsible for this as well because it is synthesized whenever an organism is under stress (Díaz et al., 2014).

6.1.1 Main findings

Chapter 2: This chapter highlights the methods of rooting phylogenetic trees and the reasons as to why it is important to accurately locate the root of the phylogenetic tree. Since our work revolves around making inference on the phylogenetic tree this chapter prepares the ground for the deeper analysis carried out in this thesis. Incorrectly rooted phylogenetic trees might mislead evolutionary and taxonomic inferences (Graham et al., 2002). We demonstrated one of the common methods for rooting phylogenetic trees, the outgroup method, using a modified *B. tabaci* dataset from Boykin et al. (2013). This method is accurate but it can be problematic and lead to long branch attraction problems if the outgroup is distantly related to the ingroup (Felsenstein, 1978). In our dataset

rooted with *Bemisia atriplex*, we did not have the long branch attraction problem and in other chapters of the thesis we rooted our phylogenetic trees *B. afer*.

The other methods discussed in this chapter include: midpoint rooting, which is a fast method; however, it is dependent on clocklike data and its doesn't work well with imbalanced tree topologies (Swofford et al., 1996). The molecular clock rooting method assumes a constant rate of evolution for sequences of interest (Yang and Rannala, 2012) and is problematic for distantly related species, as well as computationally intensive. Finally the Bayesian molecular clock method provides support for the location of the root on any branch of the ingroup topology (Huelsenbeck et al., 2002). It also helps users to evaluate alternative rooting, which is not the case with other rooting methods. Therefore it is especially important when the actual root of the tree is unknown. .

Chapter 3: There is evidence that recombination is driving the evolution of *B. tabaci* species in East Africa (Cangi et al., 2016) because we detected recombinant events in the HSP90 gene of *B. tabaci* species. Recombination might have influenced *B. tabaci* species to adapt to high temperatures since we found whiteflies surviving at high temperatures and in large numbers in the East African region. However, more biological experiments are needed to evaluate our results of the recombination signal detected in the HSP90 gene of the members of *B. tabaci* species.

Apart from recombination, the HSP90 gene is also known for thermal adaption of species (Diaz et al., 2015; Hoffmann et al., 2003). With the rising temperature in the world (Pereira, 2017) one would expect whiteflies to be wiped out, but this is not the case; instead, superabundant populations have been seen on cassava plants in hot seasons in East Africa. This means that whiteflies might have developed a molecular

response to tolerate heat stress and this is the typical role of the HSP90 gene (Hoffmann et al., 2003; Fasolo and Krebs, 2004). The population density of whiteflies may depend on many factors, including whitefly species. Studies have shown that some whitefly species have a wider range of temperature compared to others; for example, MED species are adapted to a wider range of temperature than MEAM1 (Jiang et al., 2017). Unfortunately none of these studies have been carried out on the SSA species.

Chapter 4: We observed that the Bayesian estimate of the mtCOI gene tree was not concordant with the Bayesian estimate of the nuclear gene tree topologies (Figure 2 and Figure 3, chapter 4). This is usually due to evolutionary processes such as gene duplication, incomplete lineage sorting, sampling error and horizontal gene transfer, and is hence the reason gene trees are not necessarily equal to species trees (Maddison, 1997; Wiens et al., 2010). Gene flow in MED and MEAM1 species has been reported by Elfekih et al. (2018). Tajebe et al. (2015) reported hybridisation between SSA2 and SSA1 in Tanzania and suggested that SSA1 was responsible for the cassava virus pandemic. The genetic diversity of the members of the *B. tabaci* species complex has always been based on the mtCOI gene (Boykin et al., 2007; De Barro, 2012); however, we know that using the mtCOI gene alone might not accurately represent all the members of the *B. tabaci* complex (Wosula et al., 2017). We used the two common nuclear genes (Shaker cognate W and RNA polymerase II); we further noted that most of the species in these trees were not well resolved, thus the two genes were not informative enough.

We formulated a species tree of the two nuclear genes (Figure 4, chapter 4) and compared it with mtCOI gene tree topology; we noted that the species tree topology from the two nuclear genes had low clade support. This showed the need to increase the

number of nuclear genes to improve the resolution of the phylogenetic signal for comparison with the mtCOI gene tree topology. Therefore, with a few nuclear genes, mtCOI gene tree topology remains robust in delimiting members of the *B. tabaci* species complex. At this point we recommend that a large number of genes would have enough information for a reliable estimate of the *B. tabaci* species tree (Maddison and Knowles, 2006) that would be comparable with the mtCOI tree topology. This introduces us to chapter 5, where we used 3000 orthologous nuclear genes to estimate the *B. tabaci* species tree.

Chapter 5: To date there is no published *B. tabaci* species tree, and the genetic diversity of *B. tabaci* species has been mostly based on the partial mtCOI gene (Dinsdale et al., 2010; Boykin et al., 2007; De Barro, 2012). Wosula et al. (2017) pointed out that the mtCOI gene alone might not be an efficient marker for delimiting cassava colonising *B. tabaci* species. Therefore, in this study we estimated the *B. tabaci* species tree from 3000 orthologous genes generated from the transcriptomes of 49 individual whiteflies representing 12 genetic groups. My results show that the placement of *B. tabaci* species Uganda was not consistent with its usual placement in the conventional mtCOI gene tree topology (Boykin et al., 2007; Mugerwa et al., 2012; Sseruwagi et al., 2005), while the other species placements from the 11 genetic groups were consistant with the mtCOI tree topology. My results show that using multi-loci datasets improves the phylogenetic signal more than the resolution from a single locus. Maddison and Knowles (2006) suggest that using multi-loci data provides a reliable estimate of the phylogenetic tree.

The *B. tabaci* Uganda species were first recognised by Boykin et al. (2007) as a special group that was genetically different from all other groups, and since *B. tabaci* Uganda and the other SSA species were at the base of her phylogeny, it was hypothesised that

Africa was the origin of the members of the *B. tabaci* species complex. Dinsdale et al. (2010) also identified *B. tabaci* Uganda, called them an unusual group, and suggested that the incorporation of Uganda into the phylogeny was the cause of some conflict as to its placement, but its removal omitted the problem. Therefore, the authors (Dinsdale et al., 2010) omitted the Ugandan species in the analysis, but in the phylogeny they indicated what would have been its position with dashed lines. Unfortunately the annual review (De Barro et al., 2011a) carried on the phylogeny from Dinsdale et al. (2010) with a contested location of *B. tabaci* Uganda. Therefore, we might not be the first ones to have seen this pattern; previous researchers saw the pattern and ignored it. The Ugandan species of *B. tabaci* is placed as sister to the invasive species (MED, MEAM1) and Indian Ocean species of *B. tabaci* (Figure 2 and Figure 3, chapter 5). In other research publications for mtCOI phylogeny, *B. tabaci* Uganda has been seen at the base of the tree (Mugerwa et al., 2012; Sseruwagi et al., 2005; Boykin et al., 2007).

Previously the genetic diversity of the *B. tabaci* species was based on a few genes; for example, the partial mtCOI gene (Boykin et al., 2007; Dinsdale et al., 2010). The major reason for this was that the available technology was the Sanger sequencing (Sanger et al., 1977; Sanger and Coulson, 1975). However, the advent of next generation sequencing technology (NGS) has revolutionised biology – it is now relatively easy to sequence thousands of genes from an individual whitefly with the recent transcriptome protocols developed by Sseruwagi et al. (2017). Therefore, with NGS technology we can sequence a large number of genes. This has equipped researchers with enormous amounts of data and consequently improved the methods of data analysis. For example, there are a lot of new phylogenetic analysis methods for multi-loci data that have been developed to catch up with the technology; for example, SVDQuartets (Chifman and Kubatko, 2014), Agalma (Dunn et al., 2013), ASTRAL (Mirarab et al., 2014; Zhang et

al., 2018) and others. The increase in new sequencing technology coming on the market – for example the portable DNA sequencer (Oxford Nanopore) – is going to revolutionise the way whiteflies and viruses are detected across the borders of East African countries. This will ultimately improve biosecurity measures through the screening of all agricultural products imported and exported between East African countries in real time and improve virus tracking and *B. tabaci* management.

How is this information going to help farmers and researchers?

This robust species tree will help agricultural research institutes, field extension officers and researchers in the whitefly community to correctly identify the right whitefly species that is infesting farmers' fields globally. In East Africa, the cassava breeders will take advantage by breeding new cassava varieties that are resistant to the virus and the right whitefly species vectoring viruses in the region. The ultimate goal is for farmers to have clean seeds and virus-free cassava in order to eradicate hunger.

6.2 Concluding remarks and future research needed:

This was the first attempt to generate a species tree for the *Bemisia tabaci* species complex. The species trees were estimated from 3000 nuclear genes and these species trees have improved our understanding of the *B. tabaci* systematics. We have provided a robust and improved systematics framework of the *B. tabaci* species, which is largely congruent with the traditional mtCOI phylogeny. The major difference between our species tree and the mtCOI trees is the placement of the Uganda species, which is now sister to the MED, MEAM1 and Indian Ocean species, versus at the base of the phylogeny as seen in the previous mtCOI tree. We determined the mtCOI locus is good for rapid identification of species but it might not reflect evolutionary processes such as

hybridisation. We suggest adding more species to our robust species tree beyond the twelve species included here. In the process of trying to unravel the rates of evolution for the various genes used in the species tree estimation we uncovered that recombination events detected in the HSP90 gene of the members of the *B. tabaci* species complex might be driving the adaptive evolution of whiteflies in East Africa. However, more research is needed to evaluate these results with biological experiments. It has been well documented that the HSP90 gene is responsible for thermal adaptation of whitefly species but unfortunately no gene expression study among the HSP90 gene has been carried out on the SSA *B. tabaci* species. We suggest this as a way forward in unravelling the phenomenon of superabundance. Beyond adding more species to our species tree by generating more whole genome data for global whiteflies, we urgently need to test all of the various *B. tabaci* species' ability to transmit viruses and their response to the myriad of insecticides being used in farmers' fields. Once these key elements are determined, field-based diagnostics need to be rolled out to help identify the species quickly – so farmers can take action.

Bibliography

ALABI, O. J., KUMAR, P. L. & NAIDU, R. A. 2011. Cassava mosaic disease: a curse to food security in subSaharan Africa. *APSnet Features*.

ALICAI, T., NDUNGURU, J., SSERUWAGI, P., TAIRO, F., OKAO-OKUJA, G., NANVUBYA, R., KIIZA, L., KUBATKO, L., KEHOE, M. A. & BOYKIN, L. M. 2016. Cassava brown streak virus has a rapidly evolving genome: implications for virus speciation, variability, diagnosis and host resistance. *Scientific reports,* 6.

ALICAI, T., OMONGO, C., MARUTHI, M., HILLOCKS, R., BAGUMA, Y., KAWUKI, R., BUA, A., OTIM-NAPE, G. & COLVIN, J. 2007. Re-emergence of cassava brown streak disease in Uganda. *Plant Disease,* 91**,** 24-29.

ANDREWS, S. 2010. FastQC: a quality control tool for high throughput sequence data.

AREGBESOLA, O. Z., LEGG, J. P., SIGSGAARD, L., LUND, O. S. & RAPISARDA, C. 2018. Potential impact of climate change on whiteflies and implications for the spread of vectored viruses. *Journal of Pest Science***,** 1-12.

ATEKA, E., ALICAI, T., NDUNGURU, J., TAIRO, F., SSERUWAGI, P., KIARIE, S., MAKORI, T., KEHOE, M. A. & BOYKIN, L. M. 2017. Unusual occurrence of a DAG motif in the Ipomovirus Cassava brown streak virus and implications for its vector transmission. *PloS one,* 12**,** e0187883.

BANKS, G., BEDFORD, I., BEITIA, F., RODRIGUEZ-CEREZO, E. & MARKHAM, P. 1999. A novel geminivirus of Ipomoea indica (Convolvulacae) from Southern Spain. *Plant Disease,* 83**,** 486-486.

BENNETT, B. 2015. Guest editorial: smallholder cassava production and the cassava processing sector in Africa. Practical Action Publishing.

BERRY, S. D., FONDONG, V. N., REY, C., ROGAN, D., FAUQUET, C. M. & BROWN, J. K. 2004. Molecular evidence for five distinct Bemisia tabaci (Homoptera: Aleyrodidae) geographic haplotypes associated with cassava plants

in sub-Saharan Africa. *Annals of the Entomological Society of America,* 97**,** 852-859.

BIRD, J. 1957. A whitefly-transmitted mosaic of Jatropha gossypifolia.

BONATO, O., LURETTE, A., VIDAL, C. & FARGUES, J. 2007. Modelling temperature ‐ dependent bionomics of Bemisia tabaci (Q ‐ biotype). *Physiological Entomology,* 32**,** 50-55.

BOYKIN, L. M. 2014. Bemisia tabaci nomenclature: lessons learned. *Pest management science,* 70**,** 1454-1459.

BOYKIN, L. M., ARMSTRONG, K. F., KUBATKO, L. & DE BARRO, P. 2012. Species delimitation and global biosecurity. *Evolutionary bioinformatics online,* 8**,** 1.

BOYKIN, L. M., BELL, C. D., EVANS, G., SMALL, I. & DE BARRO, P. J. 2013. Is agriculture driving the diversification of the Bemisia tabaci species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)?: Dating, diversification and biogeographic evidence revealed. *BMC Evol Biol,* 13**,** 228.

BOYKIN, L. M. & DE BARRO, P. J. 2014. A practical guide to identifying members of the Bemisia tabaci species complex: and other morphologically identical species. *Frontiers in Ecology and Evolution,* 2**,** 45.

BOYKIN, L. M., KINENE, T., WAINAINA, J. M., SAVILL, A., SEAL, S., MUGERWA, H., MACFADYEN, S., TAY, W. T., DE BARRO, P. & KUBATKO, L. 2018. Review and guide to a future naming system of African Bemisia tabaci species. *Systematic Entomology,* 43**,** 427-433.

BOYKIN, L. M., SAVILL, A. & DE BARRO, P. 2017. Updated mtCOI reference dataset for the Bemisia tabaci species complex. *F1000Research,* 6.

BOYKIN, L. M., SHATTERS, R. G., JR., ROSELL, R. C., MCKENZIE, C. L., BAGNALL, R. A., DE BARRO, P. & FROHLICH, D. R. 2007. Global

relationships of Bemisia tabaci (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol Phylogenet Evol,* 44**,** 1306-19.

BROWN, J., COATS, S., BEDFORD, I., MARKHAM, P., BIRD, J. & FROHLICH, D. 1995a. Characterization and distribution of esterase electromorphs in the whitefly, Bemisia tabaci (Genn.)(Homoptera: Aleyrodidae). *Biochemical genetics,* 33**,** 205-214.

BROWN, J. K. & CZOSNEK, H. 2002. Whitefly transmission of plant viruses.

BROWN, J. K., FROHLICH, D. & ROSELL, R. C. 1995b. The sweetpotato or silverleaf whiteflies: biotypes of Bemisia tabaci or a species complex? *Annual review of entomology,* 40**,** 511-534.

BRUMIN, M., KONTSEDALOV, S. & GHANIM, M. 2011. Rickettsia influences thermotolerance in the whitefly Bemisia tabaci B biotype. *Insect Science,* 18**,** 57-66.

BUTLER JR, G., HENNEBERRY, T. & CLAYTON, T. 1983. Bemisia tabaci (Homoptera: Aleyrodidae): development, oviposition, and longevity in relation to temperature. *Annals of the Entomological Society of America,* 76**,** 310-313.

BYRNE, D. & BELLOWS, J. 1991. TS *Whitefly biology. annu. rev. entomol,* 36**,** 431-57.

BYRNE, D. N. & MILLER, W. B. 1990. Carbohydrate and amino acid composition of phloem sap and honeydew produced by Bemisia tabaci. *Journal of Insect Physiology,* 36**,** 433-439.

CAHILL, M., BYRNE, F. J., GORMAN, K., DENHOLM, I. & DEVONSHIRE, A. L. 1995. Pyrethroid and organophosphate resistance in the tobacco whitefly Bemisia tabaci (Homoptera: Aleyrodidae). *Bulletin of Entomological Research,* 85**,** 181-187.

CAMPBELL, B., STEFFEN-CAMPBELL, J. & GILL, R. 1996. Origin and radiation of whiteflies: an initial molecular phylogenetic assessment. *Bemisia: 1995, taxonomy, biology, damage, control and management*.

CANGI, N., GORDON, J. L., BOURNEZ, L., PINARELLO, V., APRELON, R., HUBER, K., LEFRANÇOIS, T., NEVES, L., MEYER, D. F. & VACHIÉRY, N. 2016. Recombination is a major driving force of genetic diversity in the anaplasmataceae Ehrlichia ruminantium. *Frontiers in cellular and infection microbiology,* 6**,** 111.

CASTILLO-RAMÍREZ, S., LIU, L., PEARL, D. & EDWARDS, S. 2010. Bayesian estimation of species trees: a practical guide to optimal sampling and analysis. *Estimating species trees: practical and theoretical aspects***,** 15-33.

