OPINION

Review and guide to a future naming system of African Bemisia tabaci species

Laura M. Boykin, Tonny Kine NE, James M. Wainaina, Anders Savill, Susan Seal, Habibu Mugerwa, Sarina Macfadyen, Wee Tek Tay, Paul de Barro, Laura Kubatko, Titus Alica, Christopher A. Omongo, Fred Tairo, Joseph Ndunguru and Peter Sseruwagi

Introduction

Once a pest has been correctly identified, its genus and species name can provide a link to valuable indications of its ecology, biology and life history that are critical for developing control strategies. Importantly, this link should exist even when the pest was known under other names (synonyms), or was not considered a pest at all (National Research Council, 1968). Many examples have shown that incorrect identification or classification of a pest has led to fruitless searches for biocontrol agents in the native range, incorrect assignments as disease vectors, and costly, yet misdirected, suppression measures. As new approaches for delimiting species based on molecular information become more widely used, the process of correctly identifying a species has become even more complex. Fortunately, we have good systematic frameworks and nomenclatural systems that are able to cope with these challenges. Here we review challenges associated with classification and identification within the Bemisia tabaci (Gennadius) species complex. These pests and the viruses they transmit have emerged in the past few decades as among the most damaging to food and fibre crops globally (Varma & Malathi, 2003; Pimental et al., 2005; Seal et al., 2006), especially in sub-Saharan Africa (SSA). The systematics of the B. tabaci species group has been a highly debated topic for years (Boykin, 2014). Putative species are indistinguishable morphologically, so other biological data have been collected to investigate the species in the complex. Based on genetic differences (Colvin et al., 2004; Sseruwagi et al., 2005; Boykin et al., 2007; Boykin et al., 2013; Hsieh et al., 2014) and mating incompatibility (Colvin et al., 2004; Liu et al., 2007; Xu et al., 2010), B. tabaci is now recognized as a species complex that consists of at least 34 putative species (Boykin et al., 2012). The rapid discovery of significant species diversity has led to many changes in the informal names used over the last 10 years (Boykin, 2014), creating confusion in the literature.

Bemisia tabaci is known to cause three main types of plant damage. The first is due to direct feeding on host plants by both nymphs and adults. Second, indirect damage results from the excretion of honeydew onto the surfaces of leaves and fruit (Byrne & Bellows Jr, 1991); this acts as a substrate for the growth of sooty mould, interfering with photosynthesis and reducing the quality of fruit and fibre. The third, and often the most damaging, are the hundreds of plant viruses (> 350 species) (Polston et al., 2014) that are transmitted, which include members of the genera Begomovirus (Geminiviridae), Crinivirus (Closteroviridae), Ipomovirus (Potyviridae), Carlavirus (Betaflexiviridae) and Torraviridus (Secoviridae). These viruses have diverse viroin morphologies, genome organization and modes of transmission (Navas-Castillo et al., 2011). Most economically important plant viruses are begomoviruses, a group recognized as the most important emerging plant virus group affecting vegetable, tuber and fibre crops in subtropical and tropical regions (Varma & Malathi, 2003; Seal et al., 2006; Sseruwagi et al., 2006; Navas-Castillo et al., 2011; Polston et al., 2014).

