Doctoral Thesis

A STUDY OF THE SUPER-ABUNDANT *BEMISIA TABACI* (GENNADIUS) SPECIES COMPLEX (HEMIPTERA: ALEYRODIDAE) IN CASSAVA MOSAIC DISEASE PANDEMIC AREAS IN TANZANIA

LENSA SEFERA TAJEBE

Principal supervisor: Professor Carmelo Rapisarda (University of Catania, Italy)
Co-supervisor: Associate Professor Ole Søgaard Lund (University of Copenhagen, Denmark)
Field supervisor: Dr. James P. Legg (International Institute of Tropical Agriculture, Tanzania)

AgTraIn – Agricultural Transformation by Innovation

2014

(Page intentionally left blank)

A study of the super-abundant *Bemisia tabaci* (Gennadius) species complex (Hemiptera: Aleyrodidae) in cassava mosaic disease pandemic areas in Tanzania

LENSA SEFERA TAJEBE

A thesis submitted to
the Department of Agriculture, Food and Environment,
University of Catania and the Department of Plant and Environmental Sciences,
University of Copenhagen in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Principal supervisor: Professor Carmelo Rapisarda (University of Catania, Italy)
Co-supervisor: Associate Professor Ole Søgaard Lund (University of Copenhagen, Denmark)
Field supervisor: Dr. James P. Legg (International Institute of Tropical Agriculture, Tanzania)

2014

Declaration

I hereby declare that this thesis is my own, that it has never been submitted to any other University and that all sources have been duly acknowledged.

SIGNED: LENSA SEFERA TAJEBE

A study of the super-abundant *Bemisia tabaci* (Gennadius) species complex (Hemiptera: Aleyrodidae) in cassava mosaic disease pandemic areas in Tanzania

PhD candidate: Lensa Sefera Tajebe[1,2]
Principal supervisor: Professor Carmelo Rapisarda[1]
Co-supervisor: Associate Professor Ole Søgaard Lund[2]
Field supervisor: Dr. James P. Legg[3]

[1]Dipartimento di Agricoltura, Alimentazione e Ambiente (D*i*3A), Sezione Entomologia applicata, Università degli Studi, Via Santa Sofia 100, 95123 Catania, Italy
[2]Department of Plant and Environmental Science, University of Copenhagen, Højbakkegård Allé 13, DK-2630 Taastrup, Denmark
[3]International Institute of Tropical Agriculture, P.O. Box 34441, Dar es Salaam, Tanzania

Summary

Bemisia tabaci (Gennadius) (Hemiptera: Aleyrodidae) is a cryptic species complex comprising at least 35 morphologically indistinguishable species that exhibit high genetic variation. *B. tabaci* is the vector of cassava mosaic geminiviruses (CMGs) causing cassava mosaic disease (CMD) to cassava (*Manihot esculenta* Crantz) – one of the most important food security crops in sub-Saharan Africa. After the outbreak and spread of a severe form of CMD from Uganda to several East and African countries, the production of cassava has been impaired causing severe yield loss resulting in food shortages and forcing some farmers to abandon the cultivation of the crop. The severe CMD pandemic is mainly characterized by high severity and incidence of the disease dominated by whitefly-borne infection and super-abundant populations of *B. tabaci*. All *Bemisia tabaci* individuals harbour a primary bacterial symbiont (*Portiera aleyrodidarum*) essential for their survival, and many also harbour non-essential secondary symbionts.
In an attempt to understand the cause of the super-abundant populations, it was suggested that a distinct genotype cluster is associated with the pandemic, however, there was no definitive proof for this or other potential causes of super-abundance. Moreover, these super-abundant populations, that occur in pandemic areas, remain unstudied in terms of the endosymbiotic bacteria they harbour, their population structure, and the potential for gene flow and hybridization. Therefore, this research through the use of field survey and molecular approaches aimed to fill the knowledge gap to this effect and sought (i) to assess the genetic diversity and distribution of cassava-associated *B. tabaci* in CMD pandemic-affected areas in Tanzania and to establish the association of a distinct *B. tabaci* genotype with the pandemic; (ii) to establish the prevalence of endosymbionts among *B. tabaci* populations on cassava and nearby plants, to assess if there is a unique endosymbiont that is associated with super-abundant whitefly populations and to provide baseline information on endosymbionts of whiteflies for Tanzania; and (iii) to investigate the population structure of and the possibilities of gene flow between *B. tabaci* populations

found in CMD pandemic affected and not yet affected parts of Tanzania. Hence, several field surveys were conducted in CMD pandemic-affected and unaffected areas in Tanzania and whiteflies were collected from cassava plants, intercrops and weeds in and surrounding cassava fields. Insects collected in this way were subjected to molecular analyses using mitochondrial and nuclear markers.

The four *B. tabaci* putative species found to be present in the study area, based on mitochondrial data, were Sub-Saharan Africa 1 (SSA1), Mediterranean (MED), Indian Ocean (IO) and East Africa 1 (EA1). Moreover, a group of unknown whitefly species was also identified. There were four sub-groups under SSA1 which included SSA1-SG1, SSA1-SG2, SSA1-SG1/2 and SSA1-SG3.

The findings of this study also established SSA1-SG1 to be the pandemic-associated *B. tabaci* in Tanzania, being found predominantly in the pandemic affected north-western part of the country. *B. tabaci* abundance and mean CMD severity values were highest for north-western Tanzania, matching the characteristic of the pandemic. The SSA1-SG2 was shown to be present in the pandemic-unaffected central part of Tanzania whereas the coastal sub-group that is SSA1-SG3 was exclusively found confined to the coastal parts and was absent from the central and north-western parts.

Furthermore, microsatellite data also identified SSA1, MED and IO putative species from the *B. tabaci* species complex as well as *B. afer*. It further revealed two sub-clusters within SSA1, which were each dominating different geographic areas across the sampling zones. One type predominated in the north-western pandemic affected part of Tanzania but diminished in frequency towards the south-east, whilst the other became increasingly common until it becomes almost the only type present in the coastal area. The central part was a mix of the two clusters. Some evidence of gene flow and hybridization was observed between the two sub-clusters where several hybrids were prevalent with no strong assignment to any of the clusters.

The CMD pandemic front was estimated to lie in Geita Region located in north-western Tanzania. The pandemic is spreading from the north-west to south-east of the country at an approximate rate of 26 km per year. The pandemic-associated SSA1-SG1 was also found extending up to 180 km to the south east of the pandemic front and very recently also in the central parts of Tanzania – an area not yet affected by the pandemic.

In this study cassava was noted to be colonized only by SSA1 putative species and was never seen to be colonized by the other putative species suggesting that cassava in Africa may only be colonized by cassava types.

The different sub-groups and putative species in the study exhibited contrasting endosymbionts profiles. *Arsenophonus* was the most commonly infecting secondary symbiont among the pandemic-associated SSA1-SG1 populations. While multiple infection was infrequent in this group, *Arsenophonus*/*Cardinium* was the most common double infection type. Although no particular endosymbiont was found to be associated only with the pandemic-associated SSA1-SG1, the most remarkable feature was the absence of secondary symbiont infection for more than a third of the population and being the only sub-group comprising a significant number of individuals with no secondary symbiont infection. Multiple infection was very common in the remaining SSA1 sub-groups; SG2 and SG1/2, mainly occurring in the non-pandemic central parts of Tanzania, frequently harboured *Arsenophonus*/*Cardinium* or *Hamiltonella*/*Arsenophonus*/*Cardinium* infections.

The coastal type sub-group, SSA1-SG3, often showed multiple infection by *Rickettsia*/*Hamiltonella*/*Arsenophonus*/*Cardinium*. Multiple infections were common in whiteflies from MED which harboured combinations of *Rickettsia*/*Hamiltonella*/*Wolbachia* with *Arsenophonus* or *Cardinium*. Similarly, double and multiple infections were also common in the IO and EA1 putative species where EA1 individuals commonly harboured *Cardinium*/*Wolbachia* with several additional combinations of *Rickettsia*, *Hamiltonella* or *Arsenophonus*. The secondary endosymbiont *Fritschea* was never recorded in this study.

Finally, the extended presence of the pandemic-associated SSA1-SG1 in areas of up to 180 km to the south-east of the pandemic front is not entirely surprising, as changes in whitefly populations precede changes in disease status. However, the advancement of this pandemic-associated sub-group further into the non-pandemic central part coupled with the possibilities of gene flow and hybridization with other sub-groups poses an on-going threat to cassava production in the region as it implicates a further spread of the pandemic to previously unaffected areas. This pandemic-associated sub-group is also distinct from the others in terms of its secondary endosymbiont profile, being the only sub-group in this study exhibiting no infection in more than one third of the population.

Therefore, future studies should focus towards understanding further the *B. tabaci* species complex in the area in general and in understanding this super-abundant pandemic-associated population in particular. Specific studies to assess the presence of other secondary endosymbionts, antibiotic treatments to eliminate certain endosymbionts or to artificially introduce endosymbionts are paramount to fully understand their exact effects on these insects in the wider region of East and Central Africa. Moreover, mating studies involving the pandemic-associated and the other sub-groups are necessary to exactly determine if the pandemic-associated type is taking over through hybridization and if the hybrids are conferred with increased fitness. Only by understanding the vector better can we design integrated pest management strategies to stop the spread of the severe CMD pandemic that continues to threaten the food security of several millions of people in sub-Saharan Africa.

This work was the first study that confirmed the association of a distinct genetic sub group to the severe CMD pandemic in Tanzania more than a decade after similar associations were first made in Uganda in 2002. Moreover, the current study showed for the first time some evidence of gene flow and hybridization among the sub-clusters of the SSA1 putative species. The study was also the first work that provided baseline information on the contrasting endosymbiont patterns of whiteflies for a part of East Africa which until now had not been studied.

Key words: cassava, East Africa, whitefly, genetic groups, pandemic front, secondary symbionts, gene flow, hybridization, microsatellite, population structure

Research activities

List of publications in peer-reviewed journals

- **Tajebe LS**, Boni SB, Guastella G, Cavalieri V, Lund OS, Rugumamu CP, Rapisarda C, Legg JP. 2014. Abundance, diversity and geographic distribution of cassava mosaic disease pandemic associated *Bemisia tabaci* in Tanzania. J. Appl. Entomol. (Accepted Article) doi: 10.1111/jen.12197
- **Tajebe LS**, Guastella D, Cavalieri V, Kelly SE, Hunter MS, Lund OS, Legg JP, Rapisarda C. 2014. Diversity of symbiotic bacteria associated with *Bemisia tabaci* (Homoptera: Aleyrodidae) in cassava mosaic disease pandemic areas of Tanzania. Ann. Appl. Biol. (Accepted Article) doi:10.1111/aab.12183
- **Tajebe LS**, Simiand C, Lund OS, Reynaud B, Rapisarda C, Legg JP, Delatte H. Microsatellites reveal population genetic structure of *Bemisia tabaci* species complex associated with cassava in Tanzania (Submitted Manuscript)
- Guastella D, **Tajebe LS**, Evans G, Fovo FP, Rapisarda C, Legg JP, 2014. First Record of *Aleuroclava psidii* (Singh) and *Aleurotrachelus tuberculatus* Singh (Hemiptera: Aleyrodidae) in East Africa. Afr Entomol. 22, 437-440. DOI: http://dx.doi.org/10.4001/003.022.0206
- Legg JP, Shirima R, **Tajebe LS**, Guastella D, Boniface S, Jeremiah S, Nsami E, Chikoti P, Rapisarda C, 2014. Biology and management of *Bemisia* whitefly vectors of cassava virus pandemics in Africa. Pest Manag Sci. 70, 1446-1453.
- Guastella D, Lulah H, **Tajebe LS**, Cavalieri V, Evans GA, Pedata PA, Rapisarda C, Legg JP, 2014. Survey on whiteflies and their parasitoids in cassava mosaic pandemic areas of

Tanzania using morphological and molecular techniques. Pest Manag Sci. DOI: 10.1002/ps.3810. [Epub ahead of print]

Selected list of seminars and conferences attended

- Fourth European Whitefly Symposium in Rehovot, Israel (September, 2011)
- Conference on Integrated Control in Protected Crops, Mediterranean Climate (IOBC), in Catania, Italy (October, 2012)
- First International Whitefly Symposium in Kolymbari Crete, Greece (May 2013)
- Seminar on "Understanding and Managing Cassava Virus Pandemics and Whitefly Vectors" in Catania, Italy (May, 2013)
- Workshop on "Author workshop Springer" organized by Springer in Catania, Italy (June, 2013)

Collaborations with institutions other than host universities

- International Institute of Tropical Agriculture (IITA), Dar es Salaam, Tanzania
- Department of Entomology, University of Arizona, Tucson, USA
- Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), UMR PVBMT, La Réunion, France

Dedication

This work is dedicated to my loving and caring family for their continued love and support.

Acknowledgement

I would like to thank GOD for loving, guiding and supporting me throughout my doctoral period!

This doctoral work was funded by the European Union within the educational initiative – Agricultural Transformation by Innovation (AgTraIn), a Joint Doctorate Program in the frame of the Erasmus Mundus II program.

I would like to state my sincere appreciation and deepest gratitude to my principal supervisor Professor Carmelo Rapisarda from the University of Catania, Italy and to my co-supervisor Associate Professor Ole Søgaard Lund from the University of Copenhagen, Denmark for sharing their expert knowledge and experience in whiteflies and viruses and for the support and guidance they rendered me throughout my doctoral period. I honestly appreciate their constructive advices, motivations and problem solving nature that made my doctoral journey a very smooth and fruitful one. I am very thankful to both of them as they not only nurtured me intellectually and scientifically but also helped greatly in the bureaucratic and administrative matters whenever needed. This work could have not been done without their support. Grazie/Tak! Special thanks also goes to the staff at the two universities for facilitating my double degree enrollment and making my stay a conducive one.

I am highly indebted to my field supervisor Dr. James P. Legg from the International Institute of Tropical Agriculture – IITA, TZ, for his invaluable comments, unlimited support and for sharing his vast knowledge and experience in cassava virus diseases and whitefly vectors. I am also very thankful to him for advising and linking me to make my two study/research trips to University of Arizona, USA and CIRAD, UMR PVBMT La Réunion, France. I extend my heartfelt thanks to IITA-TZ staff for making all arrangements to obtain the required official documents for my stay in Tanzania and for maintaining a friendly and constructive work environment throughout my stay.

A very special thanks goes to Professor Martha (Molly) Hunter from the University of Arizona, USA, her amazing lab manager Suzanne Kelly and her entire lab team for accepting me as a visiting scholar and for the very productive collaboration we had. I am sincerely thankful to Professor Martha (Molly) Hunter, Suzanne Kelly and Dr. Anna Himler for giving professional advice on my day-to-day research activities. I would also like to acknowledge that the laboratory costs during my stay were covered by the National Science Foundation grant DEB-1020460 to MSH.

My greatest appreciations also goes to Dr. Hélène Delatte from CIRAD, UMR PVBMT, Pôle de Protection des Plantes, La Réunion, France for accepting me as a visiting doctoral trainee and for covering all related costs. I would like to specially thank her for sharing her deep knowledge in population genetics in such a detailed and friendly manner. All our discussions and her continuous guidance both during my stay and after my departure through to the write up of the work have been immensely useful to strengthen my understanding of the subject matter. Thank you! The excellent technical assistance and amazing patience of Cristophe Simiand (Lulu) is highly appreciated. Lulu and the lab team really made my work at CIRAD easy, enjoyable and fun. Thank you all!

I extend my special appreciation to Dr. Vincenzo Cavalieri from University of Catania, Italy for sharing his experience and extending his assistance in the molecular analysis of some samples. I also appreciate colleagues and department staff at my two universities for making my doctoral journey not to be only about science but also about life by sharing valuable and fun moments. Moreover, I am thankful to Dr. Vincenzo Cavalieri, Dr. Sarra Ben Attia, Dr. Salvo Bella, Dr. Alessandro Vitale and Dr. Alessandra La Pergola for the good moments we shared.

My warmest and heartfelt gratitude goes to my beloved family and fiancè Devid for their prayers, love, motivation and support all the way through my doctoral study. I will always be grateful for having them!

Finally, I thank everybody else whom I did not mention here, for their help.

Preface

Bemisia tabaci is a cryptic species complex of morphologically indistinguishable but genetically variable putative species. In situations like that of *B. tabaci* where morphological means of whitefly identification are not enough to distinguish the species complex, molecular approaches have widely been employed. The thesis reported here used molecular approaches together with field surveys to investigate the genetic diversity, population structure, geographic distribution and endosymbiont profiles of cassava-associated *Bemisia tabaci* in Tanzania. The first part of the thesis presents background information (chapter 1) on cassava, cassava mosaic disease (CMD), and cassava mosaic geminiviruses (CMGs) and their vector *Bemisia tabaci*. The subsequent three chapters report the main studies of the doctoral work, each with their own Introduction, Material and Methods, Results and Discussions sections. The first of these studies (Chapter 2) presents genetic diversity (based on mitochondrial marker) and geographic distribution of cassava associated *B. tabaci* describing a distinct CMD pandemic-associated *B. tabaci*. The second study (Chapter 3) focuses on endosymbiotic bacterial communities harboured by *B. tabaci* in Tanzania giving detailed information on their contrasting prevalence and distribution among the various *B. tabaci* putative species. The last study (Chapter 4) describes population structure and genetic variability of *B. tabaci* by using both mitochondrial and nuclear markers. It further discusses the possibilities of gene flow and hybridization among sub-populations. Finally, this thesis is concluded with general conclusions and recommendations.

CONTENTS

List of Figures

List of Tables

(Page intentionally left blank)

1. Introduction

1.1 Background on Cassava

1.1.1 Cassava: Its introduction and importance in Africa

Cassava (*Manihot esculenta* Crantz) is believed to have originated in South America (Allem, 2002). Its domestication dates back to 5000-7000 years BC (Lathrap, 1970) in the Amazonian basin (Olsen and Schaal, 1999). The Portuguese were responsible for its introduction to several areas in the tropics and sub-tropics. In Africa, cassava was first introduced in the 16th century by Portuguese navigators who brought it from Brazil to West Africa (Jones, 1959). It was not until the end of the 18th century that cassava was grown in East Africa. It was recorded in Zanzibar in 1799 (Purseglove, 1968) and in Uganda in 1878 (Hillocks, 2002). Trade routes and colonial powers have contributed to the widespread cultivation of cassava in Africa during the 20th century (Hillocks, 2002). Soon after its introduction, the fact that it can tolerate and give acceptable yields in marginal environments where cereals and other crops fail gave cassava quick popularity among smallholder farmers and it became a staple food crop in several tropical and sub-tropical places (FAO, 2004).

At the beginning of the current millennium, cassava was the second most important staple crop in Africa. It was the main source of calories for two out of every five Africans (Nweke *et al.*, 2002). Globally cassava ranks sixth in production following sugarcane, maize, rice, wheat and potato (FAOSTAT, 2014).

Cassava is mainly grown for its starchy roots which can be consumed boiled, fried or processed to make flour or other products such as cassava chips and crisps. The carbohydrate rich roots constitute an important part of the dietary energy for several millions of people in the tropics (FAO, 2013). The leaves are good source of protein and are consumed as a vegetable, supplementing the starch-dominated diet of resource-poor households in the tropics (Lancaster and Brooks, 1983). Cassava roots are highly perishable and deteriorate from 24 to 72 hours after harvest (Alves, 2002). However, the roots can be stored in the field for several months without harvesting and can be dug up when necessary. This makes it possible to delay harvesting according to market demand and according to household needs. Thus cassava is considered a key food security crop (FAO, 2004). In addition to subsistence farming, the crop can be commercialized for processing, such as starch production. The importance of cassava is expected to grow even more in the future with increasing population, declining soil fertility and changing climate (Thresh and Cooter, 2005; Jarvis *et al*., 2012).

1.1.2 Cassava production in the world and in Africa

In the last two decades, on average, more than half of the world's cassava production came from Africa. In 2013 alone, Africa produced around 158 million tonnes of cassava, amounting to ca 57% of the world's production of which 54 million tonnes were from Nigeria – the most important producer of cassava both in Africa and in the world (FAOSTAT, 2014).

At present, cassava is the leading crop produced in Africa followed by sugarcane and maize. Tanzania was the sixth largest producer in Africa in 2013 with a total production of 5.4 million tonnes (FAOSTAT, 2014). Although over the years the production of cassava has increased in Africa, it is mainly attributed to increased production area rather than improved yields (Hillocks, 2002).

1.1.3 Constraints of cassava production in Africa

Despite the importance of cassava, little research was directed towards the improvement of the crop in the past (FAO, 2004). However, that changed nearly four decades ago with the establishment of research institutes in tropical agriculture, particularly the founding of International Institute of Tropical Agriculture (IITA) in Africa and the International Center for Tropical Agriculture (CIAT) in South America (Ceballos *et al.*, 2004). Subsequently several research programs in breeding, improvement of cultural practices and processing were initiated (Cock, 1985; Jennings and Iglesias, 2002).

Although Africa's cassava production constituted 57% of the global total in 2013, this came from 68% of the total area cultivated with cassava in the world, indicating the lower productivity in Africa in comparison with other parts of the world. For several years, discrepancies between production quantity and cultivated area in Africa have been attributed to poor yields and pest and diseases (Legg and Thresh, UD).

Legg and Thresh (UD) state the pests and diseases found in Africa are more damaging even if less diverse compared to those in Latin America. This is true as many of the pests are

exotic being introduced to Africa, just like the crop, from South America, hence incurring more damage in Africa due to absence of natural enemies, unlike in South America where the natural enemies have co-evolved with the pests (Hillocks, 2002; Legg and Thresh, UD). In Africa, the cassava green mite [*Mononychellus tanajoa* (Bondar)], the cassava mealybug (*Phenacoccus manihoti* Matile-Ferrero) and the whitefly *Bemisia tabaci* (Gennadius) are among the economically important pests on cassava. The first two have been under control after successful introduction of their natural enemies through classical biological control programs (CGIAR, 1997). Cassava bacterial blight, caused by *Xanthomonas axonopodis* pv. *manihotis* (Bondar) (Vauterin *et al*., 1995; Bull *et al*., 2010), and the two viral diseases, cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) caused by cassava mosaic geminiviruses (CMGs) and cassava brown streak viruses (CBSVs), respectively, constitute some of the most harmful diseases in cassava production in Africa (Legg and Thresh, UD; Thresh and Cooter, 2005).

For several decades CMD has been a main constraint to cassava production in Africa. It causes several million tonnes of crop loss in Africa amounting to more than a billion United States dollars (USD) every year (Thresh *et al.*, 1997; Legg and Fauquet, 2004; Legg *et al*., 2006).

1.2 Cassava Mosaic Disease

1.2.1 Historical overview

At the end of the 19th century, the symptoms of CMD were first reported from East Africa particularly from north-eastern part of what is now Tanzania (Warburg, 1894). At the

beginning of the 20th century, with no visible pathogen present, the disease was assumed to be caused by a virus (Zimmerman, 1906).

In the 1920s and 1930s the disease had become widespread and reports of its occurrence came from several cassava growing countries in sub-Saharan Africa (SSA), such as Sierra Leone (Deighton, 1926), Uganda (Hall, 1928), Cameroon (Dufrenoy and Hedin, 1929), Ghana (Dade, 1930), Ivory Coast (Hedin, 1931), Nigeria (Golding, 1936) and Madagascar (Francois, 1937).

Nearly three to four decades after the study by Zimmerman (1906), some of the earliest work on epidemiology, virus description, and transmission mechanism came from Amani Research Station (Storey 1936, 1938; Storey and Nichols, 1938). At the later years of the 20th century studies by Bock *et al*. (1978) and Bock and Woods (1983), from the Kenyan virology research project, gave the proof for the viral aetiology of CMD.

1.2.2 From an epidemic to a regional pandemic

Although the disease was prevalent in several countries across Africa in the early decades of the 20th century, it was not until the end of the 1980s that there was a large-scale outbreak of a severe form of the disease. Until that time the disease was managed using resistant cassava cultivars and phytosanitation practices (Legg, 1999). However, in 1988 an unusually severe type of CMD began spreading in north-central Uganda. From there, the epidemic continued to expand to other cassava growing regions of the country, devastating cassava production in several fields (Otim-Nape, 1988; Otim-Nape *et al*., 1997). The

epidemic together with drought in 1993-1994 forced farmers to abandon cultivating the crop and caused food shortages and deaths due to hunger and famine (Thresh and Otim-Nape, 1994).

The severe epidemic subsequently spread to neighbouring countries becoming a pandemic in the wider region of East and Central Africa (Otim-Nape *et al*., 1997). In the late 1990s and the beginning of 2000s it had already reached Kenya, Sudan, Democratic Republic of Congo (DRC), Tanzania and Rwanda (Otim-Nape *et al*., 1997; Legg, 1999). By 2005, an area of almost three millions square kilometers of East and Central Africa had been affected by the pandemic, covering large parts of nine countries (Legg *et al*., 2006).

In an extensive review of the losses caused by the pandemic, Legg *et al.* (2006) estimated mean losses from CMD-affected plants in areas not yet affected by the severe CMD pandemic to be 35%, whereas in areas already affected by the pandemic, the estimation rose to more than double that of the unaffected areas – amounting to 72%.

1.2.3 Cassava mosaic geminiviruses

After the discovery of the viral aetiology of CMD (Bock *et al*., 1978; Bock and Woods, 1983), the first isolate was named cassava latent virus (CLV) and later *African cassava mosaic virus* (ACMV) (*Geminiviridae*: *Begomovirus*) (Bock and Woods, 1983). DNA sequences contributed to the description of the ACMV and two other viruses - *East African cassava mosaic virus* (EACMV) (Bock and Woods, 1983) and *Indian cassava mosaic virus* (ICMV) (Hong *et al*., 1993).

After the outbreak of the severe CMD pandemic in Uganda, the reports of unusually severe symptoms led to the sequencing of virus isolates from severely symptomatic plants which revealed the occurrence of a recombinant virus between ACMV and EACMV termed the Uganda Variant (UgV) (Zhou *et al*., 1997) and later as EACMV-UG as it was considered to be a strain of EACMV (Deng *et al*., 1997). The coming years witnessed the discoveries of more CMGs such as *East African cassava mosaic Malawi virus* (EACMMV) (Zhou *et al.*, 1998), *East African cassava mosaic Cameroon virus* (EACMCV) (Fondong *et al*., 2000), *South African cassava mosaic virus* (SACMV) (Berrie *et al.,* 2001), *East African cassava mosaic Zanzibar virus* (EACMZV) (Maruthi *et al*., 2004), *East African cassava mosaic Kenya virus* (EACMKV) (Bull *et al*., 2006) and *Sri-Lankan cassava mosaic virus* (SLCMV) (Fauquet and Stanley, 2003).

Experiments by Kufferath and Ghesquière (1932) showed that viruses causing CMD are transmitted by a whitefly belonging to the genus *Bemisia* which was later confirmed by Storey and colleagues (Storey and Nichols, 1938). They also established that the viruses can be transmitted through grafts (Storey, 1936, 1938) and indicated that the viruses can be disseminated through the use of infected stem cuttings as planting materials (Bock and Woods, 1983). Chant (1958) and Dubern (1979, 1994) gave the final proof that *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) is the vector of CMGs.

1.2.4 Symptoms and characteristics of the severe CMD pandemic

Although CMD symptoms vary according to susceptibility of the cassava variety, virus species or strain and environmental conditions, the most common ones include yellow or yellow–green chlorotic mosaic of the leaves, malformation and distortion of leaves, and stunted growth (Storey, 1936; Fauquet and Fargette, 1990). In severe symptoms, as in the case of infection by EACMV-UG2 or even in more severe symptoms as caused by double infections by ACMV and EACMV-UG2 (Harrison *et al*., 1997; Pita *et al*., 2001), the petioles may even fall down (Legg and Thresh, UD) (Figure 1).

The pandemic of severe CMD is characterized by i) greatly increased populations of the whitefly vector (Legg and Ogwal, 1998) being referred to as 'super-abundant' (Legg, 2010); ii) high CMD severity and incidence dominated by whitefly-borne infection (Otim-Nape *et al*., 1997); and iii) a disease front – an area of disease expansion lying between the pandemic zone and the non-pandemic zone – where CMD incidence is moderate with mixed high and low CMD severities and high *B. tabaci* abundance (Otim-Nape *et al*., 1997; Otim-Nape and Thresh, 1998; Legg, 1999; Otim-Nape *et al*., 2000). Evidence has also been presented to suggest that satellite molecules may enhance the severity of CMD symptoms, although this association has not been geographically linked with the severe CMD pandemic (Ndunguru, 2005).

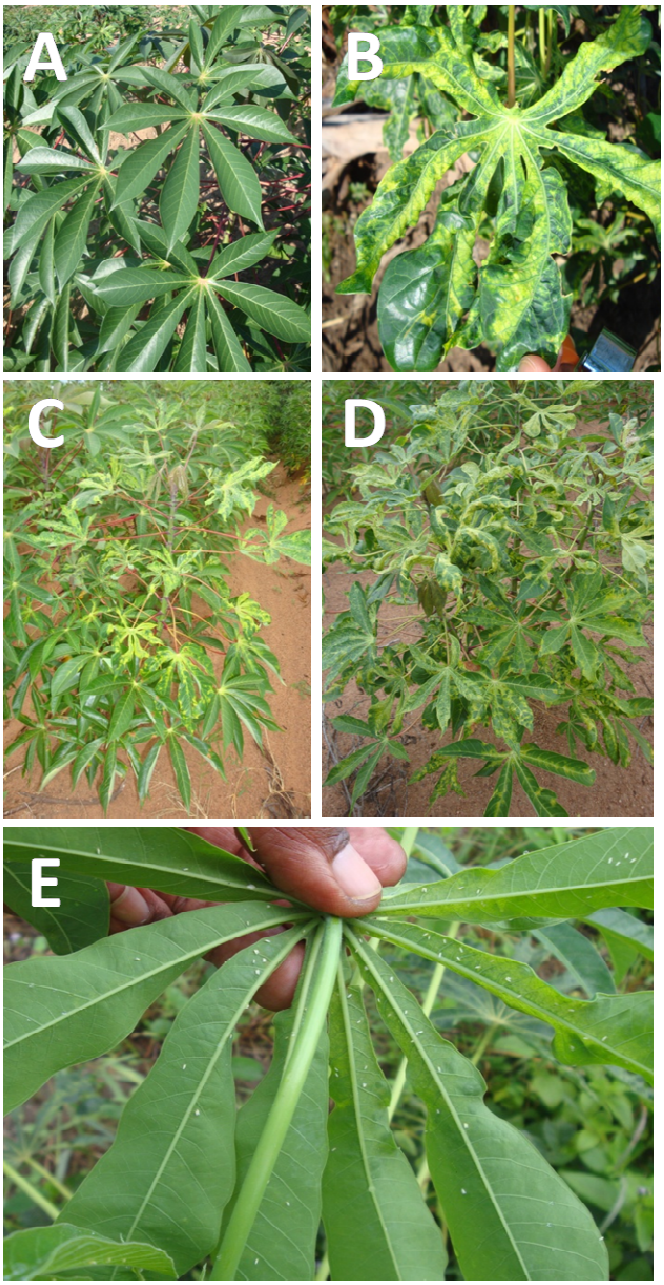


Figure 1 (A) Healthy cassava plant; (B) cassava leaf showing CMD symptoms; (C) diseased cassava plant showing severe whitefly-borne infection symptoms (symptoms appear from the top leaves); (D) diseased cassava plant exhibiting severe cutting-borne infection symptoms (symptoms manifest throughout); (E) cassava leaf highly infested with *Bemisia tabaci*.

1.3 *Bemisia tabaci* species complex

1.3.1 Historical background

Within the order Hemiptera, whiteflies belong to the suborder Sternorrhyncha which also contains aphids, psyllids, scale insects and mealybugs (Gullan and Martin, 2003). The whitefly *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) was first described in 1889 in Greece by Gennadius (1889) as *Aleyrodes tabaci* – a pest of tobacco. A quarter of a century later, the genus *Bemisia* was described by Quaintance and Baker (1914). However, *Aleyrodes tabaci* could not be placed in the new genus *Bemisia* due to insufficient information. The following five decades witnessed discovery and description of several whiteflies, many of which were younger synonyms of *Bemisia tabaci* (Russel, 1957). It was not until nearly fifty years after its discovery by Gennadius that *A. tabaci* was placed in the genus *Bemisia* by Takahashi under the name we now know: *B. tabaci* (Gennadius) (Takahashi, 1936). After puzzling scientists for five decades through synonymization events, *Bemisia tabaci* - now more than a century after its discovery - still continues to puzzle researchers with regards to its species status, which will be discussed in section 1.3.3.

1.3.2 A damaging cosmopolitan pest in global agriculture

Until recently *Bemisia tabaci* was known to be a pest in the tropics and sub-tropics whereas in temperate regions the greenhouse whitefly *Trialeurodes vaporariorum* (Westwood) was the most common and damaging (Stansly and Naranjo, 2010). In many countries, *B. tabaci* was only regarded as a secondary pest (Patti and Rapisarda, 1981).

However, in the 1980s the pest status of *B. tabaci* changed tremendously owing to global movement of plants and the explosion of protected agriculture which led to the increased pest status of the insect in temperate parts of the world as in the tropics and sub-tropics (Brown, 1990; Brown *et al.,* 1995a; De Barro *et al*., 2000; Martin *et al.,* 2000). Eventually *B. tabaci* was reported from all continents except Antarctica (Global Invasive Species Database, 2014). *B. tabaci* causes significant damage to agriculture both in protected environments and in open field cultivations (Brown *et al.*, 1995a; De Barro *et al.*, 2000; Martin *et al*., 2000) through phloem sap feeding, excretion of honeydew, induction of physiological disorders, and transmission of plant viruses (Brown *et al*., 1995a; Martin, 1999). Now *B. tabaci* is one of the most economically important insect pests.

Before *B. tabaci* attained global status, it had been reported to cause serious damage mainly to cotton production in several countries of which India, Sudan, Iran, Mexico, Brazil and Israel are a few examples (Misra and Lamba, 1929; Husain and Trehan, 1933; Basu, 1995). After the 1980s, severe infestations with *B. tabaci* were favoured by conducive climate coupled with intensive agriculture that meant the high use of insecticides and domination of monocultures, allowing for the development of resistance within certain *B. tabaci* populations. Hence it became a major constraint of food and fiber crops production worldwide (Brown, 1994). Likewise, in Mediterranean areas, severe damages in vegetables and ornamentals caused by *B. tabaci* and *T. vaporariorum* have been reported both in terms of physical sucking damage as well as through the transmission of viruses (Rapisarda, 1990; Gerling and Mayer, 1996). In the United States, losses incurred by *B. tabaci* infestations were estimated to be about 700 million USD in 1991 and 1992. In California

alone, an annual loss of more than 100 USD was incurred from 1991-1995 (Oliveira *et al*., 2001). At a social level, significant losses in personal income worth more than a million USD and loss of jobs have been recorded (Oliveira *et al*., 2001).

During 1991-1992, losses of more than 33 million USD were recorded in Mexico (Oliveira *et al*., 2001) and accumulated losses of 5 billion USD have been incurred in Brazil between 1995 and 2000 (Lima *et al*., 2000), resulting in several thousands of jobs being lost in the tomato industry (Oliveira *et al*., 2001). Pakistan was no exception to this situation where the cotton industry suffered losses of nearly 5 billion USD from 1994-1999 (Mansoor *et al*., 1999).