CHAPMAN, R. F. 1998. *The insects: structure and function*, Cambridge university press.

CHIFMAN, J. & KUBATKO, L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics***,** btu530.

CHIKOTI, P. C., NDUNGURU, J., MELIS, R., TAIRO, F., SHANAHAN, P. & SSERUWAGI, P. 2013. Cassava mosaic disease and associated viruses in Zambia: occurrence and distribution. *International Journal of Pest Management,* 59**,** 63-72.

CHU, D., WAN, F. H., ZHANG, Y. J. & BROWN, J. K. 2010. Change in the biotype composition of Bemisia tabaci in Shandong Province of China from 2005 to 2008. *Environmental entomology,* 39**,** 1028-1036.

COLVIN, J., OMONGO, C., MARUTHI, M., OTIM‐NAPE, G. & THRESH, J. 2004. Dual begomovirus infections and high Bemisia tabaci populations: two factors driving the spread of a cassava mosaic disease pandemic. *Plant Pathology,* 53**,** 577-584.

CONESA, A., MADRIGAL, P., TARAZONA, S., GOMEZ-CABRERO, D., CERVERA, A., MCPHERSON, A., SZCZEŚNIAK, M. W., GAFFNEY, D. J., ELO, L. L. & ZHANG, X. 2016. A survey of best practices for RNA-seq data analysis. *Genome biology,* 17**,** 13.

COSSINS, A. & BOWLER, K. 1987. Rate compensations and capacity adaptations. *Temperature biology of animals.* Springer.

COSTA, A. & RUSSELL, L. 1975. Failure of Bemisia tabaci to breed on cassava plants in Brazil (Homoptera: Aleyrodidae). *Cienc. Cult,* 27**,** 388-390.

COSTA, H. & BROWN, J. 1990. Variability in biological characteristics, isozyme patterns and virus transmission among populations of Bemisia tabaci in Arizona. *Phytopathology,* 80**,** 763-766.

COSTA, H. & BROWN, J. K. 1991. Variation in biological characteristics and esterase patterns among populations of Bemisia tabaci, and the association of one population with silverleaf symptom induction. *Entomologia experimentalis et applicata,* 61**,** 211-219.

CUI, X., WAN, F., XIE, M. & LIU, T. 2008. Effects of heat shock on survival and reproduction of two whitefly species, Trialeurodes vaporariorum and Bemisia tabaci biotype B. *Journal of Insect Science,* 8**,** 1-10.

CURNUTTE, L. B., SIMMONS, A. M. & ABD-RABOU, S. 2014. Climate change and Bemisia tabaci (Hemiptera: Aleyrodidae): impacts of temperature and carbon dioxide on life history. *Annals of the Entomological Society of America,* 107**,** 933-943.

DARRIBA, D., TABOADA, G. L., DOALLO, R. & POSADA, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods,* 9**,** 772-772.

DE BARRO, P. J. 2012. The Bemisia tabaci species complex: questions to guide future research. *Journal of Integrative Agriculture,* 11**,** 187-196.

DE BARRO, P. J., LIU, S.-S., BOYKIN, L. M. & DINSDALE, A. B. 2011a. Bemisia tabaci: a statement of species status. *Annual review of entomology,* 56**,** 1-19.

DE BARRO, P. J., LIU, S. S., BOYKIN, L. M. & DINSDALE, A. B. 2011b. Bemisia tabaci: a statement of species status. *Annu Rev Entomol,* 56**,** 1-19.

DE BON, H., HUAT, J., PARROT, L., SINZOGAN, A., MARTIN, T., MALEZIEUX, E. & VAYSSIERES, J.-F. 2014. Pesticide risks from fruit and vegetable pest management by small farmers in sub-Saharan Africa. A review. *Agronomy for Sustainable Development,* 34**,** 723-736.

DEGNAN, J. H. & ROSENBERG, N. A. 2006. Discordance of species trees with their most likely gene trees. *PLoS genetics,* 2**,** e68.

DELATTE, H., HOLOTA, H., WARREN, B. H., BECKER, N., THIERRY, M. & REYNAUD, B. 2011. Genetic diversity, geographical range and origin of Bemisia tabaci (Hemiptera: Aleyrodidae) Indian Ocean Ms. *Bulletin of entomological research,* 101**,** 487-497.

DESAI, M. M. & FISHER, D. S. 2007. Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics*.

DÍAZ, F., MUÑOZ‐VALENCIA, V., JUVINAO‐QUINTERO, D., MANZANO‐MARTÍNEZ, M., TORO‐PEREA, N., CÁRDENAS‐HENAO, H. & HOFFMANN, A. 2014. Evidence for adaptive divergence of thermal responses among B emisia tabaci populations from tropical C olombia following a recent invasion. *Journal of evolutionary biology,* 27**,** 1160-1171.

DIAZ, F., OROBIO, R. F., CHAVARRIAGA, P. & TORO-PEREA, N. 2015. Differential expression patterns among heat-shock protein genes and thermal

responses in the whitefly Bemisia tabaci (MEAM 1). *Journal of thermal biology,* 52**,** 199-207.

DINSDALE, A., COOK, L., RIGINOS, C., BUCKLEY, Y. & DE BARRO, P. 2010. Refined global analysis of Bemisia tabaci (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Annals of the Entomological Society of America,* 103**,** 196-208.

DITTRICH, V., HASSAN, S. & ERNST, G. 1986. Development of a new primary pest of cotton in the Sudan: Bemisia tabaci, the whitefly. *Agriculture, ecosystems & environment,* 17**,** 137-142.

DIXON, A. G., BANDYOPADHYAY, R., COYNE, D., FERGUSON, M., FERRIS, R. S. B., HANNA, R., HUGHES, J., INGELBRECHT, I., LEGG, J. & MAHUNGU, N. 2003. Cassava: from poor farmers' crop to pacesetter of African rural development. *Chronica Horticulturae,* 43**,** 8-15.

DOWELL, R. V. 1981. Biology, plant-insect relations, and control of the citrus blackfly. *Florida. Agricultural Experiment Station. Bulletin (USA)*.

DUNN, C. W., HOWISON, M. & ZAPATA, F. 2013. Agalma: an automated phylogenomics workflow. *BMC bioinformatics,* 14**,** 330.

EDWARDS, S. V., LIU, L. & PEARL, D. K. 2007. High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences,* 104**,** 5936-5941.

ELBAZ, M., LAHAV, N. & MORIN, S. 2010. Evidence for pre-zygotic reproductive barrier between the B and Q biotypes of Bemisia tabaci (Hemiptera: Aleyrodidae). *Bulletin of entomological research,* 100**,** 581-590.

ELFEKIH, S., ETTER, P., TAY, W., FUMAGALLI, M., GORDON, K., JOHNSON, E. & DE BARRO, P. 2018. Genome-wide analyses of the Bemisia tabaci species

complex reveal contrasting patterns of admixture and complex demographic histories. *PloS one,* 13**,** e0190555.

FASOLO, A. G. & KREBS, R. A. 2004. A comparison of behavioural change in Drosophila during exposure to thermal stress. *Biological Journal of the Linnean Society,* 83**,** 197-205.

FAUQUET, C., BISARO, D., BRIDDON, R., BROWN, J., HARRISON, B., RYBICKI, E., STENGER, D. & STANLEY, J. 2003. Virology division news: revision of taxonomic criteria for species demarcation in the family Geminiviridae, and an updated list of begomovirus species. *Archives of virology,* 148**,** 405-421.

FAUQUET, C., BRIDDON, R., BROWN, J., MORIONES, E., STANLEY, J., ZERBINI, M. & ZHOU, X. 2008. Geminivirus strain demarcation and nomenclature. *Archives of virology,* 153**,** 783-821.

FAUQUET, C. & FARGETTE, D. 1990. African cassava mosaic virus: etiology, epidemiology and control. *Plant Dis,* 74**,** 404-411.

FISHPOOL, L. & BURBAN, C. 1994. Bemisia tabaci: the whitefly vector of African cassava mosaic geminivirus. *Tropical Science,* 34**,** 55-72.

GELMAN, D. B., BLACKBURN, M. B., HU, J. S. & GERLING, D. 2002. The nymphal‐adult molt of the silverleaf whitefly (Bemisia argentifolii): Timing, regulation, and progress. *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America,* 51**,** 67-79.

GIBBS, M. J., ARMSTRONG, J. S. & GIBBS, A. J. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics,* 16**,** 573-582.

GILL, R. 1990. The morphology of whiteflies. *Whiteflies: their bionomics, pest status and management*, 13-46.

GILL, R. J. & BROWN, J. K. 2009. Systematics of Bemisia and Bemisia relatives: can molecular techniques solve the Bemisia tabaci complex conundrum–a taxonomist's viewpoint. *Bemisia: bionomics and management of a global pest.* Springer.

GOLD, C. S., ALTIERI, M. A. & BELLOTTI, A. C. 1990. Direct and residual effects of short duration intercrops on the cassava whiteflies Aleurotrachelus socialis and Trialeurodes variabilis (Homoptera: Aleyrodidae) in Colombia. *Agriculture, Ecosystems & Environment,* 32, 57-67.

GOROVITS, R. & CZOSNEK, H. 2017. The involvement of heat shock proteins in the establishment of Tomato yellow leaf curl virus infection. *Frontiers in Plant Science,* 8.

HASEGAWA, D. K., CHEN, W., ZHENG, Y., KAUR, N., WINTERMANTEL, W. M., SIMMONS, A. M., FEI, Z. & LING, K.-S. 2018. Comparative transcriptome analysis reveals networks of genes activated in the whitefly, Bemisia tabaci when fed on tomato plants infected with Tomato yellow leaf curl virus. *Virology,* 513, 52-64.

HELED, J. & DRUMMOND, A. J. 2009. Bayesian inference of species trees from multilocus data. *Molecular biology and evolution,* 27, 570-580.

HILJE, L., COSTA, H. S. & STANSLY, P. A. 2001. Cultural practices for managing Bemisia tabaci and associated viral diseases. *Crop protection,* 20, 801-812.

HODDLE, M. 1999. The biology and management of silverleaf whitefly, Bemisia argentifolii Bellows and Perring (Homoptera: Aleyrodidae) on greenhouse grown ornamentals. *Department of Entomology, University of California, Riverside. USA. Applied Biological Control Research*.

HOELMER, K. 1996. Whitefly parasitoids: can they control field populations of Bemisia? *Bemisia: 1995, taxonomy, biology, damage, control and management*.

HOFFMANN, A. A., SØRENSEN, J. G. & LOESCHCKE, V. 2003. Adaptation of Drosophila to temperature extremes: bringing together quantitative and molecular approaches. *Journal of Thermal Biology,* 28**,** 175-216.

HOWELER, R. L., THOMAS, N., ARRIETA BOLAÑOS, S., OCHOA HENRÍQUEZ, J. M., ODIO ECHEVERRÍA, F., MORENO, V., SIMONETTI, L. L. A., DARDÓN, O., RAMIROMEJÍA, F. S. C. & PREISSING, X. 2013. *Save and grow: cassava. A guide to sustainable production intensificationProduire plus avec moins Ahorrar para crecer*, FAO, Roma (Italia).

HSIEH, C.-H., KO, C.-C., CHUNG, C.-H. & WANG, H.-Y. 2014. Multilocus approach to clarify species status and the divergence history of the Bemisia tabaci (Hemiptera: Aleyrodidae) species complex. *Molecular phylogenetics and evolution,* 76**,** 172-180.

HUANG, Y. W., HU, C. C., LIOU, M. R., CHANG, B. Y., TSAI, C. H., MENG, M., LIN, N. S. & HSU, Y. H. 2012. Hsp90 interacts specifically with viral RNA and differentially regulates replication initiation of Bamboo mosaic virus and associated satellite RNA. *PLoS pathogens,* 8**,** e1002726.

HWANG, B., LEE, J. H. & BANG, D. 2018. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & molecular medicine,* 50**,** 96.

ISLAMI, T., GURITNO, B. & UTOMO, W. H. 2011. Performance of cassava (Manihot esculenta Crantz) based cropping systems and associated soil quality changes in the degraded tropical uplands of East Java, Indonesia. *Journal of Tropical Agriculture,* 49**,** 31-39.

JIANG, R., QI, L.-D., DU, Y.-Z. & LI, Y.-X. 2017. Thermotolerance and Heat-Shock Protein Gene Expression Patterns in Bemisia tabaci (Hemiptera: Aleyrodidae)

Mediterranean in Relation to Developmental Stage. *Journal of economic entomology,* 110**,** 2190-2198.

KATOH, K., MISAWA, K., KUMA, K. I. & MIYATA, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research,* 30**,** 3059-3066.

KAUR, N., CHEN, W., ZHENG, Y., HASEGAWA, D. K., LING, K.-S., FEI, Z. & WINTERMANTEL, W. M. 2017. Transcriptome analysis of the whitefly, Bemisia tabaci MEAM1 during feeding on tomato infected with the crinivirus, Tomato chlorosis virus, identifies a temporal shift in gene expression and differential regulation of novel orphan genes. *BMC genomics,* 18**,** 370.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S. & DURAN, C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-1649.

KELLERMANN, V., OVERGAARD, J., HOFFMANN, A. A., FLØJGAARD, C., SVENNING, J.-C. & LOESCHCKE, V. 2012. Upper thermal limits of Drosophila are linked to species distributions and strongly constrained phylogenetically. *Proceedings of the National Academy of Sciences,* 109**,** 16228-16233.

KINENE, T., LUBOOBI, L. S., NANNYONGA, B. & MWANGA, G. G. 2015. A mathematical model for the dynamics and cost effectiveness of the current controls of cassava brown streak disease in Uganda. *Journal of Mathematical and Computational Science,* 5**,** 567.

KINENE, T., WAINAINA, J., MAINA, S. & BOYKIN, L. 2016. Rooting trees, methods for.

KINGMAN, J. F. C. 1982. The coalescent. *Stochastic processes and their applications,* 13**,** 235-248.

KUBATKO, L. S., CARSTENS, B. C. & KNOWLES, L. L. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics,* 25**,** 971-973.

KUBATKO, L. S. & DEGNAN, J. H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology,* 56**,** 17-24.

KUBATKO, L. S., GIBBS, H. L. & BLOOMQUIST, E. W. 2011. Inferring species-level phylogenies and taxonomic distinctiveness using multilocus data in Sistrurus rattlesnakes. *Systematic Biology,* 60**,** 393-409.

LEACHÉ, A. D. & FUJITA, M. K. 2010. Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings of the Royal Society of London B: Biological Sciences***,** rspb20100662.

LEGG, J. 1996. Host‐associated strains within Ugandan populations of the whitefly Bemisia tabaci (Genn.),(Hom., Aleyrodidae). *Journal of Applied Entomology,* 120**,** 523-527.

LEGG, J., GERLING, D. & NEUENSCHWANDER, P. 2003. Biological control of whiteflies in sub-Saharan Africa. *Biological control in IPM systems in Africa***,** 87-100.

LEGG, J., JEREMIAH, S., OBIERO, H., MARUTHI, M., NDYETABULA, I., OKAO-OKUJA, G., BOUWMEESTER, H., BIGIRIMANA, S., TATA-HANGY, W. & GASHAKA, G. 2011. Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Research,* 159**,** 161-170.

LEGG, J., OWOR, B., SSERUWAGI, P. & NDUNGURU, J. 2006. Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Advances in virus research,* 67**,** 355-418.

LEGG, J. P., SHIRIMA, R., TAJEBE, L. S., GUASTELLA, D., BONIFACE, S., JEREMIAH, S., NSAMI, E., CHIKOTI, P. & RAPISARDA, C. 2014a. Biology and management of Bemisia whitefly vectors of cassava virus pandemics in Africa. *Pest management science,* 70**,** 1446-1453.

LEGG, J. P., SSERUWAGI, P., BONIFACE, S., OKAO-OKUJA, G., SHIRIMA, R., BIGIRIMANA, S., GASHAKA, G., HERRMANN, H.-W., JEREMIAH, S. & OBIERO, H. 2014b. Spatio-temporal patterns of genetic change amongst populations of cassava Bemisia tabaci whiteflies driving virus pandemics in East and Central Africa. *Virus research,* 186**,** 61-75.

LIU, L. & PEARL, D. K. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic biology,* 56**,** 504-514.

LIU, L., YU, L. & EDWARDS, S. V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC evolutionary biology,* 10**,** 302.

LIU, L., YU, L., KUBATKO, L., PEARL, D. K. & EDWARDS, S. V. 2009. Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution,* 53**,** 320-328.

LIU, S.-S., COLVIN, J. & DE BARRO, P. J. 2012. Species concepts as applied to the whitefly Bemisia tabaci systematics: how many species are there? *Journal of Integrative Agriculture,* 11**,** 176-186.

LÜ, Z.-C. & WAN, F.-H. 2011. Using double-stranded RNA to explore the role of heat shock protein genes in heat tolerance in Bemisia tabaci (Gennadius). *Journal of Experimental Biology,* 214**,** 764-769.

MA, F.-Z., LÜ, Z.-C., WANG, R. & WAN, F.-H. 2014. Heritability and evolutionary potential in thermal tolerance traits in the invasive Mediterranean cryptic species of Bemisia tabaci (Hemiptera: Aleyrodidae). *PloS one,* 9**,** e103279.