One very important food security crop on which B. tabaci species feed is cassava (Manihot esculenta Crantz). Cassava is a major source of carbohydrate in sub-Saharan Africa. Over 800 million people globally rely on cassava for their daily calories in the tropical and subtropical regions (Howler et al., 2007; Boykin et al., 2013; Hsieh et al., 2014).
In East Africa, *B. tabaci* species, referred to as *Bemisia tabaci* SSA (sub-Saharan Africa) in the literature, are vectors of begomoviruses causing cassava mosaic disease (CMD) (Chant, 1958; Maruthi et al., 2002), and ipomoviruses causing cassava brown streak disease (CBSD; Maruthi et al., 2013). In East Africa, millions of smallholder farmers struggle to produce enough cassava to feed their families due to whitefly species in Africa has in the past decade become quite extensive, and several studies have analysed *B. tabaci* species within a phylogenetic framework (Maruthi et al., 2002; Abdullahi et al., 2003; Berry et al., 2004; Colvin et al., 2004; Sseruwagi et al., 2005; Maruthi et al., 2005a; Sseruwagi et al., 2006; Mugerwa et al., 2012; Legg et al., 2014). However, inconsistent naming (Table 1) has made it very difficult to compare studies of ecological and biological importance. For example, the characterization of the Ugandan cassava-associated populations using partial mitochondrial cytochrome oxidase I (mtCOI) DNA sequence data was originally reported to reveal two genotypic clusters, Uganda 1 (Ug1) and Uganda 2 (Ug2), which had ∼7.8% sequence dissimilarity (Legg et al., 2002). In 2004, new nomenclature was proposed to classify these and other sub-Saharan African samples as sub-Saharan 1–V (Berry et al., 2004). Sseruwagi et al. (2005) subsequently described eight genotypic clusters, Ug1–8, further adding to the confusion. The *B. tabaci* SSA1 species, which includes ‘Ug1’ of Legg et al. (2002), was also further subdivided into both SSA1 subclade I (Mugerwa et al., 2012) and subgroup 1 (Legg et al., 2014). Mugerwa et al. (2012) reported that members of the SSA subclade I are widely distributed throughout the region, while SSA subclade II is restricted to the coast. Confusion is further illustrated by the description of Ug2 by Sseruwagi et al. (2005) as the same as SSA2 (Mugerwa et al., 2012; Legg et al., 2014). SSA2 has not only been reported from sub-Saharan Africa, but an SSA2 population has been reported from the Mediterranean Basin, including Spain, France and northern Africa (Banks et al., 1999; Hadjistylli et al., 2015; Laarif et al., 2015). Banks et al. (1999) named this population the S biotype, and De la Rúa et al. (2006) placed them (along with several sub-Saharan samples) in a new subgroup (i.e. sub-Saharan subgroup VI) following the nomenclature proposed by Berry et al. (2004).

The confusion involving the SSA1 species was caused by several factors: (i) inadequate sequence sampling, since these studies failed to include all reported *B. tabaci* COI sequences in their phylogenetic analyses; (ii) lack of in-depth sequence analyses; and (iii) lack of effective communication between scientists. For example, Sseruwagi et al. (2005) used ‘Ug1′, Legg et al. (2014) used ‘SSA1 subgroups 1–4′ and Mugerwa et al. (2012) used SSA1 subclades I–II, which all appear to represent the same entity. A thorough investigation of naming across all SSA species, especially SSA1, is needed so that identities can be clarified and consolidated under a common terminology.

### Why naming matters

An accurate naming system provides the ability to use all published data to develop species-specific management strategies.

### Table 1. Naming systems used in previous published literature for the *Bemisia tabaci* SSA species.

<table>
<thead>
<tr>
<th>Publication and naming used</th>
<th>GenBank reference sequence</th>
<th>Different names for the same entity</th>
<th>Putative species name in Whiteflybase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legg et al. (2014) SSA1 subgroup 1</td>
<td>AY057180</td>
<td>SSA1 subclade I (Mugerwa et al., 2012)</td>
<td>SubSahAf1</td>
</tr>
<tr>
<td>SSA1 subgroup 2</td>
<td>AY057184</td>
<td>n/a</td>
<td>SubSahAf1</td>
</tr>
<tr>
<td>SSA1 subgroup 3</td>
<td>AF344278</td>
<td>SSA1 sub-clade II (Mugerwa et al., 2012)</td>
<td>SubSahAf1</td>
</tr>
<tr>
<td>SSA2</td>
<td>AY057175</td>
<td>‘Invader’ or ‘super-abundant’</td>
<td>SubSahAf2</td>
</tr>
<tr>
<td>SSA3</td>
<td>AF344257</td>
<td></td>
<td>SubSahAf3</td>
</tr>
<tr>
<td>East Africa 1</td>
<td>KF425620</td>
<td></td>
<td>Not included</td>
</tr>
<tr>
<td>Mugerwa et al. (2012) SSA1 subclade I</td>
<td>AY057185</td>
<td>SSA1 subgroup 1 (Legg et al., 2014)</td>
<td>SubSahAf1</td>
</tr>
<tr>
<td>SSA1 subclade II</td>
<td>AF344278</td>
<td>SSA1 subgroup 3 (Legg et al., 2014)</td>
<td>SubSahAf1</td>
</tr>
<tr>
<td>SSA2</td>
<td>AY057194</td>
<td></td>
<td>SubSahAf2</td>
</tr>
<tr>
<td>SSA3</td>
<td>AF344257</td>
<td></td>
<td>SubSahAf3</td>
</tr>
<tr>
<td>SSA4</td>
<td>AF344254</td>
<td></td>
<td>SubSahAf4</td>
</tr>
<tr>
<td>Uganda</td>
<td>AF418665</td>
<td></td>
<td>Uganda</td>
</tr>
<tr>
<td>Sseruwagi et al. (2005) Ug1</td>
<td>AY057171</td>
<td>SSA1</td>
<td>SubSahAf1</td>
</tr>
<tr>
<td>Ug2</td>
<td>AY057158</td>
<td>SSA2</td>
<td>SubSahAf2</td>
</tr>
<tr>
<td>Ug3 (mint)</td>
<td>AY013659</td>
<td>SSA6</td>
<td>SubSahAf6</td>
</tr>
<tr>
<td>Ug4 (okra)</td>
<td>AY093533</td>
<td>MED</td>
<td>Mediterranean</td>
</tr>
<tr>
<td>Ug5</td>
<td>AY093578</td>
<td>MED</td>
<td>Mediterranean</td>
</tr>
<tr>
<td>Ug6</td>
<td>AY093569</td>
<td>MEAM</td>
<td>MidEastAm1</td>
</tr>
<tr>
<td>Ug7</td>
<td>AY093521</td>
<td>IO</td>
<td>Indian Ocean</td>
</tr>
<tr>
<td>Ug8 (sweetpotato)</td>
<td>AY093579</td>
<td>Uganda</td>
<td>Uganda</td>
</tr>
</tbody>
</table>