Similarly, in Africa, several million tonnes of crops are lost annually due to diseases caused by *B. tabaci* transmitted viruses. Losses for cassava have been estimated to be more than a billion USD every year (Thresh *et al.*, 1997; Legg and Fauquet 2004; Legg *et al*., 2006). In 2003 alone, a continent-wide loss of 19 to 27 million tonnes of cassava was incurred. Assuming 100 USD per tonne, Legg and Fauquet (2004) calculated the related financial loss to be from 1.9 to 2.7 billion USD.

1.3.3 From host race and biotype to putative species complex

Host range and host plant utilization variations in *B. tabaci* were first noted by Bird in the late 1950s where he reported the occurrence in Puerto Rico of two populations of *B. tabaci* (Bird, 1957). One of the populations was primarily found on *Jatropha gossypifolia* L. and was termed the Jatropha race whereas the other population was found on several plants

and was called the Sida race (Bird, 1957) introducing the host race concept to *Bemisia tabaci*.

Later studies like Costa and Russel (1975), Brown *et al.* (1992) and Burban *et al*. (1992), started to use the word biotype to differentiate populations. The concept of biotype became more pronounced when different esterase profiles were observed for the invasive biotype B that displaced the indigenous biotype A in southwestern USA (Brown *et al*., 1992; Brown *et al*., 1995b). The description of the okra and cassava biotypes from Ivory Coast by Burban *et al*. (1992) was followed by an explosion of various biotype descriptions throughout the world that are reviewed in Perring (2001) and De Barro *et al*. (2011).

In the last decade, much work has been done to clarify the taxonomic relationships between members of this cryptic species complex (Dinsdale *et al*., 2010; Xu *et al*., 2010; De Barro *et al*., 2011; Wang *et al*., 2011; Boykin *et al*., 2012; Liu *et al*., 2012). In 2010, Dinsdale *et al*. demonstrated the existence of 11 well defined high level genetic groups and at least 24 morphologically indistinguishable species based on a 3.5% genetic distance of the mitochondrial cytochrome oxidase I gene (mtCOI). Until this time there was no quantifiable boundary to differentiate *B. tabaci* (De Barro *et al.*, 2011). Based on (i) studies of reproductive isolation (Xu *et al*., 2010); (ii) a review of existing mating studies (conducted under laboratory conditions due to the fact that many of the *Bemisia* biotypes are allopatric) that revealed low or no proportion of female progenies in addition to being sterile or less fecund (examples include Costa *et al*., 1993; Byrne *et al*., 1995; Ronda *et al*., 1999; De Barro and Hart, 2000; Maruthi *et al*., 2004; Liu *et al*., 2007); and (iii) the 3.5%

genetic distance identified as the genetic divergence limit for the different *B. tabaci* biotypes (Dinsdale *et al*., 2010), De Barro *et al*. (2011) suggested that the *Bemisia* scientific community discontinues the use of 'Biotype' and moves on to accepting that *Bemisia tabaci* is a complex of at least 24 morphologically indistinguishable species.

Subsequently, further studies added more putative species and indicated that the *B. tabaci* species complex comprises at least 35 morphologically indistinguishable species (Hu *et al*., 2011; Alemandri *et al*., 2012; Chowda-Reddy *et al*., 2012; Parrella *et al*., 2012; Lee *et al*., 2013; Legg *et al*., 2013). Among the 35 morphologically indistinguishable species, sub-Saharan Africa 1 (SSA1) has recently been further divided into five sub-groups which include SSA1-SG1, SSA1-SG2, SSA1-SG3, SSA1-SG4 and an intermediate group between SG1 and SG2 termed SSA1-SG1/2 (Legg *et al.,* 2013). These morphologically indistinguishable species vary in terms of their host utilization, fecundity, insecticide resistance and virus transmission ability (Bird, 1957; Costa *et al.,*1993; Bedford *et al.,*1994; Liu *et al.,*2010; Xu *et al.,* 2011).

1.3.4 Some molecular markers used in *Bemisia tabaci* species complex analysis

Morphological identification of whiteflies is exclusively based on the fourth instar nymph (also called as pupal-case) (OEPP/EPPO, 2004) and *Bemisia* is no exception. Nevertheless, the fact that members of the *B. tabaci* species complex are morphologically indistinguishable has necessitated the deployment of molecular approaches to identify them and to study their genetic variation. Some examples include the use of isozymes such

as esterases, Random Amplified Polymorphic DNA (RAPD), mitochondrial 16S and mitochondrial cytochrome oxidase I (mtCOI), and microsatellites and molecular markers such as the internal transcribed spacer I (ITS-1) (Costa and Brown, 1991; Simon *et al*., 1994; Brown *et al*., 1995b; Frohlich *et al*., 1999; De Barro *et al*., 2000; Legg *et al*., 2002; Maruthi *et al*., 2004; Bosco *et al*., 2006; Delatte *et al*., 2005, 2006; Sseruwagi *et al*., 2006; Dinsdale *et al*., 2010; Mugerwa *et al*., 2012; Parrella *et al*., 2012). However, some drawbacks, such as poor reproducibility of esterase and RAPDs analyses have been observed (Brown, 2007). Moreover, Frohlich *et al*. (1999) showed that less divergence was exhibited by the 16S sequences than mtCOI sequences which revealed higher nucleotide divergence. Thus, the mtCOI coding region has been considered highly informative and many studies have used it to assess genetic variation and study phylogenetic relationships in the *B. tabaci* species complex (Frohlich *et al*., 1999; Legg *et al*., 2002; Maruthi *et al*., 2004; Delatte *et al*., 2005; Sseruwagi *et al*., 2006; Dinsdale *et al*., 2010; Mugerwa *et al*., 2012; Parrella *et al*., 2012). However, species distinctions among the *B. tabaci* species complex remain unresolved (Boykin *et al*., 2007).

In the last decade, the utilization of microsatellites became common for studying the genetic and population structure of the *Bemisia tabaci* species complex (De Barro, 2005; Delatte *et al*., 2006, 2011; Simon *et al.*, 2007; Tsagkarakou *et al.*, 2007; Dalmon *et al.*, 2008; Tsagkarakou *et al.*, 2012). Such an approach is particularly interesting for assessing and evaluating the evolutionary aspects of cryptic species such as *Bemisia tabaci* whose molecular phylogeny has been more or less dominated by the use of a single marker (mtCOI) (Hadjistylli, 2010).

1.3.5 Association of *Bemisia tabaci* with cassava in Africa

Within the *B. tabaci* species complex, the cassava types are distinct from other *B. tabaci* populations found on non-cassava plants, and are restricted to cassava (Abdullahi *et al.,* 2003). However, the cassava *B. tabaci* can colonize and complete their life cycle on non-cassava hosts (Gachoka *et al*., 2005; Sseruwagi *et al*., 2006). By contrast, non-cassava *B. tabaci* do not colonize cassava even if they are polyphagous (Abdullahi *et al.,* 2003).

In a geminiviruses transmission study using *Bemisia tabaci* from different geographic regions, Bedford *et al*. (1994) reported that 'B' and some other 'non-B biotypes' transmit several geminiviruses with varying efficiencies. However, *African cassava mosaic virus* (ACMV) was shown not to be transmitted back to cassava in that study which suggested that in areas such as Africa and Asia indigenous *B. tabaci* populations transmit endemic geminiviruses. This was further supported by Maruthi *et al*. (2002) who noted that in cassava crops, viruses co-adapt with their *Bemisia tabaci* vector. In that study, CMGs from Africa and from India were readily transmitted by their respective African and Indian *B. tabaci* populations. Conversely, in a reciprocal transmission trial, both the African and Indian *B. tabaci* populations transmitted their opposing continent viruses in very low proportions (Maruthi *et al*., 2002) showing the strict association of *Bemisia tabaci* with cassava.

In addition to *B. tabaci*, several other whiteflies such as *Aleurodicus disperses* (Russell), *Bemisia afer* (Priesner and Hosny) and *Trialeurodes vaporariorum* Westwood can be found

in cassava crops. In a recent study eight whitefly species including the above four as well as an additional four including *Aleurothrixus floccosus* (Mask), *Bemisia* sp. (formerly *Asterobemisia* sp.), *Dialeurodes citri* (Ashmead) and *Paraleyrodes bondari* Peracchi were found colonizing cassava naturally in the field (Guastella *et al*., 2014).

1.3.6 Super-abundance of *Bemisia tabaci* in sub-Saharan Africa

At the end of the 1980s and the beginning of the 1990s clouds of whiteflies were easily seen flying over farms and destroying crops in southwestern USA (Gill and Brown, 2010). About the same time, *Bemisia tabaci* was also undergoing similar changes in numbers in Africa. In the late 1980s, at the outbreak of the severe CMD pandemic, increases in *B. tabaci* populations were prominent in pandemic affected areas in East and Central Africa. In Uganda, for example, Legg and Ogwal (1998) observed higher abundances in the northern disease-affected areas than in the southern part not yet affected by the severe CMD.

As the pandemic continued to advance, some years later, hundred-fold increases were observed in whitefly abundance in pandemic affected areas, and the unusually high populations of *B. tabaci* in these areas became known as 'super-abundant' (Legg *et al.,* 2006; Legg, 2010). Super-abundant *B. tabaci* populations, in addition to virus transmission, are very damaging as they cause tremendous feeding damage and excrete honeydew which facilitates the development of sooty mould on the leaves (Legg *et al.,* 2006), in both cases reducing the photosynthetic area.

In seeking a causal factor behind the emergence of super-abundant *B. tabaci* populations in East and Central Africa, two hypotheses have been suggested. The first suggests that a distinct genotype cluster is associated with the severe CMD pandemic (Legg *et al*., 2002) whereas the second suggests that the *B. tabaci* population explosion is the consequence of a synergistic interaction of the vector with CMG-infected host plants (Colvin *et al*., 2004, 2006). However, no definitive proof has yet been presented in support of either hypothesis.

1.3.7 Endosymbionts of *Bemisia tabaci*

Several studies in the second half of the 20th century have showed symbiotic associations between bacteria and insects in nature (Buchner, 1965; Dasch *et al*., 1984; Douglas, 1989; Baumann and Moran, 1997). There is a general assumption that endosymbionts are usually found in insects with imbalanced diet (Buchner, 1965) and their role as a result is to synthesize the missing nutrients that are needed by their host (Baumann *et al*., 2006). The carbohydrate-rich plant phloem sap which whiteflies and many other insects feed on is nutritionally unbalanced and poor in essential amino acids that are useful for survival. Insects supplement this amino acid poor diet with the help of the endosymbiotic bacteria they harbour (Buchner, 1965; Douglas, 1998; Baumann, 2005; Singh *et al*., 2012). In fact, the ratio of essential to nonessential amino acids in the phloem sap is estimated to range from 1:4 to 1:20 showing the very limited presence of essential amino acids in the carbohydrate rich sap. Interestingly, this ratio becomes roughly 1:1 in the insect body indicating the presence of some mechanism in the insect body that converts the nonessential amino acids to essential ones (Liadouze *et al*., 1995; Douglas, 2006). Several studies have presented evidence that endosymbionts compensate for the essential amino

acids that are missing in the phloem sap, such as *Buchnera* in aphids (Buchner, 1965; Baumann *et al*., 1995; Liadouze *et al*., 1995; Douglas, 2006).

Symbiotic bacteria can either be obligate (primary) or facultative (secondary) (Buchner, 1965). Primary symbionts are essential for the host's survival and development and are located within specialized cells called bacteriocytes or mycetocytes that make up a larger structure called the bacteriome or mycetome (Buchner, 1965; Baumann, 2005). On the other hand, facultative or secondary symbionts are not essential for the host's survival, development or reproduction and neither are they necessarily found in the bacteriocytes or mycetocytes (Buchner, 1965). Although the exact role of secondary symbionts is not clearly known, they are more diverse than primary symbionts and are thought, among other things, to assist their insect hosts in adaptation to new host plants, in pesticide resistance and tolerance to heat shock (Campbell, 1989, 1993; Nardon and Grenier, 1991; Costa *et al*., 1995; Zchori-Fein and Brown, 2002; Brumin *et al*., 2011).

One of the earliest works in *Bemisia* endosymbionts was conducted on laboratory reared whiteflies by Costa *et al.* (1995). The study described morphological variations between three endsymbionts, two of which were shared commonly by all the populations tested. One of these was later identified to be the primary endosymbionts – *Portiera aleyrodidarum* (Thao and Baumann, 2004) – harboured by all *B. tabaci* tested from the various locations. Variations in the endosymbionts harboured by the different *Bemisia* populations tested were also noted, and based on such observations the authors suggested that different *B. tabaci* populations may have endosymbionts unique to them (Costa *et al*.,

1995). The remaining endosymbionts were identified to be secondary endosymbionts. Another study further confirmed the presence of the primary endosymbionts in all the *B. tabaci* populations they tested from various host plants world-wide (Zchori-Fein and Brown, 2002). However, not all populations contained secondary symbionts indicating that not all individuals harbour secondary symbionts. Subsequently more secondary endosymbionts were described. To date *B. tabaci* is believed to harbour seven secondary symbionts that include *Rickettsia, Hamiltonella, Arsenophonus, Cardinium, Wolbachia*, *Fritschea* and *Hemipteriphilus* (Zchori-Fein and Brown, 2002; Weeks *et al.*, 2003; Thao and Baumann, 2004; Everett *et al.*, 2005; Moran *et al.,* 2005; Gottlieb *et al.*, 2006; Bing *et al.*, 2013) (Table 1).

Table 1 Primary and secondary symbionts found in *Bemisia tabaci* (adopted and slightly modified from Rosell *et al.*, 2010)

Type	Symbiont name	Reference
Primary	*Portiera aleyrodidarum*	Thao and Baumann, 2004
Secondary	*Hamiltonella*	Moran *et al.*, 2005; Chiel *et al.*, 2007
"	*Arsenophonus*	Thao and Baumann, 2004a; Zchori-Fein *et al.,* 2004; Chiel *et al.,* 2007
"	*Cardinium*	Baumann *et al.*, 2004; Zchori-Fein *et al.*, 2004; Chiel *et al.*, 2007
"	*Fritschea*	Zchori-Fein and Brown, 2002; Thao *et al.*, 2003; Weeks *et al.*, 2003; Everett *et al.,* 2005
"	*Rickettsia*	Gottlieb *et al.,* 2006; Chiel *et al.*, 2007
"	*Wolbachia*	Zchori-Fein and Brown, 2002; Nirgianaki *et al.*, 2003; Li *et al.,* 2007; Chiel *et al.,* 2007
"	*Hemipteriphilus*	Bing *et al.,* 2013

1.4 Problem statement and objectives of the study

After the outbreak, around 25 years ago, of the unusually severe CMD in East and Central Africa, several studies have focused mainly on understanding the viruses and the pandemic. Advances have been made in identifying the viruses involved including the discovery of recombinant viruses. Studies have also indicated that the CMD pandemic in East and Central Africa is mainly characterised by super-abundant populations of the whitefly vector *B. tabaci* and a high CMD severity and incidence dominated by whitefly-borne infection (Gibson *et al*., 1996; Otim-Nape *et al*., 1997). Super-abundance of *B. tabaci* populations has been of even greater concern since the mid-2000s, as a new outbreak of cassava brown streak viruses (CBSVs) has been reported from parts of pandemic-affected East Africa (Alicai *et al*., 2007) and it appears that the pandemic may now have become a yet more devastating dual pandemic of CMD and cassava brown streak disease (CBSD) (Legg *et al*., 2011).

In seeking a causal factor for the emergence of super-abundant *B. tabaci* populations in East and Central Africa, preliminary suggestions have been made that a distinct genotype cluster is associated with the pandemic (Legg *et al*., 2002). It has also been suggested that the *B. tabaci* population explosion is the consequence of a synergistic interaction with CMG-infected host plants (Colvin *et al*., 1999, 2004). However, no definitive proof has yet been presented in support of either hypothesis.

To date the prevalence and distribution of endosymbiotic bacteria among the *B. tabaci* species complex in the pandemic affected East and Central African countries remains

unknown. Moreover, population structure, possibilities of gene flow and occurrence of hybridization events are unstudied for the *B. tabaci* populations occurring in pandemic areas.

It is widely agreed that understanding *B. tabaci* is crucial in controlling and managing the severe CMD pandemic that threatens the food security crop of several millions of people in sub-Saharan Africa. Despite this fact, much remains poorly understood about the vector in CMD pandemic affected areas.

Therefore, the present study was initiated to fill the gap in this regard. The aim was to study the super-abundant *B. tabaci* species complex associated with cassava crops in CMD pandemic-affected areas of Tanzania. The three main objectives were:

1. To assess the genetic diversity and distribution of cassava associated *B. tabaci* in CMD pandemic-affected areas in Tanzania and to establish the association of a distinct *B. tabaci* genotype with the pandemic

2. To establish the prevalence of endosymbionts among *B. tabaci* populations on cassava and nearby plants, to assess if there is a unique endosymbiont that is associated with super-abundant whitefly populations and to provide baseline information on endosymbionts of whiteflies for Tanzania

3. To investigate the population structure of and the possibilities of gene flow between *B. tabaci* populations found in CMD pandemic affected and not yet affected parts of Tanzania.

2. Abundance, diversity and geographic distribution of cassava mosaic disease pandemic associated *Bemisia tabaci* in Tanzania[1]

2.1 Introduction

Bemisia tabaci (Gennadius) (Hemiptera: Aleyrodidae) is a cryptic species complex (Brown *et al.*, 1995; Frohlich *et al.*, 1999) showing high genetic variation and comprising morphologically indistinguishable populations (Bedford *et al.*, 1994; Brown *et al.*, 1995). Several of these indistinguishable populations are known for causing serious damage to agricultural crops in tropical and sub-tropical parts of the world (Brown, 1990; Brown *et al.*, 1995; De Barro *et al.*, 2000; Martin *et al.*, 2000) of which damages caused by the transmission of plant viruses are severe. From the numerous plant viruses transmitted by *B. tabaci*, the cassava mosaic geminiviruses (CMGs) are considered among the most important, causing one of the most economically important diseases to cassava – cassava mosaic disease (CMD) (Legg and Fauquet, 2004).

The severe CMD that originated from Uganda in the late 1980s and spread to neighbouring countries was shown to be associated with the spread of a recombinant virus, which contained elements of two CMG species: *African cassava mosaic virus* (ACMV) and *East African cassava mosaic virus* (EACMV) and was referred to as the Uganda variant (UgV) (Zhou *et al.*, 1997) or EACMV-UG (Deng *et al.*, 1997).

[1] Tajebe LS, Boni SB, Guastella G, Cavalieri V, Lund OS, Rugumamu CP, Rapisarda C, Legg JP. 2014. Abundance, diversity and geographic distribution of cassava mosaic disease pandemic associated *Bemisia tabaci* in Tanzania. J. Appl. Entomol. (Accepted Article) doi: 10.1111/jen.12197

Legg (1999) predicted a 300 km wide spread of the CMD pandemic from Uganda to neighbouring countries between 2000 and 2010, with an average rate of spread of approximately 30 km/yr. Since that time, severe CMD, and more recently cassava brown streak disease (CBSD) have spread widely throughout East and Central Africa. CBSD is also caused by viruses transmitted by *B. tabaci* (Maruthi *et al*., 2005).

The association between a distinct genotype group of *B. tabaci* and the severe CMD pandemic was first reported for southern Uganda in the late 1990s (Legg *et al*., 2002). In this situation, the putative species sub-Saharan Africa 2 (SSA2) was shown to predominate in the pandemic-affected parts of Uganda. More recently, a broader regional association has been demonstrated between a sub-group of sub-Saharan Africa 1 *B. tabaci* (SSA1-SG1) and the spread of severe CMD through East and Central Africa (Mugerwa *et al*., 2012; Legg *et al*., 2013). In neither of these two studies, however, were whitefly data linked directly to CMD. The study presented here aimed to address this gap by combining whitefly and CMD datasets in a way that would allow a more precise association to be defined between *B. tabaci* genetic groups and the CMD pandemic. Knowledge of this association and likely future patterns of vector-driven disease spread will be vitally important for the development of effective and sustainable approaches to managing cassava whiteflies in Tanzania and the wider East and Central African region.

2.2 Materials and Methods

2.2.1 Study area and whitefly collection

Adult whitefly samples were collected during four surveys carried out in Tanzania from 2010 to 2013. All samples were labeled and stored in 95% ethanol prior to laboratory analysis. The surveys were organized in a manner to cover the cassava-growing areas of the country and including both CMD pandemic and non-pandemic areas. To this effect two transect surveys were conducted from the coast (including Zanzibar Island) running through the central part and extending to the north-west of the country (Figure 2). The remaining two surveys focused on north-western Tanzania, in an area bounded by the coordinates 3º and 5º (South) latitude and 32º and 34º (East) longitude. This was the zone within which the severe CMD pandemic was expanding from north-west to south-east (Legg, 2010). In general, young cassava fields of less than one year old with an approximate average size of 1500 m^2 were assessed.

Survey 1: In 2010 eighty-nine whitefly populations were collected on cassava plants from 103 cassava fields in a transect survey that extended from Kagera Region, bordering Uganda to Coastal Region, including Zanzibar Island.

Survey 2: In a field survey conducted in north-western Tanzania during 2012, in the area bounded by the coordinates 3º - 5º S latitude and 32º - 34º E longitude, 88 whitefly populations from 80 cassava fields and surrounding plants were collected.

Survey 3: In 2013, a follow-up survey was conducted in north-western Tanzania in the same area as that of Survey 2. In this case, 15 whitefly populations were collected from 19 cassava fields.

Survey 4: An additional transect survey was carried out in 2013 covering parts of six Regions. In this survey, 45 whitefly populations were collected from 56 cassava fields both from the CMD pandemic and non-pandemic areas (Table 2).

The surveyed fields were divided into three groups of districts in each of the three main regional zones: the Lake Zone (group 1), Central Zone (group 2) and Eastern Zone (group 3) (Figure 2). GPS coordinates were recorded for each field assessed and were used with DIVA-GIS software (Version 7.5.0.0) to construct maps for CMD parameters and the distribution of *B. tabaci* putative species.

2.2.2 Host plants

Host plants from which whiteflies were collected included cassava (*Manihot esculenta* Crantz), pumpkin (*Cucurbita pepo* L.), sweet potato (*Ipomoea batatas* (L.) Lam.), a Lamiaceae weed - klip dagga or lion's ear (*Leonotis nepetifolia* (L.) R. Br.), okra (*Abelmoschus esculentus* (L.) Moench), tomato (*Solanum lycopersicum* L.), cotton (*Gossypium hirsutum* L.) and sunflower (*Helianthus annuus* L.) (Table 2).

2.2.3 CMD incidence and severity, pandemic front determination and whitefly abundance

Data for CMD incidence, severity and infection type were collected on thirty cassava plants selected at regular intervals along an 'X' diagonal across each field during Survey 1. CMD incidences and severity levels were determined as per Sseruwagi *et al*. (2004) where incidence was calculated as the number of diseased cassava plants out of the total plants assessed and severity as the extent of expression of the disease symptoms based on a 1-5 scale where 1 represents plants that are healthy with no symptoms and 5 represents plants that are severely diseased (Hahn *et al*., 1980). The type of infection for diseased cassava plants was also recorded either as cutting-borne (C) or whitefly-borne (W). Cutting-borne infections are distinguished from whitefly-borne infections as they cause symptoms on the lowest first-formed leaves whereas in the latter case symptoms appear only on the upper-most leaves (Sseruwagi *et al*., 2004). The calculated means of the parameters: CMD incidence and severity were compared with the CMD pandemic characteristics in accordance with previous studies (Otim-Nape *et al*., 1997; Otim-Nape and Thresh, 1998; Legg, 1999; Legg and Fauquet, 2004).

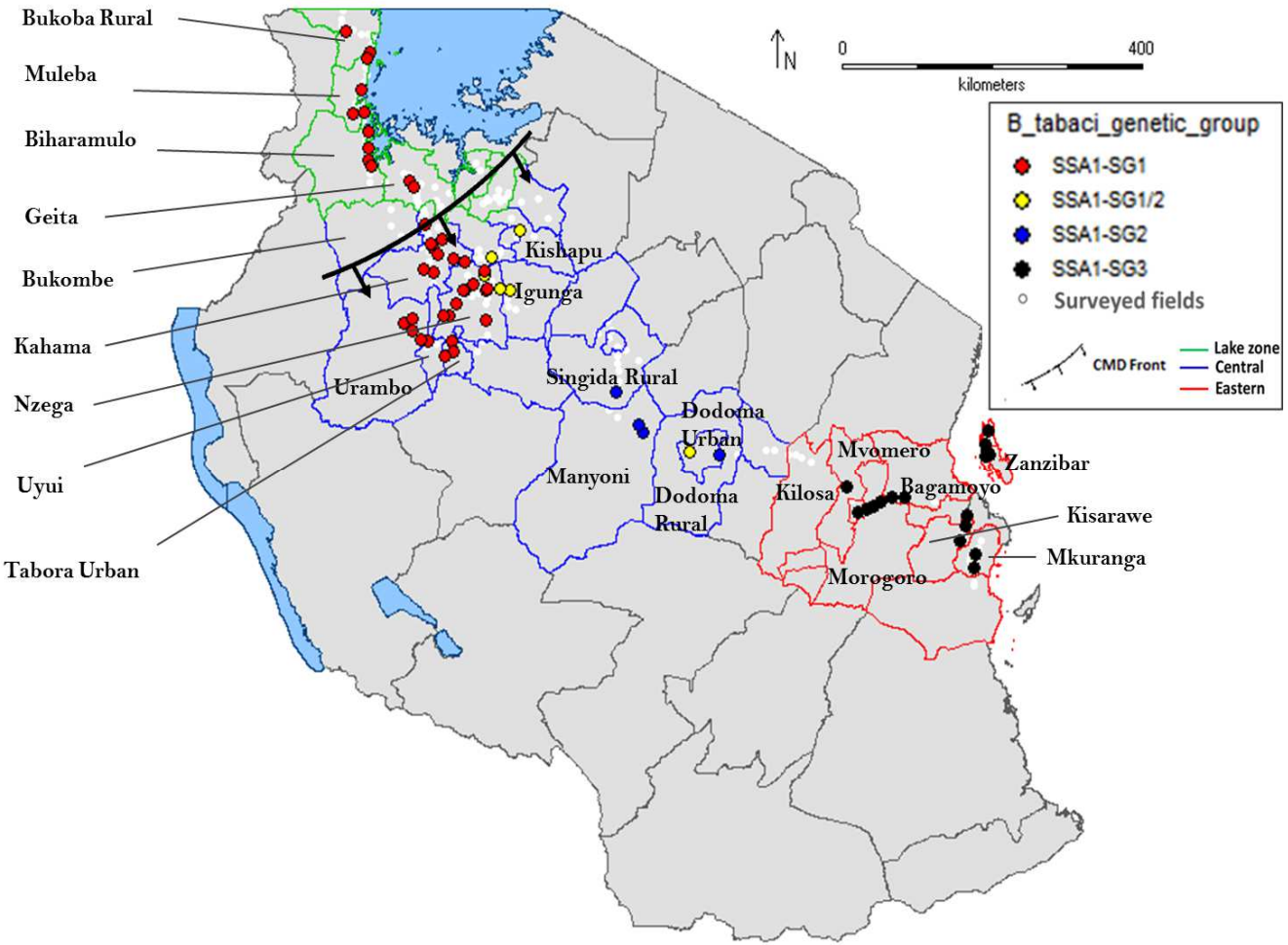


Figure 2 A north-west to coast transect of surveyed farmers' cassava fields and the location and direction of the CMD pandemic front and of the identified Bemisia tabaci SSA1 sub-groups in Tanzania. Each point represents a surveyed field and their colours differentiate the different SSA1 sub-groups. SSA1 stands for sub-Saharan Africa 1, SG for sub-group, and CMD for cassava mosaic disease. The three district zones are indicated by different coloured lines where green indicates the Lake zone, blue the Central zone and red the Eastern zone.

Table 2 Host plants and number of fields surveyed during 2010, 2012 and 2013 in Tanzania

Survey 1: Year 2010			Survey 2: Year 2012			Survey 3: Year 2013			Survey 4: Year 2013		
Region	Number of fields surveyed	Host plants surveyed	Region	Number of fields surveyed	Host plants surveyed	Region	Number of fields surveyed	Host plants surveyed	Region	Number of fields surveyed	Host plants surveyed
Coast	14	Cassava	Mwanza	8	Cassava	Geita	1	Cassava	Coast	5	Cassava
Dodoma	4	”			Pumpkin	Shinyanga	10	”	Dodoma	1	”
Kagera	30	”			Cassava	Tabora	8	”	Geita	2	”
Morogoro	10	”			Cotton	**Subtotal**	**19**		Mwanza	4	”
Mwanza	7	”			Klip dagga				Shinyanga	13	”
Shinyanga	10	”	Shinyanga	33	Okra				Singida	14	”
Singida	4	”			Pumpkin				Tabora	17	”
Tabora	8	”			Sweet potato				**Subtotal**	**56**	
Zanzibar	16	”			Sunflower						
Subtotal	**103**				Tomato						
					Cassava						
			Tabora	39	Klip dagga						
					Pumpkin						
					Sweet potato						
			Subtotal	**80**							
Total	**258**										

The CMD pandemic front was determined based on the levels of whitefly abundance, as well as CMD incidence, severity and infection type for sampled farms in districts around the location of the previously estimated position for the front (Legg, 2010). The linear distance between the CMD pandemic front location identified from this study and that recorded by Legg (1999) was measured on a GIS map and the rate of expansion of the CMD pandemic calculated. The location of the CMD pandemic front determined along the transect was extrapolated to both east and west in order to illustrate its proposed position on the map.

During all four surveys, whitefly abundance was recorded for cassava by counting adult *B. tabaci* on the undersides of the young fully-opened topmost five leaves of the thirty cassava plants selected in a similar 'X' diagonal manner as stated above for CMD data collection.

2.2.4 DNA extraction, polymerase chain reaction (PCR) and mtCOI sequencing

From the total of 237 whitefly samples collected during the four surveys, 155 samples were randomly selected based on districts surveyed and up to three female individuals from each were used for DNA extraction making a total of 269 whitefly individuals analyzed. DNA was extracted using a standard Chelex extraction method modified from White *et al*. (2009). Amplification of mitochondrial DNA fragments (mtCOI) from whitefly DNA was done using the forward primer C1-J-2195 (5' TTGATTTTTTGGTCATCCAGAAGT 3') and the reverse primer L2-N-3014 (5' TCCAATGCACTAATCTGCCATATTA 3') (Simon *et al*., 1994). PCR was performed using 2µl aliquots of the DNA extract in a total reaction volume of

10.5µl containing 0.8x FailSafe premix, 0.48 µM of each primer and 1.25 units *Taq* DNA polymerase (Invitrogen™). Thermocycler conditions were as follows: initial denaturation for 3 min at 94°C followed by 35 cycles of 1 min denaturation at 94°C, 1 min annealing at 52°C and 1:20 min amplification at 72°C followed by 10 min final extension at 72°C. Aliquots of 4 µl PCR products were separated on a 1.6% agarose gel and visualized under UV light using SYBR® Safe (Invitrogen). Samples that gave bands of the expected size, ~850 bp, were sequenced at Macrogen Inc. (USA) or BMR Genomics (Italy).

2.2.5 Sequence alignment and phylogenetic analysis

A total of 146 sequences were obtained ranging from 592 - 817 bp. The sequences were cleaned manually and ends trimmed using Chromas Lite (version 2.1) and aligned using the Clustal Omega option of SeaView (version 4.4.2) (Galtier *et al*., 1996; Gouy *et al*., 2010). Sequence identities were first confirmed using NCBI-BLAST and the whitefly mtCOI sequences produced from this study were analyzed together with selected reference sequences from the global *Bemisia* dataset (release version 31_Dec_2012) (De Barro and Boykin, 2013) and from GenBank. A jModel test (jModelTest version 2.1.3) was performed and HKY+I+G was used as the best fit nucleotide substitution model (Guindon and Gascuel, 2003; Posada, 2008; Darriba *et al*., 2012). MrBayes (version 3.2) was used to construct phylogenetic trees (Huelsenbeck and Ronquist, 2001). It was run for 10 million generations and trees were sampled every 1000 generations. All runs reached a plateau in likelihood score by a standard deviation of split frequencies of 0.003 with a potential scale reduction factor equal to one. A phylogenetic tree was first constructed comprising all 146 sequences produced in this study, several other reference sequences from the global *Bemisia* dataset

and a set of various out-groups. A representative set of sequences from all were selected to construct the final phylogenetic tree. This analysis comprised 42 sequences with a standard length of 672 bp. The outgroup used for the final tree was *Trialeurodes vaporariorum* (Westwood), GenBank code: JX841216. The trees were summarized using Fig Tree (version 1.4.0 2012). FastTree was used to infer approximate maximum likelihood trees (Price *et al*., 2010) for the group of unidentified whiteflies.

2.2.6 Statistical analyses

The data collected for the number of whiteflies per plant was statistically analyzed using GraphPad Prism® 6 software comparing means of whiteflies per region. Data for CMD incidence and severity were analyzed using SigmaPlot® software in One Way ANOVA on ranks using Dunn's test by comparing their means from the three district groups. Chi-square statistics were used to compare frequencies of severity scores between district groups.

2.3 Results

2.3.1 Phylogenetic analysis of whiteflies studied

A phylogenetic tree based on mtCOI sequences of adult *B. tabaci* revealed five distinct groups which fall under the Sub Sahara Africa 1 (SSA1), Mediterranean (MED), Indian Ocean (IO) and East Africa 1 group (EA1) as well as a group of unidentified whiteflies (Figure 3). The SSA1 genetic group was further divided into three distinct and one intermediate sub-groups: SSA1 sub-group 1 (SSA1-SG1), SSA1 sub-group 2 (SSA1-SG2),

SSA1 sub-group 3 (SSA1-SG3) and SSA1-SG1/2 - a group that was intermediate between SSA1-SG1 and SSA1-SG2 (Figure 3).

SSA1-SG1 was *ca* 1.6% divergent from SSA1-SG2 and 1.4 - 1.6% from SSA1-SG3 showing a close relationship to a previously published haplotype AY057180 from Uganda (Legg *et al.*, 2002). SSA1-SG2 diverged by *ca* 1.4% from SSA1-SG3. Within cluster divergence of 0.0% - 0.2% was observed for individuals in SSA1-SG3 which were also very close to the previously published haplotype AF418667 from Tanzania (Maruthi *et al.*, 2004). The intermediate group SSA1-SG1/2 diverged by 0.7% from SSA1-SG1 and by 0.9% from SSA1-SG2.