MACFADYEN, S., PAULL, C., BOYKIN, L., DE BARRO, P., MARUTHI, M., OTIM, M., KALYEBI, A., VASSÃO, D., SSERUWAGI, P. & TAY, W. T. 2018. Cassava whitefly, Bemisia tabaci (Gennadius)(Hemiptera: Aleyrodidae) in East African farming landscapes: a review of the factors determining abundance. *Bulletin of entomological research***,** 1-18.

MADDISON, W. P. 1997. Gene trees in species trees. *Systematic biology,* 46**,** 523-536.

MADDISON, W. P. & KNOWLES, L. L. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic biology,* 55**,** 21-30.

MAHADAV, A., KONTSEDALOV, S., CZOSNEK, H. & GHANIM, M. 2009. Thermotolerance and gene expression following heat stress in the whitefly Bemisia tabaci B and Q biotypes. *Insect biochemistry and molecular biology,* 39**,** 668-676.

MARTIN, D., POSADA, D., CRANDALL, K. & WILLIAMSON, C. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research & Human Retroviruses,* 21**,** 98-102.

MARTIN, D. & RYBICKI, E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics,* 16**,** 562-563.

MARTIN, D. P., MURRELL, B., GOLDEN, M., KHOOSAL, A. & MUHIRE, B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus evolution,* 1.

MARTIN, J. H. 2004. Whiteflies of Belize (Hemiptera: Aleyrodidae). Part 1-introduction and account of the subfamily Aleurodicinae Quaintance & Baker. Moscas blancas de Belice (Hemiptera: Aleyrodidae). Parte 1-introducción y descripción de la subfamilia Aleurodicinae Quaintance & Baker. *Zootaxa.*, 1-119.

MARUTHI, M., COLVIN, J., SEAL, S., GIBSON, G. & COOPER, J. 2002. Co-adaptation between cassava mosaic geminiviruses and their local vector populations. *Virus Research,* 86**,** 71-85.

MARUTHI, M., COLVIN, J., THWAITES, R. M., BANKS, G. K., GIBSON, G. & SEAL, S. E. 2004. Reproductive incompatibility and cytochrome oxidase I gene sequence variability amongst host‐adapted and geographically separate Bemisia tabaci populations (Hemiptera: Aleyrodidae). *Systematic Entomology,* 29**,** 560-568.

MARUTHI, M., HILLOCKS, R., MTUNDA, K., RAYA, M., MUHANNA, M., KIOZIA, H., REKHA, A., COLVIN, J. & THRESH, J. 2005. Transmission of Cassava brown streak virus by Bemisia tabaci (Gennadius). *Journal of Phytopathology,* 153**,** 307-312.

MASTERS, B. C., FAN, V. & ROSS, H. A. 2011. Species delimitation–a geneious plugin for the exploration of species boundaries. *Molecular Ecology Resources,* 11**,** 154-157.

MCAUSLANE, H. J. 2002. *Sweetpotato Whitefly B Biotype of Silverleaf Whitefly, Bemisia tabaci (Gennadius) or Bemisia argentifolii Bellows & Perring (Insecta: Homoptera: Aleyrodidae)*, University of Florida Cooperative Extension Service, Institute of Food and ….

MCKAY, B. D., MAYS JR, H. L., WU, Y., LI, H., YAO, C. T., NISHIUMI, I. & ZOU, F. 2013. An empirical comparison of character‐based and coalescent‐based

approaches to species delimitation in a young avian complex. *Molecular Ecology,* 22**,** 4943-4957.

MILENOVIC, M., WOSULA, E., RAPISARDA, C. & LEGG, J. P. 2019. Impact of host plant species and whitefly species on feeding behavior of Bemisia tabaci. *Frontiers in Plant Science,* 10**,** 1.

MIRARAB, S., REAZ, R., BAYZID, M. S., ZIMMERMANN, T., SWENSON, M. S. & WARNOW, T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics,* 30**,** i541-i548.

MUGERWA, H., REY, M. E., ALICAI, T., ATEKA, E., ATUNCHA, H., NDUNGURU, J. & SSERUWAGI, P. 2012. Genetic diversity and geographic distribution of Bemisia tabaci (Gennadius)(Hemiptera: Aleyrodidae) genotypes associated with cassava in East Africa. *Ecology and evolution,* 2**,** 2749-2762.

MUGERWA, H., SEAL, S., WANG, H.-L., PATEL, M. V., KABAALU, R., OMONGO, C. A., ALICAI, T., TAIRO, F., NDUNGURU, J. & SSERUWAGI, P. 2018. African ancestry of New World, Bemisia tabaci-whitefly species. *Scientific reports,* 8**,** 2734.

NDUNGURU, J., SSERUWAGI, P., TAIRO, F., STOMEO, F., MAINA, S., DJINKENG, A., KEHOE, M. & BOYKIN, L. M. 2015. 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,* 10**,** e0139321.

OLSEN, K. M. & SCHAAL, B. A. 1999. Evidence on the origin of cassava: phylogeography of Manihot esculenta. *Proceedings of the National Academy of Sciences,* 96**,** 5586-5591.

OMONGO, C. A., KAWUKI, R., BELLOTTI, A. C., ALICAI, T., BAGUMA, Y., MARUTHI, M., BUA, A. & COLVIN, J. 2012. African cassava whitefly,

Bemisia tabaci, resistance in African and South American cassava genotypes. *Journal of integrative agriculture,* 11**,** 327-336.

PADIDAM, M., SAWYER, S. & FAUQUET, C. M. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology,* 265**,** 218-225.

PARSELL, D. & LINDQUIST, S. 1993. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annual review of genetics,* 27**,** 437-496.

PATIL, B. L., LEGG, J. P., KANJU, E. & FAUQUET, C. M. 2015. Cassava brown streak disease: a threat to food security in Africa. *Journal of General Virology,* 96**,** 956-968.

PEREIRA, L. 2017. Climate change impacts on agriculture across Africa. *Oxford Research Encyclopedia of Environmental Science*.

POLSTON, J. E. & CAPOBIANCO, H. 2013. Transmitting plant viruses using whiteflies. *Journal of visualized experiments: JoVE*.

POLSTON, J. E., DE BARRO, P. & BOYKIN, L. M. 2014. Transmission specificities of plant viruses with the newly identified species of the Bemisia tabaci species complex. *Pest Manag Sci,* 70**,** 1547-52.

POSADA, D. & CRANDALL, K. A. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences,* 98**,** 13757-13762.

POSADA, D. & CRANDALL, K. A. 2002. The effect of recombination on the accuracy of phylogeny estimation. *Journal of molecular evolution,* 54**,** 396-402.

PRICE, J., SCHUSTER, D. & SHORT, D. 1987. Managing sweetpotato whitefly. *Greenhouse Grower,* 35**,** 55-57.

PUSAG, J. C. A., JAHAN, S. H., LEE, K.-S., LEE, S. & LEE, K.-Y. 2012. Upregulation of temperature susceptibility in Bemisia tabaci upon acquisition of

Tomato yellow leaf curl virus (TYLCV). *Journal of insect physiology,* 58**,** 1343-1348.

RAMBAUT, A., SUCHARD, M. A., XIE, D. & DRUMMOND, A. J. 2015. Tracer v1. 6. 2014.

RANNALA, B. & YANG, Z. 2013. Improved reversible jump algorithms for Bayesian species delimitation. *Genetics***,** genetics. 112.149039.

RODRIGO, A., BERTELS, F., HELED, J., NODER, R., SHEARMAN, H. & TSAI, P. 2008. The perils of plenty: what are we going to do with all these genes? *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 363**,** 3893-3902.

ROKAS, A., KING, N., FINNERTY, J. & CARROLL, S. B. 2003. Conflicting phylogenetic signals at the base of the metazoan tree. *Evolution & development,* 5**,** 346-359.

RONQUIST, F. & HUELSENBECK, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19**,** 1572-1574.

RONQUIST, F., TESLENKO, M., VAN DER MARK, P., AYRES, D. L., DARLING, A., HÖHNA, S., LARGET, B., LIU, L., SUCHARD, M. A. & HUELSENBECK, J. P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology,* 61**,** 539-542.

ROSENBERG, N. A. 2007. Statistical tests for taxonomic distinctiveness from observations of monophyly. *Evolution,* 61**,** 317-323.

ROY, A., KUCUKURAL, A. & ZHANG, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols,* 5**,** 725.

SALVUCCI, M. E., STECHER, D. S. & HENNEBERRY, T. J. 2000. Heat shock proteins in whiteflies, an insect that accumulates sorbitol in response to heat stress. *Journal of Thermal Biology,* 25**,** 363-371.

SANGER, F. & COULSON, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology,* 94**,** 441-448.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences,* 74**,** 5463-5467.

SCHIERUP, M. H. & HEIN, J. 2000. Consequences of recombination on traditional phylogenetic analysis. *Genetics,* 156**,** 879-891.

SCORNAVACCA, C. & GALTIER, N. 2017. Incomplete lineage sorting in mammalian phylogenomics. *Systematic biology,* 66**,** 112-120.

SHARAF, N. & BATTA, Y. 1985. Effect of some factors on the relationship between the whitefly Bemisia tabaci Genn.(Homopt., Aleyrodidea) and the parasitoid Eretmocerus mundus Mercet (Hymenopt., Aphelinidae) 1. *Zeitschrift für Angewandte Entomologie,* 99**,** 267-276.

SHATTERS JR, R. G., POWELL, C. A., BOYKIN, L. M., LIANSHENG, H. & MCKENZIE, C. 2009. Improved DNA barcoding method for Bemisia tabaci and related Aleyrodidae: development of universal and Bemisia tabaci biotype-specific mitochondrial cytochrome c oxidase I polymerase chain reaction primers. *Journal of Economic Entomology,* 102**,** 750-758.

SHAW, K. L. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences,* 99**,** 16122-16127.

SIMON, C., FRATI, F., BECKENBACH, A., CRESPI, B., LIU, H. & FLOOK, P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the entomological Society of America,* 87**,** 651-701.

SMITH, J. M. 1992. Analyzing the mosaic structure of genes. *Journal of molecular evolution,* 34**,** 126-129.

SOMERO, G. N. 1995. Proteins and temperature. *Annual review of physiology,* 57**,** 43-68.

SSERUWAGI, P., LEGG, J., MARUTHI, M., COLVIN, J., REY, M. & BROWN, J. K. 2005. Genetic diversity of Bemisia tabaci (Gennadius)(Hemiptera: Aleyrodidae) populations and presence of the B biotype and a non‐B biotype that can induce silverleaf symptoms in squash, in Uganda. *Annals of Applied Biology,* 147**,** 253-265.

SSERUWAGI, P., MARUTHI, M., COLVIN, J., REY, M., BROWN, J. K. & LEGG, J. 2006. Colonization of non‐cassava plant species by cassava whiteflies (Bemisia tabaci) in Uganda. *Entomologia experimentalis et applicata,* 119**,** 145-153.

SSERUWAGI, P., SSERUBOMBWE, W., LEGG, J., NDUNGURU, J. & THRESH, J. 2004. Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: a review. *Virus research,* 100**,** 129-142.

SSERUWAGI, P., WAINAINA, J., NDUNGURU, J., TUMUHIMBISE, R., TAIRO, F., GUO, J.-Y., VRIELINK, A., BLYTHE, A., KINENE, T. & DE MARCHI, B. 2017. The first transcriptomes from field-collected individual whiteflies (Bemisia tabaci, Hemiptera: Aleyrodidae). *Gates Open Research,* 1.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-1313.

STANLEY, J. & GAY, M. R. 1983. Nucleotide sequence of cassava latent virus DNA.

SUN, D.-B., LIU, Y.-Q., QIN, L., XU, J., LI, F.-F. & LIU, S.-S. 2013. Competitive displacement between two invasive whiteflies: insecticide application and host plant effects. *Bulletin of entomological research,* 103**,** 344-353.

TAJUBA, P. 2016. *Gulu, Nebbi record hottest temperatures* [Online]. Daily Monitor. Available: http://www.monitor.co.ug/News/National/Gulu--Nebbi-record-hottest-temperatures/688334-3137104-x401ub/index.html [Accessed 18 December 2017].

THRESH, J. & COOTER, R. 2005. Strategies for controlling cassava mosaic virus disease in Africa. *Plant pathology,* 54**,** 587-614.

TOCKO-MARABENA, B. K., SILLA, S., SIMIAND, C., ZINGA, I., LEGG, J., REYNAUD, B. & DELATTE, H. 2017. Genetic diversity of Bemisia tabaci species colonizing cassava in Central African Republic characterized by analysis of cytochrome c oxidase subunit I. *PloS one,* 12**,** e0182749.

TODD, E. V., BLACK, M. A. & GEMMELL, N. J. 2016. The power and promise of RNA‐seq in ecology and evolution. *Molecular ecology,* 25**,** 1224-1241.

TRACY, C. 2017. *A climate in crisis* [Online]. Oxfam. Available: https://www.oxfam.org/en/research/climate-crisis [Accessed 18 December 2017].

TSUEDA, H. & TSUCHIDA, K. 2011. Reproductive differences between Q and B whiteflies, Bemisia tabaci, on three host plants and negative interactions in mixed cohorts. *Entomologia Experimentalis et Applicata,* 141**,** 197-207.

WANG, P., RUAN, Y. M. & LIU, S. S. 2010. Crossing experiments and behavioral observations reveal reproductive incompatibility among three putative species of the whitefly Bemisia tabaci. *Insect Science,* 17**,** 508-516.

WANG, P., SUN, D. B., QIU, B. L. & LIU, S. S. 2011. The presence of six cryptic species of the whitefly Bemisia tabaci complex in China as revealed by crossing experiments. *Insect Science,* 18**,** 67-77.

WASHBURN, J. D., SCHNABLE, J. C., CONANT, G. C., BRUTNELL, T. P., SHAO, Y., ZHANG, Y., LUDWIG, M., DAVIDSE, G. & PIRES, J. C. 2017. Genome-guided phylo-transcriptomic methods and the nuclear phylogentic tree of the Paniceae grasses. *Scientific Reports,* 7**,** 13528.

WIENS, J. J., KUCZYNSKI, C. A. & STEPHENS, P. R. 2010. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. *Biological Journal of the Linnean Society,* 99**,** 445-461.

WINTER, S., KOERBLER, M., STEIN, B., PIETRUSZKA, A., PAAPE, M. & BUTGEREITT, A. 2010. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. *Journal of General Virology,* 91**,** 1365-1372.

WOSULA, E. N., CHEN, W., FEI, Z. & LEGG, J. P. 2017. Unravelling the genetic diversity among cassava Bemisia tabaci whiteflies using NextRAD sequencing. *Genome biology and evolution,* 9**,** 2958-2973.

XIA, W.-Q., WANG, X.-R., LIANG, Y., LIU, S.-S. & WANG, X.-W. 2017. Transcriptome analyses suggest a novel hypothesis for whitefly adaptation to tobacco. *Scientific reports,* 7**,** 12102.

XIAO, N., PAN, L.-L., ZHANG, C.-R., SHAN, H.-W. & LIU, S.-S. 2016. Differential tolerance capacity to unfavourable low and high temperatures between two invasive whiteflies. *Scientific reports,* 6**,** 24306.

XU, J., DE BARRO, P. & LIU, S. 2010. Reproductive incompatibility among genetic groups of Bemisia tabaci supports the proposition that the whitefly is a cryptic species complex. *Bulletin of entomological research,* 100**,** 359-366.

YANG, J., YAN, R., ROY, A., XU, D., POISSON, J. & ZHANG, Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nature methods,* 12**,** 7.

YANG, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences: CABIOS,* 13**,** 555-556.

YANG, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution,* 24**,** 1586-1591.

YANG, Z. 2015. The BPP program for species tree estimation and species delimitation. *Current Zoology,* 61**,** 854-865.

YANG, Z. & RANNALA, B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences***,** 200913022.

YU, H. & WAN, F. H. 2009. Cloning and expression of heat shock protein genes in two whitefly species in response to thermal stress. *Journal of Applied Entomology,* 133**,** 602-614.

ZHANG, C., RABIEE, M., SAYYARI, E. & MIRARAB, S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics,* 19**,** 153.

ZHANG, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics,* 9**,** 40.

Appendix

A 1. Varying rates of evolution in the Heat Shock Protein 90 (HSP90) gene of the *Bemisia tabaci* species complex

This paper was withdrawn from BMC Evolutionary Biology after detection of recombination events in the dataset. Therefore, the focus was changed to recombination analysis in Chapter 3.

Varying rates of evolution in the Heat Shock Protein 90 (HSP90) gene of the *Bemisia tabaci* species complex

Tonny Kinene*[1], Laura Kubatko[2], Livingstone S. Luboobi[3], James Mbora Wainaina[1], Bruno Rossito De Marchi[4], Ian Small[1], Titus Alicai[5], Christopher Abu Omongo[5], Anders Savill[1, 6] and Laura M. Boykin[1].