© 2018 The Authors. *Systematic Entomology* published by John Wiley & Sons Ltd on behalf of Royal Entomological Society. doi: 10.1111/syen.12294
There is a long and detailed history of research on *B. tabaci* species and the viruses vectored by these species in East Africa. Some fundamental information has been generated about life history, rates of growth under different temperatures, host plant use and natural enemies (Macfadyen et al., 2017). This type of information is needed to underpin many of the management strategies that are likely to be adopted by smallholder farmers and to provide a foundation for the future development of locally appropriate integrated pest management strategies (Naranjo & Ellsworth, 2009). However, confusion over which species was used in published experiments or recorded in historical field surveys means that we cannot assume that the conclusions of those studies hold for the species causing problems in cassava today. The identity, which is the fundamental link to the scientific knowledge base, is questionable. Without confidence in taxonomic identity, instead of building a comprehensive knowledge base onto which new details are added, the initial work must be repeated.

It is known that species within the *B. tabaci* complex respond very differently to management interventions (Roditakis et al., 2009; Gnankine et al., 2013; Frewin et al., 2014; Naveen et al., 2017). Therefore, new management interventions implemented by farmers may have undesired consequences due to limited appreciation of the species being managed. For example, evidence from other parts of the world suggests that species of this complex differ in their susceptibility to insecticides (Wang et al., 2010; Horowitz & Ishaaya, 2014), their use of host plants (Iida et al., 2009; Sun et al., 2011; Tsueda & Tsuchida, 2011), and their capacity and specificity in virus transmission (Polston et al., 2014; Wei et al., 2014). Furthermore, the impact of different natural enemies on *B. tabaci* may change depending on pest species identity. For some specialist parasitoids, understanding host species identity is essential as they can oviposit and complete their life cycle only in certain host species. For other parasitoids, they are able to use multiple host species, but perform better on certain species. For example, He et al. (2017) compared the performance of the parasitoid *Encarsia formosa* Gahan on *B. tabaci* MEAM1 versus *B. tabaci* AsiaI1. They found that the parasitoid could use both species as a host resource but that host species identity did display subtle differences in oviposition and emergence rate (He et al., 2017). Furthermore, pre-imaginal experience (i.e. what host species they were reared on) influenced the choice of *B. tabaci* species to oviposit on. Clarity regarding which *B. tabaci* species smallholder farmers in Africa are dealing with now is essential to devising efficient management strategies for these pests in cassava production.

Identity underpins our understanding of virus transmission and epidemiology

To date, many virus transmission studies using *B. tabaci* as a vector have failed to indicate which species of the complex they used, so it is not possible to explore variation in vector competence across the complex. However, where this has been done, an increasing number of examples demonstrates differences in vector competence across the complex. For example, MED was a more efficient vector than MEAM1 of particular viruses in the *Tomato yellow leaf curl virus* group. Specifically, MED transmits TYLCV-Is more efficiently than TYLCV-Sr and thus, as the abundance of MED increases, so does the prevalence of TYLCV-Is (Sanchez-Campos et al., 1999). Similarly, AsiaI1 was not as efficient a vector as MED or MEAM1 (Li et al., 2010). In addition, different transmission rates and host specificity of the begomovirus *Tomato yellow leaf curl Sardinia virus* by colonies of different *B. tabaci* species of MEAM1, MED and SSA2 were observed (Jiang et al., 2004). In Africa, CMD and CBSD are caused by many different strains and species of begomoviruses and ipomoviruses (Brown et al., 2015; Ndunguru et al., 2015; Alicai et al., 2016). We are only just starting to understand the complexity of viruses transmitted to cassava by members of the *B. tabaci* complex in Africa, and thus able to clearly communicate the identity of the vector and relate it to epidemiology as an essential component of an effective disease management strategy.