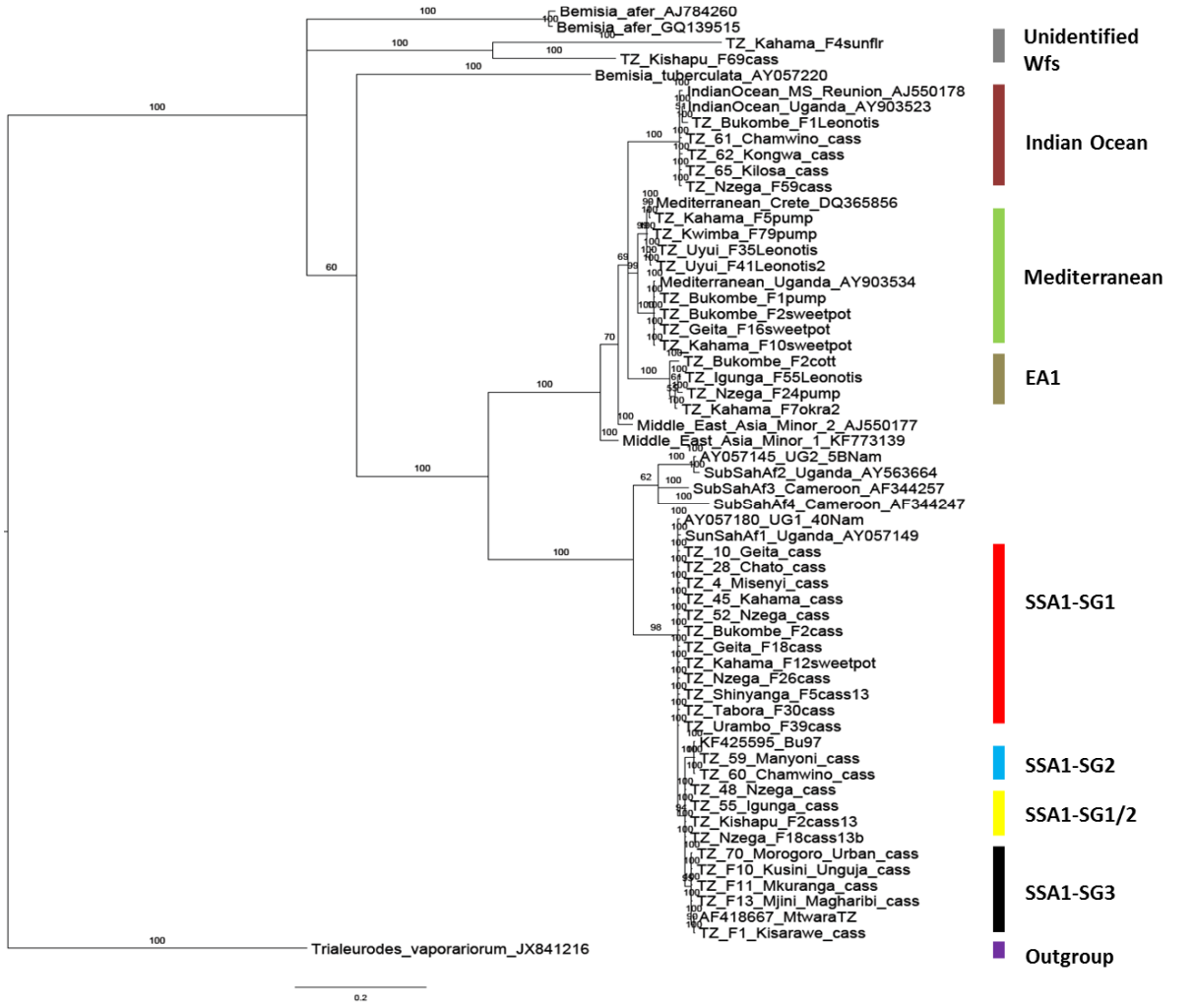


Figure 3 Phylogenetic tree based on bayesian inference of mtCOI sequences of adult whiteflies. Sequences that start with 'TZ_' are produced from this study whereas all other remaining sequences are reference sequences from GenBank or global *Bemisia* dataset. SSA1 stands for sub-Saharan Africa 1, SG for sub-group, EA1 for East Africa 1 genetic group, and WFs for whiteflies.

Bemisia tabaci from Tanzania in the MED genetic group diverged by at least 5% from those in the IO genetic group and by more than 16% from the SSA1 group. Individuals in the EA1 genetic group diverged by at least 4.3% and 4.6% from reference sequences of the MED and Middle East Asia Minor (MEAM) genetic groups, respectively. The individuals identified as EA1 in this study were similar to the recently published haplotype KF425620 from

north-western Tanzania (Legg *et al*., 2013) but diverged by more than 15% from SSA1 sequences.

The two unidentified whitefly individuals were highly divergent from the *B. tabaci* species complex (Figure 4). They differed by 21% from the *B. tabaci* species complex, 15% or more from *Bemisia afer* (Priesner & Hosny) and more than 22% from *Bemisia tuberculata* Bondar. These two individuals diverged by 18.2% from each other. These are both likely to be previously undescribed species of *Bemisia* or *Bemisia* species for which no reference sequence has been deposited in GenBank.

Out of the 146 individuals sequenced, the majority belonged to SSA1 (71.9%) followed by MED (13%), EA1 (9.6%) and Indian Ocean (4.1%). From the four sub-groups identified under the SSA1 genetic group, SSA1-SG1 accounted for 56.8% followed by SSA1-SG3 (9.6%), the intermediate sub-group SSA1-SG1/2 (4.1%) and SSA1-SG2 (1.4%). The two unidentified whitefly species accounted for 1.4% of the whiteflies sequenced.

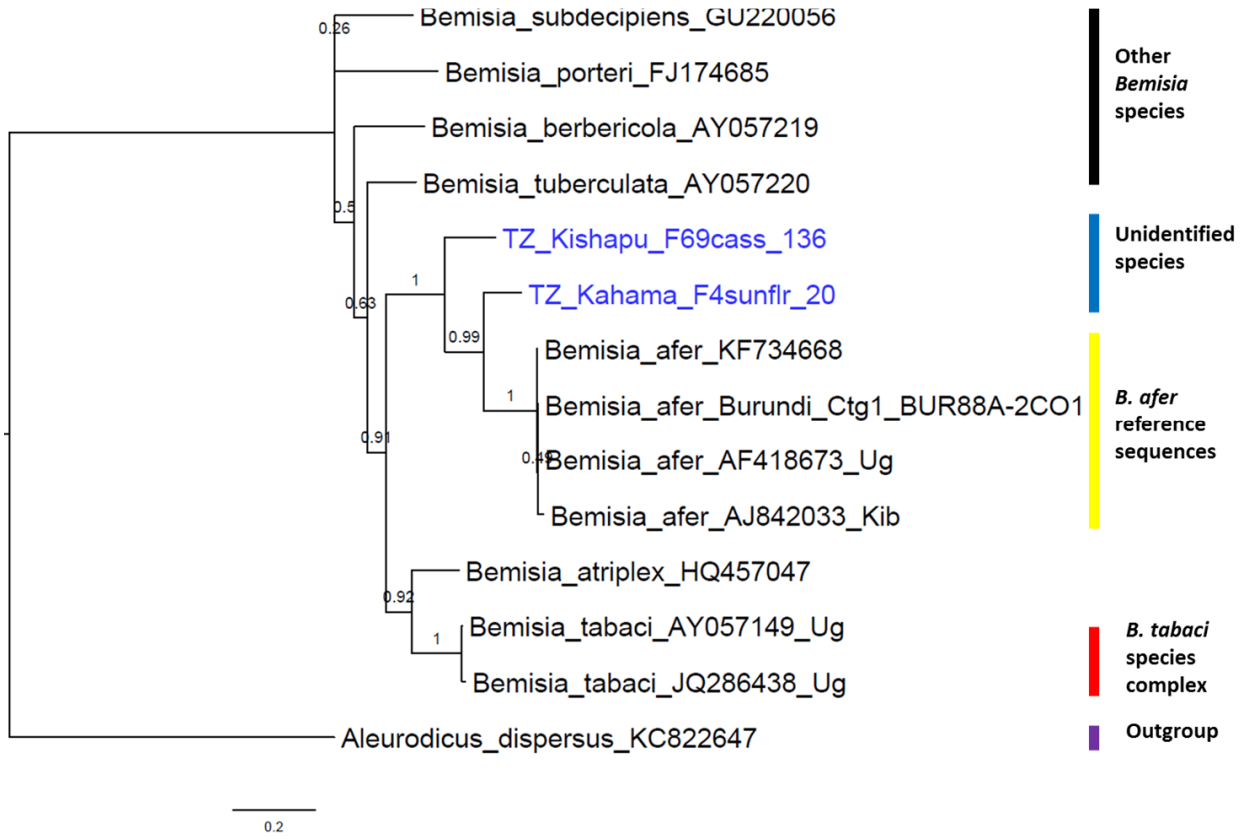


Figure 4 Phylogenetic tree based on approximate-maximum-likelihood inference of mtCOI sequences of the unidentified adult whiteflies using FastTree Version 2.1.7. The CAT-based log-likelihood value was -3700.436. Sequences 'TZ_Kishapu_F69cass_136' and 'TZ_Kahama_F4sunflr_20' are produced from this study whereas all other sequences are reference sequences from GenBank.

2.3.2 Host range of whiteflies collected

The great majority of whiteflies collected on cassava (94.4%) belonged to the SSA1 genetic group whereas those collected on sweet potato (84.6%) and pumpkin (62.5%) mainly belonged to the MED genetic group. All whiteflies on okra and cotton and the majority of the whiteflies on the weed plant klip dagga (63.6%) belonged to the recently designated EA1 genetic group. Generally, whiteflies belonging to SSA1 were very rarely (2.9%) encountered on non-cassava host plants. In addition, whiteflies from genetic groups other

than SSA1 were uncommon on cassava with the exception of few Indian Ocean (IO) individuals (4.6%) and an unidentified whitefly species (<1%) (Table 3). The other unidentified whitefly individual was recorded from sunflower.

Table 3 *Bemisia tabaci* genetic groups and the host plants on which they were encountered

Plant host (No. of WF[a])	*B. tabaci* genetic groups							
	SSA1				Mediterranean	Indian Ocean	EA1	Unidentified WF
	SG1	SG2	SG1/2	SG3				
Cassava (n[b]=108)	80	2	6	14	-	5	-	1
Sweet potato (n=13)	2	-	-	-	11	-	-	-
Pumpkin (n=8)	-	-	-	-	5	-	3	-
Okra (n=3)	-	-	-	-	-	-	3	-
Cotton (n=1)	-	-	-	-	-	-	1	-
Sunflower (n=1)	-	-	-	-	-	-	-	1
Tomato (n=1)	1	-	-	-	-	-	-	-
Klip dagga (n=11)	-	-	-	-	3	1	7	-
Total (n=146)	83	2	6	14	19	6	14	2

[a]WF=whiteflies, [b]n=number

2.3.3 Abundance of whiteflies

The abundance of adult whiteflies per plant in the surveyed cassava fields varied among the Regions, the highest being in Kagera Region (76.7) and the lowest in Geita, Singida and Shinyanga Regions (< 1), and differences between Regions were statistically significant ($F(9,244) = 5.49$, $P<0.0001$) (Table 4). Tukey's multiple comparison test between the Regions

revealed a statistically significant difference between the Kagera vs. Zanzibar, Morogoro, Singida, Tabora, Shinyanga and Mwanza Regions.

Table 4 ANOVA table for whitefly abundance on cassava among Regions

ANOVA table	SS[a]	df[b]	MS[c]	F[d]	[e]P value
Between Regions	157592	9	17510	6.05	P < 0,0001
Within Regions	706373	244	2895		
Total	863965	253			

[a]SS=sum of squares, [b]df=degree of freedom, [c]MS=the mean sum of squares, [d]F= *F*-statistic, [e]P=*P*-value

2.3.4 CMD pandemic: incidence, severity and disease front

CMD incidences due to whitefly-borne infection in the three district groups were in the order: group 3 > group 1 > group 2 ($P < 0.001$), whereas CMD incidences due to cutting-borne infection were in the order: group 1 > group 3 > group 2 ($P = 0.002$) and total CMD incidence was in the order: group 1 > group 3 > group 2 ($P < 0.001$) (Table 5). In pairwise multiple comparisons, district groups 1 and 3 each had significantly greater incidences of total CMD than district group 2 (P = 0.05). Based on a chi-squared analysis considering all symptomatic plants, there were significant differences in CMD severity between the three groups ($\chi^2 = 23.2$, df = 6, P < 0.001*).* Mean CMD severity values were in the order: group 1 > group 3 > group 2).

Table 5 CMD incidence, severity and whitefly abundance on cassava plants surveyed in a transect from north-western border to the coast of Tanzania

District group	Zones	CMD[a] incidence (%)			CMD[a] severity	Whitefly abundance/ plant
		Whitefly-borne	Cutting-borne	Total		
1	Lake	8.92 ± 0.02	28.20 ± 0.05	37.12 ± 0.05	3.00 ± 0.10	71.99 ± 22.07
2	Central	0.89 ± 0.00	10.33 ± 0.04	11.22 ± 0.04	2.76 ± 0.19	2.35 ± 0.86
3	Eastern	18.15 ± 0.03	12.78 ± 0.02	30.93 ± 0.04	2.85 ± 0.09	25.30 ± 6.46

[a]CMD= cassava mosaic disease

The CMD pandemic front was determined to lie across two villages in Geita Region: Buyagu (latitude 2.98193°S, longitude 32.23045°E, altitude 1362m.a.s.l.) and Nyarwasaja (latitude 3.04901°S, longitude 32.28747°E, altitude 1355m.a.s.l.); where CMD status (whitefly-borne incidence 33.3% vs. 3.3% of cutting-borne, whitefly abundance of 51 insects/plant for Buyagu and whitefly-borne incidence 26.7% vs. 1.7% of cutting borne, whitefly abundance of 64 insects/plant for Nyarwasaja with severity of 2 and 3 in Buyagu and Nyarwasaja respectively) were comparable with the characteristics of the CMD pandemic front as described in 2.2.3 above (Legg, 1999). Pandemic-associated *B. tabaci* SSA1-SG1 (Figure 1) were present in the two villages, although they were also present significantly ahead of the front, being recorded down to Nzega District in Tabora Region. In general, districts in Group 1 had disease and whitefly characteristics that were most closely comparable to those typical of the CMD pandemic (Table 5). The distance from the current estimated location of the CMD pandemic front to the position determined for 2000 (Legg, 1999) was measured using GIS software to be around 260 km. Based on this distance and the 10 years

between the two estimates, the rate of CMD pandemic spread is estimated as 26 km/yr, a value that approximates very closely with the predicted spread rate of 30 km/yr (Legg, 1999).

2.3.5 Geographic distribution and association of *Bemisia tabaci* genetic groups with the CMD pandemic

To understand the geographic distribution of *B. tabaci* and assess the association with the CMD pandemic, individuals from the SSA1 genetic group were mapped (Figure 1). The map revealed discrete zones of occurrence for each of the SSA1 sub-groups. SSA1-SG1 was the only group recorded from the CMD pandemic-affected parts of north-western Tanzania, although SSA1-SG1 also predominated in a large area up to 180km ahead of the estimated front of the pandemic. SSA1-SG2 was present in central parts of the country, mainly in Dodoma and Singida Regions. The intermediate group SSA1-SG1/2 was found together with SSA1-SG1 in sites located along the area extending south-east from the Lake Zone to central regions of the country, as well as in one field in Dodoma. SSA1-SG3 was found only in the eastern parts of the country and along the coast (Figure 1). In the current study virus diagnosis was not carried out, however, a previous extensive study on the diversity of cassava begomoviruses in Tanzania by Ndunguru *et al.* (2005) has shown EACMV-UG to be the predominant virus in the pandemic-affected zone of north-western Tanzania. A similar study has also indicated the presence of EACMV-UG in the area (Legg and Fauquet, 2004).

2.4 Discussion

This study assessed the genetic diversity of *B. tabaci* collected from parts of Tanzania that were either affected or not yet affected by the pandemic of severe CMD (Legg *et al*., 2006; Legg, 2010). *B. tabaci* individuals were identified from four genetic groups: SSA1, MED, IO and EA1. The first three were designated by Dinsdale *et al*. (2010) where as EA1 has recently been described by Legg *et al*. (2013) based on an individual that was found in north-western Tanzania showing 5% divergence from recognized major *B. tabaci* genetic groups. In our study several EA1 individuals were encountered in north-western parts of Tanzania, and sequence data confirmed the designation of EA1 as a new genetic group under the *Bemisia* species complex. Although the first record of EA1 was from cassava (Legg *et al*., 2013), all EA1 individuals identified in the current study were collected from non-cassava hosts, and it seems likely that these annual plants are the preferred hosts of EA1. SSA1 individuals grouped into three main phylogenetically distinct clusters, which corresponded to the sub-groups SSA1-SG1, SSA1-SG2 and SSA1-SG3 (Mugerwa *et al*., 2012; Legg *et al*., 2013). Additionally, several individuals were encountered that were intermediate between SG1 and SG2 (SSA1-SG1/2). All of these sub-groups share more than 98% homology in MtCO1 sequences, and are therefore not likely to be distinct species. Generally, the different genetic groups showed contrasting host plant associations, supporting the findings of earlier studies which have indicated that *B. tabaci* on cassava are distinct from those found on other host plants (Burban *et al*., 1992; Legg *et al*., 2002; Abdullahi *et al*., 2003). Additionally, two unidentified whitefly species were recorded. Although these both appeared to be *Bemisia* species, they were strongly divergent from

their closest relative (*B. afer*, > 15%) and from each other. This result highlights the diversity of *Bemisia* in this part of East Africa.

The current study revealed SSA1-SG1 to be the only genetic group found in the part of Tanzania affected by the CMD pandemic and SSA1-SG2 as the group found ahead of the disease front in areas unaffected by the pandemic. SSA1-SG3 was only found in coastal parts of Tanzania. SSA1-SG1 was found to be identical to the previously recorded *B. tabaci* genotype UG1 (Legg *et al*., 2002) which was reported to be associated with CMD pandemic. Although SSA2 was initially observed to be the predominant genetic group in areas of Uganda affected by severe CMD in 1997 (Legg *et al*., 2002), the frequency of occurrence of this group declined over subsequent years, as SSA1-SG1 appeared to replace it as the group associated with the CMD pandemic (Legg *et al*., 2002; Legg *et al*., 2013). SSA2 was not recorded at all in the current study, a result also reported by Mugerwa *et al*. (2012).

It is possible to think that 'non-pandemic' sub-group - SSA1-SG2 - could be an indigenous cassava whitefly in north-western and central Tanzania which is being displaced by SSA1-SG1. The apparent 'switch' from SSA2 to SSA1-SG1 as the haplotype associated with the CMD pandemic could also suggest the occurrence of hybridization, a suggestion that has been supported by successful mating experiments between SSA1 and SSA2 (Maruthi *et al*., 2004). However, further studies are strongly recommended to confirm this hypothesis through studies of gene flow and hybridization between the various putative species and sub-groups.

Elsewhere, invasive *B. tabaci* biotypes have been shown to displace local biotypes. Examples include the displacement of 'biotype A' (New world - Dinsdale *et al*., 2010) by 'biotype B' (MEAM1 - Dinsdale *et al*., 2010) in southwestern USA (Costa *et al*., 1993; Perring *et al*., 1993). In China the invasive MEAM1 that was dominant across the country was recently replaced by the similarly invasive MED (Hu *et al*., 2011). The apparently unique feature of the dynamics of cassava colonizing *B. tabaci* populations in East Africa, however, is that the invasive populations appear to do so through hybridization with local populations rather than via their displacement.

In the current study, CMD pandemic characteristics, as described by earlier studies (Otim-Nape *et al*., 1997; Otim-Nape and Thresh, 1998; Legg, 1999; Legg and Fauquet, 2004), were evident between Misenyi District, which borders Uganda in the northernmost part of Tanzania, and Geita District, to the south of Lake Victoria. The clearest pandemic-associated features were the high populations of whiteflies and severe CMD. Moreover, the CMD pandemic front seems to run through Geita District, and most likely extends westwards and eastwards into Chato and Sengerema Districts respectively, moving south-east at a rate of 26 km per annum which is within the 20 to 30 km per annum range that was estimated in 1999 (Legg, 1999). This suggests that the rate of expansion of CMD pandemic has remained almost constant for 10 years. The fact that SSA1-SG1 predominated in areas up to 180 km to the south-east of the pandemic front is not entirely surprising, as changes in whitefly situations, which could be in terms of the relative frequencies of genetic groups or population abundance, precede changes in disease status. Whitefly abundance was greatest in the Lake Victoria Zone, corresponding to the area

which has been affected by the CMD pandemic and the area in which SSA1-SG1 occurs. The unusually high abundance of these whiteflies is considered to be the key factor in the expansion of both the severe CMD and CBSD pandemics. The potential of any invasive whitefly genotype to spread further into central Tanzania, however, is likely to be constrained by environmental factors and geographic barriers (Guastella, 2014). Relatively little cassava is grown in central Tanzania, in part due to the low annual rainfall in this region. Conditions for the continued spread of invasive populations of *B. tabaci* are more favourable further to the west, along the shores of Lake Tanganyika in both Tanzania (to the east) and the Democratic Republic of Congo (DRC) (to the west). In fact, there is preliminary evidence that SSA-SG1 populations are spreading southwards along the shoreline of Lake Tanganyika in DRC (Shirima R, 2014, pers. comm.). Recently super-abundant *B. tabaci* populations have also been recorded from Zambia where high numbers of whiteflies have been associated with increased CMD incidence (Chikoti *et al*., 2013).

Although the current study suggests that genetic factors are associated with the super-abundant phenotype of some populations of *B. tabaci*, the precise nature of these genetic factors remains to be determined. Future studies will therefore be required to assess endosymbiotic bacteria that are known to have modifying effects on the biology of insects – including *Bemisia* – as well as to investigate sets of genes associated with host adaptation (known to be difficult for cassava) and overall fitness. An important first step with such studies will be the assessment of the fitness on cassava of the various putative species of *B. tabaci* and their sub-groups that are known to colonize cassava. The long-term goal of such studies will be the development of targeted and effective *Bemisia* management strategies.

Such control measures are urgently required to mitigate the devastating impacts on the food security of African smallholder farmers that result from the virus pandemics driven by these insects.

3. Diversity of symbiotic bacteria associated with *Bemisia tabaci* (Homoptera: Aleyrodidae) in cassava mosaic disease pandemic areas of Tanzania[2]

3.1 Introduction

All whiteflies harbour an obligate (primary) endosymbiotic bacterium essential for the host's survival and development – *Portiera aleyrodidarum* (Baumann *et al.,* 2004; Sloan and Moran, 2012). In addition, many harbour facultative (secondary) symbionts (Buchner, 1965). Different genetically distinct species or groups of *B. tabaci* host different complexes of several secondary symbionts that may include *Rickettsia, Hamiltonella, Arsenophonus, Cardinium, Wolbachia*, *Fritschea* and *Hemipteriphilus* (Zchori-Fein and Brown, 2002; Weeks *et al.*, 2003; Thao and Baumann, 2004; Everett *et al.*, 2005; Moran *et al.,* 2005; Gottlieb *et al.*, 2006; Bing *et al*., 2013). The effect of many of these secondary symbionts in whiteflies is not well known, although *Rickettsia* has been shown to increase the tolerance to heat shock in one population (Brumin *et al.,* 2011) and to induce fitness benefits and female bias in another (Himler *et al.*, 2011). In other hosts, secondary symbionts have been shown to affect their hosts in several ways, such as: manipulation of reproduction, (Zchori-Fein *et al.*, 2001; Hunter *et al.*, 2003; Duron *et al.,* 2008; Werren *et al.,* 2008), conferring resistance to parasitism, fungal attack or virus (Oliver *et al,.* 2003; Hedges *et al.,* 2008; Lukasik *et al.,* 2013) or increasing their tolerance to temperature (Russell and Moran, 2006; Henry *et al.,* 2013).

[2] Tajebe LS, Guastella D, Cavalieri V, Kelly SE, Hunter MS, Lund OS, Legg JP, Rapisarda C. 2014. Diversity of symbiotic bacteria associated with *Bemisia tabaci* (Homoptera: Aleyrodidae) in cassava mosaic disease pandemic areas of Tanzania. Ann. Appl. Biol. (Accepted Article) doi:10.1111/aab.12183

Within the cryptic species complex of *B. tabaci*, major genetic groups are generally reproductively isolated (Xu *et al.,* 2010; Sun *et al.,* 2011) and can vary in both biological traits as well as the endosymbiotic community they harbour (Brown *et al.,* 1995a; Costa *et al.,* 1995; Brown, 2000; Gueguen *et al*., 2010). For example, Chiel *et al.* (2007) demonstrated that *Arsenophonus* and *Wolbachia* were found only in the Mediterranean (MED) genetic group (Q biotype) and *Hamiltonella* was found only in the Middle-East Asia Mionr 1 (MEAM 1) genetic group (B biotype) in Israel where these whiteflies are sympatric.

Recently, the association of symbiotic bacteria with the various genetic groups of *B. tabaci* was published for some West African countries (Gnankine *et al.,* 2012). The study revealed an association between symbiotic bacterial communities and *B. tabaci* genetic groups. However, such associations of *B. tabaci* genetic groups and their symbionts remain unknown for East Africa. Understanding the symbiotic bacterial community hosted by *B. tabaci* may be important given that symbiotic bacteria may influence virus transmission capabilities of whiteflies (Gottlieb *et al.,* 2010). These associations may be very problematic if they are influencing the rapid distribution and progression of the virus disease pandemics in cassava in East and Central Africa that are associated with unusually high populations of *B. tabaci* (Legg *et al*., 2006; 2011).

This study was initiated in order to a) assess associations between the various whitefly genetic groups on cassava and nearby plant hosts and the endosymbionts they harbour, b) determine if there is a unique endosymbiont that is associated with super-abundant whitefly populations and c) provide baseline information on endosymbionts of whiteflies

for a part of East Africa (Tanzania). Here we report the first description of the diversity of the symbiotic bacterial community hosted by various genetic groups of *Bemisia tabaci* in Tanzania, with a special focus on the CMD pandemic affected areas.

3.2 Materials and Methods

3.2.1 Study area and whitefly collection

In a survey conducted during 2012 and 2013 in Tanzania, covering parts of six regions (Mwanza, Shinyanga, Tabora, Singida, Dodoma and Pwani) (Figure 5), 148 *B. tabaci* adult samples were collected from 155 farmers' cassava fields (more than 500 individuals total). Some of the regions sampled include the area of cassava cultivation through which the pandemic of severe CMD is currently expanding, from northwest to central Tanzania. Previous studies have indicated the likely direction of the severe CMD pandemic (Legg, 2010).

Sampling was done in farmers' cassava fields and in the immediate area surrounding cassava, including non-cassava crops and common whitefly-bearing weeds. Fields were sampled every 10 - 15 km along main and secondary roads. Samples collected were immediately preserved in 1.5 ml tubes containing 95% ethanol and stored at -20^{o}C until DNA extraction. Host plants included cassava (*Manihot esculenta* Crantz), pumpkin (*Cucurbita pepo* L.), sweet potato (*Ipomoea batatas* (L.) Lam.), a Lamiaceae weed - klip dagga or lion's ear (*Leonotis nepetifolia* (L.) R. Br.), okra (*Abelmoschus esculentus* (L.) Moench), tomato (*Solanum lycopersicum* L.) and cotton (*Gossypium hirsutum* L.).

3.2.2 DNA extraction and PCR of *Bemisia tabaci*

From 77 randomly selected whitefly samples, 210 individuals were subjected to molecular analyses to identify their genetic group and to assess their symbiotic bacterial communities (Table 6). DNA was extracted using a standard Chelex extraction method. Individual insects were first washed with distilled water to remove ethanol, homogenized under a microscope in a 3 µl drop of proteinase-K (20 mg/ml dH_2O) on a piece of parafilm, and then placed in a 0.5 ml microfuge tube containing 50 µl of 10% (w/v) Chelex solution kept on ice. The tubes were vortexed and centrifuged briefly before, during and at the end of a 1 h incubation period at 37°C. The tubes were then incubated at 96°C for 8 min to deactivate the proteinase-K, vortexed and centrifuged briefly before being stored in the freezer at -20°C until further use (White *et al.,* 2009).

Figure 5 The six sampling Regions of the current study - shaded in grey.

Table 6 Localities, host plants and genetic groups of the whitefly individuals used for molecular analyses with the most frequent secondary symbiont they harboured. R represents *Rickettsia*; H, *Hamiltonella*; A, *Arsenophonus*; C, *Cardinium*; W, *Wolbachia*; F, *Fritschea*; None, no infection by secondary symbionts; SSA1, Sub-Saharan Africa 1; MED; EA1, East Africa 1; IO, Indian ocean genetic group. The numbers in brackets represent the number of individuals on each host plant and belonging to the various genetic groups

Localities		Host Plants	Genetic group	Most frequent secondary symbiont infection
Region	District			
Pwani	Mkuranga	Cassava (n=148)	SSA1 (n=145)	A, AC, None
Dodoma	Dodoma Urban			
Singida	Manyoni, Ikungi, Singida Urban			
Tabora	Igunga, Nzega, Tabora, Urambo, Uyui		IO (n=3)	AC
Shinyanga	Bukombe, Kahama, Maswa, Shinyanga, Shinyanga Rural, Ushirombo			
Mwanza	Misungwi, Nyang'hwale, Geita			
Tabora	Nzega, Urambo	Pumpkin (n=23)	MED (n=14)	RHCW, no A
Shinyanga	Bukombe, Kahama			
Mwanza	Kwimba		EA1 (n=9)	RCW, A
Shinyanga	Bukombe, Kahama	Sweet potato (n=19)	MED (n=18)	RHW
Mwanza	Geita		SSA1 (n=1)	A
Tabora	Nzega, Uyui, Igunga	Klip dagga (n=14)	EA1 (n=7)	HCW
			MED (n=5)	only H
Shinyanga	Bukombe, Shinyanga Rural		IO (n=2)	RHC
Shinyanga	Kahama	Okra (n=3)	EA1 (n=3)	RACW
Mwanza	Geita			
Shinyanga	Kahama	Tomato (n=1)	SSA1 (n=1)	A
Shinyanga	Bukombe	Cotton (n=2)	EA1 (n=2)	RACW
		Total n=210		

Tubes were vortexed again and centrifuged briefly before using the supernatant as template DNA in PCR. Mitochondrial DNA fragments (*mtCOI*) from whitefly DNA were amplified using the forward primer C1-J-2195 and the reverse primer L2-N-3014 (Table 7, Simon *et al.*, 1994). PCR was performed using 2 µl aliquots of the DNA extract in a total reaction volume of 10.5µl containing 0.8x FailSafe premix, 0.48 µM of each primer and 1.25 units *Taq* DNA polymerase (Invitrogen™). Initial DNA denaturation was conducted for 3 min at 94°C followed by 35 cycles of 1 min denaturation at 94°C, 1 min annealing at 52°C and 1:20 min amplification at 72°C. Final extension was done at 72°C for 10 minutes. PCR products (aliquots of 4µl) were separated on a 1% agarose gel and visualized with Gel Red under UV light. For a subset of individuals, PCR products of the expected size (~850 bp) were sent to be sequenced either at Macrogen Inc. (USA) or BMR Genomics (Italy).

3.2.3 Detection and identification of endosymbionts

The quality of DNA extraction was first checked by detecting the primary endosymbiont *Portiera aleyrodidarum* using the forward and reverse primers PortF and PortR, respectively (Thierry *et al.,* 2011). Secondary endosymbiont infection was then assessed using specific PCR primers targeting the *16S* rDNA genes for *Rickettsia*, *Hamiltonella*, *Arsenophonus*, *Wolbachia, Cardinium* and *Fritschea.* Specific endosymbiont primers and annealing temperatures are given in Table 7. PCR was performed using 1 to 2 µl of the DNA extract in a total reaction volume of 10.5µl containing 0.8x FailSafe premix, 0.48 – 0.76 µM of each primer and 1.25 units *Taq* DNA polymerase (Invitrogen™). PCR products were separated and visualized under UV light. Selected bands of expected sizes were sequenced either at Macrogen Inc. (USA) or BMR Genomics (Italy).

Table 7 List of primers and PCR conditions used in the study

Targeted organism and/or gene	Name of Primer	Primer sequences (5' - 3')	Annealing temp. (°C) / Product size (bp)	Reference
B. tabaci	C1-J-2195	TTGATTTTTTGGTCATCCAGAAGT	52°/~850	Simon *et al.*, 1994
mtCOI	L2-N-3014	TCCAATGCACTAATCTGCCATATTA		
Portiera	Port-F	GGAAACGTACGCTAATAC	59°/~900	Thierry *et al.,* 2011
16S rDNA	Port-R	TGACGACAGCCATGCAGCAC		
Rickettsia	Rick 16S 528F	ACTAATCTAGAGTGTAGTAGGGGATGATGG	60°/~900	Chiel *et al.,* 2009
16S rDNA	Rick 16S 1044R	GTTTTCTTATAGTTCCTGGCATTACCC		
Hamiltonella	Ham-F	TGAGTAAAGTCTGGAATCTGG	60°/~700	Zchori-Fein & Brown, 2002
16S rDNA	Ham-R	AGTTCAAGACCGCAACCTC		
Arsenophonus	CAIf	GCCTGATGCAGCCATGCCGCGTGTATG	65°/~500	Dale *et al.*, 2006
16S rDNA	CAIr	GTCATCCCCACCTTCC		
Wolbachia	V1	TTGTAGCCTGCTATGGTATAACT	52°/~900	O'Neill *et al.*, 1992
16S rRNA	V6	GAATAGGTATGATTTTCATGT		
Cardinium	CLO F	GCGGTGTAAAATGAGCGTG	57°/~500	Weeks & Breeuwer, 2003
16S rDNA	CLO R	ACCTMTTCTTAACTCAAGCCT		
Fritschea	U23F	GATGCCTTGGCATTGATAGGCGATGAAGGA	55°/~600	Everett *et al.,* 2005
16S rDNA	23SIGR	TGGCTCATCATGCAAAAGGCA		

3.2.4 Phylogenetic analysis of *Bemisia tabaci* and endosymbionts

Out of the 210 individuals, 96 were sequenced for a portion of the *mitochondrial COI* gene. Sequences were cleaned manually and ends trimmed using Chromas Lite version 2.1 (2012) and aligned using the clustalO option of SeaView (version 4.4.2) (Galtier *et al*., 1996; Gouy *et al*., 2010). DNA sequence identities were confirmed using NCBI-BLAST. After alignment, a standard 657 bp *mtCOI* fragment was used to generate the phylogenetic tree. Selected sequences were used from the global *Bemisia* dataset (release version 31_Dec_2012) (De Barro and Boykin, 2013) and from GenBank® for comparison with sequences from this study. Maximum likelihood phylogenetic trees were constructed with 1000 bootstrap replicates using MEGA 6 (Tamura *et al.,* 2013). A jModel test (jModelTest version 8 2.1.3) was performed and TrN+I+G was used as the best fit nucleotide substitution model. The outgroup used for *B. tabaci* analysis was *Trialeurodes vaporariorum* (Westwood), GenBank code: JF693935. The trees were summarized using Fig Tree (version 1.4.0) (Rambaut, 2009).

Phylogenetic analysis for representative endosymbiont sequences was performed in a similar manner, sequences were compared to known sequences in the GenBank® database to verify their identity and phylogenetic trees were constructed (Appendix 1a-f).

3.2.5 Statistical analysis of association of secondary symbionts in whitefly genetic groups and host plants

To assess if the apparent differences in secondary symbionts infection status of our samples are randomly occurring by chance or if there is some relation between secondary symbionts and genetic group as well as host plants, a Chi square test was performed. Since we assume there could be sampling bias towards cassava we tested secondary symbiont infected and uninfected whitefly individuals collected only from cassava and belonging to SSA1-SG1 on one hand and belonging to other groups on the other hand. This was done to see if the striking feature of high numbers of uninfected SSA1-SG1 *B. tabaci* individuals could be expected by chance.

3.3 Results

3.3.1 Phylogenetic analysis of *Bemisia tabaci*

The phylogenetic analysis of *mtCOI* sequences revealed four putative species of the *B. tabaci* species complex, which included: SSA1, Mediterranean (MED), Indian Ocean (IO) and East Africa 1 (EA1) (Figure 6). Four sub-groups were also detected under the SSA1 putative species, namely, SSA1-SG1, SSA1-SG2, SSA1-SG3 and SSA1-SG1/2.

