A high-throughput amplicon sequencing approach for population-wide species diversity and composition survey

WT Tay[1,†], LN Court[1], S Macfadyen[1], F Jacomb[1], S Vyskočilová[1,2], J Colvin[2], PJ De Barro[3]

1. CSIRO Black Mountain Laboratories, Clunies Ross Street, ACT 2601, Australia
2. Natural Resources Institute, University of Greenwich, Central Avenue, Chatham, Maritime Kent ME4 4TB, United Kingdom
3. CSIRO Brisbane, Queensland 4001, Australia

† Author for correspondence: **E-mail:** weetek.tay@csiro.au **Tel:** +61-2-6246 4286

Abstract

Management of agricultural pests requires an understanding of pest species diversity, their interactions with beneficial insects and spatial-temporal patterns of pest abundance. Invasive and agriculturally important insect pests can build up very high populations, especially in cropping landscapes. Traditionally, sampling effort for species identification involves small sample sizes and is labour intensive. Here, we describe a high throughput sequencing (HTS) PCR amplicon method and associated analytical workflow for a rapid, intensive, high-volume survey of pest species compositions. We demonstrate our method using the *Bemisia* pest cryptic species complex as examples. The whiteflies *Bemisia* including the '*tabaci*' species are agriculturally important capable of vectoring diverse plant viruses that cause diseases and crop losses. We use our HTS amplicon platform to simultaneously process high volumes of whitefly individuals, with efficiency to detect rare (i.e., 1%) test-species and beneficial hymenopteran parasitoid species. Field-testing our HTS amplicon method across the Tanzania, Uganda and Malawi cassava cultivation landscapes, we identified the sub-Saharan Africa 1 *Bemisia* putative species as the dominant pest species, with other cryptic *Bemisia* species being detected at various abundances. We also provide evidence that *Bemisia* species compositions can be affected by sampling techniques that target either nymphs or adults. Our method is particularly suitable to molecular diagnostic surveys of cryptic insect species with high population densities. Our approach can be adopted to understand species biodiversity across landscapes, with broad implications for improving trans-boundary biosecurity preparedness, thus contributing

to molecular ecological knowledge and the development of control strategies for high-density, cryptic, pest-species complexes.

Key Words

Bemisia; African cassava whitefly; molecular diagnostics; population dynamics; Aphelinidae, parasitoids

Introduction

The ecology of a species is linked by definition to its species status. It underpins our understanding of species diversity in conservation genetics, biosecurity preparedness and developing effective pest management strategies. The correct identification of cryptic species can be challenging and often involves the use of molecular approaches. Timely identification and characterisation of pest species will greatly assist the development of effective management strategies to minimise impacts of crop losses due to pests on food security. Furthermore, techniques that enable the rapid screening and accurate identification of large numbers of individuals can accelerate research on a target pest species and lead to important insights into species ecology and biology.

The use of the maternally-inherited mitochondrial cytochrome oxidase subunit I (mtCOI) partial gene in 'DNA barcoding' via the Sanger sequencing method has contributed to molecular diagnostics of species (e.g., Berry et al., 2004; Boykin et al., 2007; Dinsdale, Cook, Riginos, Buckley, & De Barro, 2010; Elfekih, Tay, Gordon, Court, & De Barro, 2018; Tay, Evans, Boykin, & De Barro, 2012), and provided insights into the potential species diversity where cryptic species may co-exist (e.g., Hebert, Cywinska, Ball, & DeWaard, 2003; Rao, Liew, Yow, & Ratnayeke, 2018; Tay, Beckett, & De Barro, 2016). Cryptic species such as the hemipteran whitefly *Bemisia tabaci* and related 'non-*tabaci*' species are excellent examples where putative species identification can be achieved via molecular characterisation of the partial mtCOI gene (e.g., Lee, Park, Lee, Lee, & Akimoto, 2013; Mugerwa et al., 2018), provided that PCR artefacts, such as NUMTs and/or poor quality sequence trace files and/or gene regions, were first removed (e.g., Elfekih et al., 2019; Kunz, Tay, Elfekih, Gordon, & De Barro, 2019; Tay, Elfekih, Court, et al., 2017; Vyskočilová, Tay, van Brunschot, Seal, & Colvin, 2018). Within the *B. tabaci* cryptic species complex, over 40 putative *B. tabac*i cryptic species have been reported (Kunz, Tay, Elfekih, et al., 2019; Lee et al., 2013; Mugerwa et al., 2018; Vyskočilová et al., 2018). The cryptic species status of the *B. tabaci* complex reflected the significant challenges associated with the correct identification of species status (reviewed by Boykin et al., 2018; De Barro, Liu, Boykin, & Dinsdale, 2011; Tay & Gordon, 2019). Species of the *B. tabaci* complex are notorious around the world as effective vectors of viruses that cause significant diseases of crop plants (Polston, De Barro, & Boykin, 2014). In east Africa, *B. tabaci* species transmit diverse viruses that cause diseases in cassava. Cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) have led to significant crop losses and loss of production by smallholder farmers (Legg et al., 2011; Legg et al., 2014).

Understanding the *Bemisia* species present in east Africa, and their patterns of abundance and distribution in different farming landscapes, is a necessary first step to managing this pest problem (Macfadyen et al., 2018).

In African nations such as Tanzania, Malawi and Uganda, cassava is typically grown by smallholder farming families, surrounded by non-cassava crops such as tomatoes, sweet potatoes, groundnuts, soybean, and maize. Native ecosystems that include unmanaged grazing lands, weeds, and shrubs, are also present and contain plants which are capable of supporting a diversity of whitefly species. Some may be novel species, which are yet to be characterised by morphological, behavioural and molecular approaches (e.g., Mugerwa et al., 2018; Vyskocilova, Seal, & Colvin, 2019; Vyskočilová et al., 2018). Globally, cassava is an important carbohydrate source for over 500 million people (ARC 2014, accessed 21-April-2019), and is planted in Africa, South America, and Asia/South East Asia. Thirteen of the top 20 global cassava producers are African countries (i.e., Nigeria, Angola, Ghana, Democratic Republic of Congo, Malawi, United Republic of Tanzania, Cameroon, Mozambique, Benin, Sierra Leone, Madagascar, Uganda and Rwanda (<www.worldatlas.com> up-dated 25-April-2017; accessed 21-April-2019)). In other countries such as Thailand, Vietnam and Cambodia, it is grown as a source of high-quality starch and is used in diverse food products (Delaquis et al., 2018; Gotz & Winter, 2016; Graziosi et al., 2016; Kumar, Chang, Narayanan, & Ramasamy, 2017). Therefore, having techniques that enable rapid and cost-effective surveillance to investigate pest species compositions especially in cassava (and other food crops, i.e., tomatoes; other cash crops, i.e., cotton, ornamental plants) production landscapes are incredibly valuable. While identifying cryptic species diversity can best be achieved through molecular characterisation of appropriate genomic regions, adapting the concept of high-throughput sequencing (HTS) platforms to answering questions around species diversity and evolutionary interactions can be challenging.

The HTS platform offers a possible solution for surveillance of insect pest species, including cryptic species that reach high population densities, provided that methods are sufficiently flexible to accommodate study designs tailored to address targeted research questions. For example, understanding host-parasitoid relationships, cryptic species diversity, host-plant relationships, and identifying keystone species are fundamental ecological questions common to diverse ecosystems. Addressing these ecological and evolutionary questions can be challenging and requires scaling up of the data collection processes. HTS platforms represent a solution when combined with appropriate species diversity databases, flexible bioinformatics and associated analytical pipelines, and optimised molecular methods.

88 The HTS concept aims to combine individuals within each population for species identification, and is
89 particularly susceptible to sub-optimal primer efficacies (Elfekih, Tay, et al., 2018; Mugerwa et al., 2018;
90 Shatters, Powell, Boykin, Liansheng, & McKenzie, 2009; Tay, Elfekih, Court, et al., 2017) that are often
91 encountered especially in the *B. tabaci* cryptic species complex. Partial or inconsistent Polymerase Chain
92 Reaction (PCR) amplification, especially when targeting co-extracted DNA of individual cryptic species that may
93 be found together, can lead to inaccurate conclusion of species composition within the surveyed
94 agroecosystems. For example, a minor species may become over-represented within a population if co-amplified
95 within a pool of major species but enjoys superior primer efficacies compared with the major species. Also, given
96 the time, effort and cost necessarily associated with the preparation of HTS DNA libraries, especially where
97 accurate input genomic DNA concentration is needed, the high-throughput amplicon sequencing method should
98 incorporate a 'back-up' step. Such step may involve a second set of primers, ensuring the amplification of the
99 same population, should the first set of primers fail and thereby avoiding a total PCR failure. Such a back-up
100 design also enables capturing of species that are missed by the first primer pairs, and vice versa. Provided that
101 the primer pairs developed were sufficiently robust, within a pest insect species complex such as in the whitefly
102 cryptic species, beneficial insects such as parasitoid species can also be investigated with appropriate sampling
103 of pest-insect life stages (e.g., whitefly nymphs).