Addresses

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

[2]The Ohio State University, 154W 12th Avenue, Columbus, Ohio 43210, USA

[3]Department of Mathematics, Makerere University, P.O. Box 7062, Kampala, Uganda

[4]São Paulo State University (UNESP), School of Agriculture, Dept. of Plant Protection, CEP 18610-307, Botucatu (SP), Brazil

[5]National Crops Resources Research Institute, P.O. Box 7084, Kampala, Uganda

[6]Pawsey Supercomputing Centre Technology Park, 26 Dick Perry Ave Kensington 6151, Australia

*Author for correspondence: Tonny Kinene, Email: kinenetonny@gmail.com; tonny.kinene@research.uwa.edu.au

Authors' email addresses:

Laura M. Boykin: laura.boykin@uwa.edu.au
Laura Kubatko: lkubatko@stat.osu.edu
Livingstone Serwadda Luboobi: luboobi@gmail.com
Bruno Rossito De Marchi: bruno_dmarchi@hotmail.com
James Mbora Wainaina: james.wainaina@research.uwa.edu.au
Ian Small: ian.small@uwa.edu.au
Christopher Abu Omongo: chrisomongo@yahoo.com
Anders Savill: anders.savill@csiro.au
Titus Alicai: talicai@hotmail.com

Abstract

Background

Bemisia tabaci species (whiteflies) are a global pest causing billions of dollars in damage each year, leaving farmers with low yields. In East Africa, whiteflies are superabundant and present on cassava plants throughout the year. Whiteflies do not decrease in number in the hot dry seasons in East Africa, therefore, it has been suggested that the synthesis of Heat Shock Protein (HSP) may protect the whitefly from heat stress and other biotic factors. In this study we used codon models to investigate conservation of the HSP90 gene in members of the *B. tabaci* species complex.

Results

The branch-site model and site models used in this study to test for sites under adaptive evolution revealed that the HSP90 gene was dominated by purifying selection. Therefore, the majority of the point mutations were not altering the function of the protein, as is typical of conserved genes. However, in the Sub-Saharan Africa (SSA) lineages we identified four sites that were under positive selection. These could be advantageous mutations that became fixed in the population faster than those mutations with minor advantage. This might be the reason for the increased chance of survival and reproductive capabilities of whiteflies in hot weather conditions in East Africa. Thus, there is variable selective pressure on the HSP90 gene

Conclusion

This is the first study on the evolution of the HSP90 gene in SSA whitefly species. The results can be used as the basis for future gene expression studies on SSA species and the response of whiteflies to hot weather conditions in East Africa.

Keywords: Evolution, HSP90 gene, Climate change, *Bemisia tabaci.* Food security

BACKGROUND

The *Bemisia tabaci* species complex (whiteflies) is a group of small phloem sap-feeding insects capable of causing extensive damage globally (Brown et al., 1995a, Hoddle, 1999). The species complex is composed of at least 34 species that are morphologically indistinct (Boykin et al., 2012, Boykin et al., 2013). These species cause extensive damage to plants through direct feeding by both juvenile and adult stages of the whitefly (Boykin, 2014, Boykin and De Barro, 2014, Liu et al., 2012). They also cause indirect damage through excretion of honeydew, which covers the leaf and fruit surfaces promoting the growth of black sooty mold. This mold interferes with plant photosynthesis, affects plant growth, and leads to poor quality fruits (Omongo et al., 2012). Whiteflies also transmit plant viruses in the process of feeding from one plant to another (Brown and Czosnek, 2002). Yield losses due to direct feeding and black sooty mold are estimated at 40% (Omongo et al., 2012). In Africa, whiteflies are a threat to food security (Patil et al., 2015).

Many smallholder farmers in East Africa rely heavily on cassava as a food security crop and cash crop because of its tolerance to low fertility soils, low rainfall, low labour intensity and long harvesting window (Alicai et al., 2007, Colvin et al., 2004, Legg et al., 2014b). Cassava roots are eaten and are rich in carbohydrates, and the leaves are rich in protein and are eaten as vegetables. Cassava can be processed into other commercial products such as flour, cakes and alcohol. Unfortunately, cassava is highly vulnerable to both the whiteflies and the viruses they transmit. Attempts to protect cassava from whiteflies with chemical pesticides have been ineffective and costly (Cahill et al., 1995, Howeler et al., 2013, Kinene et al., 2015). The honeydew excreted

by superabundant populations of whiteflies in East Africa interferes with photosynthesis and reduces the effectiveness of insecticides (Boykin, 2014, Brown et al., 1995a, Liu et al., 2012, Omongo et al., 2012). The whiteflies transmit two viral diseases: cassava brown streak disease (CBSD) and cassava mosaic disease (CMD) (Alicai et al., 2016, Colvin et al., 2004, Maruthi et al., 2005, Sseruwagi et al., 2004, Ateka et al., 2017).

Agriculture development in Africa has been constrained by rising temperatures due to global warming. Impacts are economically significant in countries in East Africa that are heavily reliant on natural resources, such as rainwater, for agriculture. Increasing temperature has resulted in prolonged drought seasons, which have left most plants dry. However, the drought tolerant plants like cassava, which save farmers from famine, are now heavily infested with whiteflies. This has caused loss of food and income for farmers, due to the viral diseases associated with whiteflies that infect their crops. The high numbers of whiteflies seen in cassava fields in East Africa, termed as superabundant populations, have increased with increasing temperatures over time which may be due to the fact that temperature determines the geographical distribution and reproductive abundance of species (Cossins and Bowler, 1987, Hoffmann et al., 2003, Kellermann et al., 2012).

The capacity of Sub-Saharan Africa (SSA) whitefly species to breed in large numbers in the increased temperatures in East Africa indicates that SSA species might have developed a molecular response to tolerate heat stress. This is because high temperatures in the environment can cause organisms to respond with physiological, biochemical or behavioural traits that enhance their chances of survival (Fasolo and Krebs, 2004, Hoffmann et al., 2003). During stressful conditions the cells of an organism produce a set of conserved proteins called heat shock proteins (HSPs). HSPs

are molecular chaperones that bind to and stabilize unfolded proteins and are found in nearly all living organisms (Parsell and Lindquist, 1993, Somero, 1995). HSPs are named according to their molecular weight (for example, HSP100, 90, 70, 60, and 40) and not all are associated with the heat shock response (Parsell and Lindquist, 1993). HSP70 and HSP90 are the most widely-studied families and are synthesized when the environmental temperature exceeds the optimal temperature of an organism (Gorovits and Czosnek, 2017, Parsell and Lindquist, 1993).

Gorovits & Czosnek (2017) revealed that HSP70 and HSP90 in plants and vectors are very important for efficient virus infection. HSP90 is also considered to promote virus replication by interacting with virus replicase in *Bambo mosaic* virus (Huang et al., 2012). Differential expression studies on HSP on Middle East Asia Minor 1 (MEAM1), one of the whitefly species in the complex, indicated that heat shock proteins (HSP20, 70 & 90) were up-regulated when whiteflies were subjected to a high temperature of 40 ^{0}C, and this increased the fitness of whiteflies following heat stress (Diaz et al., 2015). Other studies indicate that up-regulation of HSP could be key in the determination of the natural geographical distribution of natural populations by seasonal dynamics (Yu and Wan, 2009). HSP response was examined on silver leaf whitefly and it was found that HSP70 and HSP90 were the major polypeptides synthesized by whiteflies in response to heat stress (Salvucci et al., 2000). There is more evidence that HSPs are up-regulated by stress conditions in *B. tabaci* (Cui et al., 2008, Lü and Wan, 2011, Mahadav et al., 2009). However, none of these studies have been carried out on SSA whitefly species.

We investigated whether there were variable selective pressures on the HSP90 gene of the various *B. tabaci* species. Focusing on SSA species, we examined the rates of

evolution of the HSP90 gene to determine whether SSA species are evolving rapidly compared to other species.

RESULTS

A total of 21 samples were sequenced, these included several genetic groups of *B. tabaci*, i.e. 8 SSA1, 5 SSA2, 2 Australia I (AUSI), New world Africa (NWAfrica), B. *after*, Uganda, Mediterranean (MED), and MEAM1. 15 of these samples were sequenced on an Illumina Hiseq 2500 on a rapid run mode, this resulted into 8,000,000 raw reads and 6 samples were sequenced on an Illumina Hiseq 4000 on a high through put mode which yielded raw reads ranging from 43,194,264 to 37, 306,246. After trimming for quality, the reads reduced to a range of 7,876,283 to 7,816,022 for the first 15 samples and 42,953,727 to 37,071,584 for the 6 samples (additional file 2). *De novo* assembly produced contigs ranging from 2530 – 30716, which were then mapped to EU934241 reference sequence. The final coding sequence of length of 2160 base pairs was obtained and used in this study as the HSP90 sequence. It consisted of a consensus between the *de novo* contig of interest and the mapped consensus.

Site models and branch model were applied to detect adaptive evolution of the HSP90 using the codeml package implemented in Phylogenetic Analysis by Maximum Likelihood (PAML) software (Yang, 2007). Table 1 shows parameter estimates under the site models

Table 1 Parameter estimates in site models

Model	ℓ	**Estimate of parameters**	**Positively selected sites**	**P- Value**
M0	-6861.31	$\widehat{\omega} = 0.0377$	None	-
M1a	-6755.30	$\hat{p}_0 = 0.9705$, $\hat{p}_1 = 0.0294$ $\widehat{\omega}_0 = 0.0120$, $\omega_1=1$	Not allowed	9.12×10^{-47}
M2a	-6748.64	$\hat{p}_0 = 0.9717$, $\hat{p}_1 = 0.0235$ $\hat{p}_2 = 0.0046$ $\widehat{\omega}_0 = 0.0128$, $\omega_1 = 1$	166 K* 198 D 333 R**	1.28×10^{-03}

		$\widehat{\omega}_2 = 3.7592$	693 M*	
M7	-6770.85	$\hat{p} = 0.0321, \hat{q} = 0.46508$	Not allowed	-
M8	-6750.61	$\hat{p}_0 = 0.9923, \hat{p}_1 = 0.0076$ $\hat{p} = 0.0352, \hat{q} = 0.8744$ $\widehat{\omega}_s = 2.8664$	166 K 198 D 333 R* 693 M*	1.62 x10^{-09}

ℓ is the log likelihood, * and ** represent positively selected sites inferred at a cut off of posterior probability $P \geq 95\%$ and $P \geq 99\%$ respectively. The species represented by each of these residues are indicated in additional file 3

In a statistical test for sites under adaptive evolution using site models, model M0 indicates that sites in HSP90 gene are under strong purifying selection with $\widehat{\omega} = 0.0377$. The LRT comparing M1a against M2a was statistically significant thus, 0.5% of the sites in the HSP90 gene were under positive selection with $\widehat{\omega} = 3.8$. Using the Bayes empirical Bayes analysis (BEB) approach implemented in M2a, we identified four codon sites under positive selection: 166, 198, 333, 693 with amino acids lysine (K), aspartate (D), arginine (R), methionine (M), respectively (Figure 2 and Figure 3). Site 166 K, 333 R and 693 M had posterior probabilities greater than 95% compared to site 198 D with posterior probability of 75%. The LRT comparing model M7 against M8 was also highly significant and confirmed the same amino acid sites under positive selection in M2a (Table2). The inferred amino acid residues under positively selection were also mapped to the 3D protein structure predicted on I-TASSER online sever (Figure 3). Pin pointing the positively selected amino acid sites might provide insights for further scientific investigations on HSP90.

In the branch model, we divided the HSP90 gene tree (Figure 1) into 4 clades i.e. A, B, C, D, based on the assumption of different geographical locations and the viruses they transmit. Species in clade A and B are from the East Africa and are efficient vectors for *Begomoviruses* and *Ipomoviruses* viruses on cassava and common beans, those in clade

C and D are from Australia, Asia and South America where the environmental conditions are different from those in E. Africa. We tested the null hypothesis that all clades of the HSP90 gene tree have equal rates of molecular evolution against the alternative hypothesis that each clade had a unique rate of molecular evolution (Table 2). All hypotheses and analyses in this study were performed on the tree estimated by MrBayes 3.2.2 (Figure 1).

The null hypothesis H_0 (the one-ratio model) assumes uniform selective pressure over all branches of the phylogenetic tree (Figure 1). The estimate $\widehat{\omega}A = 0.0377$ indicates that the evolution of HSP90 is dominated by strong purifying selection (Table 2). The aim of H_1 is to investigate the assumption of the same selective pressure for clade A, clade B and the clade containing C and D. The LRT comparing H_0 against H_1 is statically significant and suggests that the selective pressure in clades A, B and the clade containing C and D differed significantly. The parameter estimate ω indicated an increase in the rate of nonsynonymous substitutions in clade A, and strong purifying selection in the clade B followed by clade (C, D), see Table 2.

Our findings indicate that there was variable selective pressure on the HSP90 gene and purifying selection was dominating the clades in the tree.

Table 2: Parameter estimates under models of variable selective pressure among branches. Condon frequencies are calculated using the base frequency at the three-codon positions under the FMutSel model.

Models	**ωA**	**ωB**	**ωC**	**ωD**	**kappa (ts/tv)**	**Log likelihood (ℓ)**	**LRT ($2\Delta\ell$)**	***P*-value**
H_0	0.0377	=ωA	=ωA	=ωA	2.47934	-6861.31	-	-

H_1	0.0475	0.0001	0.0381	=ωC	2.47975	-6855.14	12.34	0.00044*

* Indicates significant P- value ($P < 0.05$) and also significance at the 5% level after the Bonferroni correction.

DISCUSSION

Varying rates of evolution were detected for members of the *Bemisia tabaci* species complex in the HSP90 gene. The site models and branch model used in this study to test for sites evolving under positive selection revealed that the evolution of the HSP90 gene was dominated by purifying selection (Figure 2), therefore the majority of the point mutations were not altering the function of the protein. While this is typical of conserved genes, we identified four statistically significant codon sites that were evolving under positive selection (Table 1). Thus, these could be advantageous mutations that have become fixed in the population faster than mutations with minor advantage (Desai and Fisher, 2007). This could explain the phenomenon for the increased chance of survival and reproductive abundance of whiteflies in hot conditions in East Africa.

Experimental studies on the members of the *B. tabaci* species complex have demonstrated that HSP plays an important role in thermal adaption, there is convincing evidence based on HSP90 gene indicating that MED species are adapted to a wider range of temperature than MEAM1(Jiang et al., 2017, Xiao et al., 2016, Díaz et al., 2014, Ma et al., 2014). We hypothesis that this the cause of the long branch that we have in Figure 1 with the recent MED sample collected from the tropical region of Brazil. More reports show that the cryptic MEAM1 species are being replaced by MED cryptic species in China (Xiao et al., 2016, Chu et al., 2010). We think that MEAM1 species have not yet adaptively evolved to high temperature changes, hence the low evolutionary rate (Figure 1). However, thermal adaptation of whiteflies may depend on many other factors such us,

insecticide resistance, host plants and the viruses (Sun et al., 2013). The invasive MED cryptic species are thought to have originated from North Africa and they could have a connection with the SSA species (Boykin et al., 2007).

In the branch model we observed that the SSA species were evolving at two different rates (Table 2), this could be explained by a number of factors, i.e. whitefly host plants, the viruses carried by whiteflies in E. Africa region which are different from those in other geographic regions, and endosymbiont composition of the whiteflies (Tsueda and Tsuchida, 2011, Brumin et al., 2011, Pusag et al., 2012). Further research is needed to evaluate these interesting patterns as non-has been carried out on SSA species.

We hypothesize that prolonged dry spells due to global warming have led to adaptive evolution of SSA species of whiteflies in East Africa. Table 2 shows that one of the conversed gene (HSP90) expressed under heat stress evolving at different rates for *B. tabaci* SSA species. This could be due to changes in climate over time. For example, in Uganda in 2016, the temperature went from the normal 28 ^{0}C to 32.8^{0}C in Kampala, 35.5^{0}C Wakiso, 37.5^{0}C Gulu, 38.1^{0}C Nebbi, 37.2^{0}C Tororo, 36.7^{0}C Kasese, 36.5^{0}C Soroti, 36.2^{0}C Serere, 36.0^{0}C Lira, 34.2^{0}C Jinja and 35.5^{0}C in Mabrara district (Tajuba, 2016). According to Oxfam International, there is evidence that climate change is contributing to drought in East Africa (Tracy, 2017). Temperatures across East Africa are rising and have been higher in recent years compared to the historic average over the period of 1940 to 1981 (Tracy, 2017).

Despite the increase in temperature, whitefly numbers have not reduced and we therefore hypothesize that whiteflies have adapted to these hotter conditions. The optimal temperatures for whitefly development ranges now from 28.0^{0}C to 32.0^{0}C, which are the current temperatures in the East African region (Bonato et al., 2007, Curnutte et al.,

2014). The chaperon function of HSP90 is a factor in the survival of whiteflies, as it is triggered by temperature changes (Hoffmann et al., 2003, Diaz et al., 2015). While this is not a conclusive study on climate change in the region, the findings help us hypothesise why there are high numbers of whiteflies in the cassava fields in East Africa, thus forming a basis for future studies. Therefore, we recommend that gene expression studies be carried out on SSA species in East Africa.