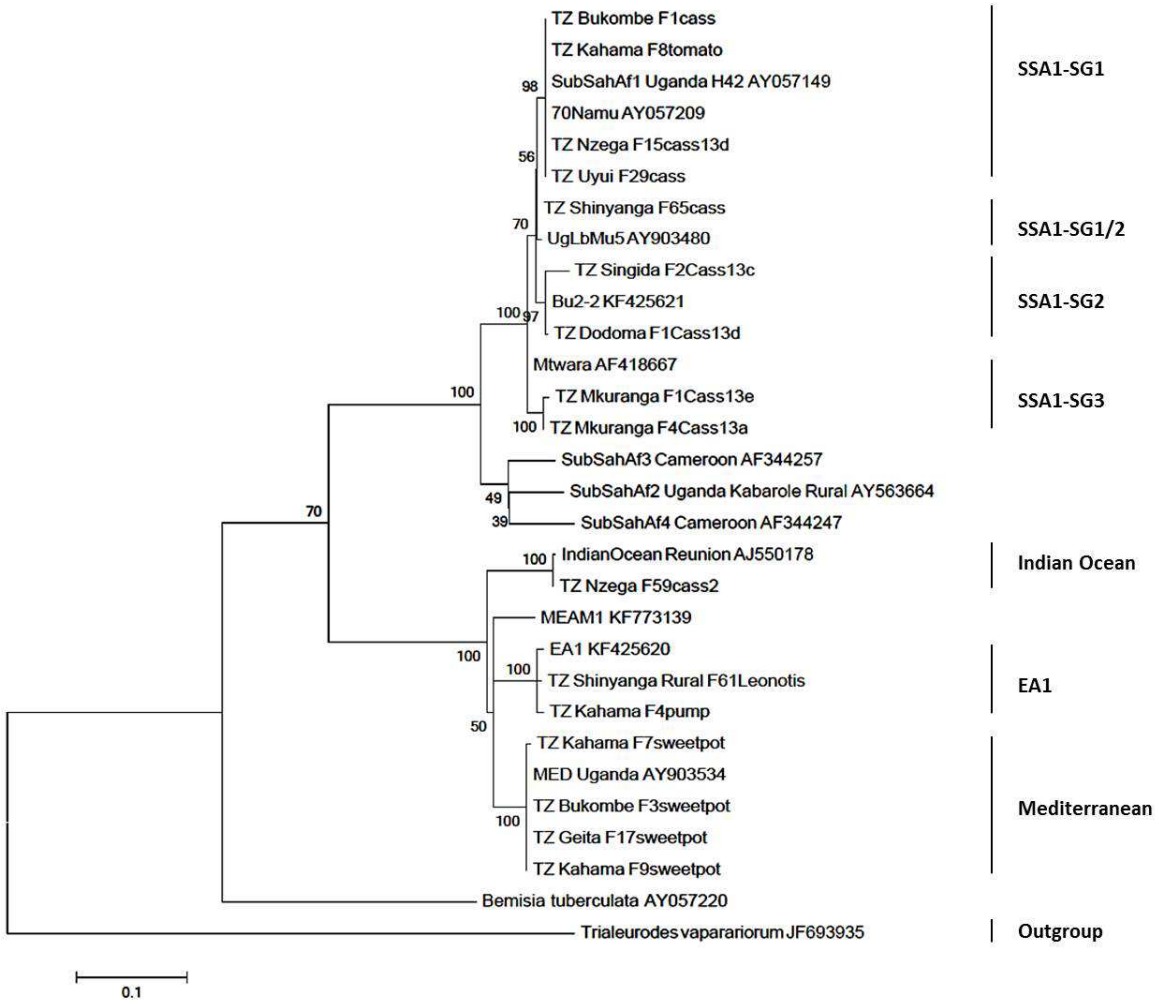


Figure 6 Maximum likelihood phylogeny of mtCOI sequences of *Bemisia tabaci* collected during 2012 – 2013 surveys with selected reference sequences from GenBank. All sequences starting with 'TZ_' are sequences produced in this study whereas all other sequences are reference sequences from GenBank. The tree is drawn to scale and the scale bar represents the number of substitutions per site. SSA1 represents Sub-Saharan Africa 1 genetic group; SG, sub-groups; and EA1, East Africa 1 genetic group.

3.3.2 *Bemisia tabaci* infection by secondary symbionts (SS) and geographic distribution of the SS

Detection of the primary symbiont *P. aleyrodidarum* in all individuals confirmed the DNA extract quality. Different patterns of double and multiple infections (infection with three or more secondary symbionts) were observed in the different genetic groups (Figure 7). *Rickettsia* (R) was always found in double or multiple infections and *Fritschea* (F) was completely absent from all of the whitefly individuals examined in this study (Figure 7, Appendix 2). We did not test for *Hemipteriphilus*, which was documented in China for the first time in 2013 (Bing *et al.,* 2013).

The majority (81%) of MED whiteflies were infected by two or more secondary symbionts. Whiteflies from this group had either single or multiple combinations of one to four symbionts, with *Hamiltonella* (H) and *Wolbachia* (W) appearing as single infections, but *Rickettsia, Cardinium* (C) and *Arsenophonus* (A) present only in multiply-infected hosts (Figure 7, Appendix 2). Double and multiple infections were also common in whiteflies belonging to the IO and EA1 putative species. The majority of EA1 individuals harboured *Cardinium* and *Wolbachia* with several additional combinations of *Rickettsia*, *Hamiltonella* and *Arsenophonus*.

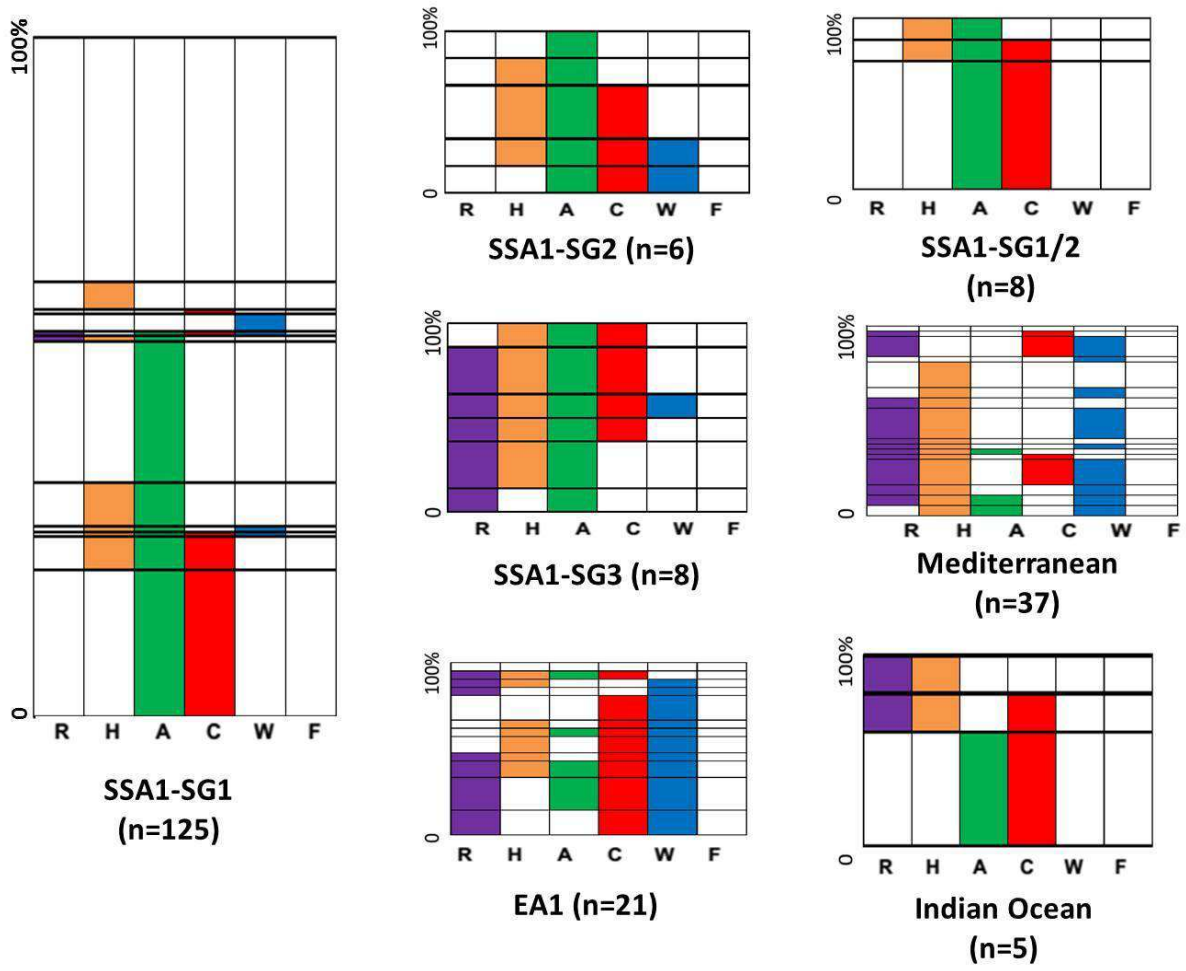


Figure 7 Secondary symbiont infection in *Bemisia tabaci* individuals belonging to the different genetic groups and sub-groups found in this study. The number of individuals tested and the genetic groups they belong to are indicated below each graph. The various colours represent the different secondary symbionts tested where R is *Rickettsia*; H, *Hamiltonella*; A, *Arsenophonus*; C, *Cardinium*; W, *Wolbachia*; and F, *Fritschea.* The combination of colours in rows represents whiteflies that shared that particular complex of symbionts. The width of the row indicates the percentage of individuals that have that combination of symbionts. SSA1 represents Sub-Saharan Africa 1 genetic group; SG, sub-groups; and EA1, East Africa 1 genetic group. The numbers in brackets indicate the number of individuals tested in each of the genetic groups and sub-groups.

Generally, *Arsenophonus* was the most commonly found secondary symbiont in all the four sub-groups of SSA1. More than one third of SSA1-SG1 individuals were free of secondary symbionts (36%), but infected individuals frequently harboured *Arsenophonus.* AC was the most common double infection type in individuals in SSA1-SG1, whilst HA and HAC infections were infrequent. In this group, multiple infection was very rare, and only observed in 8% of the examined individuals. SSA1-SG1 was the only sub-group where individuals with no secondary symbiont infection were recorded (Figure 7, Appendix 2). SSA1-SG1 is principally found in the north-west of Tanzania surrounding Lake Victoria, corresponding to the area affected by the severe CMD pandemic (Figure 8). Conversely, all SSA1-SG1/2 and SSA1-SG2 individuals, which occurred primarily in the central part of Tanzania, harboured either one or more of the secondary symbionts. Infections by AC and HAC were common in these two sub-groups. SSA1-SG3, which only occurred in coastal areas, frequently harboured R in combination with HA or HAC (Figures 7 and 8). Taking into consideration all 210 *B. tabaci* individuals screened for secondary symbionts, multiple infection was the most common infection type followed by double infection, no infection and single infection (Figure 9).

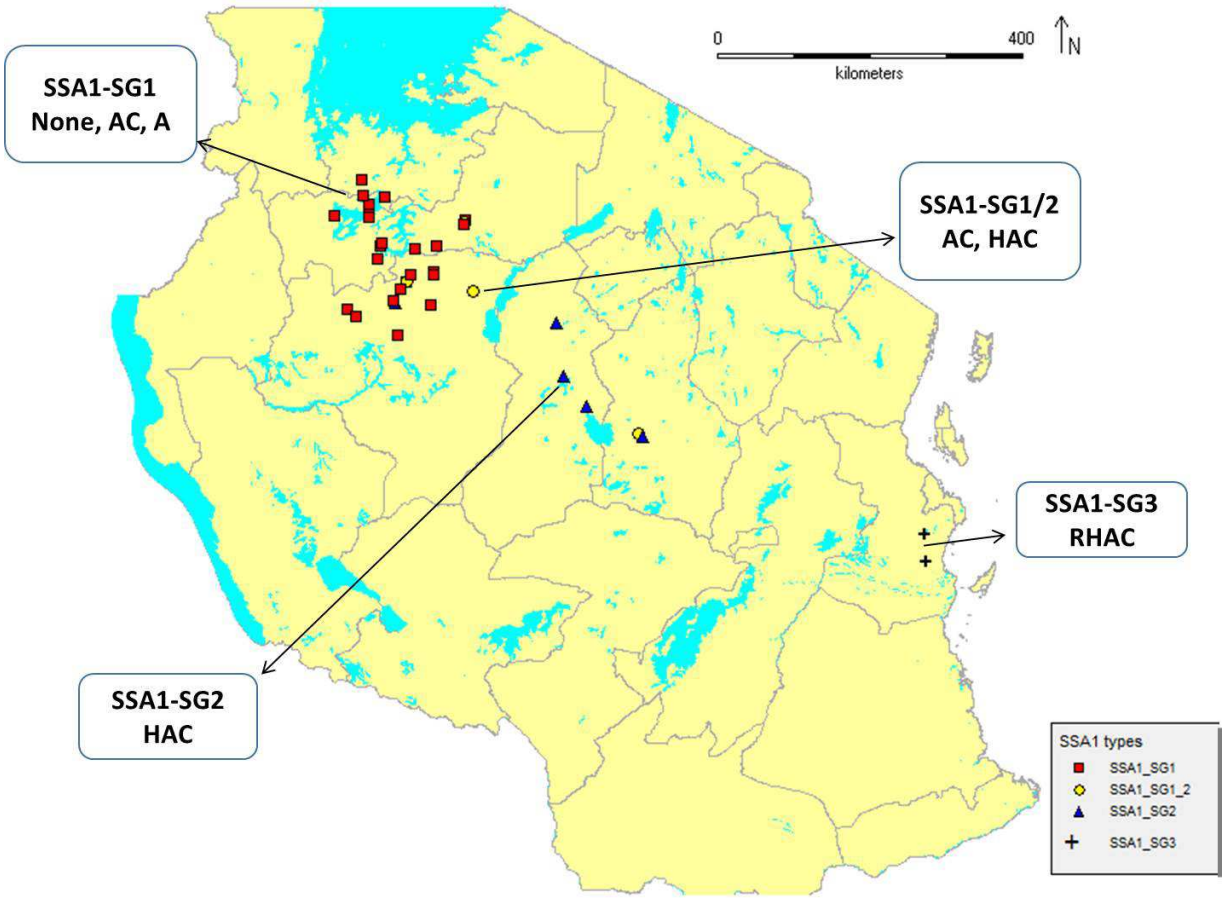


Figure 8 Geographic distribution of *Bemisia tabaci* genetic groups on cassava and their most frequent secondary symbionts. R, *Rickettsia*; H, *Hamiltonella*; A, *Arsenophonus*; C, *Cardinium*; W, *Wolbachia*; F, *Fritschea*; None, no infection by secondary symbionts; SSA1, Sub-Saharan Africa 1; SG1, Sub-group 1; SG2, Sub-group 2; SG1/2, intermediate Sub-group between 1 & 2; SG3, Sub-group 3.

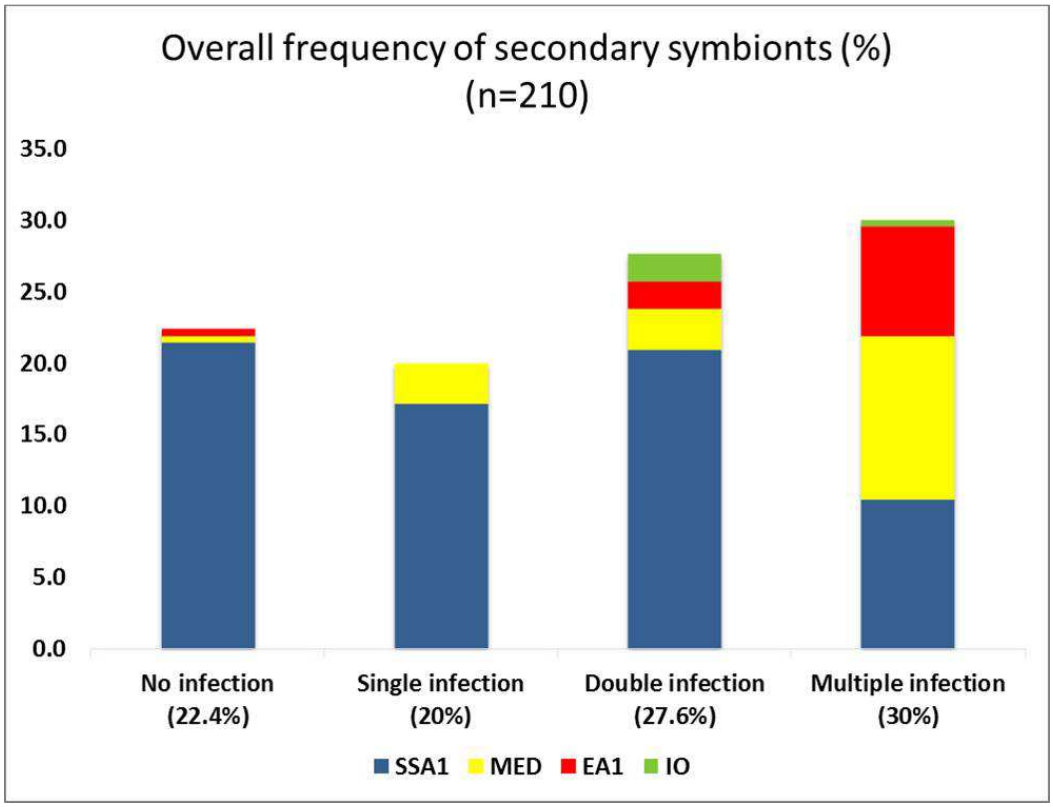


Figure 9 Overall frequency of secondary symbionts harboured by *Bemisia tabaci* individuals sampled during 2012 and 2013 in Tanzania. The various colours indicate the different genetic groups whereas SSA1 represents Sub-Saharan Africa 1; MED, Mediterranean; EA1, East Africa 1; IO, Indian Ocean genetic group. The numbers in brackets indicate the percentage of each secondary symbiont infection type among the total 210 whitefly individuals tested.

3.3.3 Host plants, *Bemisia tabaci* and secondary symbionts

Whiteflies collected on different host plants belonged to different genetic groups and harboured different secondary endosymbionts. Whiteflies on okra, tomato and cotton were very few, so we were not able to make strong conclusions on these host plants. However, it was still possible to compare their symbiont profile with whiteflies of the same genetic group collected on other host plants. Most whiteflies collected on cassava and the one individual collected on tomato were SSA1, whereas whiteflies collected on the remaining host plants mainly belonged to MED, EA1 and IO (Table 6). The whiteflies on cassava and

on tomato harboured A, AC or no secondary endosymbionts at all. Almost all whiteflies on sweet potato were MED (except one SSA1 individual with only A infection) and harboured mainly RHW. In addition, some MED whiteflies had A and a few had C (A and C not shown in Table 6 which presents only the most frequent secondary symbiont infection) whilst those on okra and cotton were all EA1 genetic group, all with four secondary symbionts (RACW). On the other hand, whiteflies on pumpkin and on the weed called klip dagga or lion's ear mainly belonged to both MED and EA1 with few on klip dagga belonging to IO (Table 6). MED whiteflies on pumpkin were infected by a different subset of four secondary symbionts: RHCW. EA1 individuals from the same host plant, besides harbouring RHCW, also had A but fewer H. On the other hand, MED whiteflies on klip dagga only had single infection by H while EA1 individuals collected from the same host plant exhibited multiple infections: mainly HCW and some also with R and A (R and A not shown in Table 6 which presents only the most frequent secondary symbiont infection). Whiteflies belonging to IO genetic group were collected on cassava (n=3) and klip dagga (n=2). Although the number of individuals was very few, those on cassava harboured AC while those on klip dagga had RH and RHC (Table 6).

Almost all of the uninfected whiteflies were collected on cassava (95.8%). Two individual whiteflies, one from okra and one from klip dagga were the only other whiteflies without secondary symbionts in the study. All other individuals harboured either one or more of the secondary symbionts in various combinations and frequencies.

3.3.4 Association of secondary symbionts in whitefly genetic groups and host plants

A χ^2 analysis of secondary symbiont infection status of whiteflies collected on cassava revealed statistically significant differences [χ^2 (1, N=148) =13.14, P<0.001] between SSA1-SG1 whiteflies and all others indicating that this group was less likely to be infected with secondary symbionts than others (Appendix 3). Similar analyses revealed that SSA-SG1 whiteflies were more likely to be free of secondary symbionts than were any other genetic group and sub-group collected in the CMD pandemic-affected area of the Lake Zone [χ^2 (1, N=190) =18.83, P<0.001] (Appendix 4).

3.3.5 Phylogenetic analysis of endosymbionts

A total of 2 - 5 partial sequences were obtained for the *16S* rDNA gene of each of the endosymbionts found to infect the whiteflies tested in the study. The amplified sequences (between 388 and 666 bp, depending on the endosymbiont) were aligned together with selected reference sequences from GenBank® and the alignment was used to reconstruct the respective phylogenetic trees (Appendix 1a-f). The phylogenetic trees revealed that the endosymbionts detected grouped with the reference sequences, confirming the identity of the sequences produced.

3.4 Discussion

Secondary symbionts and Bemisia tabaci *genetic groups*

About one quarter of the more than 1200 species of whitefly that have been described worldwide occur in Africa (Mound and Halsey, 1978). In the past years, molecular markers

have been widely used to discriminate genetic variation in morphologically indistinguishable populations of *B. tabaci* whiteflies (Brown *et al.,* 1995b; Frohlich *et al.,* 1999; Costa and Brown, 1991; Delatte *et al.,* 2005; Dinsdale *et al.,* 2010). In our study we used the *mtCOI* gene to detect the various genetic groups of *B. tabaci*. The whiteflies studied clustered into four putative species; three of which are designated in Dinsdale *et al.* (2010) as SSA1, MED and IO. The fourth was the recently designated putative species EA1 (Legg *et al.,* 2013). SSA1 further clustered into four sub-groups: SG1, SG2, SG1/2 and SG3. The extent of gene flow within subgroups of SSA1 has yet to be studied, although successful matings have been demonstrated between sympatric populations of the SSA1 and SSA2 putative species (Maruthi *et al*., 2002).

The current study assessed associations between secondary symbionts and several genetic groups of *B. tabaci.* Generally, whiteflies belonging to SSA1 were shown to frequently harbour *Arsenophonus* but almost never *Rickettsia,* similar to the results reported in West Africa (Gnankine *et al.,* 2012) for the sub-Saharan African non-Silver Leafing 1 (AnSL 1) 'biotype', which is equivalent to SSA1 (Dinsdale *et al*., 2010). However, while in this study a high frequency of uninfected individuals was detected from SSA1-SG1, in West Africa (Gnankine *et al.,* 2012) it was from SSA1-SG3. Contrary to the findings in West Africa (Gnankine *et al.,* 2012), the SSA1-SG3 individuals in our study frequently harboured *Rickettsia*. *Fritschea* was never detected in any of the samples in both studies.

Two previous studies (Gottlieb *et al.,* 2008; Gueguen *et al.,* 2010) indicated that *Hamiltonella* and *Arsenophonus* did not occur in double infections in whiteflies in the MED

genetic group. However, in our study, double infection by *Hamiltonella* and *Arsenophonus* was not uncommon in MED individuals, similar to Parrella *et al*. (2013), as well as among EA1 individuals or SSA1 group whiteflies. From the previously identified four clades under the MED genetic group (Boykin *et al*., 2007; Chu *et al*., 2008; Gueguen *et al.,* 2010), the individuals in our study belonged to Q1 clade. Earlier studies reported *Rickettsia* to be absent in Q1 individuals while *Hamiltonella* was found near fixation (Gueguen *et al.,* 2010; Parrella *et al*., 2013). Contrary to these previous studies, the MED (Q1) individuals we tested were frequently infected by *Rickettsia.* According to Parrella *et al*. (2013), *Arsenophonus* had low prevalence among the southern Italy Q1 individuals which was similar to our findings for the Q1 individuals from Tanzania. However, contrary to their findings, our Q1 individuals harboured *Wolbachia* frequently. These findings suggest that the occurrence and frequency of double and multiple infections could vary according to location, whitefly population and time.

Our results indicated a clear association between *B. tabaci* putative species and host plant groupings, as has been widely reported elsewhere (Abdullahi *et al.,* 2003; Gnankine *et al.,* 2012). SSA1 was largely confined to cassava whilst MED, IO and EA1 occurred on a diverse group of annual crop and weed hosts.

Secondary symbionts and 'super-abundance' of Bemisia tabaci

In Tanzania and other neighbouring countries in East and Central Africa, super-abundant whitefly populations on cassava crops have been reported from CMD pandemic affected areas (Legg and Ogwal, 1998; Legg, 2010; Chikoti *et al*., 2013). Our sampling covered CMD

pandemic affected areas in north-west Tanzania, where a super-abundant SSA1-SG1 *B. tabaci* seems to be associated with the pandemic (Legg *et al.,* 2013). Survey collections were also made in central and eastern parts of the country not yet affected by the severe CMD pandemic. This study was initiated partly to assess the factors behind the "super-abundance" of these whiteflies. *Rickettsia* is known to confer fitness to *B. tabaci* through increased fecundity and female bias (Himler *et al*., 2011). In our study, however, the SSA1-SG1 known to occur in increased numbers in CMD affected areas almost never harboured *Rickettsia* (only 2 out of 125 individuals). Moreover, about 38% of the whiteflies collected on cassava and belonging to SSA1-SG1 showed no infection by secondary endosymbionts. Multiple infection was also very rare in this group.

Why super-abundant whiteflies are associated with the lack of a secondary symbiont infection is puzzling. In general, for strictly vertically transmitted symbionts, theory predicts that they will not be maintained or spread in populations unless they confer fitness benefits to their hosts, or manipulate host reproduction in such a way (e.g. female-biasing) as to confer a transmission advantage to the symbiont (O'Neill *et al.*, 1997). However, symbiont benefits may not occur if there is regular horizontal transmission, in which case symbionts can persist even if pathogenic. The possibility of horizontal transmission being involved in the patterns of symbionts in whiteflies observed in the current study (Table 5) cannot be determined from these data. However, previous studies have presented evidence that certain clades or species of aphids that acquired particular secondary symbionts through horizontal transmission were able to utilize new host plants that were formerly unsuitable; clearly indicating that host plant utilization is influenced by

secondary symbionts and that secondary symbionts are important for adaptation and utilization of new host plants (Tsuchida *et al*., 2004; Tsuchida *et al*., 2011; Henry *et al.,* 2013). Therefore, determining the possibility and the frequency of horizontal transmission in *Bemisia tabaci* through detailed phylogenetic and co-evolutionary study of the whiteflies and their secondary symbionts in relation to the different host plants they utilize in different geographic locations would be very useful in helping predict the role of host plants and the multiple secondary symbionts observed in whitefly fitness. It would also be useful to improve understanding of the relationship between secondary symbionts in cassava whiteflies and vector capacity. While symbionts have been implicated in increasing transmission of plant viruses of whiteflies (Gottlieb *et al.*, 2010), symbionts in *Drosophila* and mosquitoes have conferred resistance to human and insect viruses (Hedges *et al.*, 2008; Moreira *et al.*, 2009). Future study could address the question of whether secondary symbionts of whiteflies on cassava reduce CMD vector competence.

This study reports for the first time the variable and complex symbiotic communities of the genetic groups of *B. tabaci* occurring on crop and weed hosts in a part of East Africa where whiteflies and viruses vectored by whiteflies threaten the viability of the cassava crop. In our study, no particular symbiont appeared to be associated with the super-abundant SSA1-SG1 *B. tabaci*. In contrast, a sizeable proportion of SSA1-SG1 individuals exhibited no infection by secondary symbionts. The exact association between a higher frequency of individuals with no secondary symbionts in the pandemic associated SSA1-SG1 *B. tabaci* and its apparent super-abundant nature remains to be determined and further studies are thus recommended. While future studies should be directed towards understanding the

role of these symbionts in the biology and vector competence of the various *B. tabaci* genetic groups, it is important to note that similarly high levels of uninfected individuals as in SSA1-SG1 in our study were detected from SSA1-SG3 in West Africa. Hence it seems unlikely that endosymbionts are the causal factors for the super-abundance phenomenon of SSA1-SG1 in the Lake Zone in Tanzania.

4. Microsatellites reveal population genetic structure of *Bemisia tabaci* species complex associated with cassava in Tanzania3

4.1 Introduction

To date, several biochemical and molecular techniques such as esterase, mitochondrial cytochrome oxidase I (mtCOI) and microsatellites have been utilized to assess the genetic variation of the *Bemisia tabaci* species complex which remains unresolved (Brown *et al.,* 1995; Frohlich *et al.,* 1999; De Barro *et al.,* 2000; Legg *et al.,* 2002; Delatte *et al.,* 2005; 2006; Boykin *et al.,* 2007; Dinsdale *et al.,* 2010). The fact that microsatellites exhibit a high degree of polymorphism makes them suitable for population genetics studies and particularly for intra-population level studies (Tautz, 1989; Ellegren, 2004; Hadjistylli, 2010). Microsatellites can also be used to study hybridization events, level of gene flow, occurrence of inbreeding and population differentiation. In recent years several microsatellite loci have been isolated and widely deployed in population genetics studies of *B. tabaci* species complex (De Barro *et al.*, 2003; Tsagkarakou and Roditakis, 2003; Delatte *et al.*, 2006; Simon *et al.*, 2007; Tsagkarakou *et al.*, 2007; Dalmon *et al.*, 2008; Gauthier *et al.*, 2008; Delatte *et al.*, 2011). Microsatellites have revealed genetic structure and relationships between previously unresolved genotypes of the *B. tabaci* species complex in the Asia-Pacific region as well as the presence of minimal, if any, gene flow between the genetic populations identified in the region (De Barro, 2005). In Greece microsatellites were able to discriminate between the B and Q biotypes (Middle East-Asia Minor 1 and Mediterranean genetic groups – Dinsdale *et al.*, 2010), when used in population genetic

[3] Tajebe LS, Simiand C, Lund OS, Reynaud B, Rapisarda C, Legg JP, Delatte H. Microsatellites reveal population genetic structure of *Bemisia tabaci* species complex associated with cassava in Tanzania (Submitted Manuscript)

studies (Tsagkarakou *et al.*, 2007). In the wider region of the Mediterranean Basin microsatellite variations have been used to infer genetic variation and population structure of the B and Q biotypes (Simon *et al.*, 2007). Similarly microsatellites have been useful in showing the genetic and geographic structure of a local (Ms) and an invasive biotype (B) on La Réunion island (Delatte *et al.*, 2006; Thierry *et al*., 2011).

After the outbreak of the severe CMD pandemic in the late 1980s, research efforts have been directed towards understanding the superabundant whitefly vector population that characterizes the severe CMD pandemic, among many others (Legg and Ogwal, 1998; Legg, 2010). As a result studies in the region have widely deployed mtCOI to investigate the genetic variation of *B. tabaci* populations (Legg *et al.,* 2002; Maruthi *et al.,* 2004; Sseruwagi *et al.,* 2006; Mugerwa *et al.,* 2012; Legg *et al.*, 2013). Nevertheless, the vector remains largely poorly understood in Africa with regards to population genetics and structuring. Therefore, in this study we used microsatellite markers to further investigate the genetic structure of and the possibilities of gene flow between *B. tabaci* populations found in Tanzania, taking into consideration CMD pandemic affected and unaffected areas. The study also discusses the genetic structure of the vector and its implications on cassava production and disease management in Tanzania and in the wider region of East and Central Africa.

4.2 Materials and Methods

4.2.1. Whitefly collection

Adult and nymph whitefly samples were collected in a survey conducted in Tanzania during 2012 and 2013. Sampling was done from cassava crops (*Manihot esculenta* Crantz) in 59 farmers' cassava fields. Whiteflies from one pumpkin (*Cucurbita pepo* L.) and three sweet potato (*Ipomoea batatas* (L.) Lam.) fields were also collected. In addition, screen house reared whitefly sample from IITA – Kibaha (Tanzania) and laboratory reared whitefly individuals belonging to the IO and MED obtained from CIRAD UMR PVBMT– La Réunion (France) were also included in the analyses as control (Table 8). All field-collected samples were labelled and stored in 95% ethanol until DNA extraction.

Table 8 Location and host plant information of the sample set used in the final analyses. Details are given for the 59 field collected whitefly populations from Tanzania as well as for one screen house reared whitefly sample from Kibaha (Tanzania) and two laboratory reared whitefly samples from La Réunion island (France)

					Location	
S. No	Sampling year	Sample	Number of individuals (n)	Host plant	Region	District
1	2012	F1capu12	23	Cassava (n=12) Pumpkin (n=11)	Shinyanga	Bukombe
2	2012	F2casp12	13	Cassava (n=10) Sweet potato (n=3)	Shinyanga	Bukombe
3	2012	F3sp12	11	Sweet potato	Shinyanga	Bukombe
4	2012	F10sp12	7	Sweet potato	Shinyanga	Kahama
5	2012	F17ca12	2	Cassava	Shinyanga	Geita
6	2012	F18ca12	19	Cassava	Shinyanga	Geita
7	2012	F19ca12	11	Cassava	Shinyanga	Geita
8	2012	F20ca12	10	Cassava	Shinyanga	Geita
9	2012	F24ca12	2	Cassava	Tabora	Nzega
10	2012	F28ca12	9	Cassava	Tabora	Nzega
11	2012	F29ca12	10	Cassava	Tabora	Uyui
12	2013	F35ca13	29	Cassava	Shinyanga	Kahama
13	2012	F37ca12	20	Cassava	Tabora	Urambo
14	2012	F39ca12	28	Cassava	Tabora	Urambo
15	2012	F69ca12	15	Cassava	Shinyanga	Kishapu
16	2013	F1ca13Mk	32	Cassava	Pwani	Mkuranga
17	2013	F3ca13Rf	21	Cassava	Pwani	Rufiji

18	2013	F4ca13Rf	15	Cassava	Pwani	Rufiji
19	2013	F5ca13Rf	4	Cassava	Pwani	Rufiji
20	2013	F1ca13	20	Cassava	Dodoma	Dodoma Urban
21	2013	F2ca13	32	Cassava	Singida	Manyoni
22	2013	F3ca13	1	Cassava	Singida	Manyoni
23	2013	F4ca13	8	Cassava	Singida	Ikungi
24	2013	F5ca13	15	Cassava	Singida	Ikungi
25	2013	F6ca13	2	Cassava	Singida	Ikungi
26	2013	F7ca13	1	Cassava	Singida	Ikungi
27	2013	F8ca13	4	Cassava	Singida	Ikungi
28	2013	F10ca13	2	Cassava	Singida	Ikungi
29	2013	F11ca13	11	Cassava	Singida	Ikungi
30	2013	F12ca13	17	Cassava	Singida	Ikungi
31	2013	F13ca13	34	Cassava	Singida	Singida Urban
32	2013	F9ca13	4	Cassava	Singida	Singida
33	2013	F15ca13	1	Cassava	Singida	Singida Urban
34	2013	F16ca13	1	Cassava	Tabora	Igunga
35	2013	F17ca13	9	Cassava	Tabora	Igunga
36	2013	F18ca13	1	Cassava	Tabora	Nzega
37	2013	F19ca13	14	Cassava	Tabora	Nzega
38	2013	F23ca13	12	Cassava	Tabora	Nzega
39	2013	F24ca13	8	Cassava	Tabora	Nzega
40	2013	F25ca13	33	Cassava	Tabora	Nzega
41	2013	F26ca13	35	Cassava	Tabora	Nzega
42	2013	F27ca13	11	Cassava	Tabora	Nzega
43	2013	F28ca13	34	Cassava	Tabora	Nzega
44	2013	F29ca13	3	Cassava	Tabora	Nzega
45	2013	F30ca13	17	Cassava	Tabora	Tabora
46	2013	F31ca13	41	Cassava	Tabora	Tabora
47	2013	F32ca13	6	Cassava	Tabora	Tabora Urban
48	2013	F33ca13	4	Cassava	Shinyanga	Ushitu
49	2013	F34ca13	6	Cassava	Shinyanga	Kahama
50	2013	F36ca13	15	Cassava	Shinyanga	Kahama
51	2013	F37ca13	7	Cassava	Shinyanga	Kahama
52	2013	F38ca13	5	Cassava	Shinyanga	Msalala
53	2013	F39ca13	19	Cassava	Shinyanga	Msalala
54	2013	F40ca13	9	Cassava	Shinyanga	Msalala
55	2013	F41ca13	22	Cassava	Mwanza	Nyang'hwale
56	2013	F42ca13	9	Cassava	Mwanza	Geita
57	2013	F45ca13	8	Cassava	Shinyanga	Kahama
58	2013	F46ca13	3	Cassava	Shinyanga	Shinyanga
59	2013	F51ca13	35	Cassava	Mwanza	Misungwi
IITA Kibaha Screen house reared whitefly samples						
60	2013	Kibca13	39	Cassava	Pwani	Kibaha
CIRAD Laboratory reared whitefly samples (La Réunion, France)						
61	2014	B	1			
62	2014	IO	1			
	Total		**841**			

4.2.2. DNA extraction, mtCOI amplification and sequencing

DNA was extracted from a total of 918 whitefly individuals using an extraction protocol modified from Delatte *et al.* (2005). Individual whitefly adults and nymphs were placed in a 96-well PCR plate containing 25 µl and 10 µl of extraction buffer, respectively, containing: 50 mMKCl, 10 mMTris-HCl, pH 8, 0.45% Nonidet P-40, 0.45% Tween 20, and 500 µg ml^{-1} proteinase K. The extraction plate was sealed, centrifuged briefly and incubated overnight at 37°C. The extract was then briefly re-centrifuged and incubated for 10 minutes at 90°C to deactivate the proteinase K. After the incubation 35 µl and 10 µl of pure PCR grade water was added to the adult and nymph extract, bringing the total volume to 60 µl and 20 µl respectively. The extracts were briefly centrifuged and stored at -20°C until further use.