104 Here, we describe the HTS method and primers developed for surveying the *Bemisia* cassava whitefly,
105 related *B. tabaci* cryptic species and associated parasitoid species complex, and the workflow for processing the
106 large volume of HTS amplicon sequence reads generated. Our goal was to develop a technique that enabled us
107 to determine the dominant *Bemisia* species in any cassava field quantitatively, using a robust sample size in
108 terms of numbers of individuals. We provide a verification of primer efficacies and both successful and failed
109 detection rates of the method, and confirmed the method and primer efficacies via Sanger sequencing of
110 individual field-sampled whiteflies. Our samples were collected from geographically diverse cassava cropping
111 landscapes in Uganda, Malawi and Tanzania, so test the limitations and usefulness of this approach especially
112 for countries with significant infrastructure and economic challenges for researching agricultural pests impacting
113 food security.

114

115 **Material & Methods**

116 **Samples used in PCR primer efficacy tests**

117 The DNA of whitefly cryptic species from Africa, Asia, South America, and Australia were individually
118 extracted using the Qiagen Blood and Tissue DNA extraction kit (Cat #. 69506). Extracted DNA of all samples to
119 be used for testing of PCR primer efficacies were individually eluted in 27.5 µL of EB. Whitefly species status was
120 confirmed by PCR and Sanger sequencing using the primer set wfly-COI-F1/R1 and/or wfly-COI-F2/R2 (Suppl.
121 Material I). Primer development and Sanger sequencing followed the methods of Elfekih, Tay, et al. (2018).
122 Sanger sequencing was carried out at the BRF at JCSMR ANU, Canberra, ACT. Whitefly species used to verify
123 primer efficacies and their respective GenBank accession numbers are listed in Table 1.

124 **Table 1:** *Bemisia* species used to confirm PCR primer efficacies. With the exception of MN646951, all GenBank
125 accession numbers provided represent partial mtCOI sequences of *Bemisia* species tested where 100% sequence
126 identity against publicly available sequences have been detected.

Species	**GenBank**	**Field sample identifiers**	**Country/source**
MED-ASL	MH357342[a]	60132	Malawi (Field ID M2-O3-2), crop = sweet potato Lat -12.90243413, long 34.28411598
MEAM1	KY951449[b]		Peru
B. afer	AF418673[c]	60266	Malawi, on ornamental tree cassava in township.
Bemisia sp01	MN646951[d]	30350	Field ID T1-NC1-1, Lantana weed, Tanzania, Lat -2.74227622, Long 33.02190157
SSA1	JQ286457[e]	40011	Tanzania (Field ID T2-F2), crop = cassava Lat -6.522181425, Long 38.92432862
SSA2	AY057141[f]		NRI laboratory culture
SSA3	KM377923[g]		NRI laboratory culture
NW1	MK386668[h]		Barbados
Asia II_1	MF497064[i]		Bangladesh
Asia1	MF497040[j]		Bangladesh
Lipaleyrodes (*Bemisia*)	MN056066[k]		Australia

127 **Note: a:** Vyskocilova et al. (2019); **b:** Tay, Elfekih, Court, et al. (2017), **c:** Maruthi et al. (2004); **d:** New *Bemisia*
128 sp. (this study); **e:** Mugerwa et al. (2012); **f:** Legg, French, Rogan, Okao-Okuja, & Brown (2002); **g:** Ghosh S, Gowda
129 M, Bouvaine S (GenBank unpublished); **h:** Kunz, Tay, Court, et al. (2019); **I, j:** Acharya, R and Kee, KY (GenBank
130 unpublished); **k:** Kunz, Tay, Elfekih, et al. (2019).

131 To ascertain PCR efficacies of the high-throughput amplicon sequencing primers, various hypothetical
132 field scenarios were simulated. These scenarios involved different genomic DNA (gDNA) that belonged to 11
133 *Bemisia* species (Table 2) with predetermined species status based on the standard partial mtCOI dataset (Kunz,
134 Tay, Elfekih, et al., 2019). These hypothetical scenarios included: (I) a two-species co-occurrence scenario, in
135 which there is a dominant *Bemisia* whitefly species (e.g., SSA1) and a minor species present in different

136 predetermined proportions (i.e., Table 2: Mixed-Ia (5%), Mixed-Ib (2.5%), and Mixed-Ic (1%)), and (II-IV)
137 scenarios assuming co-occurrence of multiple mixed *Bemisia* cryptic pest species at different proportions based
138 on estimates of gDNA concentration. With the exception of the test scenario 'Mixed-III' and 'Mix-IV' which were
139 each replicated twice, all remaining test scenarios were replicated three times to estimate error rates and
140 protocol robustness.

141 **Field samples survey methods**

142 Cassava fields in Uganda, Tanzania, and Malawi were surveyed for *Bemisia* cryptic species between
143 2015-2016. In total, we conducted three data collection trips across a two-year period (trip 1, 1/8/2015 -
144 26/08/2015, trip 2, 5/4/2016 - 25/4/2016, trip 3, 29/10/2016 - 9/11/2016). In each region, we selected up to 10
145 cassava fields to survey and sample for nymph and adult *Bemisia* whiteflies (for sampling details see Macfayden
146 et al. submitted). Briefly, up to *ca*. 50 adults were collected using an aspirator from the top leaves of cassava
147 plants of known variety. Nymphs were collected by selecting leaves that were lower down on the plant (and so
148 had visible nymphs in the second to forth instar age range on the underside of the leaf). The top lip of two
149 centimetre diameter circular plastic vial (5ml volume) was used to cut small leaf discs that contained nymphs.
150 These discs were placed in the vial filled with 99% ethanol. As a comparison, nymphs from other crops
151 surrounding the focal cassava field were also collected. All ethanol-preserved samples were stored in -20˚C prior
152 to being processed in the laboratory, where nymphs were dislodged from leaves and placed in fresh Eppendorf
153 (1.5mL) test tubes in batches of between 20 and 40 individuals per field (note that in some cases, fields sampled
154 for *Bemisia* contained <20 individuals). For our research question we selected nymphs that had no obvious signs
155 of parasitism by hymenoptera parasitoids in order to optimize the *B. tabaci* DNA in the sample. We randomly
156 selected 26 field-collected *Bemisia* populations from Malawi, Uganda and Tanzania for analysis to demonstrate
157 the efficacies of the HTS mtCOI amplicon approach under field conditions (Table 2).

158

159

Table 2: Test scenarios to ascertain primer efficacies and species detection sensitivity by high-throughput sequencing platform. Average ± standard deviation of detection sensitivity for each category is provided based on amplicon reads as generated. 'Others' include amplicons associated with NUMTs, unknown *Bemisia* genomic regions, parasitoids (*Encarsia*, *Eretmocerus*) partial mtCOI gene region, bacteria including secondary symbionts, fungal, and of unknown origins. '0': Species not considered in test scenarios.

Detection sensitivity	**Selected known *Bemisia* species**											Others†
	SSA1	SSA2	SSA3	*Bemisia*. Sp01 (MN646951)	MEAM1	MED	NW1	Asia1	AsiaII-1	*Bemisia* sp. (MN056066)	*B. afer*	
Mixed-Ia (1:20)	95 (88.9±0.3)	0	0	0	0	5 (6.1±0.7)	0	0	0	0	0	0 (5±0.01)
Mixed-Ib (1:40)	97.5 (94.8±0.3)	0	0	0	0	2.5 (1.6±0.1)	0	0	0	0	0	0 (3.6±0.2)
Mixed-Ic (1:100)	99 (92±0.7)	0	0	0	1 (1.2±0.4)	0	0	0	0	0	0	0 (7.5±1.3)
Mixed-II	0	0	0	0	0	0	25 (11.4±0.7)	25 (12.7±0.9)	25 (28±1)	25 (20.2±1.2)	0	0 (27.9±2.1)
Mixed-III	0	0	0	0	0	0	30 (24.1±0.6)	50 (30.3±0.4)	10 (8.8±0.5)	10 (14±0.4)	0	0 (24.1±1.4)
Mixed-IV	58 (42.5±4.9)	8.8 (12.5±1.8)	8.8 (10.5±0.7)	6 (3±1.8)	6 (7±0.6)	6 (3.1±0.7)	0	0	0	0	6 (2.9±0.6)	0 (18.2±3.1)

Note: *Bemisia* sp01 (MN646951; From Tanzania field code: 30350). Lipaleyrodes (*Bemisia*) from Queensland, Australia (MN056066). † Only the first 10-15 amplicon contigs were searched for sequence identity by blastn (i.e., optimised for somewhat similar sequences) search against the non-redundant (nr) NCBI DNA database.