CONCLUSION

Temperature is a key determinant is the geographical distribution and survival of species. Theoretical and experimental results from other studies confirm that HSP90 plays an important role in thermal adaptation but not of these has been conducted on SSA species. Therefore, in this study we observed variable selective pressure on HSP90 gene that could have enabled *B. tabaci* species to adapt to regions with varying environmental temperatures.

METHODS

Whitefly sampling

Seven whitefly samples were collected from whitefly "hot spots" in Uganda in a country-wide survey conducted in 2013 from Nakaseke, Luwero and Nakasongola districts of Uganda. Adult whiteflies from symptomatic cassava were collected from the top five leaves using an aspirator, transferred immediately into 70% alcohol in an Eppendorf tube, and then later exported to the laboratory at the University of Western Australia (UWA) where they were kept at -20^0 C before analysis. Five whiteflies from Kenya were collected from symptomatic common beans with an aspirator. Collaborators from Brazil, Australia and Malawi also sent whitefly samples already preserved in 70% ethanol and they were stored at – 20^0 C upon arrival at UWA. See additional file 1.

Total RNA extraction from individual whiteflies

RNA extractions were performed in the UWA genomics laboratory using an ARCTURUS PicoPure RNA Isolation Kit (Arcturus, CA, USA) following extraction procedures from (Sseruwagi et al., 2017). To remove contaminating DNA, DNase and divalent cations from the extracted RNA, we used the TURBO DNA free kit as described by the manufacturer (Ambion Life Technologies CA, USA). To increase the concentration of the RNA, we used a vacuum centrifuge (Eppendorf, Germany) set at room temperature for one hour then resuspended the pellet in 18 µl of RNase free water and stored the prepared RNA immediately at -80^0 C awaiting further analysis. RNA integrity was quantified by 2100 Bio-analyser (Aligent Technologies).

cDNA and Illumina library preparation

The Illumina TruSeq Stranded total RNA preparation kit was used to make cDNA libraries from the RNA of each individual whitefly as described by the manufacturer (Illumina, San Diego, CA, USA). Libraries were sent to Macrogen, Korea (www.macrogen.com), where 15 samples were sequenced on a HiSeq 2500 on a rapid run mode and 6 samples on HiSeq 4000 on a high-throughput mode, see additional file 2. Sequencing control software HCS v2.2 and HCS v3.3 were used for base calling and quality assessment respectively.

NGS data analysis (De novo sequence assembly and mapping)

All samples were trimmed using CLC Genomic workbench 8.5.1 (CLCGW) with the quality score limit set to 0.01, and the ambiguous limit set to 2. Trimmed reads were assembled into contigs using de novo assembly function of CLCGW with automatic word size, automatic bubble size, minimum contig length 500, mismatch cost 2,

insertion cost 3, deletion cost 3, length fraction 0.5 and similarity fraction 0.9 (additional file 2). Contigs were imported in Geneious 9.1.8 (Kearse et al., 2012) on Mac OS 10.6, and contigs were then mapped onto the reference sequence for HSP90 MEAM1 EU934241 from GenBank. Mapping was performed with the following settings in Geneious: minimum overlap 10%, minimum overlap identity 80%, allow gaps 10%, fine tuning set to iterate up to 10 times at custom sensitivity. The consensus contig from mapping was aligned to the *de novo* contig using MAFFT (Katoh et al., 2002). The resulting alignment consensus was manually inspected for ambiguities and gave rise to the new HSP90 sequences covering the full coding region of length 2160 base pairs. Open reading frames (ORFs) were predicted in Geneious and other HSP90 sequences used in this study were downloaded from Genbank (i.e., HM013713, HM367080, HM01370, DQ093380 and DQ093381). Alignment of the new HSP90 sequences with Genbank sequences was performed in Geneious using the MAFFT plugin (Katoh et al., 2002). A total of 27 sequences of 2160 base pairs were aligned.

Phylogenetic analysis

We ran jModelTest 2 (Darriba et al., 2012) on the final data set to select statistically the best-fit model of nucleotide substitution. The best model was GTR+I+G and this model was then used in a Bayesian phylogenetic analysis. We ran MrBayes 3.2.2 (Ronquist et al., 2012) in parallel on Magnus (Pawsey supercomputing Centre, Perth, Western Australia) for phylogenetic tree construction. Four Markov chains were run for 50 million generations, trees were sampled every 1000 generations, and 12500 sub-optimal trees were discarded as burn-in at the beginning of the MCMC run. No runs indicated a lack of convergence and the potential scale reduction factor for all parameters approached one. The effective sample size for all parameters was above 200 for each run.

We used the codeml package implemented in Phylogenetic Analysis by Maximum Likelihood (PAML) software (Yang, 2007, Yang, 1997) to test for adaptive evolution of the HSP90 gene. Random-site models and branch models were used to test for adaptive evolution of the HSP90 gene.

Identifying sites under adaptive evolution using random-site models

The site models implemented in codeml package allow the ω ratio to vary among sites. We used five of these models on our data set: M0 (one-ratio), M1a (nearly neutral), M2a (Positive selection), M7 (beta) and M8 (beta & ω). M0 is the standard model and assumes one ω for all sites. M1a was compared with M2a, and M1a was the null model for this pair and assumed two site classes in proportion p_0 and $p_1 = 1 - p_0$ with $0 < \omega_0 < 1$ and $\omega_1 = 1$ respectively (Yang, 2007). We compared M7 with M8, and M7 was the null model and assumed beta distribution for ω with shape parameters $\tilde{p}$ and $\tilde{q}$ while M8 was the alternative model with an extra class of sites with $\omega_s > 1$ for positive selection. The Likelihood Ratio Test (LRT) was used to compare the two pairs of the models when testing for positive selection. The test for both pairs had 2 degrees of freedom, and the null distributions were therefore χ_2^2.

HSP90 3D protein modelling

SSA1_70_HSP90 amino acid sequence was submitted to I-TASSER online server (Zhang, 2008, Roy et al., 2010, Yang et al., 2015) for protein structure and function prediction. Top five models were predicted by I-TASSER and the best model indicating the correct global topology was selected basing on the C- score of – 0.05. The protein structure was used to locate amino acid residues positions under positive selection in

HSP90 gene. Annotation of the positively selected sites was done in CLC Genomic workbench 8.5.1.

LIST OF ABBREVIATIONS

ASIAII; Asia II

AUSI; Australia I

CBSD; Cassava brown streak disease

CMD; Cassava mosaic disease

HSP; Heat shock protein

LRT; Likelihood ratio test

MCMC; Markov chain Monte Carlo

MEAMI; Middle East Asia Minor I

MED; Mediterranean

NWI; New world I

NWAfrica; New world Africa

SSA; Sub-Saharan Africa

UGA; Uganda

UWA; University of Western Australia

DECLARATIONS

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon request and sequence data were submitted to Genbank with accession numbers (MH383308 - MH383328).

Competing interests

Not applicable.

Funding

The Natural Resources Institute, University of Greenwich, supported sequencing, fieldwork and the first 2 years of TK's PhD from the grant provided by the Bill and Melinda Gates Foundation (Investment ID OPP1058938). The University of Western Australia Scholarship for International Research Fees (SIRF) F51416, University Postgraduate Awards (UPA) F51204 and UWA Safety Net Top Up scholarship support TK's PhD.

Author's contributions

TK carried out the evolutionary analysis with inputs from LK, LMB and LSL. TK, BRD, JMW participated in wet laboratory work, TK, JMW, BRD, CAO and TA, provided samples. TK drafted the manuscript and all authors read and approved the manuscript.

Acknowledgements

TK is supported by the University of Western Australia, Scholarship for International Research Fees (SIRF), University Postgraduate Awards (UPA) and the University of Western Australia Safety Net top up scholarship. This work forms part of his PhD. The Natural Resources Institute, University of Greenwich, supported sequencing, fieldwork

and the first 2 years of TK's PhD from the grant provided by the Bill and Melinda Gates Foundation (Investment ID OPP1058938). National Crop Resources Research Institute (NaCRRI) provided samples and facilitated sample collection. We thank Paul De Barro, JMW, BRD, and Donald Kachigamba for providing samples. Resources provided by Pawsey Supercomputing Centre with Funding from the Australia Government and the Government of Western Australia supported computation analysis. We thank Joanne Edmondston from Graduate Research School and Joe Bielawski for reviewing the earlier versions of this manuscript. Finally, we thank the farmers in East Africa for allowing us to collect samples on their farms.

References

ALABI, O. J., KUMAR, P. L. & NAIDU, R. A. 2011. Cassava mosaic disease: a curse to food security in subSaharan Africa. *APSnet Features*.

ALICAI, T., NDUNGURU, J., SSERUWAGI, P., TAIRO, F., OKAO-OKUJA, G., NANVUBYA, R., KIIZA, L., KUBATKO, L., KEHOE, M. A. & BOYKIN, L. M. 2016. Cassava brown streak virus has a rapidly evolving genome: implications for virus speciation, variability, diagnosis and host resistance. *Scientific reports,* 6.

ALICAI, T., OMONGO, C., MARUTHI, M., HILLOCKS, R., BAGUMA, Y., KAWUKI, R., BUA, A., OTIM-NAPE, G. & COLVIN, J. 2007. Re-emergence of cassava brown streak disease in Uganda. *Plant Disease,* 91**,** 24-29.

ANDREWS, S. 2010. FastQC: a quality control tool for high throughput sequence data.

AREGBESOLA, O. Z., LEGG, J. P., SIGSGAARD, L., LUND, O. S. & RAPISARDA, C. 2018. Potential impact of climate change on whiteflies and implications for the spread of vectored viruses. *Journal of Pest Science***,** 1-12.

ATEKA, E., ALICAI, T., NDUNGURU, J., TAIRO, F., SSERUWAGI, P., KIARIE, S., MAKORI, T., KEHOE, M. A. & BOYKIN, L. M. 2017. Unusual occurrence of a DAG motif in the Ipomovirus Cassava brown streak virus and implications for its vector transmission. *PloS one,* 12**,** e0187883.

BANKS, G., BEDFORD, I., BEITIA, F., RODRIGUEZ-CEREZO, E. & MARKHAM, P. 1999. A novel geminivirus of Ipomoea indica (Convolvulacae) from Southern Spain. *Plant Disease,* 83**,** 486-486.

BENNETT, B. 2015. Guest editorial: smallholder cassava production and the cassava processing sector in Africa. Practical Action Publishing.

BERRY, S. D., FONDONG, V. N., REY, C., ROGAN, D., FAUQUET, C. M. & BROWN, J. K. 2004. Molecular evidence for five distinct Bemisia tabaci (Homoptera: Aleyrodidae) geographic haplotypes associated with cassava plants in sub-Saharan Africa. *Annals of the Entomological Society of America,* 97**,** 852-859.

BIRD, J. 1957. A whitefly-transmitted mosaic of Jatropha gossypifolia.

BONATO, O., LURETTE, A., VIDAL, C. & FARGUES, J. 2007. Modelling temperature ‐ dependent bionomics of Bemisia tabaci (Q ‐ biotype). *Physiological Entomology,* 32**,** 50-55.

BOYKIN, L. M. 2014. Bemisia tabaci nomenclature: lessons learned. *Pest management science,* 70**,** 1454-1459.

BOYKIN, L. M., ARMSTRONG, K. F., KUBATKO, L. & DE BARRO, P. 2012. Species delimitation and global biosecurity. *Evolutionary bioinformatics online,* 8**,** 1.

BOYKIN, L. M., BELL, C. D., EVANS, G., SMALL, I. & DE BARRO, P. J. 2013. Is agriculture driving the diversification of the Bemisia tabaci species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)?: Dating, diversification and biogeographic evidence revealed. *BMC Evol Biol,* 13**,** 228.

BOYKIN, L. M. & DE BARRO, P. J. 2014. A practical guide to identifying members of the Bemisia tabaci species complex: and other morphologically identical species. *Frontiers in Ecology and Evolution,* 2**,** 45.

BOYKIN, L. M., KINENE, T., WAINAINA, J. M., SAVILL, A., SEAL, S., MUGERWA, H., MACFADYEN, S., TAY, W. T., DE BARRO, P. & KUBATKO, L. 2018. Review and guide to a future naming system of African Bemisia tabaci species. *Systematic Entomology,* 43**,** 427-433.

BOYKIN, L. M., SAVILL, A. & DE BARRO, P. 2017. Updated mtCOI reference dataset for the Bemisia tabaci species complex. *F1000Research,* 6.

BOYKIN, L. M., SHATTERS, R. G., JR., ROSELL, R. C., MCKENZIE, C. L., BAGNALL, R. A., DE BARRO, P. & FROHLICH, D. R. 2007. Global relationships of Bemisia tabaci (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol Phylogenet Evol,* 44**,** 1306-19.

BROWN, J., COATS, S., BEDFORD, I., MARKHAM, P., BIRD, J. & FROHLICH, D. 1995a. Characterization and distribution of esterase electromorphs in the whitefly, Bemisia tabaci (Genn.)(Homoptera: Aleyrodidae). *Biochemical genetics,* 33**,** 205-214.

BROWN, J. K. & CZOSNEK, H. 2002. Whitefly transmission of plant viruses.

BROWN, J. K., FROHLICH, D. & ROSELL, R. C. 1995b. The sweetpotato or silverleaf whiteflies: biotypes of Bemisia tabaci or a species complex? *Annual review of entomology,* 40**,** 511-534.

BRUMIN, M., KONTSEDALOV, S. & GHANIM, M. 2011. Rickettsia influences thermotolerance in the whitefly Bemisia tabaci B biotype. *Insect Science,* 18**,** 57-66.

BUTLER, J. G., HENNEBERRY, T. & CLAYTON, T. 1983. Bemisia tabaci (Homoptera: Aleyrodidae): development, oviposition, and longevity in relation to temperature. *Annals of the Entomological Society of America,* 76**,** 310-313.

BYRNE, D. & BELLOWS, J. 1991. TS *Whitefly biology. annu. rev. entomol,* 36**,** 431-57.

BYRNE, D. N. & MILLER, W. B. 1990. Carbohydrate and amino acid composition of phloem sap and honeydew produced by Bemisia tabaci. *Journal of Insect Physiology,* 36**,** 433-439.

CAHILL, M., BYRNE, F. J., GORMAN, K., DENHOLM, I. & DEVONSHIRE, A. L. 1995. Pyrethroid and organophosphate resistance in the tobacco whitefly Bemisia tabaci (Homoptera: Aleyrodidae). *Bulletin of Entomological Research,* 85**,** 181-187.

CAMPBELL, B., STEFFEN-CAMPBELL, J. & GILL, R. 1996. Origin and radiation of whiteflies: an initial molecular phylogenetic assessment. *Bemisia: 1995, taxonomy, biology, damage, control and management*.

CANGI, N., GORDON, J. L., BOURNEZ, L., PINARELLO, V., APRELON, R., HUBER, K., LEFRANÇOIS, T., NEVES, L., MEYER, D. F. & VACHIÉRY, N. 2016. Recombination is a major driving force of genetic diversity in the anaplasmataceae Ehrlichia ruminantium. *Frontiers in cellular and infection microbiology,* 6**,** 111.

CHAPMAN, R. F. 1998. *The insects: structure and function*, Cambridge university press.

CHEN, W., HASEGAWA, D. K., KAUR, N., KLIOT, A., PINHEIRO, P. V., LUAN, J., STENSMYR, M. C., ZHENG, Y., LIU, W. & SUN, H. 2016. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC biology,* 14**,** 110.

CHIFMAN, J. & KUBATKO, L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics***,** btu530.

CHIKOTI, P. C., NDUNGURU, J., MELIS, R., TAIRO, F., SHANAHAN, P. & SSERUWAGI, P. 2013. Cassava mosaic disease and associated viruses in Zambia: occurrence and distribution. *International Journal of Pest Management,* 59**,** 63-72.

CHU, D., WAN, F. H., ZHANG, Y. J. & BROWN, J. K. 2010. Change in the biotype composition of Bemisia tabaci in Shandong Province of China from 2005 to 2008. *Environmental entomology,* 39**,** 1028-1036.

COLVIN, J., OMONGO, C., MARUTHI, M., OTIM‐NAPE, G. & THRESH, J. 2004. Dual begomovirus infections and high Bemisia tabaci populations: two factors driving the spread of a cassava mosaic disease pandemic. *Plant Pathology,* 53**,** 577-584.

COSSINS, A. & BOWLER, K. 1987. Rate compensations and capacity adaptations. *Temperature biology of animals.* Springer.

COSTA, A. & RUSSELL, L. 1975. Failure of Bemisia tabaci to breed on cassava plants in Brazil (Homoptera: Aleyrodidae). *Cienc. Cult,* 27**,** 388-390.

COSTA, H. & BROWN, J. 1990. Variability in biological characteristics, isozyme patterns and virus transmission among populations of Bemisia tabaci in Arizona. *Phytopathology,* 80**,** 763-766.

COSTA, H. & BROWN, J. K. 1991. Variation in biological characteristics and esterase patterns among populations of Bemisia tabaci, and the association of one population with silverleaf symptom induction. *Entomologia experimentalis et applicata,* 61**,** 211-219.