The forward primer C1-J-2195 (5' TTGATTTTTTGGTCATCCAGAAGT 3') and the reverse primer L2-N-3014 (5' TCCAATGCACTAATCTGCCATATTA 3') (Simon *et al.*, 1994) were used to amplify the mitochondrial DNA fragments (mtCOI) from randomly selected whitefly adults and nymphs within each genetic cluster identified by STRUCTURE software v2.3.4 (Pritchard *et al.*, 2000).The total PCR volume used was20 µl containing 10 µl Type-It (QIAGEN™), 1 µl of each primer (20 µM), 6 µl water and 2 µl aliquot of the DNA extract. PCR conditions used were: initial denaturation for 5 min at 95°C followed by 35 cycles of 30sec denaturation at 95°C, 30 sec annealing at 52°C and 1 min amplification at 72°C followed by 5 min final extension at 72°C. PCR products were separated using capillary electrophoresis in QIAxcel Advanced System (QIAGEN™) and samples that gave bands of the expected size, ~850 bp, were sequenced at Macrogen Inc. (The Netherlands).

4.2.3. Phylogenetic analysis based on mtCOI sequences

The sequences obtained for *B. tabaci* adults and nymphs ranged from 761-848 bp. ChromasLite version 2.4 (2012) was used to clean each sequence. Cleaned sequences were aligned using the ClustalW option of MEGA 6 (version 6.0) and their identities were confirmed using NCBI-BLAST. A jModel test (jModelTest version 2.1.6) was conducted and TrN+I+G was used as the best fit nucleotide substitution model both for adults and nymphs (Guindon and Gascuel, 2003; Posada, 2008; Darriba *et al.,* 2012). *Trialeurodes vaporariorum* (Westwood), GenBank code: JX841216 was used as the outgroup (Roopa *et al*., 2012). A maximum likelihood tree was constructed for representative sequences along with reference sequences from GenBank. The phylogeny constituted 37 sequences with a final standard length of 724bp for adults and 28 sequences with a final standard length of 688 bp for nymphs.

4.2.4. Microsatellite markers, detection and genotyping

All adult whiteflies were sexed under a binocular microscope and their sexes noted. Only adult female whiteflies were used for estimating the Fstat due to their diploid nature. In total 918 individuals were screened using 13microsatellite markers (Table 9). Either the forward or the reverse primer of each locus was labelled with a fluorescent dye 6-FAM, VIC, NED and PET (Applied Biosystems®). Multiplex PCR was conducted in three mixes each with a total reaction volume of 15 µl using Type-It (Qiagen ©) and 2 µl DNA.Mix 1 comprised 0.1 µl of each of the forward and reverse primersP62, P41, Bem25 and 0.15 µl of primer WF2E11; Mix 2 had 0.1 µl of each of the primers P59, P7, WF2CO1, WF2H06, Ms145; and Mix 3 contained 0.1 µl of each of the primers P5, WF1B11, WF1D04 and

WF1G03. PCR conditions were 95 °C for 5 min followed by 35 cycles at 94 °C for 30 s, 56 °C for 1 min 30 s, 72 °C for 1 min 30 s, and 60 °C for 30 min.PCR products diluted at 1:20 (1 µl) were added to a mix of 11.9 µlHi-Di™ Formamide and 0.1 µl of size marker (GeneScan™ – 500 LIZ®) and loaded to ABI PRISM 3100 Genetic Analyzer (Applied Biosystems®) for genotyping. GeneScan™ –500 LIZ® size standard was used to determine allele sizes using GeneMapper® Software v. 4.0 (Life Technologies Corporation).

Table 9 Characteristics of the microsatellite markers used in the study, Na stands for number of alleles observed

Microsatellite Loci	Primers	Repeat	SSA1		IO		Med		*B. afer*		References
			Na	Observed allele range	Na	Observed allele range	Na	Observed allele range	Na	Observed allele range	
WF1B11	F: GCATTGAACATTTTTCTGCATGCGCG	$(CCTGA)_{12}$imp	2	100-105	8	100-160	3	100-110	3	100-125	Schwartz unpublished
	R: GCACACAGCTCTCCAAAAGAAAGGTC										
WF2C01	F: ATGATACCGCACGAAAAAGAGGACG	$(GTTT)_{11}$imp	13	100-148	10	100-148	2	132-148	9	96-148	Schwartz unpublished
	R: CTTGAATTACATCAAACGCAGCAGC										
WF2H06	F: TATTCGCCAATCGATTCCTT	$(TTTG)_{11}$	18	76-164	4	144-164	5	136-164	12	84-164	Schwartz unpublished
	R: CGGCGGAAATTTCGATAAA										
WF2E11	F: TCTCCAACCATAATTTTTAATCTCG	$(GATT)_{27}$imp	-	-	-	-	-	-	-	-	Schwartz unpublished
	R: GTCTGGGCAGGAAAACGAT										
WF1D04	F: GTTGTTAGGTTACAGGGTTTGTC	$(CAAA)_{16}$	15	106-182	16	106-194	3	118-138	14	94-182	Hadjistylli unpublished
	R: GTCTTTACTTCTCTTTTCCTCCG										
WF1G03	F: CTCCAAAATGGGACTTGAAC	$(GTTT)_{8}$	9	116-240	6	108-172	4	108-216	21	104-248	Hadjistylli unpublished
	R: GTAGAAGCCACACATACTAGCAC										
P62	a: CTTCCTTAGCACGGCAGAAT	8(GT)	24	152-246	18	178-252	6	202-228	22	156-244	Delatte *et al.*, 2006
	b: TTTGGCGCAATTTTTAGCGTCTGT										
Bem25	a: TGAAGAATAAGAATAAGAAGG	10(CTT)	-	-	-	-	-	-	-	-	De Barro, 2005
	b: AAGAACTAGCCAGGAGGGACAAAC										
P41	a: AATCTTAGTGCTAAAGTTTTCTT	7(AC)	-	-	-	-	-	-	-	-	Delatte *et al.*, 2006
	b: TTGTGTTAGGATGATAGGCTTGGA										
P5	a: ATTAGCCTTGCTTGGGTCCT	8(GT)	-	-	-	-	-	-	-	-	Delatte *et al.*, 2006
	b: TTTGCAAAAACAAAAGCATGTGTCAAA										
P7	a: AGGGTGTCAGGTCAGGTAGC	8(GT)	7	154-214	8	160-192	5	160-192	14	152-232	Delatte *et al.*, 2006
	b: TTTGCGTAATGGTAACATGTTTAGAAAA										
P59	a: CGGCGTTTCTCGTTTTCTT	44(T) 18(G)	17	156-214	13	166-206	8	166-214	3	164-174	Delatte *et al.*, 2006
	b: TTTGCCAACTGAAGCACATCAATCA										
Ms145	a: CCTACCCATGAGAGCGGTAA	9(AC)	-	-	-	-	-	-	-	-	Dalmon *et al.*, 2008
	b: TCAACAAACGCGTTCTTCAC										

4.2.5. Population data analysis

STRUCTURE software v2.3.4 (Pritchard *et al.*, 2000) was deployed to differentiate whitefly individuals into different populations. Structure uses a Bayesian approach to separate mixed populations into subpopulations (K) based on multi-locus allele distribution. The analysis was run on eight of the most amplified microsatellite loci (out of the thirteen). Data sets of more than 37.5% missing data were not included in the analysis. Each run had a burn in of 100,000 Markov chain Monte Carlo (MCMC) generations and a run of 1,000,000 MCMC generations. Log-likelihood estimates were calculated for K=1 to 20 with 20 replicates. The number of clusters was decided based on delta K (Evanno *et al.*, 2005). CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) was used to summarize the estimated cluster membership for each K and graphical visualization of the clusters was done using the program DISTRUCT v1.1 (Rosenberg, 2004). Hardy-Weinberg (HW) and linkage equilibrium which are normally assumed by Structure software were tested using Genepop v.4.0 (Rousset, 2008). Genetix 4.01 was used to estimate the expected heterozygosity (He) from the observed heterozygosity (Ho) (Belkhir*et al.*, 1996-2004). F_{stat} software was used to estimate the allelic richness (Goudet, 1995). The null allele frequencies for each locus and population were estimated using FreeNa software following the Expectation Maximization (EM) algorithm (Chapuis and Estoup, 2007). F_{is} and F_{st} estimates were calculated according to Weir & Cockerham (1984) using Genepop v.4.0.Analysis of molecular variance (AMOVA) was run using ARLEQUIN 3.0 (Excoffier *et al.*, 2005) to estimate the genetic variance. The relationship between the geographical distances and the genetic distances ($F_{st}/(1-F_{st})$) was assessed by computing the coefficient of Mantel (Z) and Pearson's coefficient of correlation (r) for the SSA1 *B. tabaci* putative species according to Mantel (1967), as implemented in Genetix 4.01 (Belkhir *et al.*, 1996-2004).

4.3 Results

4.3.1. Microsatellite variation and population structure

Out of the thirteen loci screened for, eight amplified well. The five that did not amplify well were excluded from further analyses (over 40% of null alleles). The eight analysed loci exhibited a number of alleles ranging from 2 at locus WF1B11 to 24 at locus P62. In total 101 alleles were detected.

From the total of 918 individuals analysed, 77 that amplified for fewer than 5 loci were removed. The remaining dataset of 841 individuals was subject to analysis using STRUCTURE software v2.3.4 (Pritchard *et al.*, 2000) which detected 2 strong structures with K=2 – referred to as cluster 1 and cluster 2 (Figure 10). Based on 70% assignment of individuals to cluster 2, a subset of 577 individuals were almost all from cassava (only 2 were from sweet potato) and were grouped together as a new dataset of cluster 2. The cluster was then re-analysed by STRUCTURE. Similarly, based on 70% assignment of individuals to cluster 1, 249 individuals were grouped together as cluster 1 for further analysis. Cluster 1 further divided into two subpopulations – cluster 1.1 (51 individuals) and 1.2 (197 individuals, all from cassava). One individual was weakly assigned to both subpopulations based on 60% assignment of individuals. Cluster 1.1 was re-analysed using STRUCTURE which revealed two sub-populations – cluster 1.1.1 (16 individuals, all from sweet potato except 1) and 1.1.2 (35 individuals, mainly from cassava and pumpkin). Similarly, cluster 2 comprising 577 individuals showed two subpopulations with K=2 in STRUCTURE analysis (Figure 10). All delta K and LN likelihood values are presented in Appendix 5. Mitochondrial COI sequences revealed the identity of some members of the clusters which are indicated in Figure 10 and presented in the following section

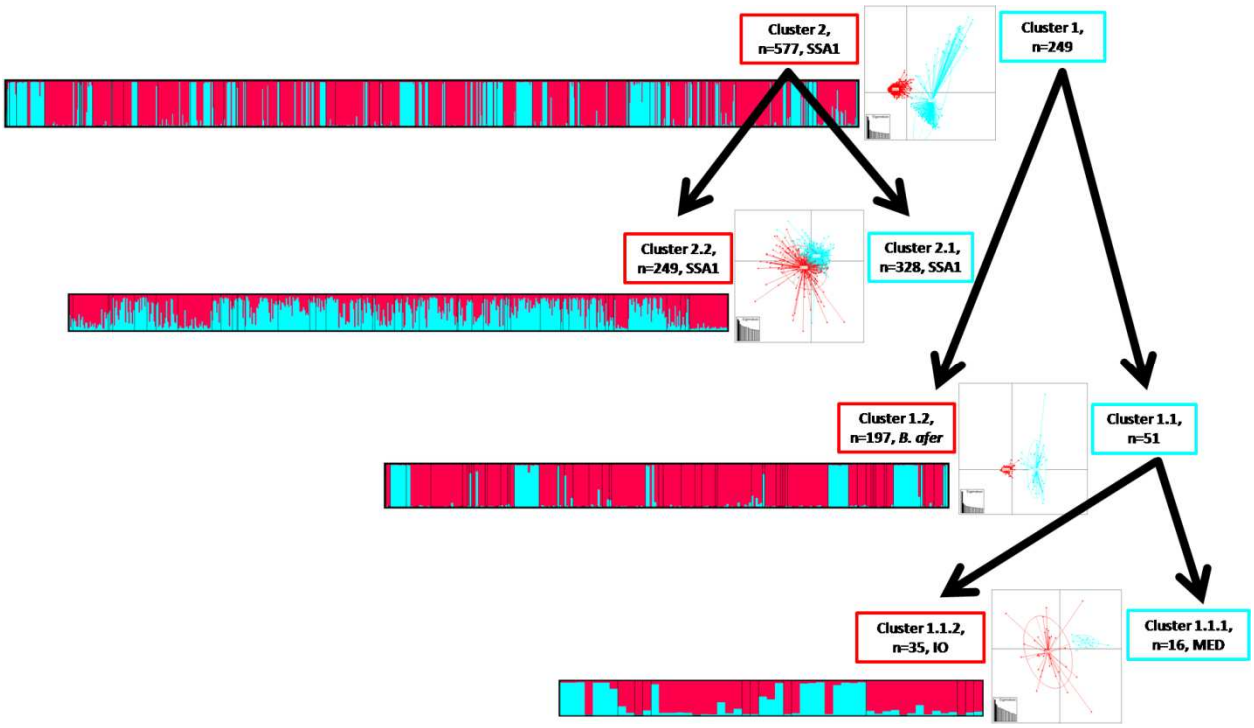


Figure 10 Clustering and sub-clustering STRUCTURE outcomes for all samples and Principal Component Analysis (PCA) corresponding to each analysis. SSA1 stands for sub-Saharan Africa 1, IO for Indian Ocean and MED for Mediterranean putative species.

4.3.2 Phylogenetic analysis

The phylogenetic analyses of mtCOI sequences revealed three clusters from the *B. tabaci* species complex - Sub Sahara Africa 1 (SSA1), Mediterranean (MED) and Indian Ocean (IO) genetic groups for the adults (Dinsdale *et al.* 2010) (Figure 11) and only SSA1 genetic group in the case of nymphs (Figure 12). The SSA1 genetic group showed similar topology both in the adults and nymphs, further dividing into three distinct and one intermediate sub-groups designated as SSA1-SG1, SSA1-SG2, SSA1-SG3 and SSA1-SG1/2as described in Legg *et al.*, 2013. Moreover, *Bemisia afer* adults were also recorded (Appendix 6).

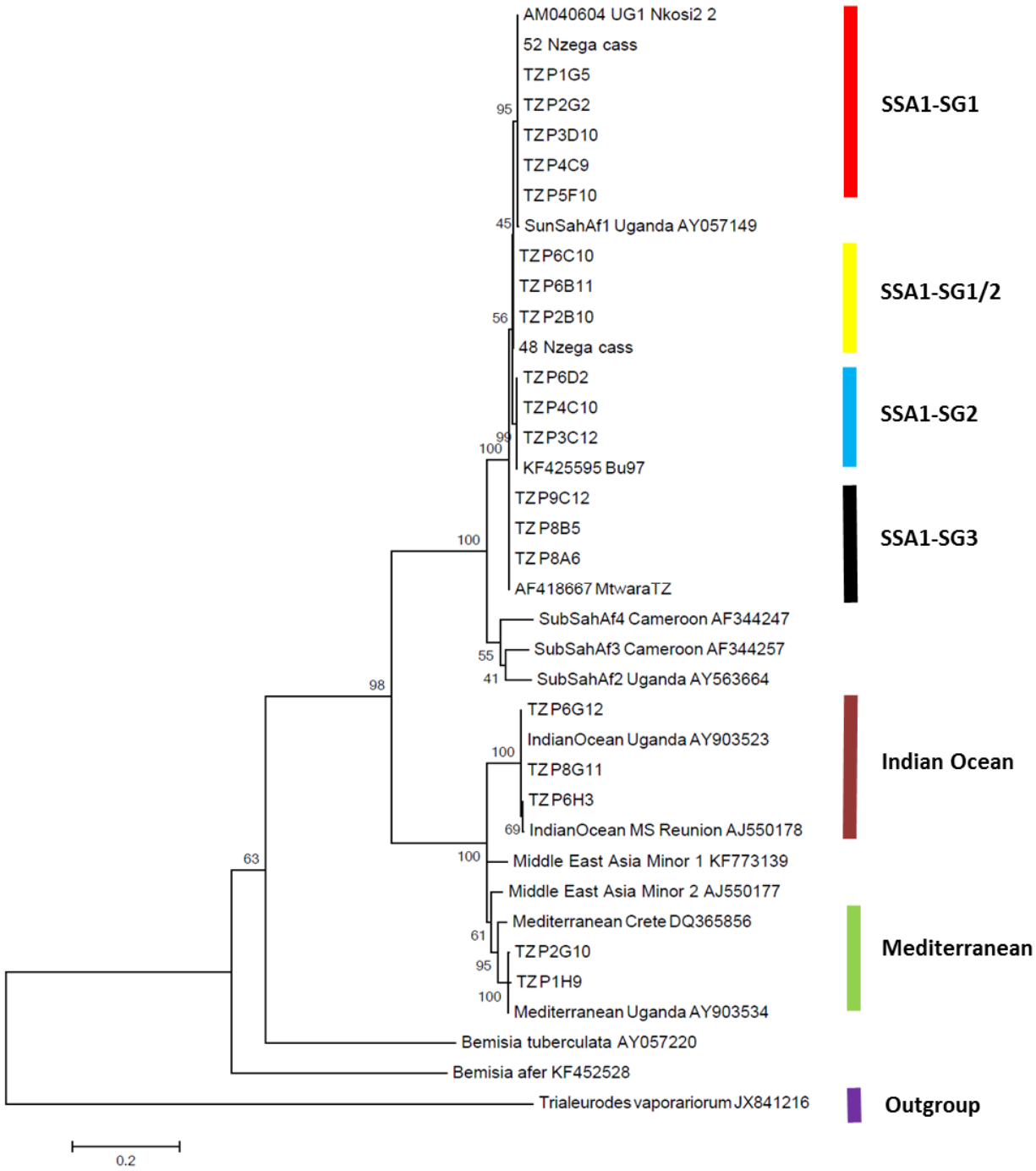


Figure 11 Maximum likelihood phylogeny of mtCOI sequences of *Bemisia tabaci* adults collected during 2012 – 2013 surveys with selected reference sequences from GenBank. All sequences starting with 'TZ' are sequences produced in this study whereas all other sequences are reference sequences from other studies or GenBank. The tree is drawn to scale and the scale bar represents the number of substitutions per site. SSA1 represents Sub-Saharan Africa 1 putative species; and SG, sub-groups.

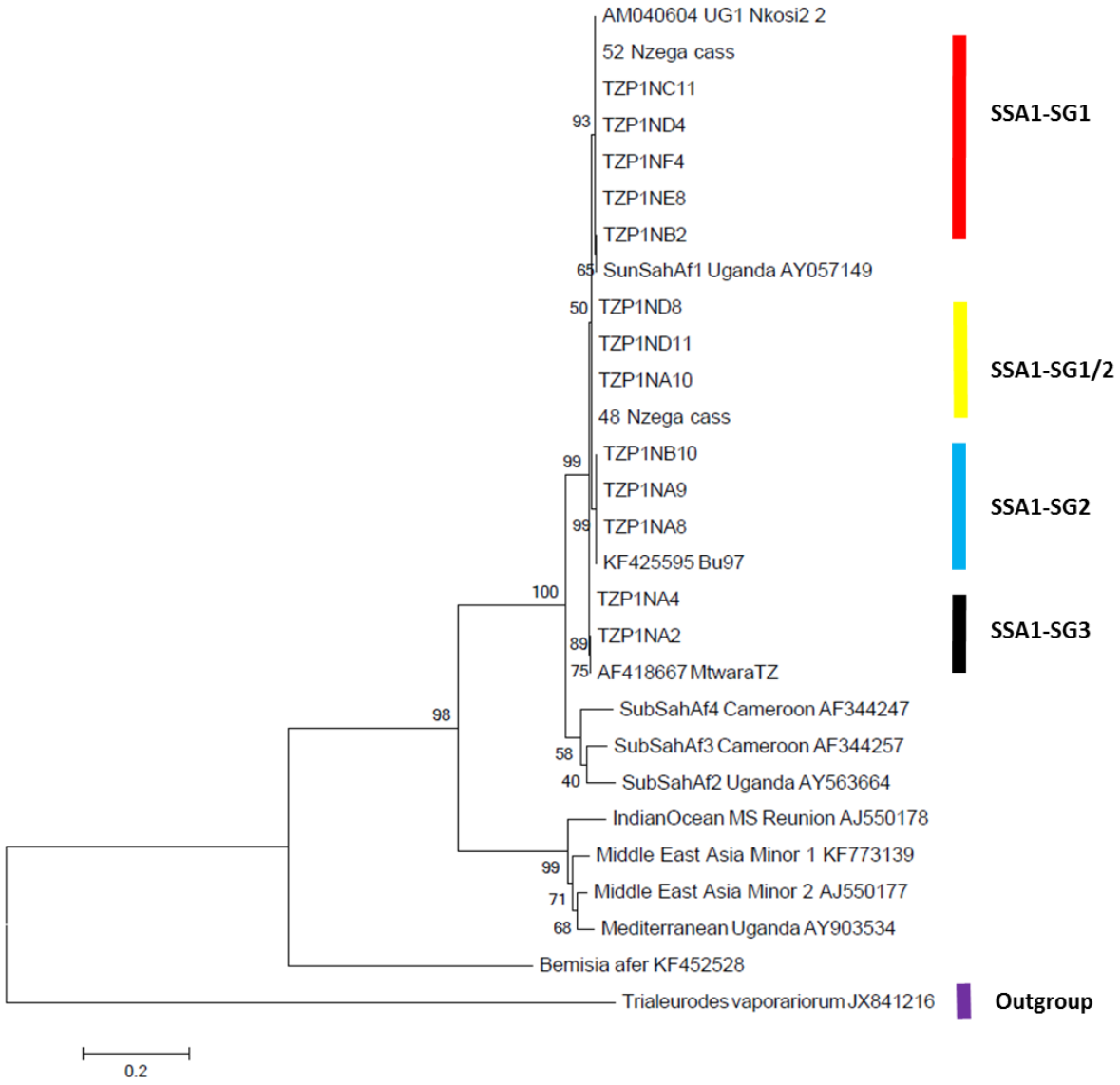


Figure 12 Maximum likelihood phylogeny of mtCOI sequences of *Bemisia tabaci* nymphs collected during 2012 – 2013 surveys with selected reference sequences from GenBank. All sequences starting with 'TZ' are sequences produced in this study whereas all other sequences are reference sequences from other studies or GenBank. The tree is drawn to scale and the scale bar represents the number of substitutions per site. SSA1 represents Sub-Saharan Africa 1 genetic group; and SG, sub-groups.

Adult whiteflies in SSA1-SG1 were *ca* 1.7% divergent fromSSA1-SG2 and 1.5% from SSA1-SG3 whereas nymph sequences diverged by 1.6% and 1.3%, respectively. All sequences in SSA1-SG1 showed a close relationship to a reference sequence in GenBank - AM040604 from Uganda (Maruthi MN, 2005, unpublished sequence). SSA1-SG2 individuals, showing close identity with a reference sequence in GenBank - KF425595 from Burundi, diverged by *ca*

1.4% (1.5% for nymphs) from SSA1-SG3 which were very close to the previously published haplotype AF418667 from Tanzania (Maruthi *et al.*, 2004). The intermediate group SSA1-SG1/2 diverged by 0.8% (0.7% for nymphs) from SSA1-SG1 and by 0.9% (0.9% for nymphs as well) from SSA1-SG2.

Adult *B. tabaci* individuals in the SSA1 genetic group diverged at least by 15% from those in the MED or IO genetic groups, which were in turn divergent by more than 6% from each other. Among the *B. tabaci* species complex, SSA1 nymphs were the only ones found colonizing cassava.

SSA1-SG1 was the dominant sub-group in north-western Tanzania (Lake zone) although it was found in two fields as far as Singida Region in the central part of the country. SSA1-SG2 was found both in the north-western as well as central part of the country. The intermediate sub-group, SSA1-SG1/2, was recorded in the central zone whereas SSA1-SG3 was confined to the eastern/coastal part of Tanzania and was the only sub-group in the region (Figure 13).

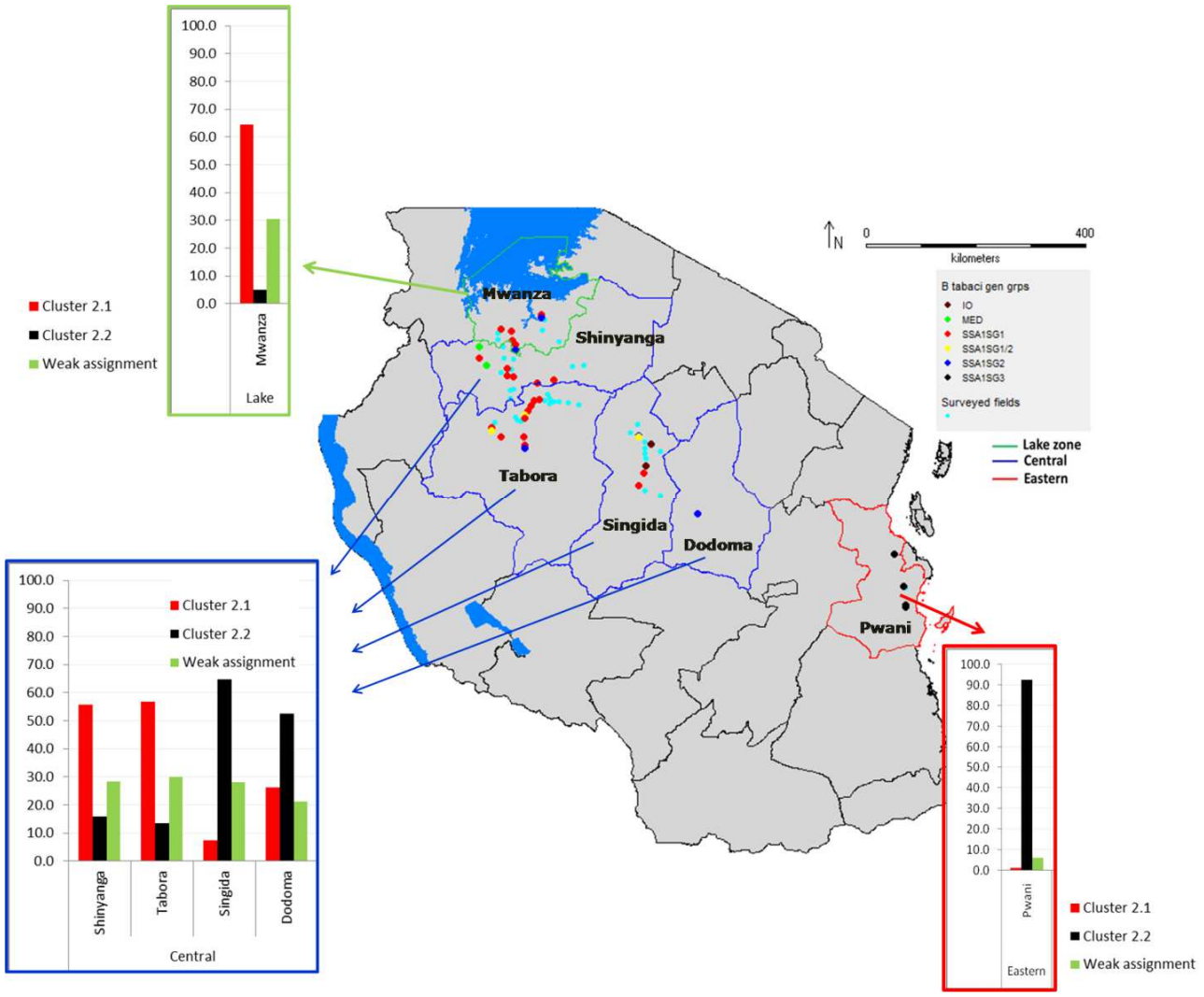


Figure 13 Distribution of *Bemisia tabaci* species complex and locations of fields surveyed during 2012 and 2013 in Tanzania. The dots represent the surveyed fields and their colours differentiate the different *Bemisia tabaci* putative species. SSA1 stands for sub-Saharan Africa 1, SG for sub-group, IO for Indian Ocean, and MED for Mediterranean putative species. The bar graphs show the dominant sub-cluster within each sampling Region and zone based on K=2 STRUCTURE cluster assignment (in percent) of *Bemisia tabaci* SSA1 individuals.

4.3.3 Genetic variation

Within the *Bemisia tabaci* species complex, significant values were observed for the average within-population (Fis) heterozygote deficiency of 23 populations (out of 31) in the SSA1 putative species, of which 9 were highly significant at P<0.0001. Average null allele frequencies were lower than 10% across the three putative species (MED=0.06, SSA1=0.047 and IO=0.043). The average allelic richness was the highest for IO followed by Med and SSA1 (Table 10). Among the two sub-clusters of SSA1, cluster 2.1 showed higher allelic richness

than cluster 2.2 (Appendix 7). Within *Bemisia afer* populations, the heterozygote deficiency was significant for 5 out of 11 populations. In all the populations, the null allele frequency was less than 5% (Table 10).

Table 10 Genetic variation of *Bemisia tabaci* species complex and *Bemisia afer*; SSA1 stands for sub-Saharan Africa 1, IO for Indian Ocean, MED for Mediterranean genetic group, SE stands for standard error, Fis for inbreeding coefficient, Hexp for expected heterozygosity, Hnb for non-biased heterozygosity, and Hobs for observed heterozygosity.

Whitefly population	Number of individuals per site (n)	Zone	Allelic Richness ± SE	Average Number of allele over all loci	Null allele frequency	Fis ± SE	Hexp.	Hn.b.	Hobs.
***Bemisia tabaci* species complex**									
SSA1									
F13ca13	30	Central	1.453±0.111	4.6	0.052	0.140±0.178*	0.445	0.453	0.428
F18ca12	19	Central	1.485±0.120	5.1	0.055	0.119±0.090***	0.472	0.486	0.405
F19ca12	11	Central	1.498±0.120	4.1	0.056	0.196±0.112**	0.474	0.498	0.378
F19ca13	8	Central	1.506±0.107	3.9	0.026	-0.02±0.119	0.475	0.506	0.516
F1ca13	18	Central	1.459±0.102	3.9	0.057	0.047±0.134**	0.447	0.460	0.402
F1capu12	8	Central	1.477±0.117	4.0	0.026	0.023±0.091*	0.446	0.477	0.446
F20ca12	7	Central	1.509±0.124	3.6	0.040	0.2±0.085**	0.467	0.510	0.396
F24ca13	6	Central	1.494±0.116	3.1	0.033	0.064±0.149	0.445	0.495	0.450
F25ca13	25	Central	1.509±0.121	5.4	0.057	0.035±0.143***	0.498	0.510	0.458
F26ca13	24	Central	1.498±0.120	5.4	0.048	0.110±0.096***	0.486	0.498	0.424
F27ca13	5	Central	1.517±0.119	2.9	0.018	-0.03±0.121	0.463	0.517	0.519
F28ca12	8	Central	1.476±0.143	3.8	0.038	0.105±0.129	0.439	0.476	0.422
F28ca13	27	Central	1.473±0.102	5.5	0.050	0.150±0.106**	0.461	0.474	0.379
F29ca12	10	Central	1.471±0.105	4.1	0.011	-0.02±0.073	0.444	0.471	0.467
F2ca13	17	Central	1.423±0.112	4.5	0.056	0.112±0.160**	0.411	0.424	0.404
F2casp12	23	Central	1.504±0.116	3.5	0.017	0.074±0.090	0.462	0.504	0.455
F30ca13	24	Central	1.421±0.130	3.0	0.033	0.130±0.126*	0.394	0.421	0.353
F31ca13	24	Central	1.5±0.104	5.4	0.081	0.251±0.170***	0.488	0.500	0.401
F35ca13	12	Central	1.474±0.128	4.4	0.045	0.086±0.122*	0.454	0.475	0.406
F36ca13	8	Central	1.461±0.130	4.1	0.020	-0.04±0.109	0.431	0.462	0.475
F37ca12	19	Central	1.483±0.124	5.3	0.068	0.190±0.098***	0.469	0.484	0.371
F39ca12	26	Central	1.519±0.114	5.9	0.048	0.068±0.103***	0.509	0.519	0.452
F39ca13	18	Central	1.483±0.106	4.8	0.008	-0.14±0.102	0.470	0.483	0.547
F40ca13	5	Central	1.550±0.094	2.9	0.107	0.354±0.198*	0.483	0.550	0.375
F5ca13	11	Central	1.451±0.096	3.8	0.077	0.145±0.169***	0.433	0.452	0.365
F1ca13Mk	30	Eastern	1.446±0.099	6.5	0.058	0.118±0.120***	0.439	0.447	0.380
F4ca13Rf	11	Eastern	1.473±0.076	4.4	0.069	0.139±0.158**	0.449	0.473	0.398
Kibca13	26	Eastern	1.452±0.131	5.9	0.028	0.066±0.061*	0.442	0.452	0.409
F41ca13	14	Lake	1.513±0.105	4.4	0.067	0.201±0.191**	0.494	0.513	0.438
F42ca13	6	Lake	1.495±0.116	3.3	0.074	0.261±0.153**	0.453	0.495	0.358
F51ca13	31	Lake	1.499±0.118	6.5	0.041	0.032±0.109***	0.491	0.499	0.460
Average							**0.459**		**0.424**

IO									
F1capu12	10	Central	5.097±0.708	5.3	0.056	0.137±0.106**	0.651	0.687	0.565
F69ca12	11	Central	5.232±0.883	5.4	0.029	-0.09±0.123	0.647	0.678	0.727
Average							**0.649**		**0.646**
MED									
F10sp12	5	Central	2.514±0.230	2.4	0.058	0.036±0.212	0.459	0.512	0.506
F3sp12	7	Central	2.639±0.273	3.3	0.078	0.058±0.230	0.488	0.528	0.509
Average							**0.474**		**0.508**
Bemisia afer									
F11ca13	8	Central	2.429±0.353	3.7	0.025	-0.00±0.184*	0.546	0.590	0.543
F12ca13	14	Central	2.263±0.318	3.9	0.025	0.028±0.199	0.512	0.536	0.497
F19ca13	6	Central	2.116±0.372	2.4	0.034	0.167±0.163	0.410	0.463	0.362
F23ca13	7	Central	2.127±0.352	2.8	0.034	0.358±0.184***	0.426	0.476	0.276
F26ca13	7	Central	2.062±0.334	2.7	0.009	0.275±0.230	0.408	0.457	0.308
F27ca13	5	Central	2.015±0.411	2.2	0.009	-0.01±0.192	0.352	0.411	0.414
F2ca13	15	Central	2.030±0.345	4.5	0.009	0.077±0.121*	0.415	0.432	0.402
F31ca13	8	Central	2.130±0.341	2.8	0.009	0.177±0.217	0.440	0.481	0.396
F35ca13	14	Central	2.328±0.382	4.5	0.020	0.218±0.187***	0.506	0.534	0.395
F37ca13	5	Central	2.493±0.360	3.2	0.011	0.148±0.215	0.531	0.631	0.529
F3ca13Rf	18	Eastern	2.273±0.380	4.7	0.011	0.067±0.186***	0.503	0.523	0.445
Average							**0.459**		**0.415**

*P<0.05; ** P<0.01; ***P<0.001. P values without '*' are non-significant; Allelic Richness for *B. afer* was calculated based on seven loci as P59amplified only for less than 20% of the population.