Table 3: Twenty-six field-collected whitefly samples from Malawi, Uganda and Tanzania analysed by high-throughput amplicon sequencing. Life stages were nymphs (N); and adults (A). Sample size (n). Detected individuals for each species were estimated from proportions of amplicons. *Bemisia* cryptic species detected were: *B. tabaci* Indian Ocean (IO), Mediterranean (MED), *B.* sp. unknown (U), *B. afer* (Baf), *B.* sp. 'Uganda' (Ug1), African cassava whitefly sub-Saharan Africa 1 (SSA1), sub-Saharan Africa 2 (SSA2), *Trialeurodes* sp. (T), *Encarsia* spp. (Enc), *Eretmocerus* spp. (Eret), Bacterial (Bac). Nuclear mitochondrial sequences/PCR artefacts (NUMTs) were detected in all cases. GenBank accession numbers of all novel *Erectmocerus* and *Encarsia* species were MN646915-MN646927 and MN646928-MN646950 respectively, GenBank accession numbers of all novel *Bemisia* and *Trialeurodes* whitefly species detected were MN660053-MN660056 and MN660057-MN660058, respectively. Raw HTS amplicon data can be obtained by request from the corresponding author.

Country	Site ID	Life stage	Host Plants (variety)	lat	long	Species composition	Total reads	Reads Identified (%)
Malawi	60109/60121	N, n= 26	Cassava (Gomani)	-12.639611255	34.17544665	**SSA1:** 12.7, **Baf:** 12.8, **Enc**: Y, **NUMTs:**	43,652	43,383 (99.35%)
	60109/60121	A, n= 7	Cassava (Gomani)	-12.639611255	34.17544665	**SSA1:** 6.4, **Baf**: 0.4, **NUMTs:** 0.01	54,392	54,135 (99.53%)
	60013/60126	N, n=22	Cassava (Gomani)	-13.03129506	34.29746802	**SSA1:** 21.7, **NUMTs:** 0.17	68,562	68,224 (99.51%)
	60013/60126	A, n= 8	Cassava (Gomani)	-13.03129506	34.29746802	**Baf:** 7.84, **NUMTs:** 0.09	44,424	44,166 (99.42%)
	60127/60250	N, n= 20	Cassava (Gomani)	-13.11325369	34.32143017	**SSA1:** 5.1, **Baf:** 14.5, **NUMTs:** 0.07	34,660	32,169 (92.81%)
	60127/60250	A, n= 20	Cassava (Gomani)	-13.11325369	34.32143017	**SSA1:** 15.9, **Baf:** 3.8, **NUMTs:** 0.13, **Bac:** 0.01	62,600	62,193 (99.35%)
	60032/60263	N, n= 18	Cassava (Manyokola)	-14.11054382	33.81826233	**SSA1:** 0.63, **Baf:** 14.7, **U:** 0.08, **NUMTs:** 1.45	43,156	40,341 (93.48%)
	60032/60263	A, n= 15	Cassava (Manyokola)	-14.11054382	33.81826233	**SSA1:** 5.3, **Baf:** 5.9, **U:** 0.56, **NUMTs:** 2.6	72,172	66,784 (92.53%)
	60010/60242	N, n= 30	Cassava (Beatrice)	-12.804086285	34.21788623	**SSAS1:** 18.7, **Baf:** 10.1, **NUMTs:** 0.91, **Enc:** Y, **Eret:** Y	63,506	62,888 (99.03%)
	60037/60265	N, n= 40	Cassava (Manyokola)	-14.1593299559021	33.8348985175844	**SSA1:** 0.6, **Baf:** 38.7, **Enc:** Y, **NUMTs:** 0.16	44,808	44,228 (98.71%)
	60039/60040	N, n= 40	Cassava (Manyokola)	-13.9888087453993	33.6391004272323	**SSA1:** 38.4, **Baf:** 0.7, **Eret:** Y, **T:** 0.2, **NUMTs:** 0.52	36,530	36,207 (99.12%)
Uganda	10109/10230	N, n= 40	Cassava (TME_14)	0.93830798	33.16688609	**SSA1:** 32.8, **SSA2:** 3.4, **Eret+Enc:** Y (0.5), **NUMTs:** 2	49, 601	47,906 (96.58%)
	20018/20019	N, n= 40	Cassava (Omongole)	-0.98750407	31.42085652	**SSA1:** 32.8, **Baf:** 2.5, **Eret+Enc:** Y (4.1), **NUMTs:** 0.3	38,082	37,752 (99.13%)
	10085/10336	N, n= 40	Cassava (TME_14)	0.87832027	33.19462426	**SSA1:** 35.6, **Baf:** 0.13, **Eret+Enc:** Y (2.8), **NUMTs:** 0.96	68,176	67,185 (98.55%)

	43130	N, n= 20	Cassava (TME_14)	2.35231436	32.93393845	**SSA1:** 9.9, **SSA2:** 3.1, **Baf:** 2.4, **Eret:** Y, **NUMTs:** 1.8	7,486	6,531 (87.24%)
	43131	N, n= 40	Cow peaCowpea	2.35246646	32.93482051	**SSA1:** 8.1, **MED:** 21.8, **U:** 1.6, **Eret:** Y, **NUMTs:** 2.1	6,674	5,883 (88.15%)
	53371	N, n= 40	Sweet potato	-0.52519054	31.6441164	**SSA1:** 1, **MED:** 26.2, **IO:** 0.4, **Ug1:** 11, **NUMTs:** 0.14, **Eret:** Y (0.1)	58,646	56,426 (96.21%)
	10362	N, n= 8	Cassava (Nase_14)	0.92017081	33.21595328	**SSA1:** 5.1, **SSA2:** 1.6, **NUMTs:** 0.14	7,305	6,202 (84.90%)
	33026	N, n= 20	other crop, Malakwang Rosella-like	0.97703406	33.04814937	**SSA1:** 0.6, **MED:** 19.2, **Eret:** Y, **NUMTs**: 0.2	54,877	54,726 (99.72%)
	33042	N, n= 20	Cassava (Nase_14)	1.00622674	32.98948722	**SSA1:** 18.4, **MED:** 0.03, **Eret+Enc:** Y (1.3), **NUMTs:** 0.18	47,014	46,812 (99.57%)
Tanzania	30145/30157	N, n= 40	Cassava (busanagulwa)	-2.388214545	33.05174421	**SSA1:** 33, **Eret+Enc:** Y (6.5), **NUMTs:** 0.27	58,826	58,407 (99.29%)
	30052/30061	N, n= 40	Sweet potato in cassava field	-2.41603623925181	33.0826151071912	**SSA1:** 39.3, **Baf:** 0.4, **Enc:** Y, **NUMTs:** 0.3	62,988	62,589 (99.37%)
	40013/40014	N, n= 40	Cassava (Kiroba)	-7.15118632229587	39.1807717492072	**SSA1:** 31.5, **Baf:** 6, **NUMTs:** 1.1	71,134	68,629 (96.48%)
	40279/40288	N, n= 40	Cassava (Kiroba)	-6.58507491183717	38.8436025425142	**SSA1:** 32.4, **Baf:** 1.1, **Eret+Enc:** Y (5.8), **NUMTs:** 0.34	49,610	49,154 (99.08%)
	30359	N, n= 12	Cassava (unknown)	-2.802858945	33.05936783	**SSA1:** 9.5, **U:** 1.2, **NUMTs:** 0.1	6,931	6,219 (89.73%)
	70120	N, n= 20	Cassava (marfaransa)	-6.59705482	38.81119587	**SSA1:** 14.2, **MED:** 0.1, **Eret+Enc:** Y, **NUMTs:** 4.3	6,029	5,912 (98.06%)