CUI, X., WAN, F., XIE, M. & LIU, T. 2008. Effects of heat shock on survival and reproduction of two whitefly species, Trialeurodes vaporariorum and Bemisia tabaci biotype B. *Journal of Insect Science,* 8**,** 1-10.

CURNUTTE, L. B., SIMMONS, A. M. & ABD-RABOU, S. 2014. Climate change and Bemisia tabaci (Hemiptera: Aleyrodidae): impacts of temperature and carbon dioxide on life history. *Annals of the Entomological Society of America,* 107**,** 933-943.

DARRIBA, D., TABOADA, G. L., DOALLO, R. & POSADA, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods,* 9**,** 772-772.

DE BARRO, P. J. 2012. The Bemisia tabaci species complex: questions to guide future research. *Journal of Integrative Agriculture,* 11**,** 187-196.

DE BARRO, P. J., LIU, S.-S., BOYKIN, L. M. & DINSDALE, A. B. 2011a. Bemisia tabaci: a statement of species status. *Annual review of entomology,* 56**,** 1-19.

DE BARRO, P. J., LIU, S. S., BOYKIN, L. M. & DINSDALE, A. B. 2011b. Bemisia tabaci: a statement of species status. *Annu Rev Entomol,* 56**,** 1-19.

DE BON, H., HUAT, J., PARROT, L., SINZOGAN, A., MARTIN, T., MALEZIEUX, E. & VAYSSIERES, J.-F. 2014. Pesticide risks from fruit and vegetable pest management by small farmers in sub-Saharan Africa. A review. *Agronomy for Sustainable Development,* 34**,** 723-736.

DEGNAN, J. H. & ROSENBERG, N. A. 2006. Discordance of species trees with their most likely gene trees. *PLoS genetics,* 2**,** e68.

DELATTE, H., HOLOTA, H., WARREN, B. H., BECKER, N., THIERRY, M. & REYNAUD, B. 2011. Genetic diversity, geographical range and origin of Bemisia tabaci (Hemiptera: Aleyrodidae) Indian Ocean Ms. *Bulletin of entomological research,* 101**,** 487-497.

DESAI, M. M. & FISHER, D. S. 2007. Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics*.

DÍAZ, F., MUÑOZ‐VALENCIA, V., JUVINAO‐QUINTERO, D., MANZANO‐MARTÍNEZ, M., TORO‐PEREA, N., CÁRDENAS‐HENAO, H. & HOFFMANN, A. 2014. Evidence for adaptive divergence of thermal responses among B emisia tabaci populations from tropical C olombia following a recent invasion. *Journal of evolutionary biology,* 27**,** 1160-1171.

DIAZ, F., OROBIO, R. F., CHAVARRIAGA, P. & TORO-PEREA, N. 2015. Differential expression patterns among heat-shock protein genes and thermal responses in the whitefly Bemisia tabaci (MEAM 1). *Journal of thermal biology,* 52**,** 199-207.

DINSDALE, A., COOK, L., RIGINOS, C., BUCKLEY, Y. & DE BARRO, P. 2010. Refined global analysis of Bemisia tabaci (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase 1 to identify species level genetic boundaries. *Annals of the Entomological Society of America,* 103**,** 196-208.

DITTRICH, V., HASSAN, S. & ERNST, G. 1986. Development of a new primary pest of cotton in the Sudan: Bemisia tabaci, the whitefly. *Agriculture, ecosystems & environment,* 17**,** 137-142.

DIXON, A. G., BANDYOPADHYAY, R., COYNE, D., FERGUSON, M., FERRIS, R. S. B., HANNA, R., HUGHES, J., INGELBRECHT, I., LEGG, J. & MAHUNGU, N. 2003. Cassava: from poor farmers' crop to pacesetter of African rural development. *Chronica Horticulturae,* 43**,** 8-15.

DOWELL, R. V. 1981. Biology, plant-insect relations, and control of the citrus blackfly. *Florida. Agricultural Experiment Station. Bulletin (USA)*.

DUNN, C. W., HOWISON, M. & ZAPATA, F. 2013. Agalma: an automated phylogenomics workflow. *BMC bioinformatics,* 14**,** 330.

EDWARDS, S. V., LIU, L. & PEARL, D. K. 2007. High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences,* 104**,** 5936-5941.

ELBAZ, M., LAHAV, N. & MORIN, S. 2010. Evidence for pre-zygotic reproductive barrier between the B and Q biotypes of Bemisia tabaci (Hemiptera: Aleyrodidae). *Bulletin of entomological research,* 100**,** 581-590.

ELFEKIH, S., ETTER, P., TAY, W., FUMAGALLI, M., GORDON, K., JOHNSON, E. & DE BARRO, P. 2018. Genome-wide analyses of the Bemisia tabaci species complex reveal contrasting patterns of admixture and complex demographic histories. *PloS one,* 13**,** e0190555.

FASOLO, A. G. & KREBS, R. A. 2004. A comparison of behavioural change in Drosophila during exposure to thermal stress. *Biological Journal of the Linnean Society,* 83**,** 197-205.

FAUQUET, C., BISARO, D., BRIDDON, R., BROWN, J., HARRISON, B., RYBICKI, E., STENGER, D. & STANLEY, J. 2003. Virology division news: revision of taxonomic criteria for species demarcation in the family Geminiviridae, and an updated list of begomovirus species. *Archives of virology,* 148**,** 405-421.

FAUQUET, C., BRIDDON, R., BROWN, J., MORIONES, E., STANLEY, J., ZERBINI, M. & ZHOU, X. 2008. Geminivirus strain demarcation and nomenclature. *Archives of virology,* 153**,** 783-821.

FAUQUET, C. & FARGETTE, D. 1990. African cassava mosaic virus: etiology, epidemiology and control. *Plant Dis,* 74**,** 404-411.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic zoology,* 27**,** 401-410.

FISHPOOL, L. & BURBAN, C. 1994. Bemisia tabaci: the whitefly vector of African cassava mosaic geminivirus. *Tropical Science,* 34**,** 55-72.

GELMAN, D. B., BLACKBURN, M. B., HU, J. S. & GERLING, D. 2002. The nymphal‐adult molt of the silverleaf whitefly (Bemisia argentifolii): Timing, regulation, and progress. *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America,* 51**,** 67-79.

GIBBS, M. J., ARMSTRONG, J. S. & GIBBS, A. J. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics,* 16**,** 573-582.

GILL, R. 1990. The morphology of whiteflies. *Whiteflies: their bionomics, pest status and management***,** 13-46.

GILL, R. J. & BROWN, J. K. 2009. Systematics of Bemisia and Bemisia relatives: can molecular techniques solve the Bemisia tabaci complex conundrum–a taxonomist's viewpoint. *Bemisia: bionomics and management of a global pest.* Springer.

GOLD, C. S., ALTIERI, M. A. & BELLOTTI, A. C. 1990. Direct and residual effects of short duration intercrops on the cassava whiteflies Aleurotrachelus socialis and Trialeurodes variabilis (Homoptera: Aleyrodidae) in Colombia. *Agriculture, Ecosystems & Environment,* 32**,** 57-67.

GOROVITS, R. & CZOSNEK, H. 2017. The involvement of heat shock proteins in the establishment of Tomato yellow leaf curl virus infection. *Frontiers in Plant Science,* 8.

GRAHAM, S. W., OLMSTEAD, R. G. & BARRETT, S. C. 2002. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Molecular biology and evolution,* 19**,** 1769-1781.

HASEGAWA, D. K., CHEN, W., ZHENG, Y., KAUR, N., WINTERMANTEL, W. M., SIMMONS, A. M., FEI, Z. & LING, K.-S. 2018. Comparative transcriptome analysis reveals networks of genes activated in the whitefly, Bemisia tabaci when fed on tomato plants infected with Tomato yellow leaf curl virus. *Virology,* 513**,** 52-64.

HELED, J. & DRUMMOND, A. J. 2009. Bayesian inference of species trees from multilocus data. *Molecular biology and evolution,* 27**,** 570-580.

HILJE, L., COSTA, H. S. & STANSLY, P. A. 2001. Cultural practices for managing Bemisia tabaci and associated viral diseases. *Crop protection,* 20**,** 801-812.

HODDLE, M. 1999. The biology and management of silverleaf whitefly, Bemisia argentifolii Bellows and Perring (Homoptera: Aleyrodidae) on greenhouse grown ornamentals. *Department of Entomology, University of California, Riverside. USA. Applied Biological Control Research*.

HOELMER, K. 1996. Whitefly parasitoids: can they control field populations of Bemisia? *Bemisia: 1995, taxonomy, biology, damage, control and management*.

HOFFMANN, A. A., SØRENSEN, J. G. & LOESCHCKE, V. 2003. Adaptation of Drosophila to temperature extremes: bringing together quantitative and molecular approaches. *Journal of Thermal Biology,* 28**,** 175-216.

HOWELER, R. L., THOMAS, N., ARRIETA BOLAÑOS, S., OCHOA HENRÍQUEZ, J. M., ODIO ECHEVERRÍA, F., MORENO, V., SIMONETTI, L. L. A., DARDÓN, O., RAMIROMEJÍA, F. S. C. & PREISSING, X. 2013. *Save and grow: cassava. A guide to sustainable production intensificationProduire plus avec moins Ahorrar para crecer*, FAO, Roma (Italia).

HSIEH, C.-H., KO, C.-C., CHUNG, C.-H. & WANG, H.-Y. 2014. Multilocus approach to clarify species status and the divergence history of the Bemisia tabaci (Hemiptera: Aleyrodidae) species complex. *Molecular phylogenetics and evolution,* 76**,** 172-180.

HUANG, Y. W., HU, C. C., LIOU, M. R., CHANG, B. Y., TSAI, C. H., MENG, M., LIN, N. S. & HSU, Y. H. 2012. Hsp90 interacts specifically with viral RNA and differentially regulates replication initiation of Bamboo mosaic virus and associated satellite RNA. *PLoS pathogens,* 8**,** e1002726.

HUELSENBECK, J. P., LARGET, B., MILLER, R. E. & RONQUIST, F. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic biology,* 51**,** 673-688.

ISLAMI, T., GURITNO, B. & UTOMO, W. H. 2011. Performance of cassava (Manihot esculenta Crantz) based cropping systems and associated soil quality changes in the degraded tropical uplands of East Java, Indonesia. *Journal of Tropical Agriculture,* 49**,** 31-39.

JIANG, R., QI, L.-D., DU, Y.-Z. & LI, Y.-X. 2017. Thermotolerance and Heat-Shock Protein Gene Expression Patterns in Bemisia tabaci (Hemiptera: Aleyrodidae) Mediterranean in Relation to Developmental Stage. *Journal of economic entomology,* 110**,** 2190-2198.

JONES, D. R. 2003. Plant viruses transmitted by whiteflies. *European Journal of Plant Pathology,* 109**,** 195-219.

KATOH, K., MISAWA, K., KUMA, K. I. & MIYATA, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research,* 30**,** 3059-3066.

KAUR, N., CHEN, W., ZHENG, Y., HASEGAWA, D. K., LING, K.-S., FEI, Z. & WINTERMANTEL, W. M. 2017. Transcriptome analysis of the whitefly, Bemisia tabaci MEAM1 during feeding on tomato infected with the crinivirus, Tomato chlorosis virus, identifies a temporal shift in gene expression and differential regulation of novel orphan genes. *BMC genomics,* 18**,** 370.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S. & DURAN, C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-1649.

KELLERMANN, V., OVERGAARD, J., HOFFMANN, A. A., FLØJGAARD, C., SVENNING, J.-C. & LOESCHCKE, V. 2012. Upper thermal limits of Drosophila are linked to species distributions and strongly constrained phylogenetically. *Proceedings of the National Academy of Sciences,* 109**,** 16228-16233.

KINENE, T., LUBOOBI, L. S., NANNYONGA, B. & MWANGA, G. G. 2015. A mathematical model for the dynamics and cost effectiveness of the current controls of cassava brown streak disease in Uganda. *Journal of Mathematical and Computational Science,* 5**,** 567.

KINENE, T., WAINAINA, J., MAINA, S. & BOYKIN, L. 2016. Rooting trees, methods for.

KUBATKO, L. S. & DEGNAN, J. H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology,* 56**,** 17-24.

KUBATKO, L. S., GIBBS, H. L. & BLOOMQUIST, E. W. 2011. Inferring species-level phylogenies and taxonomic distinctiveness using multilocus data in Sistrurus rattlesnakes. *Systematic Biology,* 60**,** 393-409.

LEACHÉ, A. D. & FUJITA, M. K. 2010. Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). *Proceedings of the Royal Society of London B: Biological Sciences***,** rspb20100662.

LEGG, J., GERLING, D. & NEUENSCHWANDER, P. 2003. Biological control of whiteflies in sub-Saharan Africa. *Biological control in IPM systems in Africa***,** 87-100.

LEGG, J., JEREMIAH, S., OBIERO, H., MARUTHI, M., NDYETABULA, I., OKAO-OKUJA, G., BOUWMEESTER, H., BIGIRIMANA, S., TATA-HANGY, W. & GASHAKA, G. 2011. Comparing the regional epidemiology of the cassava

mosaic and cassava brown streak virus pandemics in Africa. *Virus Research,* 159**,** 161-170.

LEGG, J., OWOR, B., SSERUWAGI, P. & NDUNGURU, J. 2006. Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Advances in virus research,* 67**,** 355-418.

LEGG, J. P., SHIRIMA, R., TAJEBE, L. S., GUASTELLA, D., BONIFACE, S., JEREMIAH, S., NSAMI, E., CHIKOTI, P. & RAPISARDA, C. 2014a. Biology and management of Bemisia whitefly vectors of cassava virus pandemics in Africa. *Pest management science,* 70**,** 1446-1453.

LEGG, J. P., SSERUWAGI, P., BONIFACE, S., OKAO-OKUJA, G., SHIRIMA, R., BIGIRIMANA, S., GASHAKA, G., HERRMANN, H.-W., JEREMIAH, S. & OBIERO, H. 2014b. Spatio-temporal patterns of genetic change amongst populations of cassava Bemisia tabaci whiteflies driving virus pandemics in East and Central Africa. *Virus research,* 186**,** 61-75.

LIU, L. & PEARL, D. K. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic biology,* 56**,** 504-514.

LIU, L., YU, L., KUBATKO, L., PEARL, D. K. & EDWARDS, S. V. 2009. Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution,* 53**,** 320-328.

LIU, S.-S., COLVIN, J. & DE BARRO, P. J. 2012. Species concepts as applied to the whitefly Bemisia tabaci systematics: how many species are there? *Journal of Integrative Agriculture,* 11**,** 176-186.

LÜ, Z.-C. & WAN, F.-H. 2011. Using double-stranded RNA to explore the role of heat shock protein genes in heat tolerance in Bemisia tabaci (Gennadius). *Journal of Experimental Biology,* 214**,** 764-769.

MA, F.-Z., LÜ, Z.-C., WANG, R. & WAN, F.-H. 2014. Heritability and evolutionary potential in thermal tolerance traits in the invasive Mediterranean cryptic species of Bemisia tabaci (Hemiptera: Aleyrodidae). *PloS one,* 9**,** e103279.

MACFADYEN, S., PAULL, C., BOYKIN, L., DE BARRO, P., MARUTHI, M., OTIM, M., KALYEBI, A., VASSÃO, D., SSERUWAGI, P. & TAY, W. T. 2018. Cassava whitefly, Bemisia tabaci (Gennadius)(Hemiptera: Aleyrodidae) in East African farming landscapes: a review of the factors determining abundance. *Bulletin of entomological research***,** 1-18.

MADDISON, W. P. 1997. Gene trees in species trees. *Systematic biology,* 46**,** 523-536.

MADDISON, W. P. & KNOWLES, L. L. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic biology,* 55**,** 21-30.

MAHADAV, A., KONTSEDALOV, S., CZOSNEK, H. & GHANIM, M. 2009. Thermotolerance and gene expression following heat stress in the whitefly Bemisia tabaci B and Q biotypes. *Insect biochemistry and molecular biology,* 39**,** 668-676.

MANSOOR, S., ZAFAR, Y. & BRIDDON, R. W. 2006. Geminivirus disease complexes: the threat is spreading. *Trends in plant science,* 11**,** 209-212.

MARTIN, D., POSADA, D., CRANDALL, K. & WILLIAMSON, C. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research & Human Retroviruses,* 21**,** 98-102.

MARTIN, D. & RYBICKI, E. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics,* 16**,** 562-563.

MARTIN, D. P., MURRELL, B., GOLDEN, M., KHOOSAL, A. & MUHIRE, B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus evolution,* 1.

MARTIN, J. H. 2004. Whiteflies of Belize (Hemiptera: Aleyrodidae). Part 1-introduction and account of the subfamily Aleurodicinae Quaintance & Baker. Moscas blancas de Belice (Hemiptera: Aleyrodidae). Parte 1-introducción y descripción de la subfamilia Aleurodicinae Quaintance & Baker. *Zootaxa.*, 1-119.