A pairwise Fst comparison within the SSA1 *B. tabaci* putative species showed highly significant differences between Lake and Eastern zone (F_{st}=0.093, P<0.00001) and between Central and Eastern zone (F_{st}=0.079, P<0.00001). The Fst comparison between Lake and Central zone was not significant. There were only two populations within each of IO and Med and the pairwise Fst comparison between the populations in IO revealed significant values (F_{st}=0.042, P<0.05) whereas the comparison between the two populations within Med was not significant.

The two sub-clusters within SSA1 (cluster 2.1 and cluster 2.2) had different frequency in each Region and in each sampling Zone (Figure 13). Cluster 2.2 was dominant in the eastern part of Tanzania (Coast Region, Eastern Zone) where more than 90% of the individuals belonged to

this sub-cluster and only less than 2% belonging to cluster 2.1. Conversely, cluster 2.1 was dominant in north-western part of the country (Mwanza Region, Lake Zone) with only 5.1% belonging to the other sub-cluster. The Central Zone was a mix where cluster 2.1 showed a rather diminishing pattern moving closer to the east and cluster 2.2 increased and vice versa. Individuals with weak STRUCTURE assignment to either of the sub-clusters were observed in all Regions with more than 20% frequency except in Coast Region where they had very low frequency (6.3%) (Figure 13).

Within the SSA1 putative species, ca 12% of the population were hybrids between the two sub-clusters (70 out of 577 individuals) being weakly assigned by STRUCTURE to the sub-clusters with a score ranging from 0.4 – 0.6. The PCA of this population also revealed possible hybridization and gene flow between the two sub-clusters with overlapping individuals and PCA confidence ellipses (Figure 14).

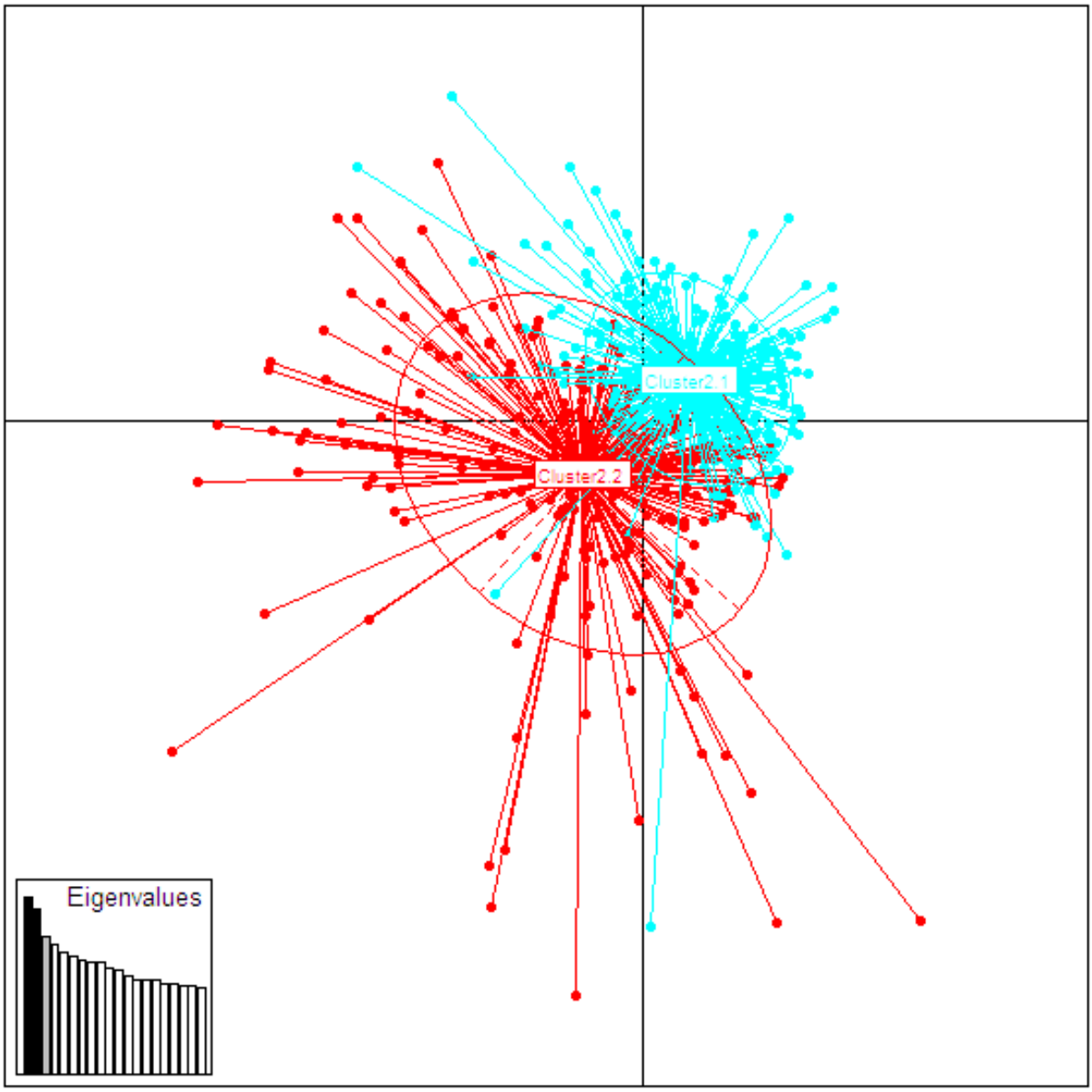


Figure 14 PCA of the two sub-clusters of SSA1 putative species as identified by K2 of STRUCTURE.

4.3.4 Geographic structuring

Since the SSA1 *B. tabaci* putative species had well distributed populations over the sampling sites compared to the other putative species, we further analysed its geographic structuring by grouping the sampling sites into three main zones: Lake, Central and Eastern/Coastal zone.

A hierarchical AMOVA test showed that the variance within individuals explained 81.38% of the total variance (F_{it}=0.19, P<0.00001) (Table 11). Highly significant differences (F_{ct}=0.04,

P<0.00001) were observed among zones and among populations within zones (F_{sc}=0.04, *P*<0.00001) although the genetic differences accounted for 3.87% and 3.76% of the total variance, respectively.

Table 11 Analysis of molecular variance and F-statistics of genetic differentiation among the three sampling zones within *Bemisia tabaci* SSA1 putative species

Source of variation											
Among zones (Lake, Central, Eastern zone)			Among populations within Zones			Among individuals within population			Within individuals		
V	%	F_{ct}	V	%	F_{sc}	V	%	F_{is}	V	%	F_{it}
0.08	3.87	0.04***	0.08	3.76	0.04***	0.23	10.99	0.12***	1.70	81.38	0.19***

The *P* value is less than 10^{-5} for all the *F-statistics*.

4.3.5 Isolation by distance

The relationship between the geographical distances and the genetic distances was assessed for the SSA1 *B. tabaci* putative species. Significant correlation was observed between geographical and genetic (F_{st}) distance within the *B. tabaci* SSA1 putative species with Pearson r=0.645. The Z value test run to assess geographical and genetic distance matrix independence according to Mantel (1967) implemented in Genetix was significant (*P* value of 0.0002), with unilateral test done on 1000 permutations, hence indicating strong relationship and isolation by distance. A strong positive relationship was observed particularly until 250 km. A graphical illustration of the correlation between molecular variation and geographical distance is presented in Appendix 8.

4.4 Discussion

In the current study, both STRCUTURE analysis and mtCOI sequencing retrieved four clusters three of which were *Bemisia tabaci* putative species – SSA1, IO, and MED - and the fourth

group, a cluster of *Bemisia afer*. Within the SSA1 putative species, further geographic and genetic structuring was revealed with gene flow. The pandemic associated SSA1-SG1 was found extending its presence further into the central part (areas where it was absent previously) in what appears to be a genetic invasion by hybridizing with the local population (Figure 13). Heterozygous deficiencies were observed for SSA1 and IO putative species. The SSA1 population showed lower allelic richness when compared to the other putative species. The lower values for allelic richness could have arisen from the fact that many of the microsatellite loci used in the study were isolated from 'B biotype' and hence more variation is generally expected in the population the loci were isolated from (the focal species) (Hutter *et al*., 1998). In their study on two *Drosophila* species, Hutter *et al*. (1998) recorded significantly higher number of alleles and heterozygosity in the focal than the nonfocal species. The occurrence of null alleles which is the failure of microsatellites to amplify could also affect heterozygous deficiencies in a population. However, in our study we observed low null allele frequencies throughout.

Many whitefly species in the genus *Bemisia* are restricted to few host plants (Mound and Halsey, 1978). However, *B. tabaci,* the most economically important species in the genus, has been recorded from several plant families including Asteraceae, Brassicaceae, Convolvulaceae, Cucurbitaceae, Euphorbiaceae, Fabaceae, Malvaceae and Solanaceae (Oliveira *et al.,* 2001). It can colonize more than 900 plant species (Hsieh *et al.*, 2006) and can adapt to originally less suitable host plants within a few generations (Basu, 1995). Nevertheless, not all hosts are equally suitable for feeding and/or reproduction. In our study although IO and Med adults were found on cassava, no IO or Med nymphs were found on cassava and cassava was found to be colonized only by SSA1 individuals. This confirmed the findings of previous studies thatcassava *B. tabaci* is strongly associated to cassava in Africa and that it is only colonized by

cassava *B. tabaci* (SSA1 genetic group) (Burban *et al.,* 1992; Legg, 1996; Abdullahi *et al.,* 2003; Sseruwagi *et al.*, 2006). According to Sseruwagi *et al.* (2006), the non-cassava adult *B. tabaci* collected from cassava are visitors in search of food or a place for oviposition, which is also supported by van Lenteren and Noldus (1990).

To our expectation, the pairwise F_{st} comparison within SSA1 revealed highest variation between Lake and Eastern zone and lowest variation between Lake and Central zone. This is further supported by two facts in our study. The first is, the presence of similar variation based on mitochondrial data. The distribution of SSA1 sub-groups based on mitochondrial data shows the coastal type (SSA1-SG3) to be the only sub-group found confined to Eastern zone. The other two sub-groups (SG1 and SG2) are shown to be found dominantly in the Lake and Central Zone, respectively. The second supporting evidence comes from STRUCTURE assignment based on K=2 into the two sub-clusters within SSA1 (cluster 2.1 and cluster 2.2) which shows the Eastern Zone (Coastal Region) to be strongly dominated by cluster 2.2 with almost no cluster 2.1 and very few hybrids (those weakly assigned any of the clusters by STRUCTURE). However, as we move further across the Central and to the Lake zone, cluster 2.2 shows lower proportions until it becomes very low in the Lake zone where cluster 2.1 takes over as the dominant type indicating clear polarized variation between the Lake and Eastern zones for the two clusters. Even though the two clusters seem to have their own geographic niches, there is clear evidence of hybridization as signalled by the presence of high proportions of STRUCTURE weakly assigned individuals.

SSA1-SG1 which is assumed to be the dominant sub-group in north-western Tanzania was found in two fields further in the central part of the country in Singida Region. Similarly, what was assumed to be the most dominant sub-group in the central part of Tanzania - SSA1-SG2,

was found in three fields in the north-western part of the country in our study. If it was due to new movements of the sub-groups to previously unrecorded areas or if it was the case that they were there even before but they went undetected due to limited sampling could not be verified from this study. However, whichever the case might be, the presence of SSA1-SG1– assumed to be the pandemic associated super-abundant sub-group – (Legg J, 2013, pers. comm.; Tajebe LS, unpublished data) further into the central parts of the country poses the risk of expansion of the CMD pandemic into areas previously unaffected by the pandemic. This is even more threatening now than ever as the two sub-clusters identified by STRUCTURE do hybridize which means that there is even a greater risk for the pandemic type (SSA1-SG1) to make its way into areas previously unaffected by the sever CMD pandemic hence possibly facilitating the quick spread of the severe pandemic. Therefore, it is crucial to direct future studies towards understanding the population structure of the *Bemisia tabaci* species complex in East and Central Africa. Moreover, it is important to conduct mating experiments involving different genetic groups in Tanzania and in the wider region of East and Central Africa to study gene flow and hybridization events between the genetic groups of the *Bemisia tabaci* species complex as understanding the vector is a big part of tackling the disease pandemic that is hindering cassava production in Africa.

5. Conclusions

The work described here combined field surveys and molecular approaches including mitochondrial and nuclear markers (microsatellites) to study the *B. tabaci* species complex in Tanzania. Four *B. tabaci* putative species including sub-Saharan Africa 1 (SSA1), Mediterranean (MED), Indian Ocean (IO) and East Africa 1 (EA1) were identified in the study. Both mitochondrial and microsatellite markers were concordant in the fact that they both were able to discriminate between the different *B. tabaci* putative species. Within the SSA1 putative species mitochondrial data further revealed four sub-groups – SG1, SG1/2, SG2 and SG3 – whereas microsatellite data identified two sub-clusters.

In the current study SSA1-SG1 was principally found in the pandemic-affected north-western parts of Tanzania, where whitefly super-abundance and high mean disease severity were evident similarly to the severe CMD pandemic in East and Central Africa. Based on these findings from the present study and the findings of earlier studies with regards to the wide prevalence,in the study area, of the recombinant virus causing the severe CMD pandemic, it was possible to conclude that SSA1-SG1is the pandemic-associated species of the *B. tabaci* group in Tanzania.

Although mitochondrial data discriminated four sub-groups and microsatellites two sub-clusters within the SSA1 putative species, both markers revealed similar trends in the specific geographic distributions observed for these sub-groups along a transect from north-western to south-eastern/coastal Tanzania. While the pandemic-associated SSA1-SG1 dominated in the north-west, SSA1-SG3 was confined to the coastal parts, being absent completely from the central and north-western parts of the country. Therefore, from both mitochondrial and microsatellite data, it was possible to conclude that the cassava-associated SSA1 putative

species shows sub-population structure, and that sub-clusters/sub-groups have distinct geographical distributions in which one sub-cluster has a more coastal distribution being either solely confined (based on mitochondrial data) or showing decreasing in frequency (based on microsatellite data) with increasing distanceinland and away from the coast in the northwesterly direction. Similarly, the other sub-cluster dominates in the north-western pandemic-affected region, extending however to some location in central Tanzania, but reducing in frequency along the transect from the North-West to the South-East and the coast. The central region was a mix of the two sub-clusters. Several individuals within the SSA1 putative species were weakly assigned to both sub-clusters identified by microsatellite data. A principal component analysis further plotted these individuals showing an overlapping pattern. This evidence suggested the occurrence of some gene flow and hybridization between the two sub-clusters.

Contrasting endosymbiont profiles were evident between the different sub-groups and putative species observed in the study. Although no particular endosymbionts were found to be associated only with the pandemic-associated SSA1-SG1, its distinctive nature was further supported by its endosymbiotic profile. Strikingly, this sub-group was the only one that included several individuals with no secondary symbiont infection. Other sub-groups frequently harboured multiple secondary symbionts. When SSA1-SG1 individuals were infected, single infection by *Arsenophonus* or double infection by *Arsenophonus*/*Cardinium* was common. The MED, IO and EA1 putative species frequently showed double or multiple infections.

This study also confirmed findings of previous studies that reported cassava to be colonized only by cassava type *B. tabaci* in Africa. It also concluded that the adults of non-SSA1 putative species that were found on cassava were just "visitors" since their nymphs were not present on cassava.

The CMD pandemic front was estimated to lie in Geita Region located in north-western Tanzania and to be spreading from the north-west to south-east of the country at ca. 26 km per year. However, the fact that there is little cassava grown in central Tanzania due to low rainfall and short growing seasons may serve as a natural barrier to the spread of the severe CMD pandemic further to the south-east.

Overall, the different *B. tabaci* putative species and the different sub-groups within SSA1 showed distinct geographic distributions and contrasting endosymbiotic profiles. SSA1-SG1 was present in areas up to 180 km to the south-east of the pandemic front, illustrating that changes in whitefly populations precede changes in disease status. SSA1-SG1 was notably present in the central part of Tanzania – an area not yet affected by the severe CMD pandemic. Finally, distinct genetic makeup, possible sub-cluster hybridization, and distinguished secondary symbionts profile could all have an impact on the current fitness of the pandemic-associated SSA1-SG1 and hence could partly explain the distinctiveness of this sub-group. However, more studies are needed to explicitly establish these supposed elements.

This doctoral work advanced considerably the existing knowledge on whiteflies giving the first thorough report on super-abundant *Bemisia tabaci* populations in Tanzania. It confirmed the association of a distinct genetic sub-group of *B. tabaci* with the severe CMD pandemic in Tanzania more than a decade after similar associations were first made in Uganda in 2002.

Some evidence of gene flow and hybridization among the sub-clusters of the SSA1 putative species was also reported for the first time in this study. This doctoral study was also the first work that provided baseline information on the complex symbiotic communities of different *B. tabaci* putative species for a part of East Africa which until now had not been studied.

6. Recommendations

In this study the genetic diversity, population structure, geographic distribution and endosymbiont profiles of the *Bemisia tabaci* species complex were investigated in Tanzania. A distinct sub-group of the SSA1 putative species was found to be the CMD pandemic-associated group. While there was no particular endosymbiont associated to it, the group comprised several individuals with no secondary symbiont infection. To establish the reason behind the observed higher frequency of individuals in the group with no secondary symbionts, future studies should first focus on determining the exact roles of secondary symbionts by trying to eliminate each secondary symbiont with an antibiotic treatment or by artificially introducing some endosymbionts and following up their subsequent effects on the whitefly individuals. Additionally, studies to assess the presence, if any, of an unknown symbiont in the pandemic-associated sub-group are important.

Cassava has a strict relationship with *B. tabaci* and is not colonized by non-cassava type *B. tabaci* in Africa. Cassava type *B. tabaci* are able to survive when transferred from cassava to other plants while non-cassava types cannot survive and are not able to colonize cassava when transferred from other plants. This could perhaps have something to do with the cassava plant which contains cyanogenic glucoside (CG) compounds. The *B. tabaci* that colonize cassava could, through co-evolution, have developed important detoxification systems that enable them colonize it. But as a first step, studies are needed to examine if CG are indeed toxic to non-cassava *B. tabaci* types and not to the cassava types. This can be done through artificial diet experiments with varying levels of CG diets to be fed to the cassava type and to different putative species known to be found on cassava ("visitors") but not able to colonize it. These investigations will shed light on the relationship of plant chemistry and

vector biology in this regard. Such insights will inform the on-going research effort to manage the disease and the pest.

It will also be important to determine the impact of temperature and climate change on the CMD pandemic and whitefly populations in order to ensure that any control strategies that are developed will be climate-robust, as the crop is expected to be even more important with climate change and erratic rainfall in East Africa.

References

- Abdullahi I, Winter S, Atiri GI, Thottappilly G, 2003. Molecular characterisation of whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae) populations infesting cassava. Bull. Entomol. Res. 93, 97–106.
- Alemandri V, De Barro P, Bejerman N, Arguello Caro EB, Dumon AD, Mattio MF, Rodriguez SM, Truol G, 2012. Species within the *Bemisia tabaci* (Hemiptera: Aleyrodidae) complex in soybean and bean crops in Argentina. J. Econ. Entomol. 105, 48–53.
- Alicai T, Omongo CA, Maruthi MN, Hillocks RJ, Baguma Y, Kawuki R, Bua A, Otim-Nape GW, Colvin J, 2007. Re-emergence of cassava brown streak disease in Uganda. Plant Dis. 91, 24-29.
- Allem AC, 2002. The Origins and Taxonomy of Cassava. In: Cassava: biology, production and utilization. Eds. Hillocks RJ, Thresh JM, Bellotti AC. CABI.
- Alves AAC, 2002. Cassava Botany and Physiology. In: Cassava: biology, production and utilization. Eds. Hillocks RJ, Thresh JM, Bellotti AC. CABI.
- Basu AN, 1995. *Bemisia tabaci* (Gennadius): Crop Pest and Principal Whitefly Vector of Plant Viruses. Westview Press, New Delhi, 183pp.
- Baumann L., Thao ML, Funk CJ, Ng JCK, Baumann P, Falk BW, 2004. Sequence analysis of DNA fragments from the genome of the primary endosymbiont of the whitefly *Bemisia tabaci*. Curr. Microbiol. 48, 77-81.
- Baumann P, 2005. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu. Rev. Microbiol. 59, 155–189.
- Baumann P, and Moran NA, 1997. Non-cultivable microorganisms from symbiotic associations of insects and other hosts. Antonie van Leeuwenhoek 72, 39–48.

- Baumann P, Baumann L, Lai C-Y, Rouhbakhsh D, Moran NA, Clark MA, 1995. Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids. Annu. Rev. Microbiol. 49, 55–94.
- Baumann P, Moran NA, Baumann L, 2006. Bacteriocyte-Associated Endosymbionts of Insects. Prokaryotes. 1, 403–438. DOI: 10.1007/0-387-30741-9_16
- Bedford ID, Briddon RW, Brown JK, Rosell RC, Markham PG, 1994. Geminivirus transmission and biological characterization of *Bemisia tabaci* biotypes from different geographic regions. Ann. Appl. Biol. 125, 311-325.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F, 1996–2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).
- Berrie LC, Rybicki EP, Rey MEC, 2001. Complete nucleotide sequence and host range of South African cassava mosaic virus: further evidence for recombination amongst geminiviruses. J. Gen. Virol. 82, 53–58.
- Best R and Henry G, 1992. In Report of the First Meeting of the International Network for Cassava Genetic Resources, eds. Roca WM and Thro AM (Cent. Int. Agric. Trop., Call, Colombia), pp. 3-11.
- Bing XL, Yang J, Zchori-Fein E, Wang XW, Liu SS, 2013. Characterization of a newly discovered symbiont in the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae). Appl. Environ. Microbiol. 79, 569–575.
- Bird J, 1957. A whitefly-transmitted mosaic of *Jatropha gossypifolia*. *Technical Paper Agricultural Experiment Station University of Puerto Rico,* **22**, 1–35.
- Bock KR, Guthrie EJ, Meredith G, 1978. Distribution, host range, properties and purification of cassava latent virus, a geminivirus. Ann. Appl. Biol. 90, 361–367.

- Bock KR, Woods RD, 1983. The etiology of African cassava mosaic disease. Plant Dis. 67, 994–995.
- Bosco D, Loria A, Sartor C, Cenis JL, 2006. PCR-RFLP identification of *Bemisia tabaci* biotypes in the Mediterranean Basin. Phytoparasitica 34, 243-251.
- Boykin LM, Armstrong KF, Kubatko L, De Barro P, 2012. Species delimitation and global biosecurity. Evol. Bioinform. 8, 1–37.
- Boykin LM, Shatters RG Jr, Rosell RC, McKenzie CL, Bagnall RA, De Barro P, Frohlich DR, 2007. Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. Mol. Phyl. Evol. 44, 1306–1319.
- Brown JK, 1990. An update on the whitefly-transmitted geminiviruses in the Americas and the Caribbean Basin. FAO Bull. 39, 5.
- Brown JK, 1994. The status of *Bemisia tabaci* (Genn.) as a pest and vector in world agroecosystems. FAO Plant Prot. Bull. 42, 3–32.
- Brown JK, 2000. Molecular markers for the identification and global tracking of whitefly vector– Begomovirus- complexes. Virus Res. 71, 233-260.
- Brown JK, 2007. The *Bemisia Tabaci* Complex: Genetic and phenotypic variation and relevance to TYLCV–vector interactions. In: Tomato Yellow Leaf Curl Virus Disease. Ed. by Czosnek H, Springer, Netherlands, 25–56.
- Brown JK, Coats SA, Bedford ID, Markham PG, Bird J, 1992. Biotypic characterization of *Bemisia tabaci* populations based on esterase profiles, DNA fingerprinting, virus transmission, and bioassay to key host plant species. Phytopathology 82, 110.
- Brown JK, Coats SA, Bedford ID, Markham PG, Bird J, Frohlich DR, 1995b. Characterization and distribution of esterase electromorphs in the whitefly, *Bemisia tabaci* (Genn.)(Homoptera: Aleyrodidae). Biochem. Genet. 33, 205-214.

- Brown JK, Frohlich DR, Rosell RC, 1995a. The sweetpotato or silverleaf whiteflies: biotypes of *Bemisia tabaci* or a species complex? Annu. Rev. Entomol. 40, 511- 534.
- Brumin M, Kontsedalov S, Ghanim M, 2011. *Rickettsia* influences thermotolerance in the whitefly *Bemisia tabaci* B biotype. Insect Sci. 18, 57–66. DOI: 10.1111/j.1744-7917.2010.01396.x.
- Buchner P, 1965. *Endosymbiosis of animals with plant microorganisms*. New York: John Wiley and Sons.
- Bull CT, De Boer SH, Denny TP, Firrao G, Fischer-Le Saux M, Saddler GS, Scortichini M, Stead DE, Takikawa Y, 2010. Comprehensive list of names of plant pathogenic bacteria, 1980-2007. J. Plant Path. 92, 551-592.
- Bull SE, Briddon RW, Sserubombwe WS, Ngugi K, Markham PG, Stanley J, 2006. Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. J. Gen. Virol. 87, 3053–3065.
- Burban C, Fishpool LDC, Fauquet C, Fargette D, Thouvenel J-C, 1992. Host-associated biotypes within West African populations of the whitefly *Bemisia tabaci* (Genn.) (Hom.,Aleyrodidae). J. Appl. Entomol. 113, 416–423.
- Byrne FJ, Cahill M, Denholm I, Devonshire AL. 1995. Biochemical identification of interbreeding between B-type and non B-type strains of tobacco whitefly *Bemisia tabaci*. Biochem. Genet. 33,13–23
- Campbell BC, 1989. On the role of microbial symbiotes in herbivorous insects. In *Insect-Plant Interactions*, Vol.I, ed. EA Bernays. Boca Raton, FL: CRC Press.
- Campbell BC, 1993. Congruent evolution between whiteflies (Homoptera: Alerodidae) and their bacterial endosymbionts based on respective 18S and 16S rDNAS. Curr. Microbiol. 26, 129–132.

- Ceballos H, Iglesias CA, Pe´rez JC, Dixon AGO, 2004. Cassava breeding: opportunities and challenges. In: Plant Molecular Biology 56, 503–516.
- CGIAR - Consultative Group on International Agricultural Research, 1997. CGIAR News: Victory Over the Cassava Green Mite. 4 (4).
- Chant SR, 1958. Studies on the transmission of cassava mosaic virus by *Bemisia* spp. (Aleyrodidae). Ann. Appl. Biolo. **46,** 210–215.
- Chapuis MP and Estoup A, 2007. Microsatellite null alleles and estimation of population differentiation. Mol. Biol. Evol. 24, 621-631.
- Chiel E, Gottlieb Y, Zchori-Fein E, 2007. Biotype-dependent secondary symbiont communities in sympatric populations of *Bemisia tabaci*. Bull. Entomol. Res. 97, 407–413.
- Chiel E, Zchori-Fein E, Inbar M, Gottlieb Y, Adachi-Hagimori T, Kelly SE, Asplen MK, Hunter MS, 2009. Almost there: transmission routes of bacterial symbionts between trophic levels. PLoS ONE, 4, e4767. DOI:10.1371/journal. pone.0004767.
- Chikoti PC, Shirma R, Legg JP, 2013. First report of super-abundant *Bemisia tabaci* on cassava in Zambia, In: First International Whitefly Symposium. Ed. by Roditakis E, Tsagkarakou A and Vontas J, Greece.
- Chowda-Reddy RV, Krankumar M, Seal SE, Muniyappa V, Valand GB, Govindappa MR, Colvin J, 2012. *Bemisia tabaci* phylogenetic groups in India and the relative transmission efficacy of tomato leaf curl Bangalore virus by an indigenous and an exotic population. J. Integrative Agr. 11, 235–248.
- Chu D, Wan F-H, Tao Y-L, Liu G-X., Fan Z-X, Bi Y-P, 2008. Genetic differentiation of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) biotype Q based on mitochondrial DNA markers. Insect Sci, 15, 115–123. DOI: 10.1111/j.1744-7917.2008.00191.x.

- Cock JH, 1985.Cassava.New Potential for a Neglected Crop. Westview Press. Boulder, CO, USA, 240 pp.
- Colvin J, Omongo CA, Govindappa MR, Stevenson PC, Maruthi MN, 2006. Host-plant viral infection effects on arthropod-vector population growth, development and behaviour: management and epidemiological implications. Adv. Virus Res. 67, 419–452.
- Colvin J, Omongo CA, Maruthi MN, Otim-Nape GW, Thresh JM, 2004. Dual begomovirus infections and high *Bemisia tabaci* populations drive the spread of a cassava mosaic disease pandemic. Plant Pathol. 53, 577–584.
- Colvin J, Otim-Nape GW, Holt J, Omongo C, Seal S, Stevenson P, Gibson R, Cooter RJ, Thresh JM, 1999. Symbiotic interactions drive epidemic of whitefly-borne cassava mosaic virus disease in Uganda. Proceedings of the Cassava mosaic disease management in smallholder cropping systems workshop, Jinja, Uganda, 7-9 December, 1998.
- Costa AS, Russell LM, 1975. Failure of *Bemisia tabaci* to breed on cassava plants in Brazil (Homoptera: Aleyroideae). Cienc. Cult. 27, 390–399.
- Costa HS, Brown JK, 1991. Variation in biological characteristics and in esterase patterns among populations of *Bemisia tabaci* (Genn) and the association of one population with silverleaf symptom development. Entomol. Exp. Appl. 61, 211–219.
- Costa HS, Brown JK, Sivasupramaniam S, Bird J, 1993. Regional distribution, insecticide resistance and reciprocal crosses between the A and B biotypes of *Bemisia tabaci*. Insect Sci. Appl. 14, 255-266.
- Costa HS, Westcot DM, Ullman E, Rosell R, Brown JK, Johnson MW, 1995. Morphological variation in *Bemisia* endosymbionts. Protoplasma, 189, 194-202.

- Dade HA, 1930. Cassava Mosaic. Paper No. XXVIII. Yearbook. Department of Agriculture, Gold Coast, pp. 245– 247.
- Dale C, Beeton M, Harbison C, Jones T, Pontes M, 2006. Isolation, pure culture, and characterization of "Candidatus *Arsenophonus arthropodicus*," an intracellular secondary endosymbiont from the hippoboscid louse fly *Pseudolynchia canariensis*. Appl. Environ. Microbiol. 72, 2997–3004.
- Dalmon A, Halkett F, Granier M, Delatte H, Peterschmitt M, 2008. Genetic structure of the invasive pest *Bemisia tabaci*: evidence of limited but persistent genetic differentiation in glasshouse populations. Heredity 100, 316–325.
- Darriba D, Taboada GL, Doallo R, Posada D, 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods. 9, 772.
- Dasch GA, Weiss E, Chang K-P, 1984. Endosymbionts of insects. Eds. N. R. Krieg and J. G. Holt Bergey's manual of systematic bacteriology. Williams and Wilkins Co. Baltimore, 1, 811–833.
- De Barro PJ, Boykin LM, 2013. *Global* Bemisia *dataset release version* 31 Dec 2012. v1. CSIRO. Data Collection. 10.4225/08/50EB54B6F1042 Available at: https://data.csiro.au/dap/landingpage?execution=e1s2 [accessed 24 May 2013]
- De Barro PJ, Driver F, Trueman JW, Curran J, 2000. Phylogenetic relationships of world populations of *Bemisia tabaci* (Gennadius) using ribosomal ITS1. Mol. Phylogenet. Evol. 16, 29–36.
- De Barro PJ, Hart PJ, 2000. Mating interactions between two biotypes of the whitefly, *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Australia. Bull. Entomol. Res. 90, 103–12
- De Barro PJ, Liu S, Boykin LM, Dinsdale A, 2011. *Bemisia tabaci*: a statement of species status. Annu. Rev. Entomol. 56, 1-19.

- De Barro PJ, Scott KD, Graham GC, Lange CL, Schutze MK, 2003. Isolation and characterization of microsatellite loci in *Bemisia tabaci*. Mol. Ecol. Notes 3, 40–43.
- De Barro PJ, Trueman JWH, Frohlich DR, 2005. *Bemisia argentifolii* is a race of *B. tabaci* (Hemiptera, Aleyrodidae), the molecular genetic differentiation of *B. tabaci* populations around the world. Bull. Entomol. Res. 95, 193–203.
- Deighton FC, 1926. Annual Report of the Lands and Forestry Department. Sierra Leone, pp. 1–2.
- Delatte H, David P, Granier M, Lett, JM, Goldbach R, Peterschmitt M, Reynaud B, 2006. Microsatellites reveal extensive geographical, ecological, and genetic contacts between invasive and indigenous whitefly biotypes in an insular environment. Genet. Res. 87, 109–124.
- 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. Bull. Entomol. Res. 101, 487–497. doi:10.1017/S0007485311000101
- Delatte H, Reynaud B, Granier M, Thornary L, Lett JM, Goldbach R, Peterschmitt M, 2005. A new silverleaf-inducing biotype Ms of *Bemisia tabaci* (Hemiptera: Aleyrodidae) indigenous to the islands of the south-west Indian Ocean. Bull. Entomol. Res. 95, 29–35.
- Deng D, Otim Nape GW, Sangare A, Ogwal S, Beachy RN, Fauquet CM, 1997. Presence of a new virus closely associated with cassava mosaic outbreak in Uganda. Afr. J. Root Tuber Crops. 2, 23-28.
- Dinsdale A, Cook L, Riginos C, Buckley YM, De Barro P, 2010. Refined global analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) mitochondrial cytochrome oxidase I to identify species level genetic boundaries. Ann. Entomol. Soc. Am. 103, 196–208.