DNA extraction and HTS library preparation

DNA extraction of pooled individuals was carried out using the Qiagen Blood and Tissue DNA extraction kit (cat. #69056) and followed manufacturer's instructions. Extracted gDNA from each pooled sample was quantified for DNA concentration and standardised to 0.5 ng/µl prior to PCR amplification of target mtCOI region. Two PCR reactions were carried out for each pooled sample using primers wfly-PCR-F1/R1 and wfly-PCR-F2/R2 for the first and second reaction, respectively. For each reaction, the PCR amplification of each pooled sample uses 3.5ng gDNA template in a 35µL PCR reaction volume and 30 PCR cycles (PCR conditions as described in Suppl. Material I). Amplicons from each of the pooled samples were ascertained for PCR amplification success on 1.25% 1x TAE agarose gel stained with GelRed and visualised on a UV-transilluminator. Amplicons were then purified using AMPure XP beads as instructed in the Illumina **16S Metagenomic Sequencing Library Preparation** (Part # 15044223 Rev. B) protocol and resuspended in 25 µL Qiagen EB (instead of the recommended 52.5µL). Purified pooled amplicons were individually quantified for DNA concentration using a Qubit v2.0 fluorometer (Life Technologies) and Qubit HS dsDNA assay kit and indexed by amplifying 5.0 ng of the purified amplicon in 50µL PCR volume over 11 PCR cycles (see Suppl. Material I). Indexed amplicons were purified using AMPure XP beads as described above and concentration ascertained using Qubit v2.0. Equal proportions of purified indexed amplicon (i.e., 100ng) from each pooled sample were mixed with 6x Loading Dye and run on 1.25% low melt agarose at 75 V for 2 hours. The F1/R1 combined amplicons and the F2/R2 combined amplicons were excised and purified using the Zymoclean Gel DNA Recovery Kit (cat. # D4001, D4002, D4007, D4008) following the manufacturer's protocol. Purified indexed amplicons were quantified for double stranded DNA concentration using Qubit and fragment size was estimated using TapeStation (Agilent Technologies). Amplicons were calculated for their nano-molarity and diluted to 4 nM (see Illumina 470-2016-007-B) for HTS run on MiSeq sequencer using the Illumina MiSeq Reagent Kit Version 3. A detailed protocol for the amplicon library preparation and related primer sequence information is provided as Suppl. Material I.

Reference sequences for identification of *Bemisia* mtCOI

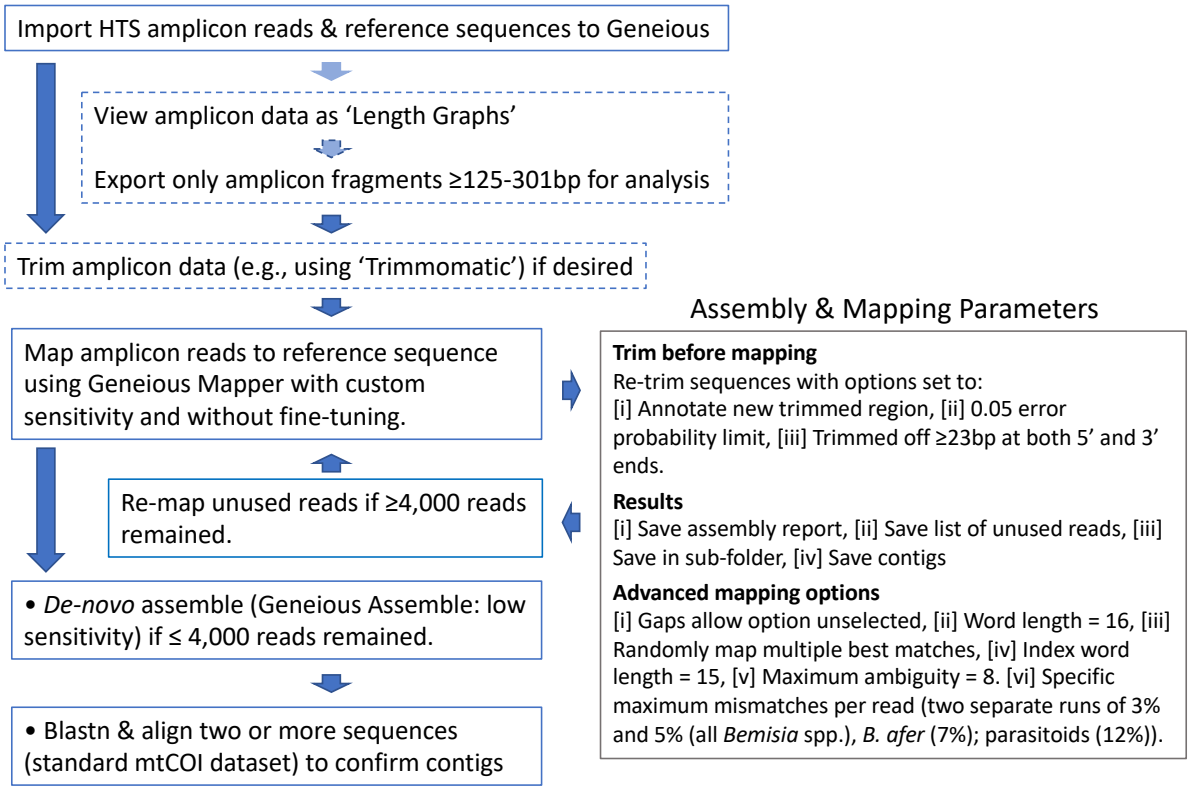
Within the African cassava cultivation landscape context, the proportion of specific *Bemisia* and related cryptic species detected was estimated as represented by the proportion of total amplicon fragments. For this purpose, we included reference sequences from the following *Bemisia* species: SSA1 (GenBank AF344264), SSA2

28 (GenBank AY057173), SSA3 (GenBank AF344257), MEAM1 (GenBank KY951449), IO (GenBank KY951448), *B. afer*
29 (GenBank KF734668), Uganda1 (GenBank KX57085), 'African' MEAM2 (GenBank KX570778), MED (GenBank,
30 KY951447), SSA6 (GenBank AY903561), SSA9 (GenBank KX570854), SSA10 (GenBank KX570837), SSA13 (GenBank
31 KX570833) and novel sub-Saharan *Bemisia* species (i.e., MN646951; MN646952) detected by Sanger sequencing
32 from field-surveyed populations (Macfadyen et al., 2019; this study). We confirmed the authenticity of these
33 reference sequences as free from NUMTs following the methods of Kunz, Tay, Elfekih, et al. (2019). A list of
34 *Bemisia* reference sequences used in this study is provided (Suppl. Material II). The reference sequences were
35 used for mapping assembly in Geneious v11.1.5 using predefined parameters (Fig. 1). Intra-specific uncorrected
36 pair-wise distance (*p*-dist) have been demonstrated to range from 0 to 2% in *B. tabaci* and 'non-*tabaci*' cryptic
37 species (Kunz, Tay, Elfekih, et al., 2019); however, this was set at 3% to also enable the detection and grouping
38 of closely related MED cryptic species complex (e.g., ASL, see Vyskočilová et al. (2018); and the Burkina Faso
39 'MED' population (see Kunz, Tay, Elfekih, et al., 2019)) that is putatively a novel species, with significant host-
40 plant range differences from the Mediterranean region MED populations (Vyskocilova et al., 2019).

41 **Parasitoids reference sequences**

42 *Eretmocerus* and *Encarsia* parasitoid mtCOI reference sequences from the sub-Saharan region
43 surrounding our field survey sites were built up (MN646919, MN646949) during the course of the analysis
44 workflow (see Fig. 1). Due to the anticipated significant whitefly hymenopteran parasitoid species diversity in
45 the sub-Saharan region, we arbitrarily determine a cut-off value of 6% within the Geneious 'Map to Reference'
46 function. The reference Aphelinidae parasitoid partial mtCOI gene sequences used in this study are provided in
47 Suppl. Material II.

48 **Fig. 1:** Workflow for analysis of MiSeq generated high-throughput sequencing of mtCOI amplicon reads for
49 estimating proportions of *Bemisia* species in the African cassava cultivation landscape, including assessment of
50 hymenopteran parasitoid species. Optional steps are indicated in dashed box depending on whether amplicon
51 libraries were gel-purified (refer to Suppl. Material I) to remove small fragments. Quality of amplicons can also
52 be improved if desired, by using an Illumina sequence trimming program such as Trimmomatic (Bolger, Lohse, &
53 Usadel, 2014). The *de-novo* assembly step enables novel *Trialeurodes*-, parasitoid- and bacterial-related
54 sequences, as well as NUMTs, to be identified.