MARUTHI, M., COLVIN, J., SEAL, S., GIBSON, G. & COOPER, J. 2002. Co-adaptation between cassava mosaic geminiviruses and their local vector populations. *Virus Research,* 86**,** 71-85.

MARUTHI, M., COLVIN, J., THWAITES, R. M., BANKS, G. K., GIBSON, G. & SEAL, S. E. 2004. Reproductive incompatibility and cytochrome oxidase I gene sequence variability amongst host‐adapted and geographically separate Bemisia tabaci populations (Hemiptera: Aleyrodidae). *Systematic Entomology,* 29**,** 560-568.

MARUTHI, M., HILLOCKS, R., MTUNDA, K., RAYA, M., MUHANNA, M., KIOZIA, H., REKHA, A., COLVIN, J. & THRESH, J. 2005. Transmission of Cassava brown streak virus by Bemisia tabaci (Gennadius). *Journal of Phytopathology,* 153**,** 307-312.

MASTERS, B. C., FAN, V. & ROSS, H. A. 2011. Species delimitation–a geneious plugin for the exploration of species boundaries. *Molecular Ecology Resources,* 11**,** 154-157.

MCAUSLANE, H. J. 2002. *Sweetpotato Whitefly B Biotype of Silverleaf Whitefly, Bemisia tabaci (Gennadius) or Bemisia argentifolii Bellows & Perring (Insecta: Homoptera: Aleyrodidae)*, University of Florida Cooperative Extension Service, Institute of Food and ….

MCKAY, B. D., MAYS JR, H. L., WU, Y., LI, H., YAO, C. T., NISHIUMI, I. & ZOU, F. 2013. An empirical comparison of character‐based and coalescent‐based

approaches to species delimitation in a young avian complex. *Molecular Ecology,* 22**,** 4943-4957.

MIRARAB, S., REAZ, R., BAYZID, M. S., ZIMMERMANN, T., SWENSON, M. S. & WARNOW, T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics,* 30**,** i541-i548.

MUGERWA, H., REY, M. E., ALICAI, T., ATEKA, E., ATUNCHA, H., NDUNGURU, J. & SSERUWAGI, P. 2012. Genetic diversity and geographic distribution of Bemisia tabaci (Gennadius)(Hemiptera: Aleyrodidae) genotypes associated with cassava in East Africa. *Ecology and evolution,* 2**,** 2749-2762.

MUGERWA, H., SEAL, S., WANG, H.-L., PATEL, M. V., KABAALU, R., OMONGO, C. A., ALICAI, T., TAIRO, F., NDUNGURU, J. & SSERUWAGI, P. 2018. African ancestry of New World, Bemisia tabaci-whitefly species. *Scientific reports,* 8**,** 2734.

MUKIIBI, D. R., ALICAI, T., KAWUKI, R., OKAO-OKUJA, G., TAIRO, F., SSERUWAGI, P., NDUNGURU, J. & ATEKA, E. M. 2019. Resistance of advanced cassava breeding clones to infection by major viruses in Uganda. *Crop Protection,* 115**,** 104-112.

NDUNGURU, J., SSERUWAGI, P., TAIRO, F., STOMEO, F., MAINA, S., DJINKENG, A., KEHOE, M. & BOYKIN, L. M. 2015. 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,* 10**,** e0139321.

OLSEN, K. M. & SCHAAL, B. A. 1999. Evidence on the origin of cassava: phylogeography of Manihot esculenta. *Proceedings of the National Academy of Sciences,* 96**,** 5586-5591.

OMONGO, C. A., KAWUKI, R., BELLOTTI, A. C., ALICAI, T., BAGUMA, Y., MARUTHI, M., BUA, A. & COLVIN, J. 2012. African cassava whitefly, Bemisia tabaci, resistance in African and South American cassava genotypes. *Journal of integrative agriculture,* 11**,** 327-336.

PADIDAM, M., SAWYER, S. & FAUQUET, C. M. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology,* 265**,** 218-225.

PARSELL, D. & LINDQUIST, S. 1993. The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annual review of genetics,* 27**,** 437-496.

PATIL, B. L., LEGG, J. P., KANJU, E. & FAUQUET, C. M. 2015. Cassava brown streak disease: a threat to food security in Africa. *Journal of General Virology,* 96**,** 956-968.

PEREIRA, L. 2017. Climate change impacts on agriculture across Africa. *Oxford Research Encyclopedia of Environmental Science*.

POLSTON, J. E. & CAPOBIANCO, H. 2013. Transmitting plant viruses using whiteflies. *Journal of visualized experiments: JoVE*.

POLSTON, J. E., DE BARRO, P. & BOYKIN, L. M. 2014. Transmission specificities of plant viruses with the newly identified species of the Bemisia tabaci species complex. *Pest Manag Sci,* 70**,** 1547-52.

POSADA, D. & CRANDALL, K. A. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences,* 98**,** 13757-13762.

POSADA, D. & CRANDALL, K. A. 2002. The effect of recombination on the accuracy of phylogeny estimation. *Journal of molecular evolution,* 54**,** 396-402.

PRICE, J., SCHUSTER, D. & SHORT, D. 1987. Managing sweetpotato whitefly. *Greenhouse Grower,* 35**,** 55-57.

PUSAG, J. C. A., JAHAN, S. H., LEE, K.-S., LEE, S. & LEE, K.-Y. 2012. Upregulation of temperature susceptibility in Bemisia tabaci upon acquisition of Tomato yellow leaf curl virus (TYLCV). *Journal of insect physiology,* 58**,** 1343-1348.

RAMBAUT, A., SUCHARD, M. A., XIE, D. & DRUMMOND, A. J. 2015. Tracer v1. 6. 2014.

RANNALA, B. & YANG, Z. 2013. Improved reversible jump algorithms for Bayesian species delimitation. *Genetics***,** genetics. 112.149039.

RODRIGO, A., BERTELS, F., HELED, J., NODER, R., SHEARMAN, H. & TSAI, P. 2008. The perils of plenty: what are we going to do with all these genes? *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 363**,** 3893-3902.

ROKAS, A., KING, N., FINNERTY, J. & CARROLL, S. B. 2003. Conflicting phylogenetic signals at the base of the metazoan tree. *Evolution & development,* 5**,** 346-359.

RONQUIST, F. & HUELSENBECK, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19**,** 1572-1574.

RONQUIST, F., TESLENKO, M., VAN DER MARK, P., AYRES, D. L., DARLING, A., HÖHNA, S., LARGET, B., LIU, L., SUCHARD, M. A. & HUELSENBECK, J. P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology,* 61**,** 539-542.

ROSENBERG, N. A. 2007. Statistical tests for taxonomic distinctiveness from observations of monophyly. *Evolution,* 61**,** 317-323.

ROY, A., KUCUKURAL, A. & ZHANG, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols,* 5**,** 725.

SALVUCCI, M. E., STECHER, D. S. & HENNEBERRY, T. J. 2000. Heat shock proteins in whiteflies, an insect that accumulates sorbitol in response to heat stress. *Journal of Thermal Biology,* 25**,** 363-371.

SANGER, F. & COULSON, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology,* 94**,** 441-448.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences,* 74**,** 5463-5467.

SCHIERUP, M. H. & HEIN, J. 2000. Consequences of recombination on traditional phylogenetic analysis. *Genetics,* 156**,** 879-891.

SCORNAVACCA, C. & GALTIER, N. 2017. Incomplete lineage sorting in mammalian phylogenomics. *Systematic biology,* 66**,** 112-120.

SHARAF, N. & BATTA, Y. 1985. Effect of some factors on the relationship between the whitefly Bemisia tabaci Genn.(Homopt., Aleyrodidea) and the parasitoid Eretmocerus mundus Mercet (Hymenopt., Aphelinidae) 1. *Zeitschrift für Angewandte Entomologie,* 99**,** 267-276.

SHATTERS JR, R. G., POWELL, C. A., BOYKIN, L. M., LIANSHENG, H. & MCKENZIE, C. 2009. Improved DNA barcoding method for Bemisia tabaci and related Aleyrodidae: development of universal and Bemisia tabaci biotype-specific mitochondrial cytochrome c oxidase I polymerase chain reaction primers. *Journal of Economic Entomology,* 102**,** 750-758.

SHAW, K. L. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences,* 99**,** 16122-16127.

SIMON, C., FRATI, F., BECKENBACH, A., CRESPI, B., LIU, H. & FLOOK, P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the entomological Society of America,* 87**,** 651-701.

SMITH, J. M. 1992. Analyzing the mosaic structure of genes. *Journal of molecular evolution,* 34**,** 126-129.

SOMERO, G. N. 1995. Proteins and temperature. *Annual review of physiology,* 57**,** 43-68.

SSERUWAGI, P., LEGG, J., MARUTHI, M., COLVIN, J., REY, M. & BROWN, J. K. 2005. Genetic diversity of Bemisia tabaci (Gennadius)(Hemiptera: Aleyrodidae) populations and presence of the B biotype and a non‐B biotype that can induce silverleaf symptoms in squash, in Uganda. *Annals of Applied Biology,* 147**,** 253-265.

SSERUWAGI, P., SSERUBOMBWE, W., LEGG, J., NDUNGURU, J. & THRESH, J. 2004. Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: a review. *Virus research,* 100**,** 129-142.

SSERUWAGI, P., WAINAINA, J., NDUNGURU, J., TUMUHIMBISE, R., TAIRO, F., GUO, J.-Y., VRIELINK, A., BLYTHE, A., KINENE, T. & DE MARCHI, B. 2017. The first transcriptomes from field-collected individual whiteflies (Bemisia tabaci, Hemiptera: Aleyrodidae). *Gates Open Research,* 1.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30**,** 1312-1313.

STANLEY, J. & GAY, M. R. 1983. Nucleotide sequence of cassava latent virus DNA.

SUN, D.-B., LIU, Y.-Q., QIN, L., XU, J., LI, F.-F. & LIU, S.-S. 2013. Competitive displacement between two invasive whiteflies: insecticide application and host plant effects. *Bulletin of entomological research,* 103**,** 344-353.

SWOFFORD, D. L. 2001. Paup*: Phylogenetic analysis using parsimony (and other methods) 4.0. B5.

SWOFFORD, D. L., OLSEN, G. J. & WADDELL, P. J. 1996. Phylogenetic inference. In "Molecular Systematics"(DM Hillis, C. Moritz, and BK Mable, eds). *Sinauer, Sunderland, Massachusetts,* 14**,** 407-5.

TAJEBE, L., BONI, S., GUASTELLA, D., CAVALIERI, V., LUND, O. S., RUGUMAMU, C., RAPISARDA, C. & LEGG, J. 2015. Abundance, diversity and geographic distribution of cassava mosaic disease pandemic‐associated Bemisia tabaci in Tanzania. *Journal of applied entomology,* 139**,** 627-637.

TAJUBA, P. 2016. *Gulu, Nebbi record hottest temperatures* [Online]. Daily Monitor. Available: http://www.monitor.co.ug/News/National/Gulu--Nebbi-record-hottest-temperatures/688334-3137104-x401ub/index.html [Accessed 18 December 2017].

TAY, W. T., ELFEKIH, S., COURT, L. N., GORDON, K. H., DELATTE, H. & DE BARRO, P. J. 2017. 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,* 9**,** 2732-2738.

THRESH, J. & COOTER, R. 2005. Strategies for controlling cassava mosaic virus disease in Africa. *Plant pathology,* 54**,** 587-614.

TOCKO-MARABENA, B. K., SILLA, S., SIMIAND, C., ZINGA, I., LEGG, J., REYNAUD, B. & DELATTE, H. 2017. Genetic diversity of Bemisia tabaci

species colonizing cassava in Central African Republic characterized by analysis of cytochrome c oxidase subunit I. *PloS one,* 12**,** e0182749.

TRACY, C. 2017. *A climate in crisis* [Online]. Oxfam. Available: https://www.oxfam.org/en/research/climate-crisis [Accessed 18 December 2017].

TSUEDA, H. & TSUCHIDA, K. 2011. Reproductive differences between Q and B whiteflies, Bemisia tabaci, on three host plants and negative interactions in mixed cohorts. *Entomologia Experimentalis et Applicata,* 141**,** 197-207.

VYSKOČILOVÁ, S., TAY, W. T., VAN BRUNSCHOT, S., SEAL, S. & COLVIN, J. 2018. An integrative approach to discovering cryptic species within the Bemisia tabaci whitefly species complex. *Scientific reports,* 8**,** 10886.

WANG, P., RUAN, Y. M. & LIU, S. S. 2010. Crossing experiments and behavioral observations reveal reproductive incompatibility among three putative species of the whitefly Bemisia tabaci. *Insect Science,* 17**,** 508-516.

WANG, P., SUN, D. B., QIU, B. L. & LIU, S. S. 2011. The presence of six cryptic species of the whitefly Bemisia tabaci complex in China as revealed by crossing experiments. *Insect Science,* 18**,** 67-77.

WASHBURN, J. D., SCHNABLE, J. C., CONANT, G. C., BRUTNELL, T. P., SHAO, Y., ZHANG, Y., LUDWIG, M., DAVIDSE, G. & PIRES, J. C. 2017. Genome-guided phylo-transcriptomic methods and the nuclear phylogentic tree of the Paniceae grasses. *Scientific Reports,* 7**,** 13528.

WIENS, J. J., KUCZYNSKI, C. A. & STEPHENS, P. R. 2010. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. *Biological Journal of the Linnean Society,* 99**,** 445-461.

WINTER, S., KOERBLER, M., STEIN, B., PIETRUSZKA, A., PAAPE, M. & BUTGEREITT, A. 2010. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. *Journal of General Virology,* 91**,** 1365-1372.

WOSULA, E. N., CHEN, W., FEI, Z. & LEGG, J. P. 2017. Unravelling the genetic diversity among cassava Bemisia tabaci whiteflies using NextRAD sequencing. *Genome biology and evolution,* 9**,** 2958-2973.

XIA, W.-Q., WANG, X.-R., LIANG, Y., LIU, S.-S. & WANG, X.-W. 2017. Transcriptome analyses suggest a novel hypothesis for whitefly adaptation to tobacco. *Scientific reports,* 7**,** 12102.

XIAO, N., PAN, L.-L., ZHANG, C.-R., SHAN, H.-W. & LIU, S.-S. 2016. Differential tolerance capacity to unfavourable low and high temperatures between two invasive whiteflies. *Scientific reports,* 6**,** 24306.

XU, J., DE BARRO, P. & LIU, S. 2010. Reproductive incompatibility among genetic groups of Bemisia tabaci supports the proposition that the whitefly is a cryptic species complex. *Bulletin of entomological research,* 100**,** 359-366.

YANG, J., YAN, R., ROY, A., XU, D., POISSON, J. & ZHANG, Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nature methods,* 12**,** 7.

YANG, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences: CABIOS,* 13**,** 555-556.

YANG, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution,* 24**,** 1586-1591.

YANG, Z. 2015. The BPP program for species tree estimation and species delimitation. *Current Zoology,* 61**,** 854-865.

YANG, Z. & RANNALA, B. 2010. Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences***,** 200913022.

YANG, Z. & RANNALA, B. 2012. Molecular phylogenetics: principles and practice. *Nature reviews genetics,* 13**,** 303.

YU, H. & WAN, F. H. 2009. Cloning and expression of heat shock protein genes in two whitefly species in response to thermal stress. *Journal of Applied Entomology,* 133**,** 602-614.

ZAIDI, S. S.-E.-A., BRIDDON, R. W. & MANSOOR, S. 2017. Engineering dual begomovirus-Bemisia tabaci resistance in plants. *Trends in plant science,* 22**,** 6-8.

ZHANG, C., RABIEE, M., SAYYARI, E. & MIRARAB, S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics,* 19**,** 153.

ZHANG, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics,* 9**,** 40.

FIGURE LEGENDS

Figure 1. Maximum clade credibility tree for the HSP90 gene of *Bemisia tabaci* species. The tree was obtained from an alignment of 27 samples of length 2160 base pairs, run in MrBayes 3.2.2 on the Pawsey supercomputer. The tree was rooted with *B. afer* for convenience but the analysis was done on the unrooted tree topology. The branch length was defined as the expected number of nucleotide substitutions per site. The tree topology was used to test various models of evolution.

Figure 2. Plot of the posterior probability that each site is from the three site classes under the M2a model, i.e. purifying selection (ω_0 <1) in grey, neutral selection (ω_1 =1) in blue and positive selection (ω_2 >1) in red. The Y-axis represents posterior probability from Bayes empirical Bayes (BEB) approach and the X-axis represents the amino acid sites of the HSP90 gene.

Figure 3. 3D protein structure of the HSP90, predicted from I – TASSER online protein structure and function predictions server. The amino acid residues identified to be under positive selection by random site modes M2a and M8 are shown as 166K at the end of the beta strand, 198D on the alpha helix, 333R and finally 693 M on the coil.