- Douglas AE, 1989. Mycetocyte symbiosis in insects. Biol. Rev. Camb. 64, 409–434.
- Douglas AE, 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. Annu. Rev. Entomol. 43, 17–37.
- Douglas AE. 2006. Phloem-sap feeding by animals: problems and solutions. J. Exp. Bot. 57, 747–754.
- Dubern J, 1979. Quelques propriétés de la mosaïque Africaine du manioc. I: La transmission. Phytopathol. Zeit. 96, 25–39.
- Dubern J, 1994. Transmission of African cassava mosaic geminivirus by the whitefly (*Bemisia tabaci*). Trop. Sci. 34, 82–91.
- Dufrenoy J and He´din L. 1929. La mosaı¨que des feuilles du manioc au Cameroun. Rev. Bot. Appl. Agric. Trop. 9, 361–365.
- Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstädter J, Husrt GD, 2008. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. BMC Biol. 6, 27. DOI:10.1186/1741-7007-6-27.
- Ellegren H, 2004. Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 5, 435–445.
- Evanno G, Regnaut S, Goudet J, 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14, 2611–2620.
- Everett KDE, Thao ML, Horn M, Dyszynski GE, Baumann P, 2005. Novel chlamydiae in whiteflies and scale insects: endosymbionts '*Candidatus Fritschea bemisiae*' strain Falk and '*Candidatus Fritschea eriococci*' strain Elm. Int. J. Syst. Evol. Micr. 55**,** 1581–1587.
- Excoffier L, Laval G, Schneidern S, 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1, 47–50. Available online at http://www.ncbi.nlm.nih.gov/pmc/ articles/PMC2658868/.

- FAO - Food and Agriculture Organization of the United Nations, 2004. The global cassava development strategy and implementation plan. Proceedings of the validation forum on the global cassava development strategy, 1. Retrieved November 14, 2014, from ftp://ftp.fao.org/docrep/fao/006/y0169e/y0169e00.pdf
- FAO - Food and Agriculture Organization of the United Nations, 2013. Save and Grow: Cassava. A Guied to Sustainable Production Intensification. Retrieved November 16, 2014, from http://www.fao.org/docrep/018/i3278e/i3278e.pdf
- FAO, 2008. Food and Agriculture Organization of the United Nations, Rome, Italy, Available at: http://www.fao.org/ag/AGP/agpc/gcds/index en.html [Accessed 12 July 2014].
- FAOSTAT - Food and Agriculture Organization of the United Nations, Statistics Division, 2014. Retrieved November 14, 2014, from http://faostat3.fao.org/browse/Q/*/E
- Fauquet C and Fargette D, 1990. African cassava mosaic virus: etiology, epidemiology, and control. Plant Dis. 74, 404–411.
- Fauquet CM and Stanley J, 2003. Geminivirus classification and nomenclature: progress and problems. Ann. Appl. Biol. 142, 165-189.
- Fondong V, Pita JS, Rey MEC, de Kochko A, Beachy RN, Fauquet CM, 2000. Evidence of synergism between African cassava mosaic virus and the new double recombinant geminivirus infecting cassava in Cameroon. J. Gen. Virol. 81, 287–297.
- Francois E, 1937. Un grave peril: la mosaı¨que du manioc. Agron. Coloniale 26, 33–38.
- Frohlich DR, Torres-Jerez I, Bedford ID, Markham PG, Brown JK, 1999. A phylogeographical analysis of the *Bemisia tabaci* species complex based on mitochondrial DNA markers. Mol. Ecol. 8, 1683-1691.

- Gachoka KK, Obeng-Ofori D, Danquah EY, 2005. Host suitability of two Ghanaian biotypes of *Bemisia tabaci* (Homoptera: Aleyrodidae) on five common tropical weeds. Int. J. Trop. Insect Sci. *25*, 236–244.
- Galtier N, Gouy M, Gautier C, 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. CABIOS 12, 543-548.
- Gauthier N, Dalleau-Clouet C, Bouvret M-E, 2008. Twelve new polymorphic microsatellite loci and PCR multiplexing in the whitefly, *Bemisia tabaci.* Mol. Ecol. Resour. 8, 1004–1007. doi: 10.1111/j.1755-0998.2008.02163.x
- Gennadius P, 1889. Disease of the tobacco plantations in the Trikonia. The aleurodid of tobacco. Ellenike Georgia. 5, 1-3
- Gerling D and Mayer RT, 1996. *Bemisia 1995: taxonomy, biology, damage, control and management*.
- Gibson RW, Legg JP, Otim-Nape GW. 1996. Unusually severe symptoms are a characteristic of the current epidemic of mosaic virus disease of cassava in Uganda. Ann. Appl. Biol. 128, 479–490.
- Gill RJ and Brown JK, 2010. Systematics of *Bemisia* and *Bemisia* Relatives: Can Molecular Techniques Solve the *Bemisia tabaci* Complex Conundrum – A Taxonomist's Viewpoint. In: *Bemisia*: Bionomics and Management of a Global Pest. Eds. PA. Stansly and SE. Naranjo.Springer Dordrecht Heidelberg London New York.
- Global Invasive Species Database, 2014. *Bemisia tabaci* (Insect). [WWW document]. URL http://www.issg.org/database/species/search.asp?sts=sss&st=sss&fr=1&sn=Bemisia+tabaci&rn=&hci=-1&ei=153&lang=EN&x=44&y=12.

- Gnankiné O, Mouton L, Henri H, Terraz G, Houndeté T, Martin T, Vavre F, Fleury F, 2012. Distribution of *Bemisia tabaci* (Homoptera: Aleyrodidae) biotypes and their associated symbiotic bacteria on host plants in West Africa. Insect Conserv. Divers. 6, 411-421. DOI: 10.1111/j.1752-4598.2012.00206.x.
- Golding FD, 1936. Cassava mosaic in southern Nigeria. Bulletin of the Department of Agriculture, Nigeria 11, 1-10.
- Gottlieb Y, Ghanim M, Chiel E, Gerling D, Portnoy V, Steinberg S, Tzuri G, Horowitz AR, Belausov E, Mozes-Daube N, Kontsedalov S, Gershon M, Gal S, Katzir N, Zchori-Fein E, 2006. Identification and localization of a *Rickettsia* sp. in *Bemisia tabaci* (Homoptera: Aleyrodidae). Appl. Environ. Microbiol. 72**,** 3646–3652.
- Gottlieb Y, Ghanim M, Gueguen G, Kontsedalov S, Vavre F, Fleury F, Zchori-Fein E, 2008. Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies. The FASEB Journal, 22, 2591-2599.
- Gottlieb Y, Zchori-Fein E, Mozes-Daube N, Kontsedalov S, Skaljac M, Brumin M, Sobol I, Czosnek H, Vavre F, Fleury F, Ghanim M, 2010. The transmission efficiency of *tomato yellow leaf curl virus* by the whitefly *Bemisia tabaci* is correlated with the presence of a specific symbiotic bacterium species. J. Virol. 84, 9310–9317.
- Goudet J, 1995. Fstat: a computer program to calculate F-statistics. J. Hered. 86, 485-486.
- Gouy M, Guindon S, Gascuel O, 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221-224.
- Guastella D, 2014. Studies on whiteflies living on herbaceous plants and their parasitoids in tropical and sub-tropical environment through combined morphological,

molecular and geostatistical approaches. Ph.D. [dissertation]. Catania (IT): University of Catania. [WWW document]. URL http://hdl.handle.net/10761/1601.

- Guastella D, Lulah H, Tajebe LS, Cavalieri V, Evans GA, Pedata PA, Rapisarda C, Legg JP, 2014. Survey on whiteflies and their parasitoids in cassava mosaic pandemic areas of Tanzania using morphological and molecular techniques. Pest Manag Sci.DOI 10.1002/ps.3810
- Gueguen G, Vavre F, Gnankiné O, Peterschmitt M, Charif D, Chiel E, Gottlieb Y, Ghanim M, Zchori-Fein E, Fleury F, 2010. Endosymbiont metacommunities, mtDNA diversity and the evolution of the *Bemisia tabaci* (Hemiptera: Aleyrodidae) species complex. Mol. Ecol. 19, 4365–4378.
- Guindon S, Gascuel O, 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Syst. Biol. 52, 696-704.
- Gullan, PJ and Martin JH, 2003. Sternorrhyncha (jumping plant-lice, whiteflies, aphids and scale insects). In: VH Resh and RT Cardé (Eds), Encyclopedia of Insects. pp. 1079-1089. Academic Press, Amsterdam.
- Hadjistylli M, 2010. Global Population Genetics and Evolution of Invasive Biotypes in the Whitefly Complex *Bemisia tabaci*.Ph.D. [dissertation]. California, Berkeley (USA): University of California, Berkeley. [WWW document]. URL http://www.escholarship.org/uc/item/6kd1q2dv
- Hahn SK, Terry ER, Leuschner K, 1980. Breeding cassava for resistance to cassava mosaic disease. Euphytica. 29, 673-683.
- Hall FW, 1928. Annual Report.Department of Agriculture, Uganda. p 35.
- Harrison BD, Zhou X, Otim-Nape GW, Liu Y, Robinson DJ, 1997. Role of a novel type of double infection in the geminivirus-induced epidemic of severe cassava mosaic in Uganda. Ann. Appl. Biol. 131, 437–448.

- He´din L, 1931. Culture du manioc enCoˆte d'Ivoire; observations comple´mentairessur la mosaı¨que.Revue de Botanique Applique´e 11, 558–563.
- Hedges LM, Brownlie JC, O'Neill SL, Johnson KN, 2008. *Wolbachia* and virus protection in insects. Science. 322: 702.
- Henry LM, Peccoud J, Simon JC, Hadfield JD, Maiden MJC, Ferrari J, Godfray HCJ, 2013. Horizontally transmitted symbionts and host colonization of ecological niches. Curr. Biol. 23, 1713–1717.
- Hillocks RJ, 2002. Cassava in Africa. In: Cassava: biology, production and utilization. Eds. Hillocks RJ, Thresh JM, Bellotti AC. CABI.
- Himler AG, Adachi-Hagimori T, Bergen JE, Kozuch A, Kelly SE, Tabashnik BE, Chiel E, Duckworth VE, Dennehy TJ, Zchori-Fein E, Hunter MS, 2011. Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. Science. 332, 254–256.
- Hong YG, Robinson DJ, Harrison BD, 1993. Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted geminiviruses in cassava. J. Gen. Virol. 74, 2437–2443.
- Hsieh CH, Wang CH, Ko CC, 2006. Analysis of *Bemisia tabaci* (Hemiptera: Aleyrodidae) Species Complex and Distribution in Eastern Asia Based on Mitochondrial DNA Markers. Systematics. 99, 768-775.
- Hu J, De Barro P, Zhao H, Wang J, Nardi F, Liu SS, 2011. An extensive field survey combined with a phylogenetic analysis reveals rapid and widespread invasion of two alien whiteflies in China. PLoS ONE. DOI: 10.1371/journal.pone.0016061
- Huelsenbeck JP, Ronquist F, 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17, 754-755.

- Hunter MS, Perlman SJ, Kelly SE, 2003. A bacterial symbiont in the *Bacteriodetes* induces cytoplasmic incompatibility in the parasitoid wasp *Encarsia pergandiella*. Proc. R. Soc. Lond. Ser. B. Biol. Sci. 207**,** 2185-2190.
- Husain MA, Trehan KN, 1933. Observations on the life-history, bionomics and control of the whitefly of cotton (*Bemisia gossypiperda*M and L). Indian J. Agric. Sci. 3, 701–753.
- Hutter CM, Schug MD, Aquadro CF, 1998. Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: A reciprocal test of the ascertainment bias hypothesis. Mol. Biol. Evol. 15, 1620-1636.
- Jakobsson M and Rosenberg N, 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 23, 1801–1806.
- Jarvis A, Ramirez-Villegas J, Campo BVH and Navarro-Racines C, 2012. Is cassava the answer to African climate change adaptation? Trop. Plant. Biol. 5, 9 – 29.
- Jennings DL and Iglesias CA, 2002. Breeding for crop improvement. In: RJ Hillocks, JM Thresh and AC Bellotti (Eds.), Cassava: Biology, Production and Utilization. CABI Publishing, pp. 149–166.
- Jones WO, 1959. Manioc in Africa. Standford University Press. Standford, Connecticut.
- Kufferath H, Ghesquière J, 1932. La mosaïque du manioc. Compte-Rendu de la Société de Biologie Belge 109, 1146–1148.
- Lancaster PA and Brooks JE, 1983. Cassava leaves as human food. Econ. Bot.. 37, 331-348.
- Lathrap DW, 1970. The Upper Amazon. Thames and Hudson, London.
- Lee W, Park J, Lee GS, Lee S, Akimoto S-i, 2013. Taxonomic status of the *Bemisia tabaci* complex (Hemiptera: Aleyrodidae) and reassessment of the number of its constituent species. PLoS ONE. DOI:10.1371/journal.pone.0063817

- Legg JP and Thresh JM, UD. Cassava virus diseases in Africa. [WWW data] URL http://old.iita.org/cms/details/virology/pdf files/517-552.pdf [Accessed on 10 November 2014]
- Legg JP, 1994. *Bemisia tabaci*: The whitefly vector of cassava mosaic geminiviruses in Africa: An ecological perspective. Afr. Crop Sci. J. 2, 437–448.
- Legg JP, 1996. Host-associated strains within Ugandan populations of the whitefly *Bemisia tabaci* (Genn.), (Hom., Aleyrodidae). J. Appl. Entomol. 120, 523–527.
- Legg JP, 1999. Emergence, spread and strategies for controlling the pandemic of cassava mosaic virus disease in east and central Africa. Crop. Prot. 18, 627–637.
- Legg JP, 2010. Epidemiology of a whitefly-transmitted cassava mosaic geminivirus pandemic in Africa, In: *Bemisia*: Bionomics and Management of a Global Pest. Ed. by Stansly PA, Naranjo SE, Springer, Dordrecht-Heidelberg-London-New York, 233–257.
- Legg JP, Fauquet CM, 2004. Cassava mosaic geminiviruses in Africa. Plant Mol. Biol. 56, 585-599.
- Legg JP, French R, Rogan D, Okao-Okuja G, Brown JK, 2002. A distinct *Bemisia tabaci* (Gennadius) (Hemiptera: Sternorrhyncha: Aleyrodidae) genotype cluster is associated with the epidemic of severe cassava mosaic virus disease in Uganda. Mol. Ecol. 11, 1219–1229.
- Legg JP, Jeremiah SC, Obiero HM, Maruthi MN, Ndyetabula I, Okao-Okuja G, Bouwmeester H, Bigirimana S, Tata-Hangy W, Gashaka G, Mkamilo G, Alicai T, Lava Kumar P, 2011. Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. Virus Res. 159**,** 161–170.
- Legg JP, Ogwal S, 1998. Changes in the incidence of African cassava mosaic geminivirus and the abundance of its whitefly vector along south-north transects in Uganda. J. Appl. Entomol. 122**,** 169–178.

- Legg JP, Owor B, Sseruwagi P, Ndunguru J, 2006. Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. Adv. Virus Res. 67, 355–418.
- Legg JP, Sseruwagi P, Boniface S, Okao-Okuja G, Shirima R, Bigirimana S, Gashaka G, Herrmann HW, Jeremiah S, Obiero H, Ndyetabula I, Tata-Hangy W, Masembe C, Brown JK, 2013. Spatio-temporal patterns of genetic change amongst populations of cassava *Bemisia tabaci* whiteflies driving virus pandemics in East and Central Africa. Virus Res. 186, 61-75.
- Li ZX, Lin HZ, Guo XP, 2007. Prevalence of *Wolbachia* infection in *Bemisia tabaci*. Curr. Microbiol. 54, 467–471.
- Liadouze I, Febvay G, Guillaud J, Bonnot G, 1995. Effect of diet on the free amino acid pools of symbiotic and aposymbiotic pea aphids, *Acyrthosiphon pisum*. J. Insect Physiol. 41, 33–40.
- Lima LHC, Navia D, Inglis PW, de Oliveira MRV, 2000. Survey of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) biotypes in Brazil using RAPD markers. Genet. Mol. Biol. 23, 781–785.
- Liu J, Li M, Li JM, Huang CJ, Zhou XP, Xu FC, Liu SS, 2010. Viral infection of tobacco plants improves performance of *Bemisia tabaci* but more so for an invasive than for an indigenous biotype of the whitefly. J Zhejiang. Univ. Sci. B. 11, 30-40.
- Liu S, Colvin J, De Barro PJ, 2012. Species concepts as applied to the whitefly *Bemisia tabaci* systematic: How many species are there? J. Integr. Agric. 11, 176-186.
- Liu SS, De Barro PJ, Xu J, Luan JB, Zang LS, Ruan Y-M, Wan F-H, 2007. Asymmetric mating interactions drive widespread invasion and displacement in a whitefly. Science. 318, 1769–72.

- Lukasik P, van Asch M, Guo HF, Ferrari J, Godfray HCJ, 2013. Unrelated facultative endosymbionts protect aphids against a fungal pathogen. Ecol. Lett. 16**,** 214–218. DOI:10.1111/ele.12031.
- Mann C, 1997. Reseeding the Green Revolution. Science. 277, 1038-1043.
- Mansoor S, Khan SH, Bashir A, Saeed M, Zafar Y, Malik KA, Briddon RW, Stanley J, Markham PG, 1999. Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. Virology. 259, 190-199.
- Mantel NA, 1967. The detection of disease clustering and a generalised regression approach. Cancer Res. 27, 209–220.
- Martin JH, Mifsud D, Rapisarda C, 2000. The whiteflies (Hemiptera: Aleyrodidae) of Europe and Mediterranean Basin. Bull. Entomol. Res. 90, 407–448.
- Martin NA, 1999. Whitefly: Biology, identification and life cycle. In: Crop and Food Research. Broadsheet No. 91, 1-8.
- Maruthi MN, Colvin J, Seal S, Gibson G, Cooper J, 2002. Co-adaptation between cassava mosaic geminiviruses and their local vector populations. Virus Res. 86, 71–85.
- Maruthi MN, Colvin J, Thwaites RM, Banks GK, Gibson G, Seal SE, 2004. Reproductive incompatibility and cytochrome oxidase I gene sequence variability amongst host-adapted and geographically separate *Bemisia tabaci* populations (Hemiptera:Aleyrodidae). Syst. Entomol. 29, 560–568.
- Maruthi MN, Hillocks RJ, Mtunda K, Raya MD, Muhanna M, Kiozia H, Rekha AR, Colvin J, Thresh JM, 2005. Transmission of Cassava brown streak virus by *Bemisia tabaci* (Gennadius). J. Phytopathol. 153, 307–312.
- Misra CS, Lamba KS, 1929. The cotton white-fly (*Bemisia gossypiperda*, n. sp.). Bull. Agric. Res. Inst. 196, 1–7.

- Moran NA, Degnan PH, Santos SR, Dunbar HE, Ochman H, 2005. The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. Proc. Natl. Acad. Sci. USA. 102, 16919–16926. DOI:10.1073/pnas.0507029102.
- Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, Rocha BC, Hall-Mendelin S, Day A, Riegler M, Hugo LE, Johnson KN, Kay BH, McGraw EA, van den Hurk AF, Ryan PA, O'Neill1 SL, 2009. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, *Chikungunya*, and *Plasmodium*. Cell. 139, 1268–1278.
- Mound LA, Halsey SH, 1978. *Whitefly of the World: A systematic catalogue of the Aleyrodidae (Homoptera) with host plant and natural enemy data.* Chichester: British Museum and John Wiley and Sons, 340pp.
- Mugerwa H, Rey ME, 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. Ecol. Evol. 2, 2749-2762.
- Nardon P, GrenierAM, 1991. Serial endosymbiosis theory and weevil evolution: the role of symbiosis. In: *Symbiosis as a Source of Evolutionary Innovation*, Ed. L Margulis, R Fester, pp. 153–169. Cambridge: Massachusetts Institute of Technology.
- Ndunguru J, 2005. Molecular characterization and dynamics of cassava mosaic geminiviruses in Tanzania. Ph.D. [dissertation]. Pretoria (SA): University of Pretoria.
- Ndunguru J, Legg JP, Aveling TAS, Thompson G, Fauquet CM, 2005. Molecular biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. Virol. J. 2, 21.

- Nirgianaki A, Banks GK, Frohlich DR, Veneti Z, Braig HR, Miller TA, Bedford ID, Markham PG, Savakis C, Bourtzis K, 2003. *Wolbachia* infections of the whitefly *Bemisia tabaci*. Curr. Microbiol. 47, 93–101.
- Nweke FI, Lynam JK, Spencer DSC, 2002. The Cassava Transformation: Africa's best-kept secret. Michigan State University Press, Michigan.
- O'Neill SL, Giordano R, Colbert AME, Karr TL, Robertson HM, 1992. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc. Natl. Acad. Sci. USA. 89, 2699-2702.
- O'Neill SL, Hoffmann AA, Werren JH, 1997. *Influential Passengers: Inherited Microorganisms and Arthropod Reproduction.* Oxford: Oxford University Press.
- OEPP/EPPO, 2004. Diagnostic protocols for regulated pests. Bulletin OEPP/EPPO. 34, 281–288. [WWW data] URL: https://www.eppo.int/QUARANTINE/insects/Bemisia_tabaci/pm7-35(1)%20BEMITA%20web.pdf [Accessed on 05 November 2014]
- Oliveira MRV, Henneberry TJ, Anderson P, 2001. History, current status and collaborative research projects for *Bemisia tabaci*. Crop Prot. 20, 709–723.
- Oliver KM, Russell J, Moran NA, Hunter M, 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proc. Natl. Acad. Sci. USA. 100, 1803–1807.
- Olsen KM and Schaal BA, 1999. Evolution Evidence on the origin of cassava: Phylogeography of *Manihot esculenta* Proc. Natl. Acad. Sci. USA. 96, 5586-5591.
- Otim-Nape GW, 1988. Cassava situation in Luwero district. A Mission Report by the Uganda Root Crops Program. Namulonge Research Station, Kampala, 8 pp.
- Otim-Nape GW, Bua A, Thresh JM, Baguma Y, Ogwal S, Semakula GN, Acola G, Byabakama B, Martin A, 1997. Cassava mosaic viruses in Uganda: The current pandemic and approaches to control. Natural Resources Institute, Chatham, UK.

- Otim-Nape GW, Bua A, Thresh JM, Baguma Y, Ogwal S, Ssemakula GN, Acola G, Byabakama B, Colvin J, Cooter RJ, Martin A, 2000. The Current Pandemic of Cassava Mosaic Virus Disease in East Africa and its Control. Natural Resources Institute, Chatham, UK.
- Otim-Nape GW, Thresh JM, 1998. The current pandemic of cassava mosaic virus disease in Uganda. In: The epidemiology of Plant Diseases. Ed. by Jones G, Kluwer, Dordrencht, Germany, 423-443.
- Parrella G, Nappo AG, Manco E, Greco B, Giorgini M, 2013, Invasion of the Q2 mitochondrial variant of Mediterranean *Bemisia tabaci* in southern Italy: possible role of bacterial endosymbionts. Pest Manage. Sci. DOI: 10.1002/ps.3686.
- Parrella G, Scassillo L, Giorgini M, 2012. Evidence for a new genetic variant in the *Bemisia tabaci* species complex and the prevalence of the biotype Q in southern Italy. J. Pest Sci. 85, 227–238.
- Patti I and Rapisarda C, 1981. Reperti morfo-biologici sugli Aleirodidi nocivi alle piante coltivate in Italia. *Bolletino di Zoologia Agraria e di Bachicoltura,* ser. II, 16, 135-190.
- Perring TM, 2001. The *Bemisia tabaci* species complex. Crop Prot. 20, 725-737.
- Perring TM, Cooper AD, Rodriguez RJ, Farrar CA, Bellows TS, 1993. Identification of a whitefly species by genomic and behavioral studies. Science. 259, 74-77.
- Pita JS, Fondong VN, Sangare A, Otim-Nape GW, Ogwal S, Fauquet CM, 2001. Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. J. Gen. Virol. 82, 655-665.
- Posada D, 2008. jModelTest: Phylogenetic Model Averaging. Mol. Biol. Evol. 25, 1253-1256.

- Price M, Dehal P, Arkin A, 2010. FastTree 2- Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 5, e9490.
- Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.
- Purseglove JW, 1968. *Tropical Crops: Dicotyledons*. Longman, London.
- Quaintance AL, Baker AC, 1914. Classification of the Aleyrodidae Part II. US Department of Agriculture. Technica Series. Bur. Entomol. 27, 95–109.
- Rambaut A, 2009. FigTree, ver. 1.4.0. [Online]. Available at: http://tree.bio.ed.ac.uk/software/figtree/ [Accessed June 05 2014].
- Rapisarda C, 1990. La *Bemisia tabaci* vettore del TYLCV [tomato yellow leaf curl virus in Sicilia]. Informatore fitopatologico, 40.
- Ronda MA, Adan A, Cifuentes D, Cenis JL, Beitia F. 1999. *Laboratory evidence of interbreeding between biotypes of* Bemisia tabaci *(Homoptera, Aleyrodidae) present in Spain*. Presented at 11th Int. Plant Virus Epidemiol. Symp. Plant Virus Epidemiol. Curr. Status Future Prospects, Aguadulce, Spain
- Roopa HK, Krishna Kumar NK, Asokan R, Rebijith KB, Mahmood R, Verghese H, 2012. Phylogenetic analysis of *Trialeurodes Spp*. (Hemiptera: Aleyrodidae) from India based on differences in mitochondrial and nuclear DNA. Fla. Entomol. 95, 1086-1094. doi: http://dx.doi.org/10.1653/024.095.0438
- Rosell RC, Blackmer JL, Czosnek H, Inbar M, 2010. Mutualistic and Dependent Relationships with Other Organisms In: *Bemisia*: Bionomics and Management of a Global Pest. Eds. PA. Stansly and SE. Naranjo.Springer Dordrecht Heidelberg London New York.
- Rosenberg N, 2004. Distruct: a program for the graphical display of population structure. Mol. Ecol. Notes 4, 137–138.

- Rousset F, 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resou. 8: 103-106.
- Russel LM, 1957. Synonyms of *Bemisia tabaci* (Gennadius) (Homoptera, Aleyrodidae). Bull. Brooklyn Entomol. Soc. 52, 122-123.
- Russell JA, Moran NA, 2006. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. Proc. Roy. Soc. Lond. B Biol. Sci. 273, 603–610.
- Sangare A, Deng D, Fauquet CM, Beachy RN, 1999. Resistance to African cassava mosaic virus conferred by a mutant of the putative NTP-binding domain of the Rep gene (AC1) in *Nicotiana benthamiana.* Mol. Biol. Rep. 5, 95–102.
- Simón B, Cenis JL, De La Rúa P, 2007. Distribution patterns of the Q and B biotypes of *Bemisia tabaci* in the Mediterranean Basin based on microsatellite variation. Entomol. Exp. Appl. 124, 327–336.
- 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. Ann. Entomol. Soc. Am. 87, 651-701.
- Sloan DB, Moran NA, 2012. Endosymbiotic bacteria as a source of carotenoids in whiteflies. Biol. Lett. 8, 986 –989. DOI:10.1098/rsbl.2012.0664.
- Sseruwagi P, Maruthi MN, Colvin J, Rey MEC, Brown JK, Legg JP, 2006. Colonisation of non-cassava plant species by cassava whiteflies (*Bemisia tabaci*) (Gennadius) (Hemiptera:Aleyrodidae) in Uganda. Entomol. Exp. Applic. 119, 145–153.
- Sseruwagi P, Sserubombwe WS, Legg JP, Ndunguru J, Thresh JM, 2004. Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: a review. Virus Res. 100, 129-142.

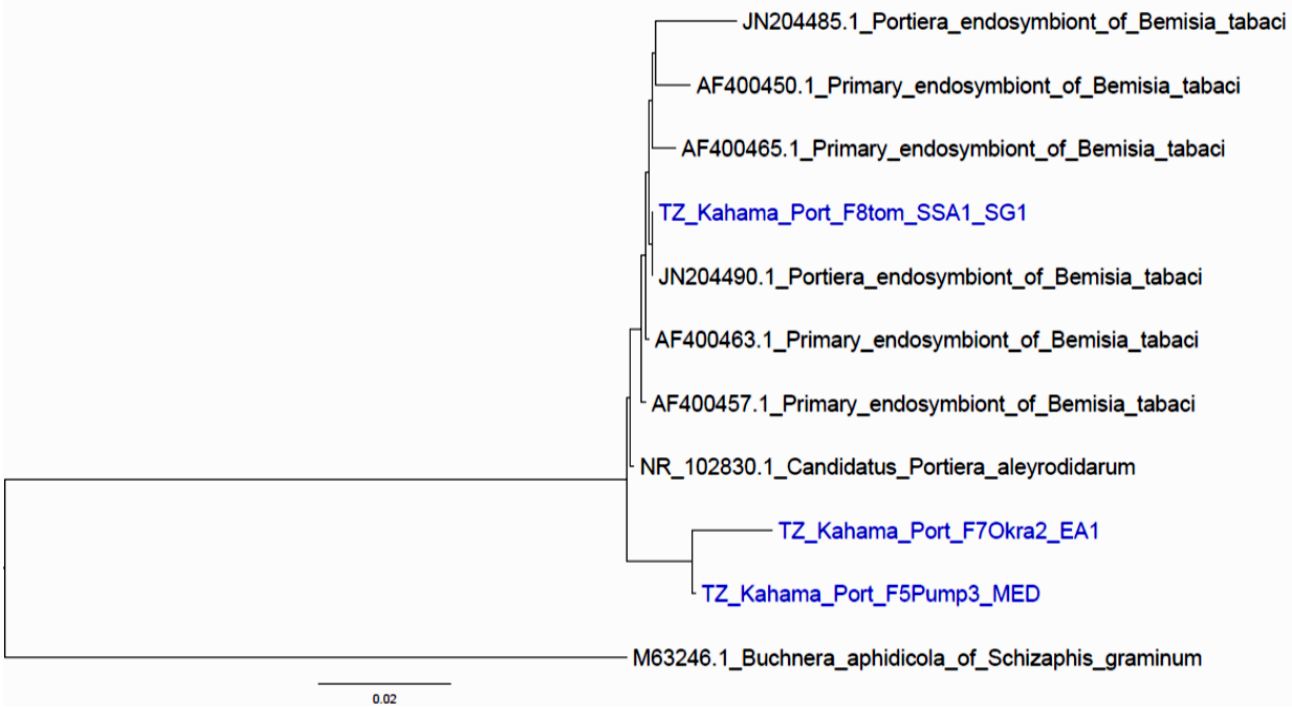
- Stansly PA and Naranjo SE, 2010. *Bemisia*: Bionomics and Management of a Global Pest. Springer Dordrecht Heidelberg London New York.
- Storey HH and Nichols RFW, 1938. Studies on the mosaic of cassava. Ann. Appl. Biol. 25, 790–806.
- Storey HH, 1936. Virus diseases of East African plants. VI-A progress report on studies of the disease of cassava. East African Agricultural Journal. 2, 34–39.
- Storey HH, 1938. Virus diseases of East African plants. VII-A field experiment in the transmission of cassava mosaic. East African Agricultural Journal 3, 446–449.
- Sun DB, Xu J, Luan JB, Liu SS, 2011. Reproductive incompatibility between the B and Q biotypes of the whitefly *Bemisia tabaci* in China: genetic and behavioural evidence. Bull. Entomol. Res. 101, 211-220. DOI:10.1017/S0007485310000416.
- Takahashi R, 1936. Some Aleyrodidae, Aphididae, Coccidae (Homoptera), and Thysanoptera from Micronesia. Tenthredo 1, 109–120.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725-2729.
- Tautz D, 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 17, 6463–6471.
- Thao ML, Baumann L, Hess JM, Falk BW, Ng JC, Gullan K, Baumann PJ, 2003. Phylogenetic evidence for two new insect-associated chlamydia of the family *Simkaniaceae*. Curr.Microbiol. 47, 46–50.
- Thao ML, Baumann P, 2004. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. Appl. Environ. Microbiol. 70, 3401–3406.
- Thao ML, Baumann P, 2004a. Evidence for multiple acquisition of *Arsenophonus* by whitefly species (Sternorrhyncha: Aleyrodidae). Curr. Microbiol. 48, 140–144.

- Thierry M, Becker N, Hajri A, Reynaud B, Lett J-M, Delatte H, 2011. Symbiont diversity and non-random hybridization among indigenous (Ms) and invasive (B) biotypes of *Bemisia tabaci*. Mol. Ecol. 20, 2172–2187.
- Thresh JM and Cooter RJ, 2005. Strategies for controlling cassava mosaic virus disease in Africa. Plant Pathol. 54, 587–614.
- Thresh JM and Otim-Nape GW, 1994. Strategies for controlling African cassava mosaic geminivirus. Adv. Dis. Vector. Res. 10, 215–236.
- Thresh JM, Otim-Nape GW, Legg JP, Fargette D, 1997. African cassava mosaic disease: the magnitude of the problem. Afr. J. Root Tuber Crops. 2, 13–19.
- Tsagkarakou A and Roditakis N, 2003. Isolation and characterization of microsatellite loci in *Bemisia tabaci* (Hemiptera, Aleyrodidae). Mol. Ecol. Notes 3, 196–198.
- Tsagkarakou A, Mouton L, Kristoffersen JB, Dokianakis E, Grispou M, Bourtzis K, 2012. Population genetic structure and secondary endosymbionts of Q *Bemisia tabaci* (Hemiptera: Aleyrodidae) from Greece. Bull. Entomol. Res. **102**, 353–365.
- Tsagkarakou A, Tsigenopoulous CS, Gorman K, Lagnel J, Bedford ID, 2007. Biotype status and genetic polymorphism of the whitefly *Bemisia tabaci* (Hemiptera: Aleyrodidae) in Greece: mitochondrial DNA and microsatellites. Bull. Entomol. Res. 97, 29–40.
- Tsuchida T, Koga R, Fukatsu T, 2004. Host plant specialization governed by facultative symbiont. Science. 303, 1989.
- Tsuchida T, Koga R, Matsumoto S, Fukatsu T, 2011. Interspecific symbiont transfection confers a novel ecological trait to the recipient insect. Biol. Lett. 7, 245–248.
- van Lenteren JC and Noldus LPJ, 1990. Whitefly-plant relationships: behavioral and ecological aspects. In: Whiteflies: Their Bionomics, Pest Status and Management, Ed. by D Gerling, pp. 47–90. Intercept, Andover, UK.