Results

Primer efficacies: The primers wfly-COI-F1/R1 and/or wfly-COI-F2/R2 were shown to be highly efficient in successful PCR amplification of diverse representative *Bemisia* species that included Australian species (Lipaleyrodes (*Bemisia*)), Asian, New World (NW), Mediterranean (MED), Middle East Asian Minor (MEAM1), sub-Saharan African (SSA) species, and 'non-*tabaci*' species (i.e., *B. afer*). Sanger sequencing of PCR amplicon from test *Bemisia* cryptic species (Table 1) partial mtCOI confirmed species identity when compared against both the up-dated standard *B. tabaci* mtCOI reference dataset and the uncorrected pairwise nucleotide distances (*p*-dist) (see Kunz, Tay, Elfekih, et al. (2019). While Sanger sequencing confirmed sequence integrity such as through conservation of amino acid residues against published *B. tabaci* cryptic species HTS-derived full mtDNA genomes (e.g., Kunz, Tay, Court, et al., 2019; Tay, Elfekih, Court, Gordon, & De Barro, 2016; Tay, Elfekih, Court, et al., 2017; Tay, Elfekih, Polaszek, et al., 2017; Vyskočilová et al., 2018), results from HTS amplicon analyses readily detected NUMTs such as similar to 'MEAM2' (Delatte et al., 2005; Tay, Elfekih, Court, et al., 2017), as well as partial mtCOI genes with INDELs. Detection of co-amplified pseudogenes/NUMTs demonstrated the complexity of the *Bemisia*

genome (Chen et al., 2016; Xie et al., 2017; Xie et al., 2018). HTS-amplicon analyses also shed light on the unexpected co-amplification of non-target whitefly genomic regions, as well as fungal, bacterial symbionts, and parasitoids genomic regions, and highlighted the complexity of trophic interactions in *Bemisia* whiteflies (e.g., Rao et al., 2018; Shamimuzzaman et al., 2019; Tay, Elfekih, Polaszek, et al., 2017).

HTS-amplicon protocol and analysis workflow: Test scenarios for detection and identification of *Bemisia* cryptic species in predetermined proportions (i.e., 20:1; 40:1; 100:1) showed that the HTS large scale sampling protocol and the analysis workflow (Fig. 1) confidently identified the dominant as well as minor species in approximate expected proportions (Table 2). Similarly, in the hypothetical mixed-species composition scenario that included different proportions of mixed species (i.e., scenarios 'Mixed-II', 'Mixed-III', 'Mixed-IV'; Table 2), the HTS protocol also successfully detected all test species while the expected proportions were affected by co-amplification NUMTs/pseudogenes, bacterial, fungal, parasitoid, and non-specific whitefly genomic regions (Table 2). The test scenarios therefore provided significant insights into potential weaknesses associated with our primers and the robustness of the developed HTS-amplicon protocol. For example, across various scenarios, detected amplicon coverage regularly fell below the expected values due to non-specific PCR co-amplification especially of nuclear genome regions.

Field application of HTS-amplicon method: Applying the HTS amplicon protocol to field-collected samples from Malawi, Tanzania, and Uganda showed that in the African cassava fields complex *Bemisia* species compositions existed, although in most cases SSA1 was the dominant species (Macfadyen et al., 2019). A total of eight whitefly species (i.e., SSA1, SSA2, IO, MED, *B. afer,* Uganda 1, two novel *Bemisia* species), and a novel *Trialeurodes* species, were detected in different field sites. Multiple novel mtCOI haplotypes for *Eretmocerus* species (MN64915-MN64927) and *Encarsia* species (MN646929-MN646950) were also detected (Table 3). Parasitoid sequences were only ever detected in nymphal samples, and this was despite the active exclusion of parasitised nymphal samples that could be visually identified prior to genomic DNA extraction. Detection of parasitoids could represent early stages of parasitism where visual identification could be difficult. The *B. tabaci* MED species complex (i.e., included either MED (e.g., Tay et al., 2012), MED-ASL (e.g., Vyskočilová et al., 2018), and/or MED-Burkina Faso (e.g., Kunz, Tay, Elfekih, et al., 2019) putative species) was present as the dominant species in Uganda from three sampling sites (field codes 33026, 43131, 53371; Table 3), and in Malawi, *B. afer* and SSA1

species dominated different field sites (Table 3). Their detection showed significant variability in species composition within the same field sites depending on the life-stage sampled (i.e., nymphal vs. adults). Field sample analyses also detected data as NUMTs and bacterial origins in the African populations, similar to our hypothetical scenarios (Figs. 2, 3), thereby demonstrating the robustness of our HTS work flow and primer-efficacy tests.

Discussion

This study successfully demonstrated the applicability, transferability and powers of the HTS platform to investigate species compositions within an agroecological context, using the highly challenging *Bemisia* whitefly cryptic pest species as a model system. This methodology also enabled concurrent survey of whitefly-nymph, parasitoid diversity, thereby further contributing to our understanding of trophic interactions between pests, beneficial insects and plant hosts across highly heterogeneous African agricultural landscapes. Our approach transforms the molecular analysis procedures that traditionally involved low volume sampling of biological material, to one that embraces PCR amplicon HTS analyses capable of handling high volume sampled biological material. This achievement is underpinned by the development of an easy-to-use tailored bioinformatic analysis workflow, and two sets of highly efficient PCR primers that have the capability to amplify the mtCOI barcoding region of *Bemisia* hosts, their associated *Eretmocerus* and *Encarsia* parasitoids, and various fungal and bacterial homologous gene regions.

DNA barcode amplicon HTS approaches have been reported for integrative taxonomy (e.g., Cruaud, Rasplus, Rodriguez, & Cruaud, 2017; Sigut et al., 2017) and their potential for invasive species surveys reviewed (Piper et al., 2019). Adoption for landscape ecological investigations including trophic interactions under field conditions between insect pest hosts and beneficial insects is now also demonstrated to be highly effective by our study. Determining which insect represents the major species within a diverse agriculture landscape requires adequate sampling of the insects' populations (Kalyebi et al., 2019; Kalyebi et al., 2018; Macfadyen et al., 2019). Conclusions based on few individuals sampled at a large number of sites are inherently problematic; there is a need, therefore, for the development of techniques to scale up data collection and processing of samples, and to develop associated bioinformatic workflows and analytical pipelines to provide robust findings at economic

124 scales. For the cassava cultivation landscape in sub-Saharan Africa, determining the dominant pest whitefly
125 species represents a significant challenge due to difficulties with species identification without the help of
126 molecular DNA markers (De Barro et al., 2011; Kunz, Tay, Elfekih, et al., 2019). Furthermore, the often-high
127 population sizes associated with individual fields, poor PCR primer efficacies (Elfekih, Tay, et al., 2018; Mugerwa
128 et al., 2018; Shatters et al., 2009; Tay, Elfekih, Court, et al., 2017) often resulting in suboptimal PCR amplification,
129 and the difficulty encountered in sequence quality control (Kunz, Tay, Elfekih, et al., 2019; Tay, Elfekih, Court, et
130 al., 2017; Vyskočilová et al., 2018), have been persistent problems in *Bemisia* molecular systematics and
131 diagnostics. In this study, we showed that by applying the DNA barcoding and the species 'genetic gaps' concepts
132 first demonstrated by Dinsdale et al. (2010; see also Kunz, Tay, Elfekih, et al., 2019), and with careful re-designing
133 of DNA markers for the HTS platform, it was possible to develop a highly efficient method for large-scale,
134 effective and economic sampling of these whitefly cryptic species to contribute to their management and control
135 in challenging agricultural landscape settings. The primers were also shown to be efficient in detecting novel
136 *Encarsia* and *Eretmocerus* species of the hymenopteran parasitoids in the African cassava and non-cassava fields,
137 thereby providing value and saving effort to understand trophic interactions between crop hosts, insect pests
138 and their potential biological control agents.