Given what we know, understanding the correct species identity and being able to communicate it clearly are essential for developing effective, sustainable management approaches. This will be essential to resolving the problems caused by whiteflies and the viruses they transmit to cassava in Africa. However, before we can create a systematic framework for the African *B. tabaci* species, we must first decode the previously published naming systems and place them within the context of our current understanding of the *B. tabaci* species complex.

The way forward with naming

We propose the following seven steps to create a robust naming system for the African *B. tabaci* species complex.

*Refrain from using subgroups/subclades when referring to SSA1 B. tabaci species until they are tested with more molecular markers and biological experiments.* The current species names (e.g. MED, MEAM1) for the *B. tabaci* species complex are already challenging for the global community to adopt consistently (Boykin, 2014), and introducing another division in the naming system without any quantitative evidence to support delimitation is not prudent. At this stage, it is advised that using the SSA intermediate naming would lead to the least confusion, with the SSA species classified merely to the point of their putative species status, e.g. SSA1. Further new putative species, based on mtCOI sequences, should only be proposed as the result of a phylogenetic analysis that uses all available sequences that have been assessed and passed for quality and that have been analysed using statistical methods for species delimitation (Boykin et al., 2012) as well as biological data (see later). It appears probable that further speciation within what is currently classified as SSA1 has occurred (Mugerwa et al., 2012; Manani et al., 2017), but we must refrain from adding further groupings to the literature until official names can be proposed based on a robust species tree and until biological data to make these designations are matched with historical museum samples.
All studies of this complex should include the mtCOI sequence of the whiteflies used. These sequences need to be deposited in an open-access database, such as GenBank, and should be accompanied by geocoding data as well as host plant.

Collect multiple genes and generate a species tree. Bemisia tabaci systematics must move beyond use of a single gene phylogeny for mtCOI to drive our understanding of the systematics of the species complex. Recent progress in phylogenetic theory has led to the development of evolutionary models and software implementing these models that, together, allow for estimation of species trees from multi-locus and even genome-scale data, along with measures of statistical support in these estimates (Liu et al., 2009; Liu et al., 2015). The value of these phylogenetic methods in the context of B. tabaci systematics is that they can be used to aid in species delimitation using data that better elucidate the evolutionary history of the complete organism (Fujita et al., 2012; Ruane et al., 2014; Rannala, 2015; Singh et al., 2015). Bringing these two resources together (genomics and methodological advances for species tree inference) provides the opportunity to estimate a robust species-level phylogeny that can be used to test hypothesized putative species groupings, while avoiding the potential pitfalls arising from subjective species delimitation assignments based on limited data (Carstens et al., 2013).

This research will provide a much more representative foundation upon which to determine the taxonomic status of putative species, and may provide the ‘map’ of taxa as a platform from which research can be undertaken to determine biological descriptors of the species; this is as yet not possible because there are no morphological distinctions upon which to base such work.

Conduct crossing experiments to confirm putative species identified using molecular data. Currently, many of the putative species have been proposed based on partial mtCOI gene nucleotide trees. As we progress with the multigene trees and species tree, it will be important to test these putative species with corresponding reciprocal crossing experiments (Liu et al., 2012). A consistent intermediate naming system for the putative species of the B. tabaci complex will provide a realistic structure against which the existence of biological species can be effectively tested (Xu et al., 2010).