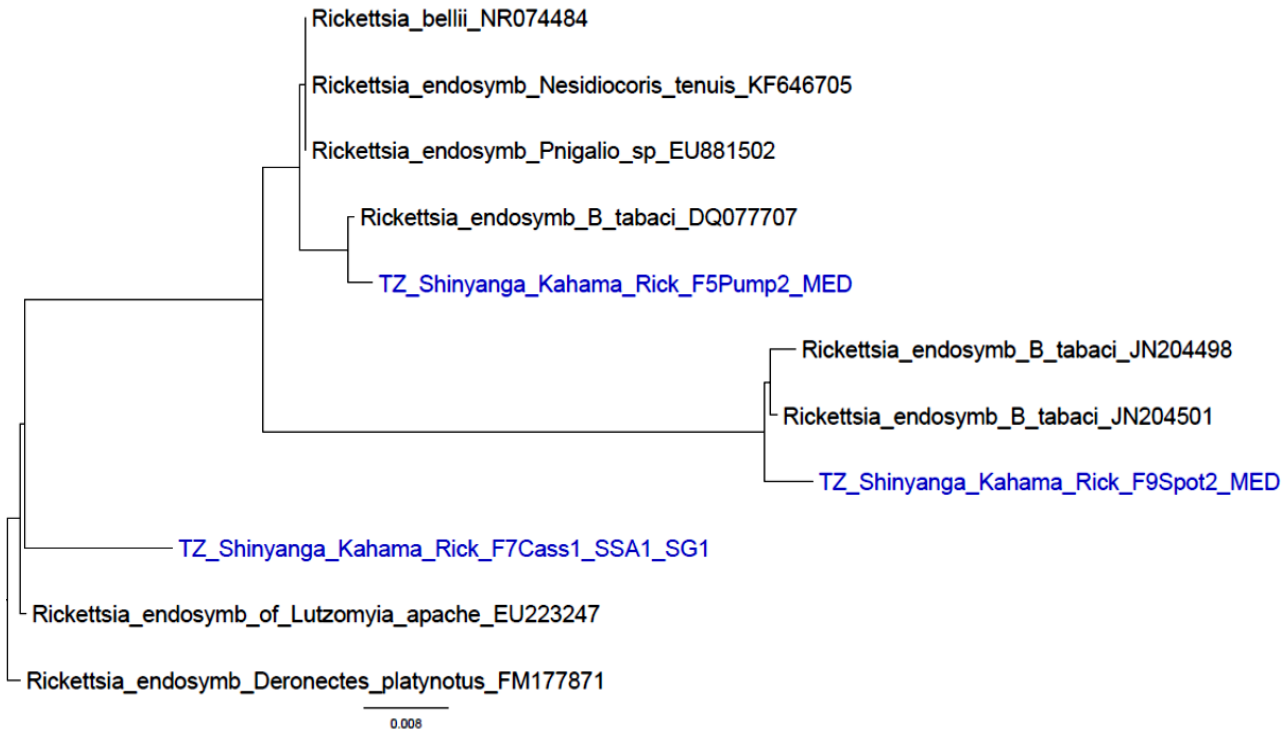
- Vauterin L, Hoste B, Kersters K, Swings J, 1995. Reclassification of *Xanthomonas.* Int. J. Syst. Bacteriol. 45, 472-489.
- Wang P, Sun D, Qiu B, Liu SS, 2011. The presence of six cryptic species of the whitefly *Bemisia tabaci* complex in China as revealed by crossing experiments. Insect Sci. 18, 67-77.
- Warburg O, 1894. Die kulturpflanzenusambaras. Mitteilungenaus den Deutschen Schutzgebieten. 7, 131.
- Weeks AR and Breeuwer JAJ, 2003. A new bacterium from the Cytophaga-Flavobacterium-Bacteroides phylum that causes sex ratio distortion. In *Insect Symbiosis II,* pp. 165-176. Eds K. Bourtzis and T. Miller. Boca Raton, Florida, USA: CRC Press.
- Weeks AR, Velten R, Stouthamer R, 2003. Incidence of a new sex-ratio-distorting endosymbiotic bacterium among arthropods. Proc. R Soc. Lond. **B** Biol. Sci. 270, 1857–1865.
- Weir BS and Cockerham CC, 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370.
- Werren JH, Baldo L, Clark ME, 2008. *Wolbachia*: master manipulators of invertebrate biology. Nat. Rev. Microbio. 6, 741–751. DOI:10.1038/nrmicro1969.
- White JA, Kelly SE, Perlman SJ, Hunter MS, 2009. Cytoplasmic incompatibility in the parasitic wasp *Encarsia inaron*: disentangling the roles of Cardinium and Wolbachia symbionts. Heredity. 102, 483-489.
- Xu J, De Barro PJ, Liu SS, 2010. Reproductive incompatibility among genetic groups of *Bemisia tabaci* supports the proposition that the whitefly is a cryptic species complex. Bull. Entomol. Res. 100, 359-366.

- Xu J, Lin KK, Liu SS, 2011. Performance on different host plants of an alien and an indigenous *Bemisia tabaci* from China. J. Appl. Entomol. 135, 771-779.
- Zchori-Fein E and Brown JK, 2002. Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae). Ann. Entomol. Soc. Am. 95**,** 711–718.
- Zchori-Fein E, Gottlieb Y, Kelly SE, Brown JK, Wilson JM, Karr TL, Hunter MS, 2001. A newly discovered bacterium associated with parthenogenesis and a change in host selection behavior in parasitoid wasps. Proc Natl Acad Sci USA. 98, 12555–12560.
- Zchori-Fein EY, Perlman SJ, Kelly SE, Katzir N, Hunter MS, 2004. Characterization of "Bacteriodetes" symbionts in *Encarsia* wasps (Hymenoptera: Aphelinidae): proposal of "*Candidatus* Cardinium hertigii". Int. J. Syst. Evol. Microbiol. 54, 961–968.
- Zhou X, Liu Y, Calvert L, Muñoz C, Otim-Nape GW, Robinson DJ, Harrison BD, 1997. Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. J. Gen. Virol. 78, 2101–2111.
- Zhou X, Robinson DJ, Harrison BD, 1998. Types of variation in DNA-A among isolates of East African cassava mosaic virus from Kenya, Malawi, Tanzania. J. Gen. Virol. 79, 2835–2840.
- Zimmermann A, 1906. Die Krauselkrankheit des Maniok. Pflanzer 2, 145.

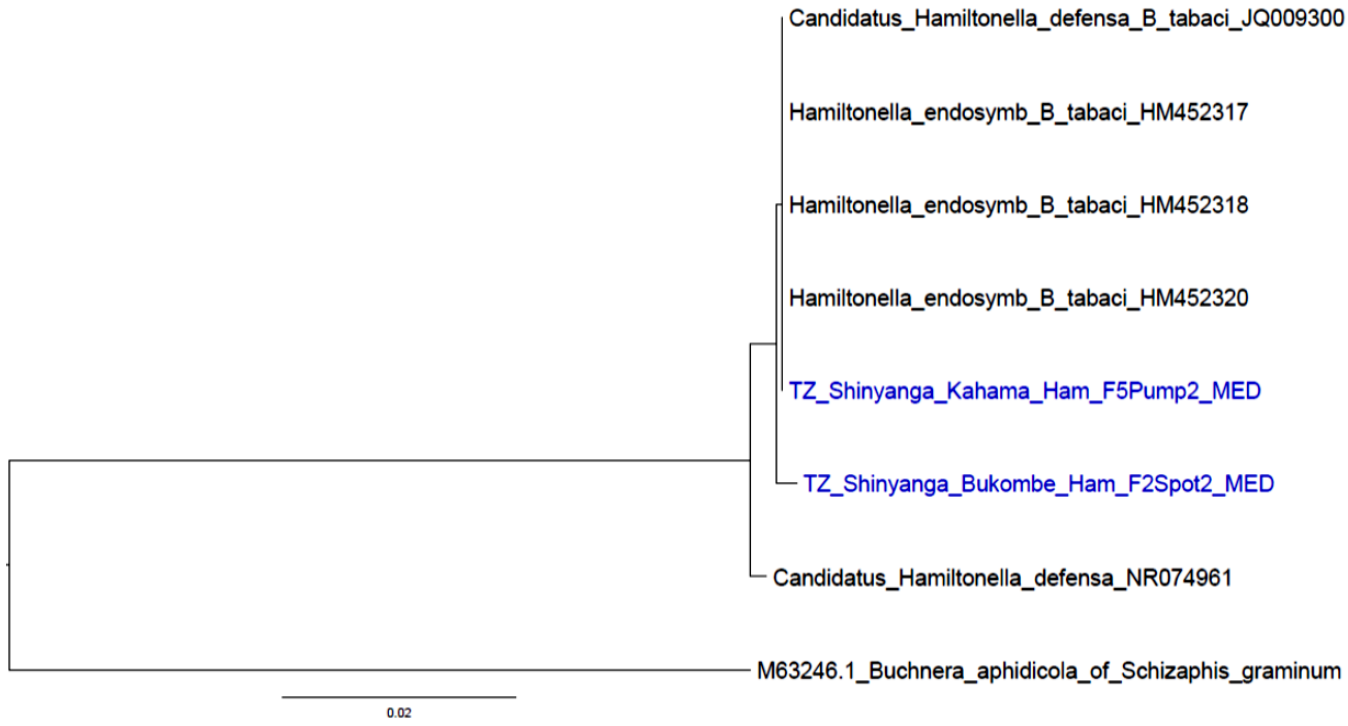
Appendix



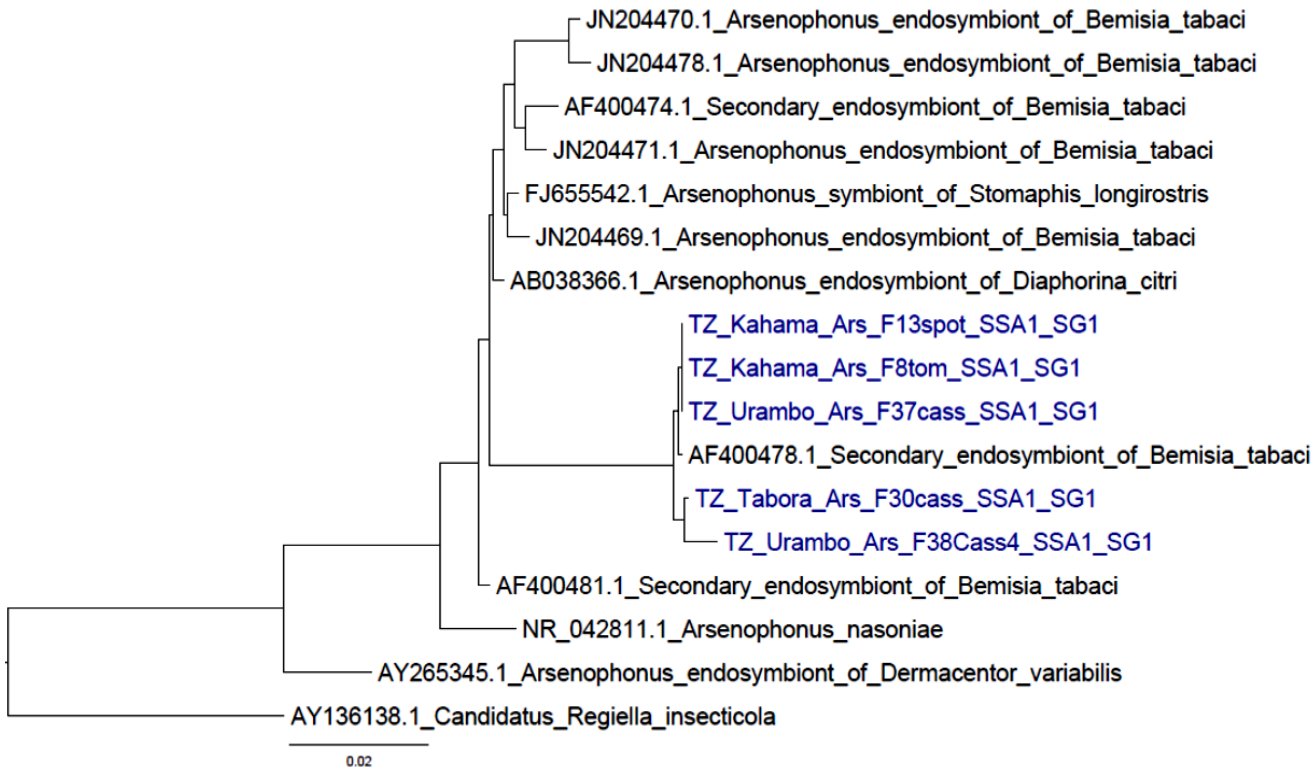
Appendix 1a. Maximum likelihood phylogeny of *Portiera aleyrodidarium* partial *16S* rDNA sequences and selected reference sequences from Genebank. Sequences starting with 'TZ_' are sequences produced in this study whereas all other sequences are reference sequences from GenBank. The scale bar represents the number of substitutions per site. SSA1-SG1 represents Sub-Saharan Africa1 sub-group 1; EA1, East Africa 1; and MED, Mediterranean genetic group.



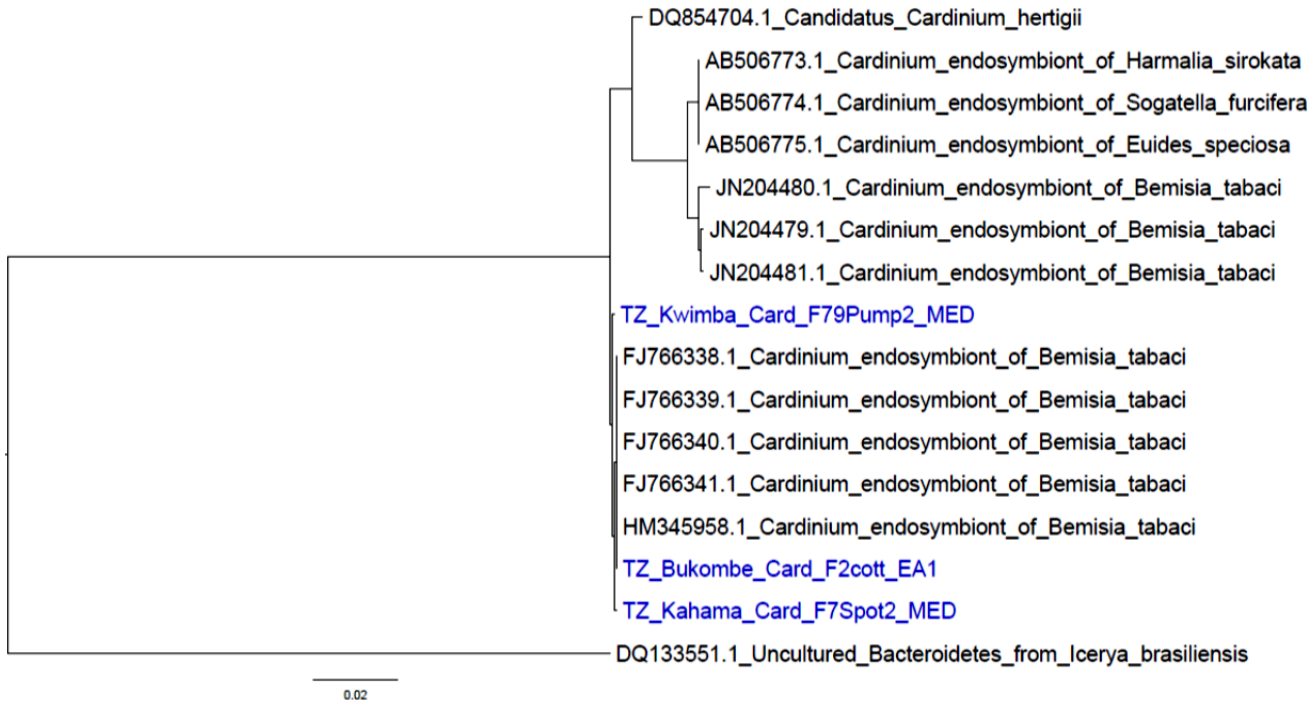
Appendix 1b. Maximum likelihood phylogeny of *Rickettsia* partial *16S* rDNA sequences and selected reference sequences from Genebank. Sequences starting with 'TZ_' are sequences produced in this study whereas all other sequences are reference sequences from GenBank. The scale bar represents the number of substitutions per site. SSA1-SG1 represents Sub-Saharan Africa1 sub-group 1 whereas MED represents Mediterranean genetic group.



Appendix 1c. Maximum likelihood phylogeny of *Hamiltonella* partial *16S* rDNA sequences and selected reference sequences from Genebank. Sequences starting with 'TZ_' are sequences produced in this study whereas all other sequences are reference sequences from GenBank. The scale bar represents the number of substitutions per site and MED represents Mediterranean genetic group.



Appendix 1d. Maximum likelihood phylogeny of *Arsenophonus* partial *16S* rDNA sequences and selected reference sequences from Genebank. Sequences starting with 'TZ_' are sequences produced in this study whereas all other sequences are reference sequences from GenBank. The scale bar represents the number of substitutions per site and SSA1-SG1 represents Sub-Saharan Africa1 sub-group 1 individuals.



Appendix 1e. Maximum likelihood phylogeny of *Cardinium* partial *16S* rDNA sequences and selected reference sequences from Genebank. Sequences starting with 'TZ_' are sequences produced in this study whereas all other sequences are reference sequences from GenBank. The scale bar represents the number of substitutions per site. MED represents Mediterranean genetic group whereas EA1 is East Africa 1 genetic group.

Appendix 1f. Maximum likelihood phylogeny of *Wolbachia* partial *16S* rDNA sequences and selected reference sequences from Genebank. Sequences starting with 'TZ_' are sequences produced in this study whereas all other sequences are reference sequences from GenBank. The scale bar represents the number of substitutions per site and MED represents Mediterranean genetic group.

Appendix 2. Percentage of whitefly single, double and multiple infections by secondary endosymbionts. SSA1 represents Sub-Saharan Africa 1; SG, Sub-group; EA1, East Africa 1; R= *Rickettsia*; H= *Hamiltonella*; A= *Arsenophonus*; C= *Cardinium*; W= *Wolbachia;* F= *Fritschea*; and '-'= 0%

	Endosymbiotic bacteria (%)																												
***B. tabaci* putative species**	R	H	A	C	W	F	None	AC	HA	HW	RC	RH	RA	CW	RW	HAW	HCW	RAC W	ACW	RCW	RHA	RHAC	RHAW	RHAC W	RHC	RHCW	RH W	HAC	HACW
SSA1-SG1 (n=125)	-	4	20.8	0.8	2.4	-	36	21.6	6.4	-	-	-	-	-	-	0.8	-	0.8	-	-	0.8	-	-	-	-	-	-	4.8	0.8
SSA1-SG2 (n=6)	-	-	16.7	-	-	-	-	-	16.7	-	-	-	-	-	-	-	-	-	16.7	-	-	-	-	-	-	-	-	33.3	16.7
SSA1-SG1/2 (n=8)	-	-	-	-	-	-	-	75	12.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12.5	-
SSA1-SG3 (n=8)	-	-	-	-	-	-	-	-	-	-	-	-	12.5	-	-	-	-	-	-	-	25	37.5	-	12.5	-	-	-	12.5	-
Mediterranean (n=37)	-	13.5	-	-	2.7	-	2.7	-	-	5.4	2.7	8.1	-	-	-	5.4	-	-	-	10.8	2.7	-	5.4	-	2.7	13.5	24.3	-	-
EA1 (n=21)	-	-	-	-	-	-	4.8	-	-	-	-	-	-	14.3	4.8	-	14.3	19	-	14.3	-	4.8	-	9.5	-	4.8	4.8	-	4.8
Indian Ocean (n=5)	-	-	-	-	-	-	-	60	-	-	-	20	-	-	-	-	-	-	-	-	-	-	-	-	20	-	-	-	-

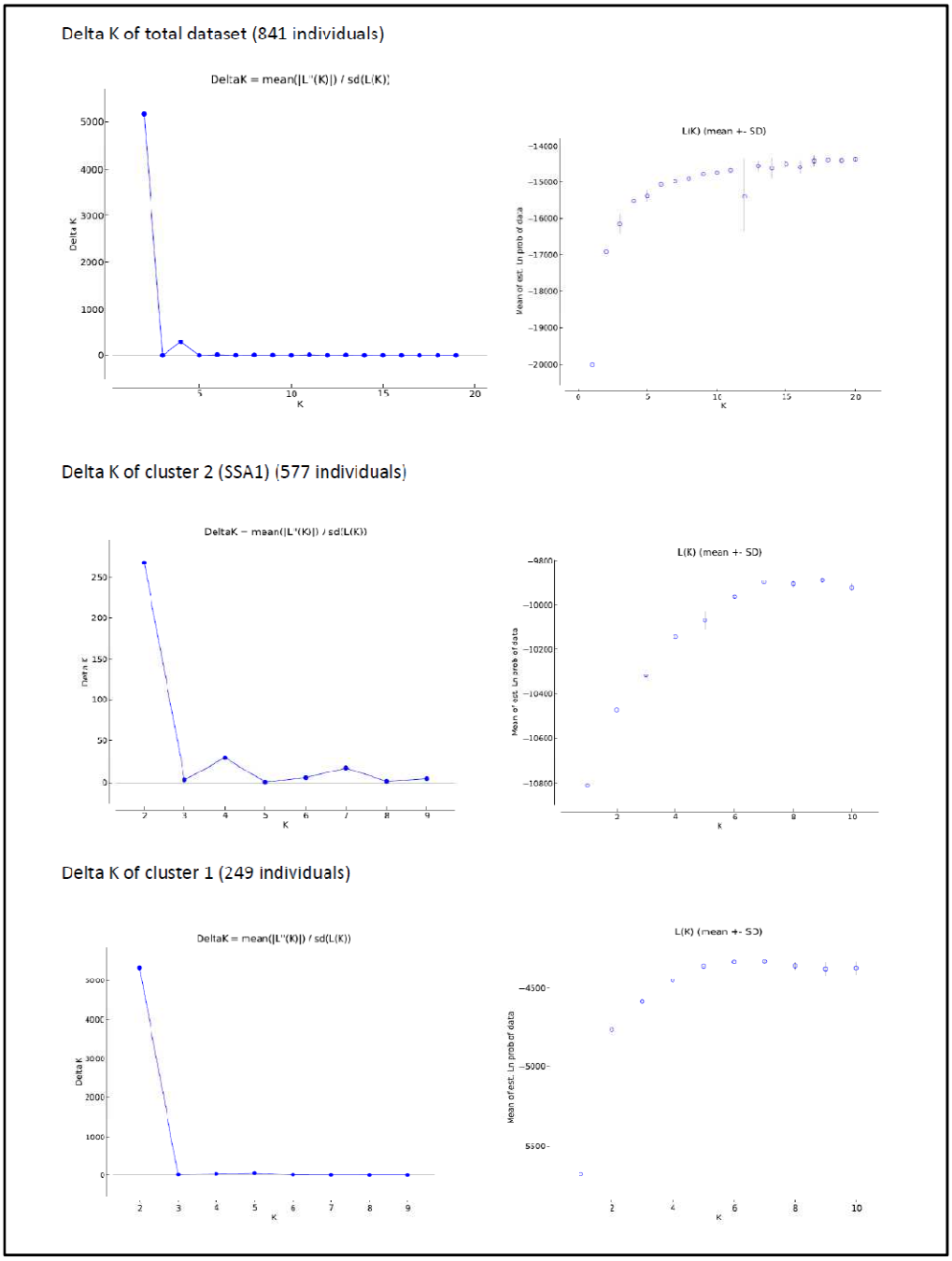
Appendix 3. Observed and expected (in brackets) secondary symbiont frequencies of whiteflies collected on cassava. SSA1-SG1 represents Sub-Saharan Africa 1 Sub-group 1 whereas Others represents all individuals other than SSA-SG1

Genetic groups/ sub-groups	secondary symbionts absent	secondary symbionts present	Total
SSA1-SG1	45 (37.4)	78 (85.6)	**123**
Others	0 (7.6)	25 (17.4)	**25**
Total	**45**	**103**	**148**

Appendix 4. Observed and expected (in brackets) secondary symbiont frequencies of whiteflies collected in the CMD pandemic area in the Lake Zone. SSA1-SG1 represents Sub-Saharan Africa 1 Sub-group 1 whereas Others represents all individuals other than SSA-SG1

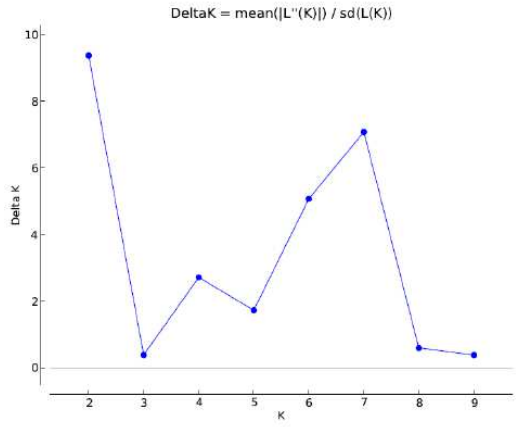
Genetic groups/ sub-groups	secondary symbionts absent	secondary symbionts present	Total
SSA1-SG1	45 (29.7)	75 (90.3)	**120**
Others	2 (17.3)	68 (52.7)	**70**
Total	**47**	**143**	**190**

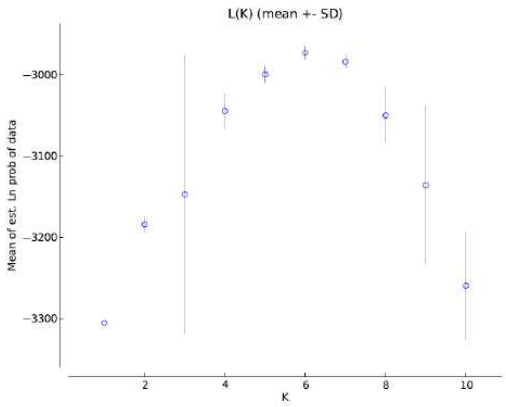
Appendix 5. Delta K values for the population groupings of the clusters identified in the study. SSA1 stands for sub-Saharan Africa 1, IO for Indian Ocean, and MED for Mediterranean genetic group.



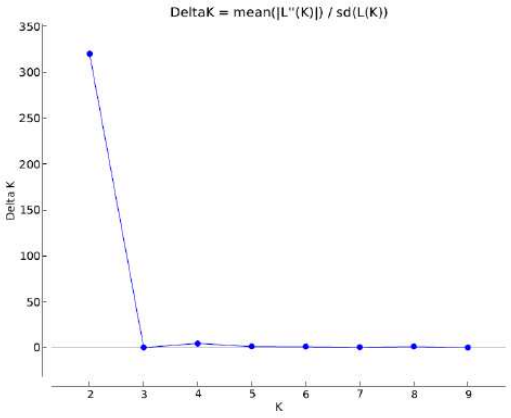
Appendix 5. *Continued.*

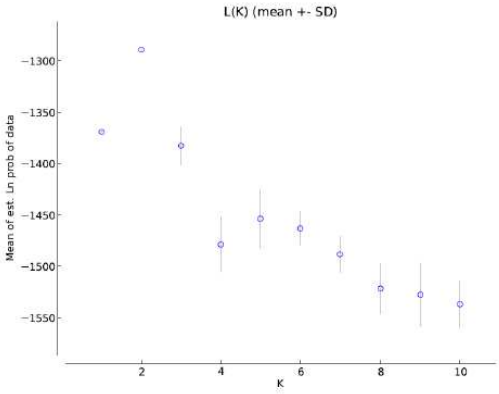
Delta K of cluster 1.2 *B. afer* (197 individuals)



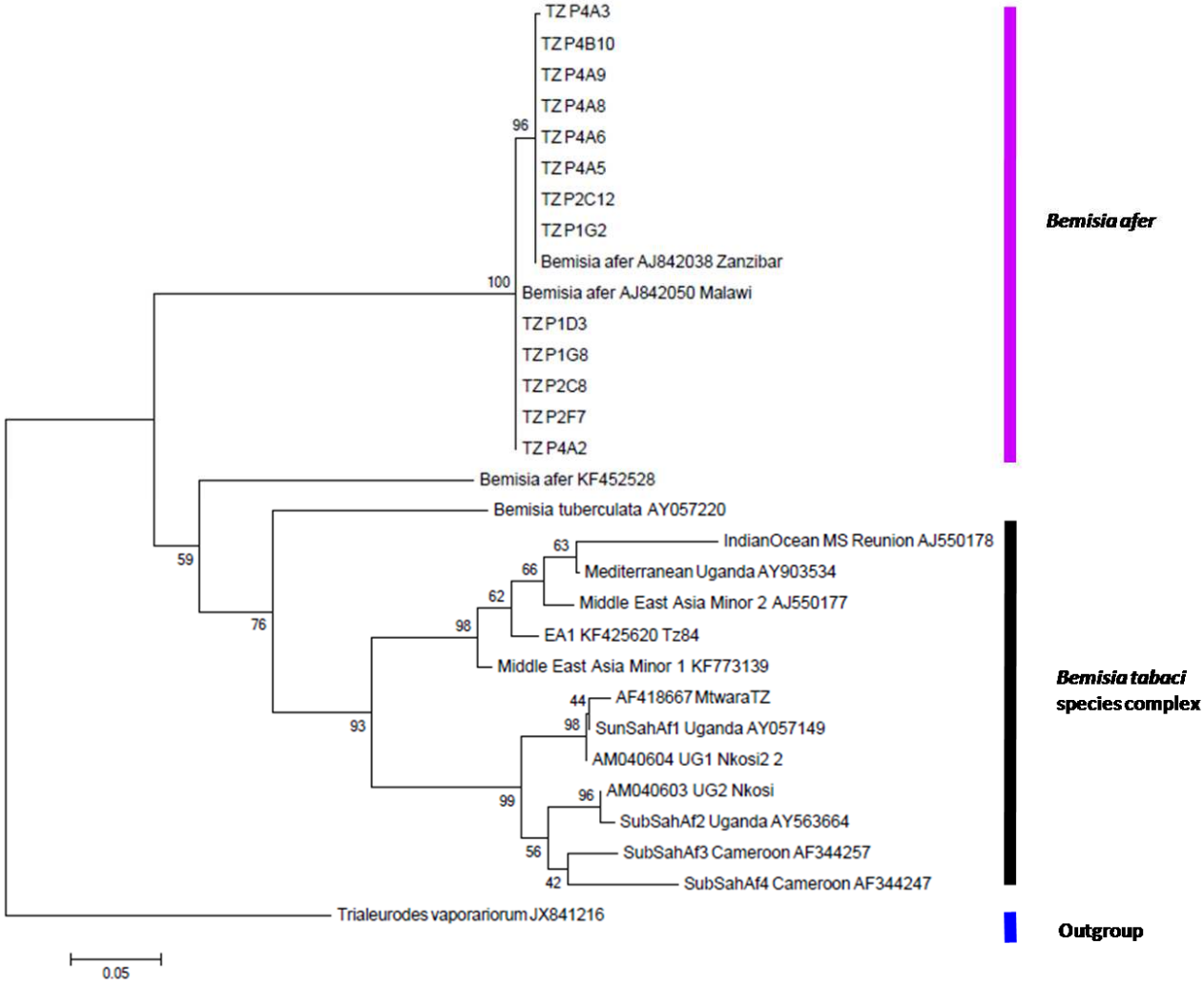


Delta K of cluster 1.1 (IO-like and MED) (51 individuals)





Appendix 6. Maximum likelihood phylogeny of mtCOI sequences of *Bemisia afer* adults collected during 2012 – 2013 surveys with selected reference sequences from GenBank. All sequences starting with 'TZ' are sequences produced in this study. The tree is drawn to scale and the scale bar represents the number of substitutions per site.



Appendix 7. Genetic variation of the two sub-clusters (as delimited by K=2 STRUCTURE) of the *Bemisia tabaci* SSA1 putative species; SSA1 stands for sub-Saharan Africa 1, SE for standard error, F_{is} for inbreeding coefficient, Hexp for expected heterozygosity, Hnb for non-biased Heterozygosity, and Hobs for observed heterozygosity

Sub-clusters within cluster 2 (SSA1)	Number of individuals per site (n)	Zone	Allelic Richness ± SE	Average Number of allele over all loci	Null allele frequency	Fis ± SE	Hexp.	Hn.b.	Hobs.
Cluster 2.1									
F18ca12	12	Central	2.175±0.284	4.1	0.050	0.13±0.097**	0.480	0.502	0.422
F19ca12	11	Central	2.172±0.299	4.0	0.056	0.196±0.112***	0.474	0.498	0.378
F1ca13	6	Central	1.886±0.221	2.4	0.038	-0.02±0.183	0.385	0.420	0.417
F1capu12	5	Central	1.999±0.251	2.9	0.028	0.058±0.119	0.405	0.453	0.413
F20ca12	6	Central	2.188±0.309	3.1	0.032	0.167±0.097*	0.460	0.509	0.413
F24ca13	5	Central	2.085±0.260	2.9	0.027	0.038±0.148	0.430	0.486	0.454
F25ca13	20	Central	2.183±0.307	4.7	0.066	0.078±0.157***	0.494	0.509	0.434
F26ca13	21	Central	2.159±0.294	4.9	0.045	0.113±0.097***	0.483	0.497	0.424
F27ca13	5	Central	2.158±0.286	2.9	0.018	-0.03±0.121	0.463	0.517	0.519
F28ca12	8	Central	2.167±0.367	3.7	0.038	0.105±0.129	0.439	0.476	0.422
F28ca13	20	Central	2.060±0.256	4.6	0.047	0.112±0.120**	0.459	0.474	0.399
F29ca12	7	Central	2.074±0.226	3.1	0.024	-0.00±0.092	0.442	0.481	0.469
F2ca13	7	Central	1.929±0.266	3.6	0.061	0.069±0.189**	0.381	0.411	0.357
F2casp12	6	Central	2.149±0.263	2.9	0.000	0.007±0.059	0.458	0.509	0.500
F30ca13	7	Central	1.963±0.309	2.6	0.035	0.148±0.133*	0.394	0.425	0.348
F31ca13	17	Central	2.091±0.277	4.1	0.067	0.168±0.199***	0.465	0.480	0.404
F35ca13	8	Central	2.129±0.322	3.4	0.032	0.071±0.104	0.441	0.475	0.419
F37ca12	13	Central	2.189±0.317	4.3	0.044	0.135±0.078**	0.474	0.497	0.416
F39ca12	23	Central	2.184±0.287	5.7	0.041	0.051±0.098***	0.500	0.512	0.457
F39ca13	17	Central	2.067±0.252	4.6	0.004	-0.15±0.098	0.466	0.480	0.550
F41ca13	12	Lake	2.113±0.248	4.0	0.050	0.154±0.194	0.478	0.500	0.456
F42ca13	6	Lake	2.123±0.293	3.1	0.074	0.261±0.153**	0.453	0.495	0.358
F51ca13	25	Lake	2.158±0.297	5.4	0.022	-0.00±0.087*	0.484	0.494	0.478
Average							**0.453**		**0.431**
Cluster 2.2									
F13ca13	28	Central	1.405±0.116	4.7	0.049	0.132±0.177*	0.442	0.450	0.430
F18ca12	7	Central	1.380±0.124	3.4	0.029	0.057±0.108*	0.392	0.428	0.375
F1ca13	12	Central	1.399±0.105	3.7	0.048	0.039±0.121*	0.437	0.456	0.396
F25ca13	5	Central	1.431±0.128	3.4	0.017	-0.19±0.140	0.440	0.489	0.550
F28ca13	7	Central	1.369±0.113	3.0	0.000	0.161±0.076	0.343	0.369	0.271
F2ca13	10	Central	1.377±0.116	3.6	0.037	0.070±0.177	0.408	0.429	0.438
F31ca13	7	Central	1.501±0.102	3.0	0.038	0.159±0.122	0.405	0.439	0.387
F36ca13	7	Central	1.413±0.152	2.9	0.026	-0.04±0.163	0.422	0.457	0.470
F37ca12	6	Central	1.394±0.121	2.9	0.079	0.299±0.179**	0.382	0.421	0.281
F4ca13Rf	11	Central	1.442±0.080	4.4	0.069	0.139±0.158**	0.449	0.473	0.398

F5ca13	9	Central	1.414±0.095	3.4	0.085	0.163±0.185***	0.443	0.469	0.361
F1ca13Mk	28	Eastern	1.386±0.100	6.6	0.048	0.116±0.130**	0.430	0.438	0.387
Kibca13	26	Eastern	1.403±0.141	4.4	0.028	0.066±0.061*	0.442	0.452	0.409
F51ca13	6	Lake	1.487±0.129	3.3	0.094	0.204±0.220***	0.458	0.501	0.379
Average							**0.421**		**0.395**

* P<0.05; ** P<0.01; *** P<0.001. P values without '*' are non-significant; Allelic Richness for Cluster 2.2 was calculated based on seven loci as locus WF2H06in one population had no genotyped individual.

Appendix 8. The relationship between geographic distance (m) and genetic distance (Fst) in the *Bemisia tabaci* Sub-Saharan Africa 1 putative species (Pearson r=0.645; P=0.0002, 1000 permutations).