139 The traditional approach (i.e., PCR and Sanger sequencing) using individual whiteflies limits our ability
140 to ascertain the common pest species within a landscape because: (i) mtCOI species identification method is
141 laborious when specimens are to be extracted individually for Sanger sequencing, and (ii) primer efficacies are
142 often poor due to the large nucleotide diversity that exists between *B. tabaci* cryptic species when applying non-
143 hemipteran 'universal' primers (Elfekih, Tay, et al., 2018). Sampling sufficiently large quantities of *B. tabaci*
144 cryptic species across diverse agricultural landscape will need to rely on a standardised sampling approach
145 (Sseruwagi, Sserubombwe, Legg, Ndunguru, & Thresh, 2004; Macfadyen et al., 2019), followed by species
146 identification via molecular characterisation of the mtCOI gene. The approach developed here is especially useful
147 for collecting nymphal data that has a strong link to host plant use, and can be further developed to examine
148 parasitism rates of *Bemisia* species. However, standardized sampling techniques for nymphs are less well-
149 developed than for adult *B. tabaci* on cassava (which are considered easier to count and collect). In reality, the
150 focus of the research question should drive the sampling approach selected (be that nymphs, parasitized

151 nymphs, or adults); however, in all cases scaling up the numbers of individuals processed is critical. An important
152 distinction of sampling nymphal *cf*. adult individuals could be seen in the Malawi samples (Table 3), where our
153 HTS species diagnostic method detected significant differences in species compositions, especially between *B*.
154 *afer* and SSA1. Factors underpinning the detected species composition differences could involve ecological and
155 biological factors despite being from the same sampling sites.

156 While our method enables efficient ascertainment of species composition at the landscape level, this
157 method should not be used to determine individual maternal lineages of mtDNA haplotypes due to the high
158 sensitivity of detecting random PCR-introduced SNPs in individual amplicon molecules. At the landscape survey
159 level, ascertaining specific maternal lineages does not represent a criterion needed to enable species identity to
160 be determined, provided that sequence quality was adequately managed to avoid misidentification of species
161 status due to mtDNA pseudogenes (Elfekih, Etter, et al., 2018; Kunz, Tay, Elfekih, et al., 2019; Tay, Elfekih, Court,
162 et al., 2017; Vyskočilová et al., 2018). Elfekih et al. (2019) have shown that NUMTs and pseudogenes could lead
163 to significant misinterpretation of *Bemisia* population structure in the African cassava landscape regardless of
164 whether partial mtCOI or genome-wide SNPs were used (see e.g., Wosula, Chen, Fei, & Legg, 2017). A significant
165 limitation on our approach is the quality of the library used to interrogate the sequences generated. Sanger
166 sequencing-generated partial mtCOI gene sequences in the public database (e.g., NCBI; see Vyskočilová et al.
167 (2018) regarding MED-related pseudogenes/NUMTs; see also Kunz, Tay, Elfekih, et al. (2019) for reanalysis of *B.*
168 *tabaci* DNA database (Boykin, Savill, & De Barro., 2017)) are gradually being recognised for containing high
169 numbers of problematic sequences (see e.g., Nacer & do Amaral (2017); Song, Buhay, Whiting, & Crandall (2008);
170 Triant & Dewoody (2007)). Such readily available public DNA global databases will need to be critically reviewed
171 and modified, given that a large scientific community relies on these sequences to assist with species
172 identification.

173 With advances in HTS to generate genome-wide SNPs data and genome resequencing, the efficacy of
174 the partial mtCOI gene as a molecular marker for species diagnostics has come under scrutiny. The inability of
175 the mtCOI gene to effectively differentiate African *Bemisia* species would substantially reduce the efficiency of
176 our HTS amplicon species identification method. Conflicting genetic signatures of African *Bemisia* SSA1, SSA2 and
177 SSA4 species status based on partial mtCOI and from genome-wide SNP patterns have been reported (Wosula et

178 al., 2017). Reanalysis of the Wosula et al. (2017) mtCOI and nextRAD data by Elfekih et al. (2019) identified sub-
179 optimal clean-up of SNP data and inclusion of NUMTs, as main factors that underpinned species status
180 discrepancies between mtCOI and genome-wide SNP markers. Removal of poor quality nextRAD data, NUMTs
181 and pseudogenes revealed that the SSA1 and SSA2 species status were in total agreement between the mtCOI
182 and nuclear markers. In addition, genome-wide SNP data provided clues as to the potential presence of sub-
183 species status/population substructure within the SSA1 (i.e., SSA1-NW and SSA1-SE, see Elfekih et al., 2019), and
184 showed SSA4 to be a misidentified species due to NUMTs, with similar genome-wide SNP signature of MEAM2
185 from MEAM1 (Elfekih, Etter, et al., 2018). The study of Elfekih et al. (2019), therefore, provided strong support
186 for the validity of our DNA-barcode HTS methods for landscape species survey purposes, including for the highly
187 challenging African cassava whitefly complex.

188 Our HTS amplicon sequencing method can be adapted to understand other pest species complexes in
189 agricultural systems and may be especially effective for invasive insect pests with small body size and where
190 population sizes can build up in large numbers, such as in invasive whiteflies, thrips, mites, and aphids. We
191 propose, therefore, to apply the high-throughput amplicon sequencing approach to: (1) greatly increase number
192 of individuals that can be sampled, processed and characterised simultaneously for their 'DNA barcoding' gene
193 including the mtCOI, 16s rRNA, 18s rRNA and 28s rRNA genes, (2) estimate individual species proportion across
194 large agricultural landscapes, (3) in nymph samples, identify parasitoid species that target the agricultural pests
195 of interest. While the high-throughput amplicon sequencing method in a species composition study in
196 agroecological landscapes represents a relatively new field of research, the use of this method to identify genes
197 of biosecurity importance has been demonstrated for understanding the spread of pyrethroid resistance genes
198 in the red-legged earth mites *Halotydeus destructor* (Edwards et al., 2018). The method described by Edwards et
199 al. (2018) can be incorporated into our high-throughput amplicon sequencing platform for rapid species
200 composition identification and resistance allele frequency estimates. This would enable a genomic approach that
201 would simultaneously allow the identification of species composition at landscape scales, the survey of
202 insecticide resistance profiles (e.g., Hopkinson et al., 2019), and assessment of the implication of chemical control
203 strategies on diversity and density of beneficial insects such as parasitoids. Such study would be especially
204 informative when temporal scales are included within the study designs.

205 Developmental stages sampled do impact *Bemisia* species compositions and the conclusions drawn
206 from a survey, as demonstrated in the Malawi cassava sampling sites for adult and nymphal *Bemisia* populations.
207 The extent of such variability in target species composition between immature and adult life stages for many
208 other biological systems can also benefit from the HTS method described here. In this study, *Bemisia* nymphs
209 that visibly showed signs of being parasitised by hymenopteran parasitoids were actively excluded prior to
210 genomic processing. Despite this, *Encarsia* and *Eretmocerus* parasitoid mtCOI sequences were detected, thereby
211 highlighting the importance of biological control of *Bemisia* pest species by parasitoids in the sub-Saharan African
212 agricultural system. While we have not optimised this workflow for addressing host-parasitoid interaction
213 research questions, we have demonstrated the power of HTS genomic approaches to document and facilitate an
214 understanding of host-parasitoid interactions. Improved understanding of the ecological impact of pest species
215 systems from genomic perspectives will further transform our understanding of sustainable agricultural
216 production. Appropriate experimental designs including careful development of sampling procedures will be
217 needed for meaningful interpretation that takes advantage of the large volume of genomic data generated by
218 the HTS platform. Our study demonstrates the use of HTS amplicon methods successfully in classical field-
219 ecological studies and for novel cryptic species discovery. The adoption of this approach more widely has the
220 potential to transform global biosecurity preparedness and thereby to impact human health and food security.

221

222 **Acknowledgements**

223 Laboratory colonies of whitefly samples (*Bemisia* SSA1, SSA2, SSA3) were provided by the NRI, University
224 of Greenwich, UK. J. Ryu (CSIRO) assisted with sample sorting, C. Paull and A. Hulthen (CSIRO) assisted with field
225 surveys of African cassava whitefly samples. We thank Dr Andy Polaszek (NHM, UK) for discussion on the
226 *Encarsisa* and *Eretmocerus* parasitoids detected in the *Bemisia* hosts. This work was supported by the Natural
227 Resources Institute, University of Greenwich from a grant provided by the Bill & Melinda Gates foundation (Grant
228 Agreement OPP1058938). We would like to thank the broader African cassava whitefly project team for their
229 enthusiasm and support throughout this study.

230

Data Availability Statement

The data that support the findings of this study are available as Supplemental Material II, and from GenBank (GenBank accession numbers cited throughout). Raw data can be obtained from the corresponding author upon reasonable request.