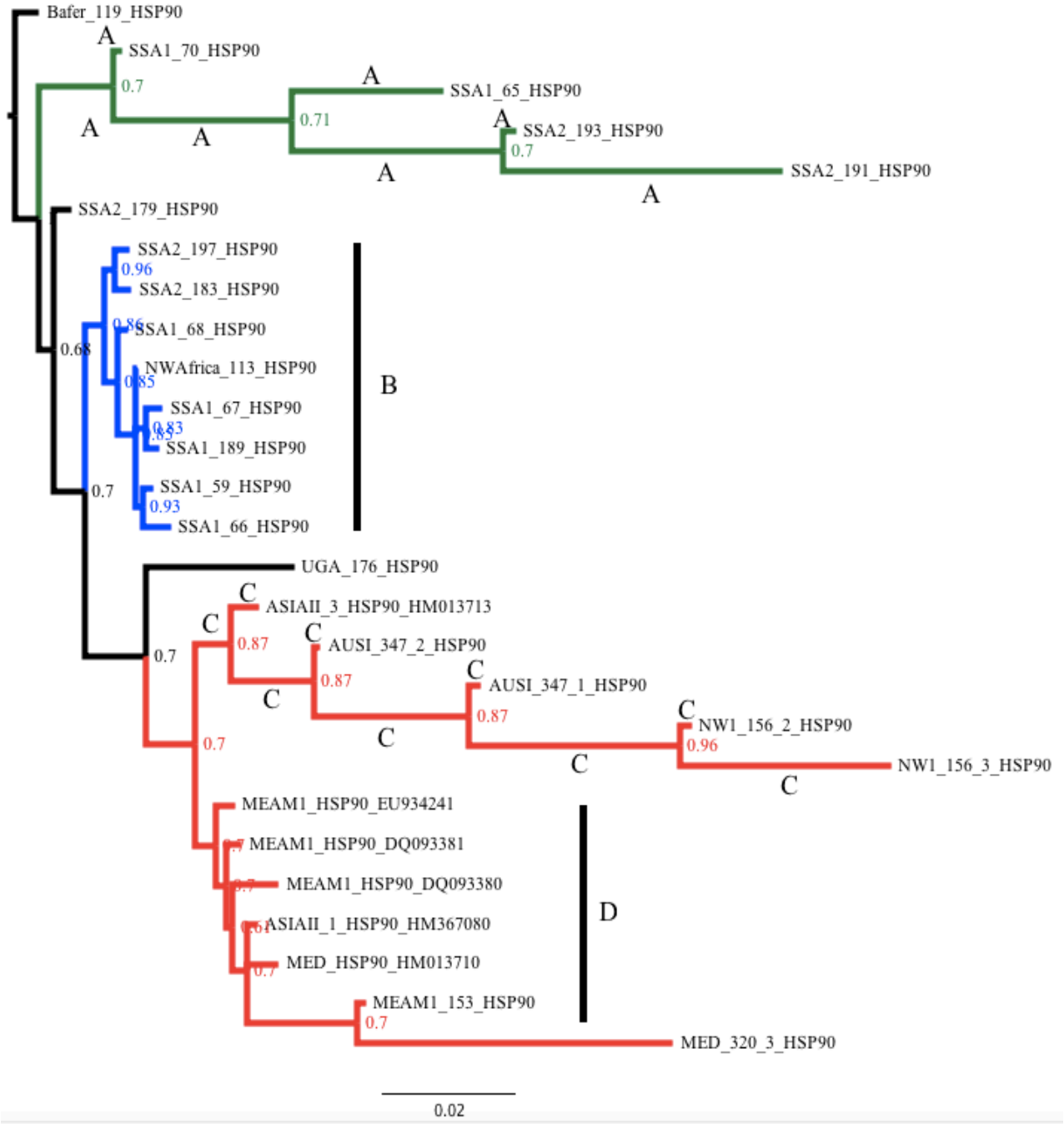


Figure 1.

Maximum clade credibility tree for the HSP90 gene of *Bemisia tabaci* species. The tree was obtained from an alignment of 27 samples of length 2160 base pairs, run in MrBayes 3.2.2 on the Pawsey supercomputer. The tree was rooted with *B. afer* for convenience but the analysis was done on the unrooted tree topology. The branch length was defined as the expected number of nucleotide substitutions per site. The tree topology was used to test various models of evolution.

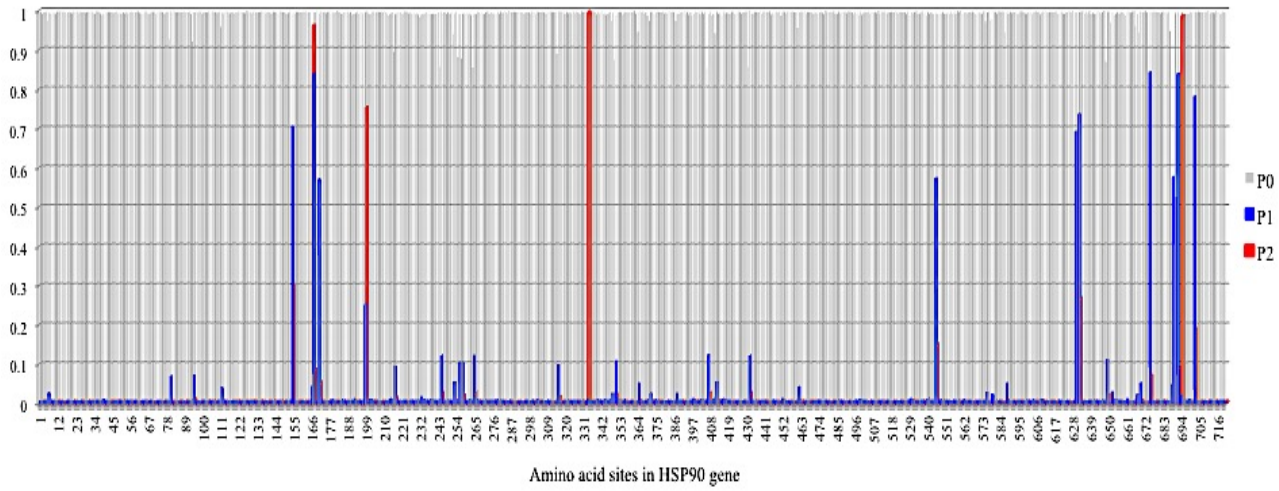


Figure 2

Plot of the posterior probability that each site is from the three site classes under the M2a model, i.e. purifying selection (ω_0 <1) in grey, neutral selection (ω_1 =1) in blue and positive selection (ω_2 >1) in red. The Y-axis represents posterior probability from Bayes empirical Bayes (BEB) approach and the X-axis represents the amino acid sites of the HSP90 gene.

Figure 3

3D protein structure of the HSP90, predicted from I – TASSER online protein structure and function predictions server. The amino acid residues identified to be under positive selection by random site modes M2a and M8 are shown as 166K at the end of the beta strand, 198D on the alpha helix, 333R and finally 693 M on the coil.

Additional file 1

Sample ID	Species ID	Genbank Accession #	Country	Latitude	Longitude	Altitude	Host plant
59	SSA1	MH383311	Uganda	0.85070	32.41920	1078	Cassava
65	SSA1	MH383326	Uganda	0.87345	32.64796	1085	Cassava
66	SSA1	MH383314	Uganda	0.86932	32.51936	1113	Cassava
67	SSA1	MH383310	Uganda	1.41966	32.43936	1064	Cassava
68	SSA1	MH383309	Uganda	1.19876	32.73804	1069	Cassava
70	SSA1	MH383308	Uganda	1.31247	32.58484	1042	Cassava
113	NWAfrica	MH383312	Uganda	0.54896	31.44420	1283	Mujjaja
119	B.*afer*	MH383318	Malawi	-12.792975	34.200964	500	Cassava
153	MEAM1	MH383320	Brazil	-22.844553	-48.435439	800	Cabbage
156_2	NWI	MH383323	Brazil	-21.980278	-50.788312	397	Euphorbia heterophylla
156_3	NWI	MH383324	Brazil	-21.980278	-50.788312	397	Euphorbia heterophylla
176	UGA	MH383319	Kenya	00.00622	034.73450	1510	Common bean
179	SSA2	MH383317	Kenya	00.02544	034.62852	1561	Common bean
189	SSA1	MH383313	Kenya	00.33262	034.25285	1322	Common bean
191	SSA2	MH383328	Kenya	00.56222	034.17064	1165	Common bean
193	SSA2	MH383327	Kenya	00.61071	037.37062	1167	Common bean
197	SSA2	MH383315	Kenya	00.59737	034.43888	1331	Common bean
320_3	MED	MH383325	Brazil	-23.01694	-048.06861		Cassava
347_1	AUSI	MH383322	Australia				
347_2	AUSI	MH383321	Australia				
183	SSA2	MH383316	Kenya	00.00622	034.55100	1561	Common bean

Additional file 2

Sample ID	Number of reads obtained	Reads after trimming	Number of Contigs	Average Contig length	N50
59	8,000,000	7,848,046	16,473	1,102	1,195
65	8,000,000	7,857,579	17,146	1,024	1,071
66	8,000,000	7,857,564	16,489	1,045	1,095
67	8,000,000	7,876,283	18,271	1,165	1,289
68	8,000,000	7,861,938	19,852	1,201	1,368
70	8,000,000	7,861,502	20,182	1,113	1,205
113	8,000,000	7,850,359	13,156	1,023	1,050
119	8,000,000	7,845,846	20,368	1,238	1,400
153	36,449,340	36,232,156	25,435	1,235	1,432
156_2	40,826,418	40,604,557	27,466	1,356	1,689
156_3	43,194,264	42,953,727	30,716	1,346	1,655
176	8,000,000	7,835,023	19,865	1,366	1,616
179	8,000,000	7,842,509	15,570	1,194	1,295
189	8,000,000	7,823,762	2,530	927	910
191	8,000,000	7,833,877	19,770	1,124	1,222
193	8,000,000	7,852,051	20,061	1,122	1,218
197	8,000,000	7,871,950	19,224	1,230	1,408
320_3	39,894,296	39,686,374	27,867	1,144	1,250
347_1	39,142,358	38,894,512	30,496	1,372	1,723
347_2	37,306,246	37,071,584	27,343	1,314	1,614
183	8,000,000	7,816,022	20,066	1,298	1,512

A 2. Presentations, international conferences and workshops attended

Primary authored manuscripts and collaborative manuscripts

De Marchi, B.R., **Kinene, T**., Wainiana, J.M., Krause-Sakate, R. and L.M.Boykin. (2018) Comparative transcriptome analysis reveals genetic diversity in the endosymbiont *Hamiltonella* between native and exotic populations of *Bemisia tabaci* from Brazil. PlosOne https://doi.org/10.1371/journal.pone.0201411

Kinene, T., Kubatko, L., Luboobi, S.L., Wainaina, J., De Marchi, B., Small, I., Alicai T., Omongo, A.C., Savill, A., and Boykin, L.M. 2018. Recombination detected in the Heat Shock Protein 90 (HSP90) gene of the Bemisia tabaci species complex. *Under review in BMC evolutionary Biologgy*

Sseruwagi, P., Wainaina, J.M., Ndunguru, J., Tumuhimbise, R., Tairo, F., Guo, J., Vrielink, A., Blythe, A., **Kinene, T.**, De Marchi, B., Kehoe, M.A., Tanz, S.K. & Boykin, L.M. (2018) The first transcriptomes from field-collected individual whiteflies (*Bemisia tabaci*, Hemiptera: Aleyrodidae): a case study of the endosymbiontcomposition. Gates Open Res 2018, 1:16 (doi: 10.12688/gatesopenres.12783.3)

Boykin, L.M., Ammar Ghalab, Bruno Rossitto De Marchi, Anders Savill, James M Wainaina, **Tonny Kinene,** Stephen Lamb, Myriam Rodrigues, Monica A Kehoe, Joseph Ndunguru, Fred Tairo, Peter Sseruwagi, Charles Kayuki, Deogratius Mark, Joel Erasto, Hilda Bachwenkizi, Titus Alicai, Geoffrey Okao-Okuja, Phillip Abidrabo, John Francis Osingada, Jimmy Akono, Elijah Ateka, Brenda Muga,

Samuel Kiarie. 2018. Real-time portable genome sequencing for global food security. F1000. https://f1000research.com/articles/7-1101/v1

Boykin, L.M., **T. Kinene**, J. Wainaina, S. Seal, H. Mugerwa, S. Macfadyen, P. De Barro, W.T. Tay, L. Kubatko, T. Alicai, C.A. Omongo, F. Tairo, J. Ndunguru and P. Sseruwagi. (2018) Review and future guide to the naming system of African *Bemisia tabaci* species. *Systematic Entomology*. Vol 43, pp. 427-433. Wiley Online Library

Kinene, T., Wainaina, J., Maina, S., and Boykin, L.M. 2016. Rooting Trees, Methods for. In: Kliman, R.M. (ed.), *Encyclopedia of Evolutionary Biology. vol. 3, pp. 489–493*. Oxford: Academic Press.

Conference Publications

Kinene, T., Kubatko, L., Small, l., Wainaina, J., Alicai, T., Omongo, C., Kachigamba, D., & Boykin, L. M**. 2016** Uncovering evolutionary patterns of the African cassava whitefly, Bemisia tabaci by analyzing rates of evolution, abstract 2nd International Whitefly Symposium, 14th -19th February 2016, Arusha Tanzania.

Boykin, L. M**.,** Sseruwagi, P., Ndunguru, J., Kehoe, M. A., Wainaina, J., **Kinene T**., De Barro, P., Tanz, S., Wang, H. L., Wang, X. W., Liu, S. S., Alicai, T., Omongo, C., Mugerwa, H., Seal, S., Colvin, J., Kachigamba, D., & Kubatko, L. 2016. Quest for robust species trees for the African cassava Bemisia tabaci species and the destructive virus diseases they transmit, abstract 2nd International Whitefly Symposium, 14th -19th February 2016, Arusha Tanzania.

Presentations and invited talks

Patterns of conflict between Mitochondrial DNA and Nuclear DNA of the *Bemisia tabaci* species complex, at the 3rd International Whitefly Symposium held at the Esplanade Hotel Fremantle, by Rydges, Western Australia from 16 -19 September 2018.

Is climate change driving the rates of molecular evolution of Heat Shock Protein 90 (HSP90) in *Bemisia tabac*i species complex? at the Science Protecting Plant Health international conference, whitefly workshop at Brisbane conventional center, Queensland, Australia. On 25 Sept 2017

Systematics of *Bemisia tabaci* species complex at XXV International Congress of Entomology (ICE 2016) held in Orlando Florida, USA on 29th Sept 2016 in a session entitled Synthesis *Sternorrhyncha* Systematics.

Food security and Supercomputing, using genomics and High performance computing (HPC) resources at the Pawsey Center to understand the evolution of African cassava whiteflies. At The University of Western Australia, Monadelphous Studio, 9th June 2016.

Uncovering the evolutionary patterns of the African cassava whitefly, *Bemisia tabaci* by analyzing rates of evolution, at the 2nd International Whitefly Symposium in Mt Meru Hotel Arusha. 16th Feb 2016.

African Cassava Whitefly *Bemisia tabaci,* systematics and rates of evolution at the School of Chemistry and Biochemistry, The University of Western Australia, The Bayliss Seminar Series, PhD proposal presentation 20th Nov 2015.

African Cassava Whitefly, *Bemisia tabaci,* systematics and rates of evolution, poster presentation at the first annual meeting of the African cassava whitefly project in Greenwich London, on 9th Dec 2015.

Workshops, conferences and courses attended

17th – 22nd June 2018. Summer school, "Taming the Beast workshop 2018" in Oberageri, Switzerland organized by the team from the Computational Evolution group at ETH Zurich.

26th – 28th Sept 2017. Science Protecting Plant Health (SPPH 2017) held at Brisbane convention and exhibition center. Queensland, Australia. CRC Plant Biosecurity funded trip.

25th – 30th Sept 2016. Attended the XXV International Congress of Entomology, (ICE 2016) in Orlando Florida, USA at the Orange County Convention Center.

17th – 27th July 2016. Workshop on molecular evolution held at Marine Biological Laboratory in Woods hole MA, USA.

30th March 2016. Mental Health First Aid training at The University of Western Australia organized by the School of Chemistry and Biochemistry.

14th – 19th Feb 2016 2nd International Whitefly Symposium, organized by International Tropical Institute of Agriculture (IITA) and Mikocheni Agricultural Research Institute (MARI) at Mt Meru Hotel in Arusha Tanzania.

6th – 11th Dec 2015. African Cassava Whitefly Project First Annual Meeting, organized by University of Greenwich under the African Cassava Whitefly Project in London.

9th -10th Nov 2015, Introduction and Intermediate supercomputing, topics included; introduction to Pawsey, Linux, Supercomputing Technology and Intermediate Supercomputing Organized by the Pawsey Supercomputing Centre, at Pod B, CSIRO ARRC, Dick Perry Avenue, Australia.

21st - 22nd Sept 2015, NECTAR Research Cloud training, introductory workshop at the University of Western Australia by Chris Bording from the Pawsey Supercomputing Centre.

21st June – July 3rd 2015, The 3rd EMBL Australia PhD course at Harry Perkins Institute of Medical Research (Topics included Bioinformatics, Introduction to NGS, Genomics and Gene expression), at the University of Western Australia.