Conduct museum matching to ground-truth names. Tay et al. (2012) demonstrated that by matching the mtCOI partial gene identity between P. Gennadius’ original 1889 collection of B. tabaci (Gennadius, 1889) and the ‘B. tabaci’ sequences available to date, the name B. tabaci was proposed originally for what is now known as the ‘MED’ species. In that study, the partial mtCOI gene in the B. tabaci specimen collected in 1889 was characterized by utilizing multiple PCR primers followed by Sanger sequencing of PCR amplicons (Tay et al., 2012). Significant advances in genomics have occurred since the study by Tay et al. (2012), notably the use of next-generation sequencing (NGS) methods to sequence the complete mtDNA genome from a single B. tabaci adult (e.g. B. tabaci Asia I) (Tay et al., 2016), as well as single individuals of MEAM1, Aus, IO and MED (Tay et al., 2017a). Access to historical museum Bemisia specimens for NGS molecular analysis therefore represents an exciting step forward to resolving the confusion surrounding identities in the B. tabaci cryptic species complex (Tay et al., 2017b). Several historical specimens are available for African species name matching, such as B. longispina (Priesner & Hosny) from Egypt, B. gossypiperda var. mosaicevuccura (Ghesquière in Mayné & Ghesquière) Democratic Republic of the Congo; B. goldingi (Corbett) and B. nigeriensis (Corbett) from Nigeria; B. rhodesiensis (Corbett) from Zimbabwe; and B. manihotis (Frappa) and B. vaxssierei (Frappa) from Madagascar. These specimens will provide important historical nomenclatural information to the current African B. tabaci species. If no historical matches are found, new nomenclature should be proposed.

Compose species descriptions in compliance with the international code of zoological nomenclature. In view of naming the different species according to the International Code of Zoological Nomenclature, species should be matched with new sequences in order to use available names with priority. When available, the type specimens of the above-mentioned Bemisia species are mostly syntypes, frequently due to the presence of numerous puparia on plant leaves for some whitefly species, theoretically available for sequencing (assuming they all belong to the same species). Still, it is controversial to sacrifice type material to obtain enough DNA for extraction product for sequencing, so nondestructive extraction methods should be favoured to keep the cleared puparium as a voucher available for further morphological studies. A protocol for nondestructive DNA extraction of parasitoids (Polaszek et al., 2014) has been used for both whitefly puparia and adults. Using specimens other than type specimens may lead to misidentifications. Formal descriptions of putative species should be forthcoming as matching older specimens continues, with new names proposed for nonmatching taxa.

Be aware of the deficiencies associated with the use of universal primers. Delimitation of species within the B. tabaci complex currently relies on the mtCOI DNA markers C1-J-2195 and TL2-N-3014 (Simon et al., 1994), which were designed as very conserved universal primers for arthropods belonging to a range of holometabolan orders, but not specifically Hemiptera. Reliance on nonspecific primers for whiteflies is adding unnecessary complexity to the analyses of this diverse cryptic species complex as they lack nuance as a tool for species identification within the B. tabaci species complex (Elfekih et al., 2018). This was recently shown by Tay et al. (2017a), who suggested that these apparent suboptimal primers were the cause of the erroneous identification of MEAM2, which was shown instead to be a pseudogene copy (NUMT).

Screen DNA extractions routinely for endosymbionts. Bemisia tabaci species harbour different primary and secondary endosymbiont communities (Gueguen et al., 2010;
Kliot et al., 2014); if a new B. tabaci species is proposed, the DNA extractions should be screened for endosymbiont diversity as well.

Conclusions

A species name can provide a link to valuable information on its ecology, biology and life history that is critical for developing control strategies. The B. tabaci species complex has suffered extensively from inconsistent naming. In recent years there has been a dramatic increase in attention paid to whiteflies from Africa, which is their continent of origin (Boykin et al., 2013) and which is believed to still contain much unexplored diversity. Bearing this in mind, until formal species descriptions of the various SSA B. tabaci species are published, it is recommended that scientists use B. tabaci SSA1, SSA2, SSA3, SSA4, SSA5 etc. in their publications and avoid using subgroup or subclade designations. New putative species should have a unique mtCOI associated with them that has been rigorously checked for quality and this should be deposited in publicly accessible sequence databases (e.g. GenBank). Also, given the emerging issue of nuclear pseudogenes (i.e. NUMTs), naming new species based on the partial mtCOI should be avoided and instead the full gene should be sequenced and investigated to confirm orthology. Formal statistical methods for species delimitation should be combined with morphological studies in order to establish true species status in the group.

Acknowledgements

This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian government and the government of Western Australia. The authors have no conflicts of interest.

References


Ateka, E., Ateka, T., Ndagur, J., Sseruwagi, P. et al. (2016) Cassava brown streak virus is outsmarting the cassava immune system with a rapidly evolving genome. Scientific Reports, 6, 36164.


© 2018 The Authors. *Systematic Entomology* published by John Wiley & Sons Ltd on behalf of Royal Entomological Society. doi: 10.1111/syen.12294


Accepted 25 January 2018