References

Berry, S. D., Fondong, V. N., Rey, C., Rogan, D., Fauquet, C. M., & Brown, J. K. (2004). Molecular evidence for five distinct *Bemisia tabaci* (Homoptera : Aleyrodidae) geographic haplotypes associated with cassava plants in sub-Saharan Africa. *Annals of the Entomological Society of America, 97*(4), 852-859. doi:Doi 10.1603/0013-8746(2004)097[0852:Meffdb]2.0.Co;2

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Boykin, L. M., Kinene, T., Wainaina, J. M., Savill, A., Seal, S., Mugerwa, H., . . . Sseruwagi, P. (2018). Review and guide to a future naming system of African *Bemisia tabaci* species. *Systematic Entomology, 43*(3), 427-433. doi:10.1111/syen.12294

Boykin, L. M., Savill, A., & De Barro, P. (2017). Updated mtCOI reference dataset for the *Bemisia tabaci* species complex *F1000Res, 6*, 1835. doi:10.12688/f1000research.12858.1

Boykin, L. M., Shatters, R. G., Jr., Rosell, R. C., McKenzie, C. L., Bagnall, R. A., De Barro, P., & Frohlich, D. R. (2007). Global relationships of *Bemisia tabaci* (Hemiptera: Aleyrodidae) revealed using Bayesian analysis of mitochondrial COI DNA sequences. *Mol Phylogenet Evol, 44*(3), 1306-1319. doi:10.1016/j.ympev.2007.04.020

Chen, W. B., Hasegawa, D. K., Kaur, N., Kliot, A., Pinheiro, P. V., Luan, J. B., . . . Fei, Z. J. (2016). The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biology, 14*. doi:ARTN 110 doi:10.1186/s12915-016-0321-y

Cruaud, P., Rasplus, J. Y., Rodriguez, L. J., & Cruaud, A. (2017). High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. *Sci Rep, 7*, 41948. doi:10.1038/srep41948

De Barro, P. J., Liu, S. S., Boykin, L. M., & Dinsdale, A. B. (2011). *Bemisia tabaci*: a statement of species status. *Annu Rev Entomol, 56*, 1-19. doi:10.1146/annurev-ento-112408-085504

Delaquis, E., Andersen, K. F., Minato, N., Cu, T. T. L., Karssenberg, M. E., Sok, S., . . . de Haan, S. (2018). Raising the Stakes: Cassava Seed Networks at Multiple Scales in Cambodia and Vietnam. . *Frontiers in Sustainable Food Systems, 2*, 73. doi:https://doi.org/10.3389/fsufs.2018.00073

Delatte, H., Reynaud, B., Granier, M., Thornary, L., Lett, J. M., 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. *Bulletin of Entomological Research, 95*(1), 29-35. doi:10.1079/Ber2004337

Dinsdale, A., Cook, L., Riginos, C., Buckley, Y. M., & De Barro, P. (2010). Refined Global Analysis of *Bemisia tabaci* (Hemiptera: Sternorrhyncha: Aleyrodoidea: Aleyrodidae) Mitochondrial Cytochrome Oxidase 1 to Identify Species Level Genetic Boundaries. *Annals of the Entomological Society of America, 103*(2), 196-208. doi:10.1603/An09061

Edwards, O. R., Walsh, T. K., Metcalfe, S., Tay, W. T., Hoffmann, A. A., Mangano, P., . . . Umina, P. A. (2018). A genomic approach to identify and monitor a novel pyrethroid resistance mutation in the redlegged earth mite, *Halotydeus destructor*. *Pestic Biochem Physiol, 144*, 83-90. doi:10.1016/j.pestbp.2017.12.002

Elfekih, S., Etter, P., Tay, W. T., Fumagalli, M., Gordon, K., Johnson, E., & De Barro, P. (2018). Genome-wide analyses of the *Bemisia tabaci* species complex reveal contrasting patterns of admixture and complex demographic histories. *PLoS ONE, 13*(1), e0190555. doi:10.1371/journal.pone.0190555

Elfekih, S., Tay, W. T., Gordon, K., Court, L. N., & De Barro, P. J. (2018). Standardized molecular diagnostic tool for the identification of cryptic species within the *Bemisia tabaci* complex. *Pest Manag Sci, 74*(1), 170-173. doi:10.1002/ps.4676

Elfekih, S., Tay, W. T., Polaszek, A., Gordon, K. H. J., Kunz, D., Macfadyen, S., . . . De Barro, P. J. (2019). On species delimitation, hybridization and population structure of cassava whitefly in Africa. *Genome Bio Evol, Subitted* doi: https://doi.org/10.1101/836072.

Gotz, M., & Winter, S. (2016). Diversity of *Bemisia tabaci* in Thailand and Vietnam and indications of species replacement. *Journal of Asia-Pacific Entomology, 19*(2), 537-543. doi:10.1016/j.aspen.2016.04.017

Graziosi, I., Minato, N., Alvarez, E., Ngo, D. T., Hoat, T. X., Aye, T. M., . . . Wyckhuys, K. A. (2016). Emerging pests and diseases of South-east Asian cassava: a comprehensive evaluation of geographic priorities, management options and research needs. *Pest Manag Sci, 72*(6), 1071-1089. doi:10.1002/ps.4250

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences, 270*(1512), 313-321. doi:10.1098/rspb.2002.2218

Hopkinson, J., Pumpa, S., van Brunschot, S., Fang, G., Frese, M., Tay, W. T., & Walsh, T. (2019). Insecticide resistance status of *Bemisia tabaci* MEAM1 (Hemiptera: Aleyrodidae) in Australia cotton production valleys. *Austral Entomology, In press*.

Kalyebi, A., Macfadyen, S., Hulthen, A., Ocitti, P., Jacomb, F., Tay, W. T., . . . De Barro, P. (2019). Within-season changes in land use impact pest abundance in smallholder African cassava production systems. *PLoS ONE, Submitted*.

Kalyebi, A., Macfadyen, S., Parry, H., Tay, W. T., De Barro, P., & Colvin, J. (2018). African cassava whitefly, *Bemisia tabaci,* cassava colonization preferences and control implications. *PLoS ONE, 13*(10), e0204862. doi:10.1371/journal.pone.0204862

Kumar, N. R., Chang, J. C., Narayanan, M. B., & Ramasamy, S. (2017). Phylogeographical structure in mitochondrial DNA of whitefly, *Bemisia tabaci* Gennadius (Hemiptera: Aleyrodidae) in southern India and Southeast Asia. *Mitochondrial DNA A DNA Mapp Seq Anal, 28*(5), 621-631. doi:10.3109/24701394.2016.1160075

Kunz, D., Tay, W. T., Court, L. N., Elfekih, S., Gordon, K. H. J., Evans, G. A., & De Barro, P. J. (2019). Draft mitochondrial DNA genome of a 1920 Barbados cryptic *Bemisia tabaci* 'New World' species (Hemiptera: Aleyrodidae). *Mitochondrial DNA Part B-Resources, 4*(1), 1183-1184. doi:10.1080/23802359.2019.1591197

Kunz, D., Tay, W. T., Elfekih, S., Gordon, K. H. J., & De Barro, P. J. (2019). Take out the rubbish - Removing NUMTs and pseudogenes from the *Bemisia tabaci* cryptic species mtCOI database. *Sci Rep, Submitted*. doi:https://doi.org/10.1101/724765

Lee, W., Park, J., Lee, G. S., Lee, S., & Akimoto, S. (2013). Taxonomic status of the *Bemisia tabaci* complex (Hemiptera: Aleyrodidae) and reassessment of the number of its constituent species. *PLoS ONE, 8*(5), e63817. doi:10.1371/journal.pone.0063817

Legg, J. P., French, R., Rogan, D., Okao-Okuja, G., & Brown, J. K. (2002). A distinct *Bemisia tabaci* (Gennadius) (Hemiptera : Sternorrhyncha : Aleyrodidae) genotype cluster is associated with the epidemic of severe cassava mosaic virus disease in Uganda. *Molecular Ecology, 11*(7), 1219-1229. doi:DOI 10.1046/j.1365-294X.2002.01514.x

Legg, J. P., Jeremiah, S. C., Obiero, H. M., Maruthi, M. N., Ndyetabula, I., Okao-Okuja, G., . . . Lava Kumar, P. (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Res, 159*(2), 161-170. doi:10.1016/j.virusres.2011.04.018

Legg, J. P., Shirima, R., Tajebe, L. S., Guastella, D., Boniface, S., Jeremiah, S., . . . Rapisarda, C. (2014). Biology and management of *Bemisia* whitefly vectors of cassava virus pandemics in Africa. *Pest Manag Sci, 70*(10), 1446-1453. doi:10.1002/ps.3793

Macfadyen, S., Paull, C., Boykin, L. M., De Barro, P., Maruthi, M. N., Otim, M., . . . Omongo, C. A. (2018). Cassava whitefly, *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) in East African farming landscapes: a review of the factors determining abundance. *Bull Entomol Res, 108*(5), 565-582. doi:10.1017/S0007485318000032

Macfadyen, S., Tay, W. T., Hulthen, A., Paull, C., Kalyebi, A., Jacomb, F., . . . De Barro, P. (2019). Landscape factors and how they influence whitefly pests in cassava fields across East Africa. *Landscape Ecology, Submitted*.

Maruthi, M. N., Colvin, J., Thwaites, R. M., Banks, G. K., Gibson, G., & Seal, S. E. (2004). Reproductive incompatibility and cytochrome oxidase I gene sequence variability amongst host-adapted and geographically separate *Bemisia tabaci* populations (Hemiptera : Aleyrodidae). *Systematic Entomology, 29*(4), 560-568. doi:DOI 10.1111/j.0307-6970.2004.00272.x

Mugerwa, H., Rey, M. E., Alicai, T., Ateka, E., Atuncha, H., Ndunguru, J., & Sseruwagi, P. (2012). Genetic diversity and geographic distribution of *Bemisia tabaci* (Gennadius) (Hemiptera: Aleyrodidae) genotypes associated with cassava in East Africa. *Ecol Evol, 2*(11), 2749-2762. doi:10.1002/ece3.379

Mugerwa, H., Seal, S., Wang, H. L., Patel, M. V., Kabaalu, R., Omongo, C. A., . . . Colvin, J. (2018). African ancestry of New World, *Bemisia tabaci*-whitefly species. *Sci Rep, 8*(1), 2734. doi:10.1038/s41598-018-20956-3

Nacer, D. F., & do Amaral, F. R. (2017). Striking pseudogenization in avian phylogenetics: Numts are large and common in falcons. *Molecular Phylogenetics and Evolution, 115*, 1-6. doi:10.1016/j.ympev.2017.07.002

Piper, A. M., Batovska, J., Cogan, N. O. I., Weiss, J., Cunningham, J. P., Rodoni, B. C., & Blacket, M. J. (2019). Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *Gigascience, 8*(8). doi:10.1093/gigascience/giz092

Polston, J. E., De Barro, P., & Boykin, L. M. (2014). Transmission specificities of plant viruses with the newly identified species of the *Bemisia tabaci* species complex. *Pest Management Science, 70*(10), 1547-1552. doi:10.1002/ps.3738

Rao, R. S., Liew, T. S., Yow, Y. Y., & Ratnayeke, S. (2018). Cryptic diversity: Two morphologically similar species of invasive apple snail in Peninsular Malaysia. *PLoS ONE, 13*(5), e0196582. doi:10.1371/journal.pone.0196582

Shamimuzzaman, M., Hasegawa, D. K., Chen, W., Simmons, A. M., Fei, Z., & Ling, K. S. (2019). Genome-wide profiling of piRNAs in the whitefly *Bemisia tabaci* reveals cluster distribution and association with begomovirus transmission. *PLoS ONE, 14*(3), e0213149. doi:10.1371/journal.pone.0213149

Shatters, R. G., Jr., Powell, C. A., Boykin, L. M., Liansheng, H., & McKenzie, C. L. (2009). Improved DNA barcoding method for *Bemisia tabaci* and related Aleyrodidae: development of universal and *Bemisia tabaci* biotype-specific mitochondrial cytochrome c oxidase I polymerase chain reaction primers. *J Econ Entomol, 102*(2), 750-758. doi:10.1603/029.102.0236

Sigut, M., Kostovcik, M., Sigutova, H., Hulcr, J., Drozd, P., & Hrcek, J. (2017). Performance of DNA metabarcoding, standard barcoding, and morphological approach in the identification of host-parasitoid interactions. *PLoS ONE, 12*(12), e0187803. doi:10.1371/journal.pone.0187803

Song, H., Buhay, J. E., Whiting, M. F., & Crandall, K. A. (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America, 105*(36), 13486-13491. doi:10.1073/pnas.0803076105

Sseruwagi, P., Sserubombwe, W. S., Legg, J. P., Ndunguru, J., & Thresh, J. M. (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*(1), 129-142. doi:10.1016/j.virusres.2003.12.021

Tay, W. T., Beckett, S. J., & De Barro, P. J. (2016). Phosphine resistance in Australian *Cryptolestes* species (Coleoptera: Laemophloeidae): perspectives from mitochondrial DNA cytochrome oxidase I analysis. *Pest Manag Sci, 72*(6), 1250-1259. doi:10.1002/ps.3805

Tay, W. T., Elfekih, S., Court, L., Gordon, K. H., & De Barro, P. J. (2016). Complete mitochondrial DNA genome of *Bemisia tabaci* cryptic pest species complex Asia I (Hemiptera: Aleyrodidae). *Mitochondrial DNA Part A, 27*(2), 972-973. doi:10.3109/19401736.2014.926511

Tay, W. T., Elfekih, S., Court, L. N., Gordon, K. H. J., Delatte, H., & De Barro, P. J. (2017). The Trouble with MEAM2: Implications of Pseudogenes on Species Delimitation in the Globally Invasive *Bemisia tabaci* (Hemiptera: Aleyrodidae) Cryptic Species Complex. *Genome Biology and Evolution, 9*(10), 2732-2738. doi:10.1093/gbe/evx173

Tay, W. T., Elfekih, S., Polaszek, A., Court, L. N., Evans, G. A., Gordon, K. H. J., & De Barro, P. J. (2017). Novel molecular approach to define pest species status and tritrophic interactions from historical *Bemisia* specimens. *Scientific Reports, 7*. doi:ARTN 429 doi:10.1038/s41598-017-00528-7

Tay, W. T., Evans, G. A., Boykin, L. M., & De Barro, P. J. (2012). Will the Real *Bemisia tabaci* Please Stand Up? *PLoS ONE, 7*(11). doi:ARTN e50550 doi:10.1371/journal.pone.0050550

Tay, W. T., & Gordon, K. H. J. (2019). Going global - genomic insights into insect invasions. *Curr Opin Insect Sci, 31*, 123-130. doi:10.1016/j.cois.2018.12.002

Tay, W. T., Court, L. N., Macfayden, S., Jacomb, F., Vyskocilova, S., Colvin, J., & De Barro, J. P. (2019). A high-throughput amplicon sequencing approach for population-wide species diversity and composition survey. Supplemental Material II. *Mol Ecol Res, Submitted.*

Triant, D. A., & Dewoody, J. A. (2007). The occurrence, detection, and avoidance of mitochondrial DNA translocations in mammalian systematics and phylogeography. *Journal of Mammalogy, 88*(4), 908-920. doi:Doi 10.1644/06-Mamm-a-204r1.1

Vyskocilova, S., Seal, S., & Colvin, J. (2019). Relative polyphagy of "Mediterranean" cryptic *Bemisia tabaci* whitefly species and global pest status implications. *Journal of Pest Science, 92*(3), 1071-1088. doi:10.1007/s10340-019-01113-9

Vyskočilová, S., Tay, W. T., van Brunschot, S., Seal, S., & Colvin, J. (2018). An integrative approach to discovering cryptic species within the *Bemisia tabaci* whitefly species complex. *Sci Rep, 8*(1), 10886. doi:10.1038/s41598-018-29305-w

Wosula, E. N., Chen, W. B., Fei, Z. J., & Legg, J. P. (2017). Unravelling the Genetic Diversity among Cassava *Bemisia tabaci* Whiteflies Using NextRAD Sequencing. *Genome Biology and Evolution, 9*(11), 2958-2973. doi:10.1093/gbe/evx219

Xie, W., Chen, C., Yang, Z., Guo, L., Yang, X., Wang, D., . . . Zhang, Y. (2017). Genome sequencing of the sweetpotato whitefly *Bemisia tabaci* MED/Q. *Gigascience, 6*(5), 1-7. doi:10.1093/gigascience/gix018

Xie, W., Yang, X., Chen, C., Yang, Z., Guo, L., Wang, D., . . . Zhang, Y. (2018). The invasive MED/Q *Bemisia tabaci* genome: a tale of gene loss and gene gain. *BMC Genomics, 19*(1), 68. doi:10.1186/s12864-018-4448-9